The Effects of Genetic-based and Swarm Intelligence-based Feature Selection on Adversarial Author Identification

By

Steve Halladay

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2022

Keywords: Author Identification, Feature Selection, Adversarial Author Identification, Genetic Algorithms, Swarm Intelligence

Approved by

Gerry Dozier, Chair, Professor of Computer Science and Software Engineering
Cheryl Seals, Professor of Computer Science and Software Engineering
David Umphress, Professor of Computer Science and Software Engineering
Jakita O. Thomas, Professor of Computer Science and Software Engineering

Abstract

Within the realm of author identification, where researchers work to classify writing samples by author, researchers are using more and diverse feature sets to try to improve classification accuracy. From a computational cost perspective, these additional feature sets become problematic. Further, adding more feature sets may inadvertently decrease classification accuracy. Therefore, selecting the appropriate subset of features is an important challenge for researchers.

However, the feature subset selection concern becomes even more challenging due to a couple of complexities. The first complexity is that different datasets require different feature sets for good identification performance. A feature set that performs well with one dataset may not perform well with another. So, it is important to customize the feature set to the characteristics of the dataset. The second complexity is that it appears that feature selection makes author identification systems more susceptible to adversarial attacks. These attacks occur when authors attempt to obfuscate their writing style or impersonate another author's writing style.

The focus of the research in this work is in this second area of complexity, namely, understanding the susceptibility of adversarial attacks on author identification systems due to feature selection. Specifically, this research investigates the susceptibility of adversarial attacks on author identification systems that use genetic-based and swarm intelligence-based feature selection. The intent of this research is to observe and characterize the factors affecting adversarial susceptibility by considering several parameters, including dataset content, dataset size and feature selection algorithm.

This work employs two datasets: the CASIS dataset, which is a collection of blog posts, and the PAN19 dataset, which is a collection of extracts from Twitter feeds and includes bot-generated writing samples. We vary the dataset sizes to ascertain the effects of a larger author pool. We also vary the bias towards minimizing the feature set. Then, we analyze the data to determine those factors that correlate with successful adversarial attacks on author identification systems both with and without feature selection.

Acknowledgments

I express sincere gratitude to Dr. Gerry Dozier, who served as my dissertation advisor, for his direction and patience. I also appreciate the time and effort of my committee members despite their grueling schedules and workloads. I also acknowledge my wonderful wife who has patiently supported me in my work.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ABCO	Artificial Bee Colony Optimization
AId	Author Identification
AIdS	Author Identification System
ASO	Ant System Optimization
EC	Evolutionary Computation
EP	Evolutionary Programming
GA	Genetic Algorithm
GEFeS	Genetic & Evolutionary Feature Selection
GSO	Glowworm Swarm Optimization
HTML	Hyper-Text Mark-up Language
LIWC	Linguistic Inquiry & Word Count
LSVM	Support Vector Machine
PSO	Particle Swarm Optimization
RAND	Random feature selection algorithm
RBF	Radial Basis Function Network
SI	Swarm Intelligence
STY	Stylometry
tf/idf	Term Frequency/Inverse Document Frequency
TM	Topic Modeling

Chapter 1

Introduction

1.1 Motivation

Author identification (AId) is the task of classifying the authorship of text samples [5, 6]. This task can be used for several purposes ranging from identifying the literary work of anonymous authors [4] to forensic applications [5]. AId has been a topic of interest since at least the late nineteenth century when Thomas Mendenhall [48] attempted to settle the controversy of the authorship of works attributed to Shakespeare, which some believed had been authored by Bacon.

The process of AId begins by extracting features from text samples with known authors [5]. Then, the process looks for patterns of features that uniquely correlate to the samples' authors. Armed with feature profiles that correspond with specific authors, the author identification system (AIdS) extracts these same features from text samples with unknown authors and matches the resulting features to the known authors' feature profiles.

The features that AIdSs extract from text samples are varied [8, 10]. The features may include simple stylometric features [4, 42], such as character and word frequencies [6], word length histograms [49], and more sophisticated features such as text sample topic classifications [26], and author sentiments or states of mind [21]. The combination of features the AIdS uses is known as a feature set.

A common measure of an AIdS is accuracy, which is the percentage of anonymous samples the system can correctly identify. Therefore, a goal of AId researchers is to improve systems' accuracy. It turns out that effectiveness of an AIdS depends heavily on the feature set employed by the system [28]. In addition, the effectiveness of a feature set varies depending on the text sample datasets [9]. AIdSs may be able to classify some datasets well with one feature set, but these same systems with the same feature set may not perform as well on a different dataset [9]. This variability causes researchers to explore a wide variety of feature sets [4, 5, 6, 7 8, 9, 10].

As the number of features increase, so to do the computational costs of extracting and processing the features. Further, additional features do not always improve the accuracy of AIdSs [13]. In some instances, they may detract from the accuracy [9]. Therefore, selecting the appropriate feature set for a given dataset is paramount to the AIdS performance.

The task of selecting the most salient features is called *feature selection* [53]. In its simplest form, feature selection amounts to trying various combination of features to see how they affect the performance of an AIdS for a given test dataset. The problem is that each additional feature considered combinatorically increases the cost of finding the best feature set. Therefore, feature selection becomes an optimization problem, to which conventional algorithms are poorly suited [3, 11, 28].

Genetic-based and Swarm-based algorithms are well suited to optimization problems [28], like feature selection, where conventional search algorithms perform poorly [11]. Feature selection approaches work to identify a feature mask, which determines which features, from the total set of features, the AIdS uses. Genetic algorithms (GA) [11], which are based on natural selection, start with a population of random feature masks, and "breed" new masks into the population by combining aspects of parents' feature masks. Following the ideas of natural selection, the population keeps those masks that are most effective, and eliminates less effective feature masks. Swarm-based feature selection algorithms [13] are based on the behavior of communities of organisms in nature and use these behaviors to search for optimal masks within the feature space. Both optimization techniques result in a feature mask tuned for optimal, or near optimal, performance for a specific feature set/dataset pairing [61].

There are situations where authors may prefer anonymity [40]. For example, authors may fear repercussions for expressing unpopular views, or may fear that their reputation may bias a conversation. In these situations, authors' interests are at odds with the goals of AId. Therefore, there is a competing body of research that seeks to thwart the effectiveness of author identification, known as *adversarial authorship* [40, 41]. There are two general strategies to adversarial authorship [40]. One strategy seeks to hide an author's identity by obfuscating the author's writing style [40]. The second strategy seeks to impersonate another author's style [40].

If adversarial authorship researchers are to be successful, they must consider the features employed by AIdSs and find ways to confuse AIdSs based on these features. Similarly, AId researchers must understand the techniques of their adversaries and work to identify and circumvent these techniques. As these two groups work to confound each other, an arms race emerges in which each group escalates the race by finding new strategies, tactics and techniques.

A recent concern has emerged within this AId arms-race with respect to feature selection [35]. Although limiting the number of features can reduce computational costs and improve identification accuracy, in some cases, feature selection creates a susceptibility, which adversarial researchers may be able to exploit [35]! Recent research has observed that, with some feature set/dataset combinations, adversarial attacks result in greater efficacy than when feature selection is not used [35]. This observation is significant to both identification researchers as well as adversarial researchers.

The question becomes, how significant is this feature selection vulnerability and what principles govern its exposure? This is the focus of this research. In this work, we quantify the effects of adversarial attacks on AIdSs, with and without feature selection. We want to consider the effects of parameters such as feature set characteristics, the dataset type, and the number of authors, on the susceptibility of feature selection.

1.2 Dissertation Overview

Motivated by the challenges outlined in Section 1.1, this work will proceed as follows. In Chapter 2, we provide a literature review, including author identification, feature selection methods, genetic-based and swarm-based algorithms. Chapter 3, describes the datasets used in this study which include subsets of the CASIS-1000 dataset and the PAN19 dataset. Chapter 4 presents a comparison of genetic and swarm intelligence-based feature selection algorithms for author identification. Chapter 5, describes the good, the bad and the ugly of using genetic-based feature selection for author identification. Chapter 6, expands on the work from the previous chapters by increasing the size of the dataset and including several feature selection algorithms. Chapter 7 presents an analysis of the effects of genetic and swarm intelligence feature selection on adversarial author identification using many authors. Chapter 8, expands the research from Chapter 7 by considering the PAN19 dataset. In Chapter 9, we present conclusions and recommend future directions.

Chapter 2

Overview: Literature Review

In this chapter, we review many areas of related research including Author Identification (AId), author profiling, feature selection methods, Evolutionary Computation (EC), Swarm Intelligence (SI), Evolutionary Feature Selection, and Genetic & Evolutionary Feature Selection (GEFeS).

2.1 Author Identification

AId identifies the authors of writing samples [6]. Author Identification (AId) is a behavioral biometric [12]. Author Identification Systems (AIdS) [10] are machine learning systems that extract features from text and use those features to classify text samples of an author [4]. AIdSs use features [5] that range from simple character n-grams [6], to psychological states of mind [7]. The number features that an AIdS uses can be large [28]. This large number of features leads to computation that is expensive both in time and space. Additionally, not all features are helpful [13]. Effective features for one dataset may not be effective in another [9]. Therefore, identifying salient features improves accuracy of AIdS and minimizes compute costs.

The theoretical foundation of AId is found in the idea that authors comport individual styles, and that those familiar with an author's works can recognize those styles [4]. These ideas reach back to the 1880's where T.C. Mendenhall suggested that authors might produce works with word-length spectrograms that are specific to the author [48]. In Mendenhall's case, he considered text style indicators such as mean word length and histograms of word lengths, which produced what he termed a "characteristic curve".

With the application of machine learning to the task of AId, we refer to the stylistic indicators we extract from text as *features* [5]. AIdSs commonly have two components that deal with features [2]. The first component extracts features from training text samples and indexes metadata about these features with the author's identity of the text sample [2]. The second component extracts features from anonymous test text sample but then attempts to match the features to those in the index so as to predict the author of the test text sample [2, 4].

In [10], Stamatatos reviews advancements in AId during the decade starting around the beginning of the millennium. Stamatatos reviews approaches used to quantify writing styles. The author also classifies, critiques, and evaluates various AId approaches, as well and identifies additional research areas. Stamatatos classifies stylometric features into categories of lexical features, character features, syntactic features, semantic features, and application specific features.

Lexical features are those features concerned with tokens, such as words, numbers and punctuation. These token occurrences may be counted to create vectors of word frequencies. This approach is simple and easy; however, Stamatatos points out that simple lexical approaches disregard information like word order.

An approach even simpler than using lexical features is using character features [4]. This approach views a text sample as a mere sequence of characters or short sequences of characters known as *n-grams* [4]. Like lexical approaches, character approaches calculate frequencies of characters or n-grams. According to [49], this approach is useful for AId. However, one issue with using n-grams is selecting a value for *n*. Smaller values (e.g., 2, 3) are computationally easier because these values produce smaller frequency tables. However, larger values would be more likely to capture sematic aspects. Also, in [10] Stamatatos points out that using n-grams is language independent, but some languages with more average characters per word (e.g., German or Greek) would benefit from larger values of n.

Syntactic features use tools to parse the text and identify parts of speech, and sentence construction, etc. [4]. This requires more computation to extract the features but may be more effective based on the assumption that authors unconsciously use identifying syntactic structures [10, 63]. Examples of syntactic features include parts of speech, dependency features, and parse tree metrics [50].

Semantic features amount to using the meaning of words to identify topics within sample text. In [10], the survey, at that time did not identify great success using sematic features, however since then, some researchers have achieved some success as in [51]. In a comparison of feature sets in [19, 46], the authors found a semantic feature set known as Topic Modeling to be most useful for a blog dataset where topics tended to be highly correlated with authors.

Application specific features include those that are applicable within a limited domain [10, 63]. For example, in an email dataset, structures such as greetings and farewells, or in HTML documents, various tags may be used as features. Also, some natural languages may lend themselves to specific features such as familiar versus formal as in French (i.e., vous vs. tu).

[10] also discusses attribution approaches and identifies two main classes (as well as a hybrid approach). The first approach, termed *profile-based* combines all training samples for an author into a single large sample. The second approach, *instance-based*, keeps the training samples separate for each author. The profile-based approach includes probability-based (e.g., naïve Bayes) methods [64], compression-based methods and common n-gram methods. The instance-based approaches include vector-spaces [65], similarity-based approaches, and meta-learning approaches.

Stamatatos [10] discusses evaluation concerns, which is really a discussion of attributes of an effective dataset for evaluating AIdSs. Stamatatos identifies various parameters, which should be considered when comparing and evaluating AIdSs such as text length, language, genre, and topic.

2.2 Author Profiling

Author profiling is like author identification, but rather than identifying a specific author, author profiling seeks to identify latent demographic features of authors [73, 74, 75, 76]. Identifiable latent traits include the author's age, gender, and native language, geographic location [73, 76], with great interest on identifying age and gender [73]. Identifying these traits may be useful for targeted marketing, personalization, and forensics [73].

The approach used to perform Author Profiling is similar to Author Identification in that researchers extract features and use these features to create a profile [75]. The features used for profiling include words, word classes, parts of speech, and n-grams [75]. [74] classifies Author Profiling features into one of two categories: style-related or content-related. Style-related features include parts of speech, function words and "blog words" (e.g., lol, ur, hyperlinks) [74, 76]. Content-related features include topic-related words (e.g., job, sports, family) [74].

2.3 Feature Selection Methods

Feature selection is the task of identifying and using a subset of the features. As shown in [52], for AId there are two potential benefits from feature selection: first, feature selection reduces the dimensionality of the problem and, therefore, the computational requirements to solve the problem; second, feature selection can improve accuracy. Many methods may be applied to feature selection [53]. [44, 28], consider EC-based approaches, while [13] focuses SI-based approaches.

Feature Selection is an optimization problem that is well suited to genetic search and SI [11, 28]. Researchers use these algorithms to find a subset of features that combines AId accuracy with feature minimization using a scalarized evaluation function. Feature selection approaches may be classified generally into three categories based on how they evaluate the feature subset: wrappers, filters, and embedded approaches [53, 13]. Wrapper approaches directly employ the classification mechanism in evaluating the feature subset. The benefit of using a wrapper approach is its accuracy. Filter approaches evaluate the feature subset using a measure that correlates with classification accuracy, thereby reducing the computational costs of feature selection. Embedded approaches also attempt to reduce the computational costs of feature selection by incorporating feature selection as part of the training process.

Chandrashekar, et al. [53] identify the stability of feature selection algorithms as a concern. Unstable feature selection occurs when adding more training data causes the feature subset to change. When the feature selection algorithm produces a different feature subset as the training data changes, then the algorithm becomes unreliable.

Chandrashekar, et al. [53] also suggest two classifiers for feature selection: Linear Support Vector Machine (LSVM) [66] and Radial Basis Function Network (RBF) [67]. The LSVM classifies by constructing a hyper-plane within the feature space, whereas the RBF is a feed-forward neural network.

Chandrashekar, et al. [53] conclude by summarizing the important considerations of feature selection as simplicity, stability, the feature subset size relative to the feature set size, the accuracy, and the computational requirements.

2.4 Evolutionary Computation

EC [54, 55, 56] is an optimization approach generally based on principles of natural selection and may employ three approaches: genetic algorithms (GAs) [11], evolutionary programming (EP) [68], and evolutionary strategies [54, 69]. EC algorithms must generate a population of individuals, evaluate the fitness of individuals, breed new individuals, and eliminate individuals from the population. The variations of how each of these algorithmic phases are performed constitute the distinction between the various EC algorithms.

According to Back, et al. [54], there are three design concerns when developing an EC algorithm including the representation of the individual, the reproduction/mutation of new individuals, and the selection mechanism. Individuals often represent solutions to an optimization problem and take the form of arrays of binary or real values, with a preference for binary representations due to its computational simplicity. The introduction of new individuals into the population is a result of reproduction and mutation. Recombination takes the form of crossover with variations in how many parents are used, how the parents are selected, etc. The selection mechanism determines which individuals remain in the population and which are removed. Various selection strategies include proportional selection, rank-based selection, and tournament selection.

2.5 Swarm Intelligence

SI algorithms [1, 2, 3, 17] are optimization algorithms, often inspired by behavior in nature, where agents interact locally in such a way as to cause a global emergent behavior. Examples of these algorithms include particle swarm optimization (PSO), bee-inspired algorithms, bacterial and ant foraging-based algorithms, and firefly and fish swarms [1, 2, 3, 17, 55].

Kennedy et al. [31], and Mavrovouniotis et al. [55] describe the PSO algorithm, where agents, represented as a location and velocity, are randomly disbursed across a search space. The location of each agent represents a potential solution to the optimization problem at hand. Agents move through the space based on the velocity, and the velocity is updated based a combination of the best position seen by all agents and the best position seen by the individual agent.

The artificial bee colony optimization (ABCO) algorithm is an optimization algorithm based on foraging activity of a hive of bees [33, 55]. In this algorithm, bees forage in random locations for nectar. The locations represent possible solutions to the optimization problem. The amount of nectar available to an individual bee corresponds to the result of the fitness function when using the foraging location as parameters to the function. During each iteration of the algorithm, bees report their locations and fitness to the (global) hive to create a probability density function for finding nectar. Onlooker bees use this density function to select areas to investigate and similarly report their results. When further investigation fails to yield more optimal solutions, bees abandon these areas and strike out in new random locations.

Ant system optimization (ASO) is different from the other main swarm intelligence algorithms in that the location of the agents (ants) does not represent an optimization solution. Instead, the path the agent pursues represents the solution [34, 55]. The ASO algorithm is a reinforcement learning algorithm, which is well suited to optimization problems like the traveling salesman problem in that ants deposit pheromones along the graph during their journeys, and the edges with the most pheromones become part of the optimal solution.

There are many other less popular swarm intelligence algorithms [55], including bacterial foraging optimization, artificial fish swarm optimization, and the firefly algorithm.

2.6 Evolutionary Computing Based Feature Selection

Feature selection is often a high dimensionality problem to which evolutionary computing is well suited [28, 44, 54, 56, 71]. The survey in [56], identifies four main characteristics for studying EC feature selection including representation, evaluation, selection approaches, and EC approaches.

In [56], de la Iglesia identifies seven representations which include binary, variable length, fixed length, decimal encoding, feature set with model parameters, rough set, and complex. There are three evaluation categories including single criteria, multicriteria, and constrained optimization. Within the area of feature selection evaluation approaches, de la Iglesia suggests that wrapper approaches are the most widely used, but de la Iglesia also addresses filter, embedded and hybrid approaches. In the survey, de la Iglesia reviews several evolutionary computation paradigms used for feature selection including GAs, ASO, PSO, GP, estimation of distribution algorithms, memetic algorithms, and multi-objective optimization algorithms. Ding et al. [71] focus on PSO and GA to optimize feature selection for Linear Support Vector Machines (LSVM).

According to de la Iglesia [56], EC feature selection is effective at reducing the number of features while improving accuracy. However, these approaches require more work when data is missing or noisy. de la Iglesia further suggests that PSO and ASO perform feature selection better than GAs.

2.7 Swarm Intelligence Based Feature Selection

Because SI is a useful optimization approach, where conventional search algorithms fail, SI has become a mainstay in feature selection [13, 29, 31, 33, 34, 35, 45, 55, 57, 58, 70]. Nguyen et al. [13] review, in detail, the three most popular SI algorithms for feature selection (PSO, ABCO, and ASO), focusing on feature representations and search mechanisms. The algorithm descriptions are consistent with those presented in section 2.3.1. In [70] Brezočnik et al. survey 64 different SI algorithms and organize them into a taxonomy of eight categories based on the algorithms' biological inspiration (insect, bacteria, bird, mammal, fish, frog, group hunting, other). Nguyen et al. [13] point out that continuous feature representations are suitable for SI feature selection, but binary representations require some algorithm extension. The common SI fitness function combines classification accuracy with minimizing the number of features, which improves overall performance of the SI system [13]. Other attempts to improve performance are algorithm specific. The PSO algorithm uses a global best position and a local best position. Therefore, attempts to improve PSO work to enhance these two positions. Since the focus of ASO is on pheromones, optimization approaches tend to consider ways to balance exploration and exploitation. Nguyen et al. suggest that ABCO is the least mature of the three main algorithms, therefore, common optimization approaches amount to combining this algorithm with others.

Brezočnik et al. [70] suggest a framework to assist researchers in developing SI feature selection approaches to specific problem types. They also identify two open concerns about SI feature selection research, specifically that researchers need to fully describe their algorithms so that results are reproducible, and the number of evaluations needs to be the terminating criteria for a common comparison approach.

Nguyen et al. [13] identifies five challenges of using SI in feature selection including representation, computational cost, feature selection bias, multi-objective feature selection, and embedded SI feature selection. Representation is challenging because most of the SI algorithm study has used continuous representations, but binary representations are more natural for feature selection and in some instances perform better. Further, features are not always independent, so a representation that articulates the dependencies would be useful. SI algorithms bare a high computational cost because they require many iterations. Feature selection bias occurs when the resulting feature set works well for the training data but does not work generally. Multi-objective feature selection becomes a challenge because, often, SI approaches equally weight accuracy and feature reduction. Arguably, accuracy is more important. Therefore, a multi-objective approach would be more effective. While embedded SI feature selection algorithms balance effectiveness and efficiency, most SI feature selection approaches are either wrapper approaches or filter approaches.

2.8 Genetic & Evolutionary Based Feature Selection

Studies using GEFeS include [14, 18, 19, 30, 46, 47]. GEFeS uses a genetic steady-state algorithm that evolves binary feature masks. On each iteration of the evolution, the algorithm uses binary tournament selection to identify two parents, then uses crossover with mutation to generate an offspring. This algorithm uses wrapper-based evaluation to assess the initial population as well as each offspring. The fitness function combines accuracy as determined by the classification function, with a reduction of features. Since the algorithm is steady state, on each iteration, after creating the offspring, the algorithm removes the worst performer from the population.

In [14, 18, 19], Gaston et al. apply multimodal machine learning [38, 39] by combining different feature sets and using Genetic & Evolutionary Feature Selection (GEFeS) to identify salient features. These multi-modal feature sets include Linguistic Inquiry and Word Count (LIWC) [20], sentiment analysis (OpinionFinder) [21, 22, 23, 24, 25], and topic modeling (Mallet) [26, 27]. In [19], Gaston et al. consider various weightings to balance AId accuracy while minimizing the number of features. Gaston et al. also improve accuracy by sequencing the AIdS preprocessing pipeline so as to perform Term Frequency-Inverse Document Frequency (tf-idf), standardization, and normalization on the various multimodal features prior to combining them. The results of Chapters 4, 5, 6, and 8, show that GEFeS generally outperforms SI feature selection algorithms in these studies.

Chapter 3

Datasets

3.1 CASIS-25

The first dataset used in this work is the CASIS-25 dataset [36, 37], which is a subset of the CASIS-1000 dataset [36, 37]. The CASIS-1000 dataset is a set of 4,000 blog post writing samples created by 1,000 authors with four samples per author. The CASIS-25 dataset consists of 25 authors' samples with four samples per author, or a total of 100 writing samples. This work forms a set of 25 adversarial texts by selecting the fourth of each author's samples and modifying the text using AuthorCAAT-V and AIM-IT [36, 43] to disguise the samples. These 25 adversarial samples are not part of the CASIS-25 dataset proper but are derivatives. When used, these adversarial texts can be used as replacements for the fourth original writing sample of each author.

AuthorCAAT-V [36, 43] is a human-assisted adversarial authoring tool that helps disguise writing samples using paraphrasing and iterative translation to obfuscate characteristics of the source author or imitate the characteristics of another author. The AuthorCAAT-V user performs these actions by using the tool to paraphrase or iteratively translate text. If the resulting text is sufficiently disguised, the user can accept the text. Otherwise, the user can use the derived text as the source for another iteration.

3.2 CASIS-50

Like the CASIS-25 dataset, the CASIS-50 dataset is a subset of the CASIS-1000 data set. However, the CASIS-50 dataset has 50 authors, with four writing samples each, for a total of 200 samples. These 50 authors include the same 25 authors in the CASIS-25 dataset as well as an additional 25 authors. The adversarial texts used for this dataset are the same 25 adversarial texts created for the CASIS-25 dataset.

3.3 CASIS-100

The CASIS-100 dataset is a super set of the CASIS-50 dataset with an additional 50 authors, again with four writing samples each. The CASIS-100 adversarial texts are the same 25 adversarial texts derived from the CASIS-25 dataset.

3.4 PAN19-25

The PAN19-25 dataset consists of a subset of 100 writing samples used in the third author profiling task at PAN 2015 [57]. This dataset consists of Twitter feed writing samples that have been classified by authors' attributes – human versus bot, and in the case of human, female versus male. We selected 25 authors (eight female, eight male and nine bots) with four samples each for a total of 100 samples.

Like the CASIS-25 adversarial samples, we selected the fourth sample from each author and derived an adversarial sample using AuthorCAAT-V.

Chapter 4

A Comparison of Genetic & Swarm Intelligence-Based Feature Selection Algorithms for Author Identification

4.1 Motivation

Researchers are moving beyond stylometric features to improve author identification systems. They are exploring non-traditional and hybrid feature sets that include areas like sentiment analysis [21, 22, 23, 24, 25] and topic models [26, 27]. This type of feature set exploration leads to the concern of determining which features are best suited for which systems and datasets. In this chapter, we compare Genetic Search [11, 30] and a number of Swarm Intelligence (SI) [15, 17, 29, 31] methods for feature selection, spcifically, Particle Swarm Optimization (PSO) [1, 2, 3, 31, 55], Artificial Bee Colony Optimization (ABCO) [33, 55], Ant Systems Optimization (ASO) [34, 55] and Glowworm Swarm Optimization (GSO) [35]. The results of the work presented in this chapter we published in [46].

4.2 The Feature Selection Algorithm Comparison Experiment

In this experiment, we compare six feature selection algorithms (GEFeS, PSO, ABCO, ASO, GSO and random sampling (RAND)) for author identification on a subset of the CASIS-25 dataset [36, 37]. Each of these feature selection algorithms generates binary feature masks where a 1 indicates applying the feature and 0 indicates ignoring the feature. The random sampling feature selection algorithm stochastically generates, and then evaluates, feature masks and chooses the best. We constructed the feature set as a concatenation of 93 LIWC [20] features, 176 OpinionFinder [21] features and 45 [26, 27] Mallet features for a total of 314 features. Mallet is a topic modeling program that does not predetermine the topics, but rather the topics emerge based on clustering. The number of features must be specified. We selected 45 topics based on the work in [19].

We evaluate each of the algorithms using a wrapper approach, which is to say we employ the AIdSs to evaluate the subset of features. We use three different AIdSs, each using a different classification algorithm, including Linear Support Vector Machine (LSVM), a Radial Basis Function (RBF) and a Multi-Level Perceptron (MLP). We compare all combinations of feature selection algorithms and AIdS classification algorithms. We gather 30 samples of each feature selection/classification combination. Also, we limit each sampled combination to 15,000 evaluations.

This experiment preprocesses the dataset features as outlined in [19], which is to say we perform tf/idf, normalization and standardization on each of the separate feature sets (produced by LIWC, OpinioFinder and Mallet) before concatenating the features into the final feature set. The experiment uses four-fold cross validation.

All feature selection algorithms share the same scalarized fitness function. The fitness function combines AId accuracy with feature minimization. Formally, the fitness function is:

$$Fitness(mask_i) = \alpha_i - \omega\beta_i \tag{4.1}$$

where $mask_i$ is the i^{th} agent in the population, α_i is the AId accuracy achieved using the mask, ω is the feature reduction weight [0, 1] that biases the impact of the feature reduction, and β_i is the ratio of features used (i.e., the number of ones in the mask divided by the mask size). We used seven values for ω (0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0). Notice that when $\omega = 0.0$, the fitness relies completely on accuracy, whereas $\omega = 1.0$ biases the fitness towards feature reduction.

The fitness function, which requires evaluating the accuracy of each member of the population with respect to the AIdS, becomes the dominant performance component of feature selection algorithms. Therefore, for this research to conduct an unbiased comparison of the feature selection algorithms, each algorithm uses the same fixed number (15,000) of evaluations as a halting criterion. While we train the algorithms according to the fitness function presented in Equation 4.1, we evaluate the algorithms based on accuracy alone.

Some of the feature selection algorithms have hyper-parameters that affect the algorithms' behavior. For example, PSO has social factor and cognitive factor hyper-parameters. To determine the optimum values for these hyper-parameters, we compared several values and found that a value of 1.3 for the social factor and 2.8 for the cognitive factor performed best. These values are consistent with [32]. We also used a Clerc constriction coefficient [16] K, where

$$K= {}^{2}\ (2)\ |2-\varphi-\surd\varphi^{2}-4\varphi|$$

$$\varphi = \varphi_{1} + \varphi_{2}\ ,$$

$$\varphi_{1} = cogintive\ factor \text{ and}$$

$$\varphi_{2} = social\ factor.$$

We determined the optimal swarm size for this experiment to be 100. Therefore, the PSO algorithm performs 149 iterations for a total of 15,000 evaluations (i.e., one evaluation of each initial particle position as well as an evaluation after each particle movement).

ASO has a population size hyper-parameter. We experimented with several sizes and found that optimum PSO population sizes are larger than population sizes for PSO. Specifically, we use a population size of 500. This means that we only perform 30 iterations of the ASO algorithm to arrive at a total of 15,000 evaluations.

ABCO has two hyper-parameters: exploration rate and an ineffective visit count. The exploration rate is the number of bits in the feature mask, which corresponds to the bee location, that change to allow exploration within a vicinity. Ineffective visit count is the number of times a bee visits a vicinity without encountering improvement. When a bee reaches the maximum ineffective count, the bee stochastically selects a different location to forage. This experiment uses values of 0.1 for the exploration rate, and 10 for the ineffective visit count. Further, the initial bee swarm size is 100 consisting of 50 worker bees and 50 onlooker bees. Since each bee performs a single evaluation on each iteration, we limit the number of iterations to 150 for a total of 15,000 evaluations.

In GSO, there are two hyper-parameters that control the luciferin levels of each glowworm: decay rate and fitness rate. The decay rate is the amount the luciferin dissipates on each iteration, and the fitness rate is the amount that the current fitness affects the luciferin levels. After some experimentation, we selected a decay rate of 0.15 and a fitness rate of 0.7. In GSO, agents update their direction based on stochastically selecting from a probability density function derived based on luciferin levels and distance. Once agents select a new direction, they move in that direction based on a step-size hyper-parameter. Exploration allowed us to select a useful step-size of 0.1. We also used a swam size of 100 which allows 150 iterations of the algorithm for a total of 15,000 evaluations.

For GEFeS algorithm, we used the work in [19] to select the hyper-parameters. Specifically, we used a population size of 100. Initial feature masks are stochastically generated with roughly 50% of features selected. We breed offspring from two parents with a mutation rate of 2%, where each parent is selected using binary tournament selection. Each iteration of the algorithm generates one offspring and removes the single least-fit individual from the population, which means we can perform 14,900 iterations (since the original population is initially evaluated) for a total of 15,000 evaluations.

The random selection algorithm serves as a reference baseline to help understand the value of the more intelligent feature selection algorithms. This algorithm stochastically generates and evaluates 15,000 feature masks to identify the most effective feature mask.

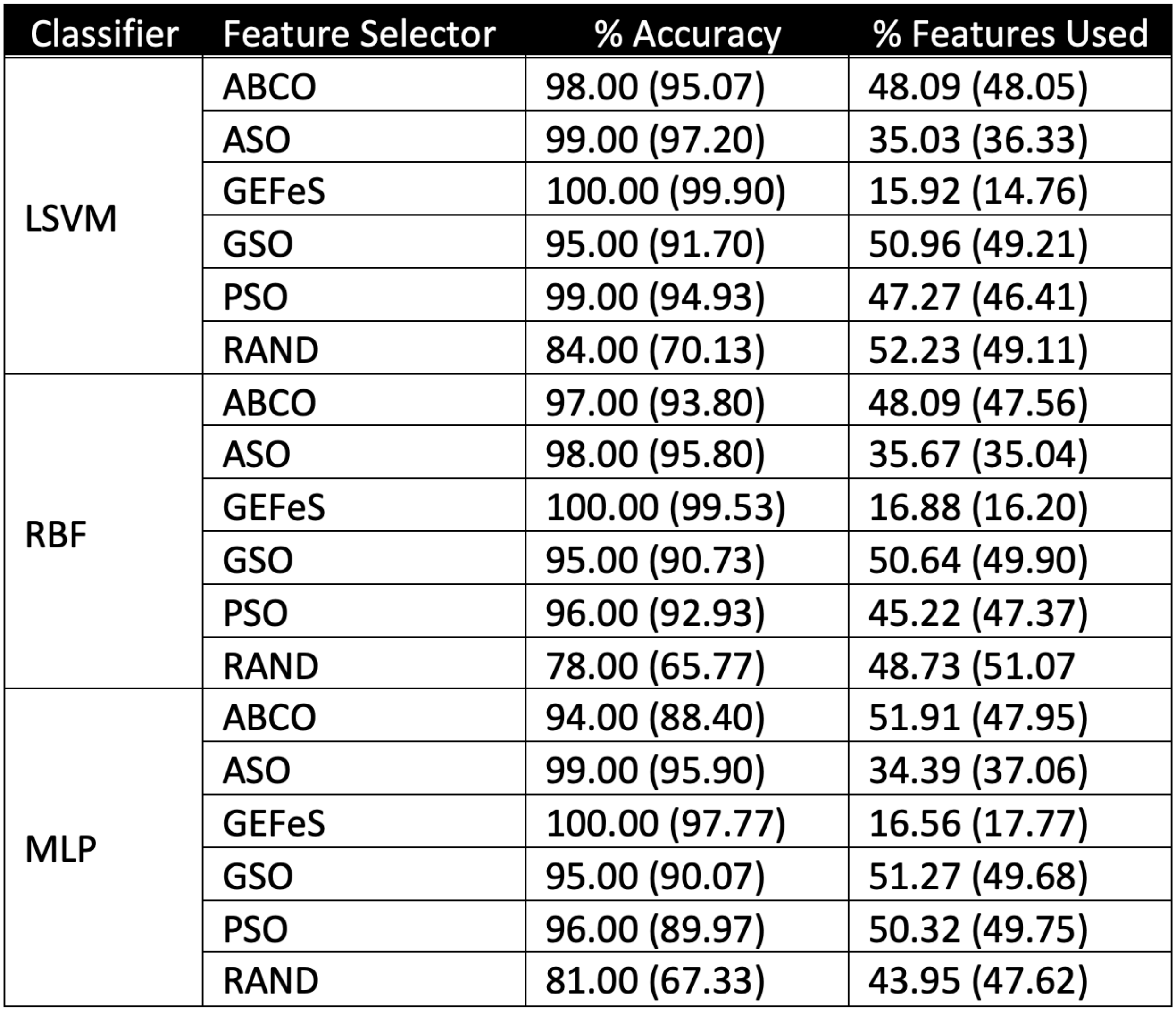
4.3 Feature Selection Algorithm Comparison Analysis

In this section, we summarize the experiment's results. See Appendix A for a review of the detailed results. Specifically, in this section we will analyze:

- The effects of varying the AIdS classification function,
- The effects of ω on accuracy,
- The effects of feature selection algorithms on both accuracy and feature reduction,
- The effects of feature sets on feature consistency.

Table 4.1 compares accuracy and feature reduction for each of the three AIdS classification algorithms (LSVM, RBF, MLP) and all six of the feature selection algorithms (ABCO, ASO, GEFeS, GSO, PSO, RAND) for ω = 0.5 as a representative value. The first column identifies the AIdS classification algorithm. The second column identifies the feature selection algorithm. The third column is the accuracy (percent of correctly identified authors). The fourth column is the feature reduction expressed as percentage of features used. The third and fourth columns contain two values; the first value is the best value of the observed 30 runs (i.e., highest accuracy and lowest features used), and the second value in the parentheses is the mean value of the 30 runs.

Table 4.1. Feature Selection Algorithm Comparison Experiment Results Summary with $\omega = 0.5$.



Classifier	Feature Selector	% Accuracy	% Features Used
LSVM	ABCO	98.00 (95.07)	48.09 (48.05)
	ASO	99.00 (97.20)	35.03 (36.33)
	GEFeS	100.00 (99.90)	15.92 (14.76)
	GSO	95.00 (91.70)	50.96 (49.21)
	PSO	99.00 (94.93)	47.27 (46.41)
	RAND	84.00 (70.13)	52.23 (49.11)
RBF	ABCO	97.00 (93.80)	48.09 (47.56)
	ASO	98.00 (95.80)	35.67 (35.04)
	GEFeS	100.00 (99.53)	16.88 (16.20)
	GSO	95.00 (90.73)	50.64 (49.90)
	PSO	96.00 (92.93)	45.22 (47.37)
	RAND	78.00 (65.77)	48.73 (51.07
MLP	ABCO	94.00 (88.40)	51.91 (47.95)
	ASO	99.00 (95.90)	34.39 (37.06)
	GEFeS	100.00 (97.77)	16.56 (17.77)
	GSO	95.00 (90.07)	51.27 (49.68)
	PSO	96.00 (89.97)	50.32 (49.75)
	RAND	81.00 (67.33)	43.95 (47.62)

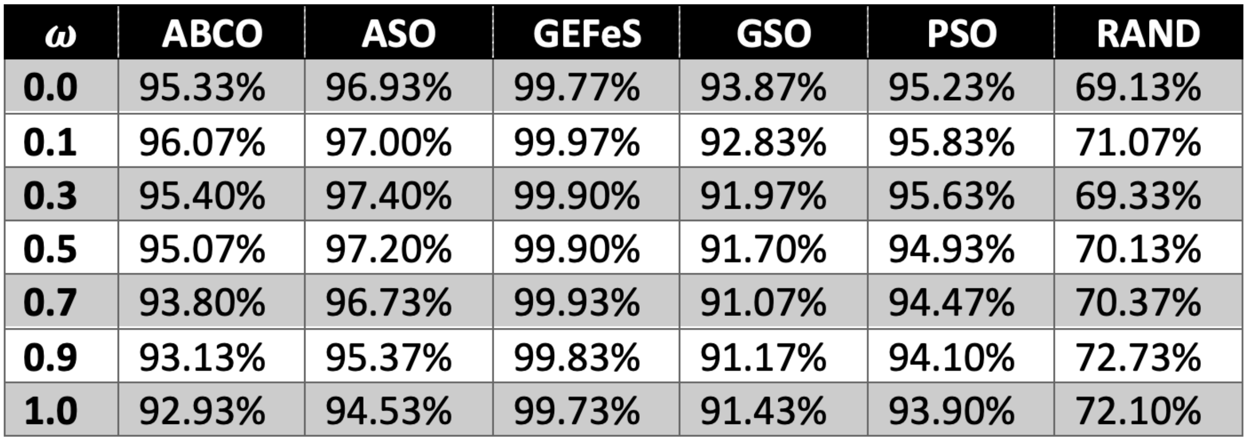
There are several points of interest in Table 4.1. First, notice that accuracy results are generally above 90% for all feature selection algorithms across all classification algorithms, with the notable exception of the RAND algorithm. This exception is evidence that the genetic-based and swarm intelligence-based algorithms are indeed intelligent.

Further, if we compare accuracies for any specific feature selection algorithm across classification algorithms, we see that the values remain relatively consistent. This is evidence that the classification algorithm is not a significant factor. Or, in other words, all three classification algorithms perform comparably. As a side note to this observation, the MLP algorithm requires much longer training computation times, which do not appear to yield superior results.

A third point of interest may be observed by considering the *Features Used* column. Notice that even though accuracies are similar, there is a significant difference between the percentage of features used based on feature selection algorithm. GEFeS clearly dominates with much lower percentage of features used. ASO is moderately lower than the other four feature selection algorithms, but the percentage of features used for ABCO, GSO, PSO and RAND all hover near 50%.

Let us turn our attention to the effects of ω on the various feature selection algorithms. Table 4.2 summarizes the mean accuracies (across the 30 samples) of all six algorithms (ABCO, ASO, GEFeS, GSO, PSO, RAND) for all seven values of ω using the LSVM classification mechanism. We only show LSVM in this table since, as previously discussed, the classification algorithm does not appear to have significant impact on accuracies.

Table 4.2. Feature Selection Algorithm Comparison Mean Accuracies for Several Values of ω using LSVM.



ω	ABCO	ASO	GEFeS	GSO	PSO	RAND
0.0	95.33%	96.93%	99.77%	93.87%	95.23%	69.13%
0.1	96.07%	97.00%	99.97%	92.83%	95.83%	71.07%
0.3	95.40%	97.40%	99.90%	91.97%	95.63%	69.33%
0.5	95.07%	97.20%	99.90%	91.70%	94.93%	70.13%
0.7	93.80%	96.73%	99.93%	91.07%	94.47%	70.37%
0.9	93.13%	95.37%	99.83%	91.17%	94.10%	72.73%
1.0	92.93%	94.53%	99.73%	91.43%	93.90%	72.10%

Again, as we look at individual rows across columns of Table 4.2, it becomes apparent that GEFeS has the best accuracies, which is consistent with our observations of Table 4.1. ASO also performs better than the four remaining algorithms, and RAND performs worst, which reinforces our belief that genetic-based and swarm-based algorithms do perform intelligently.

Now, consider individual columns of Table 4.2. Recall that we evaluate using the fitness function described by Equation 4.1, but we report raw accuracies. Therefore generally, we would expect that as we increase ω (which applies increased pressure to reduce features), we might see reduced accuracies. Indeed, this seems to be the case for four of the six feature selection algorithms (ABCO, ASO, GSO & PSO). But GEFeS and RAND do not appear to be subject to this influence. In fact, there is a slight improvement in RAND accuracies as ω increases. This is evidence that feature selection not only reduces computation, but also can improve accuracy. On the other hand, GEFeS accuracy is independent of values of ω. However, we do see a significant reduction in features as a result of increasing ω (see tables in Appendix A for supporting data).

Finally, we address feature consistency, or in other words, how frequently a feature is selected. We performed an ANOVA ranking analysis of the feature selection algorithms for each of the AIdS and for all the feature weight values. We did this by first performing an ANOVA ranking for each of the feature selection algorithms to find the best equivalence class from each algorithm. Then, we combined these top equivalence classes and performed a combined ANOVA ranking analysis. *This explains why not all feature reduction weighting values appear for all algorithms in the final ANOVA rankings.*

Table 4.3 shows results for the final rankings of this analysis. The first column, labeled "AIdS", indicates the three AIdS configurations (i.e., LSVM, RBF, and MLP) for the corresponding rows. The second column, labeled "Class List Rank", identifies the ranking of the equivalence class for that AIdS. The third column, labeled “Equivalence Class”, enumerates the feature selection algorithm(s) and feature reduction weights belonging to the equivalence class. Note that a hyphen between feature reduction weight values indicates a range of values, whereas commas indicate individual values.

In Table 4.3, the top equivalence class for LSVM includes the range of all seven values of ω, whereas for RBF the top equivalence class excludes $\omega = 0.3$ and $\omega = 0.7$, and MLP excludes $\omega = 1.0$. *These excluded values for* ω *did not make the first cut of the ANOVA analysis, so they do not appear in the final ANOVA analysis results.*

Table 4.3. Equivalency Class Lists by AIdS.

AIdS	**Class List Rank**	**Equivalence Class**
LSVM	1	GEFeS 0.0 – 1.0
	2	ASO 0.5, 0.9, 1.0
	3	ABCO 0.0,0.1, PSO 0.0 - 0.5
	4	GSO 0.1, 0.3
	5	RAND 0.0 - 1.0
RBF	1	GEFeS 0.0, 0.1, 0.5, 0.9. 1.0
	2	ASO 0.3 - 1.0
	3	PSO 0.1 - 0.5
	4	GSO 0.0 - 0.1
	5	ABCO 0.0 - 0.5
	6	RAND 0.0 - 1.0
MLP	1	GEFeS 0.1 - 0.9
	2	ASO 0.5 - 1.0
	3	PSO 0.0 - 0.5, GSO 0.0 - 0.5, ABCO 0.0 - 0.3
	4	ABCO 0.5
	5	ABCO 0.7
	6	RAND 0.0 - 1.0

For each AIdS configuration and its best performer of the corresponding top equivalence class, we considered which features consistently appeared in the feature masks. We calculated the consistency ratio by, for each feature, summing the number of times the feature was used and dividing it by the total number of observed samples (i.e., 30).

Figures 4.1-4.3 are bar charts presenting feature consistency of the best algorithms (selected by choosing the best algorithm from each of the top equivalence classes for each AIdS), where the Y axis has the feature number, and the X axis indicates the value of the consistency ratio. The first 93 features (i.e., 1-93 as shown in blue) correspond to LIWC features. The next 176 features (i.e., 94-269 as shown in red) correspond to sentiment analysis features, and the final 45 features (i.e., 270-314 as shown in gold) are topic modeling features. Figure 4.1 presents the feature consistency for LSVM with GEFeS using $\omega = 0.1$. Figure 4.2 presents feature consistency of RBF also for GEFeS where $\omega = 0.1$ and Figure 4.3 presents feature consistency of MLP for GEFeS where $\omega = 0.3$.

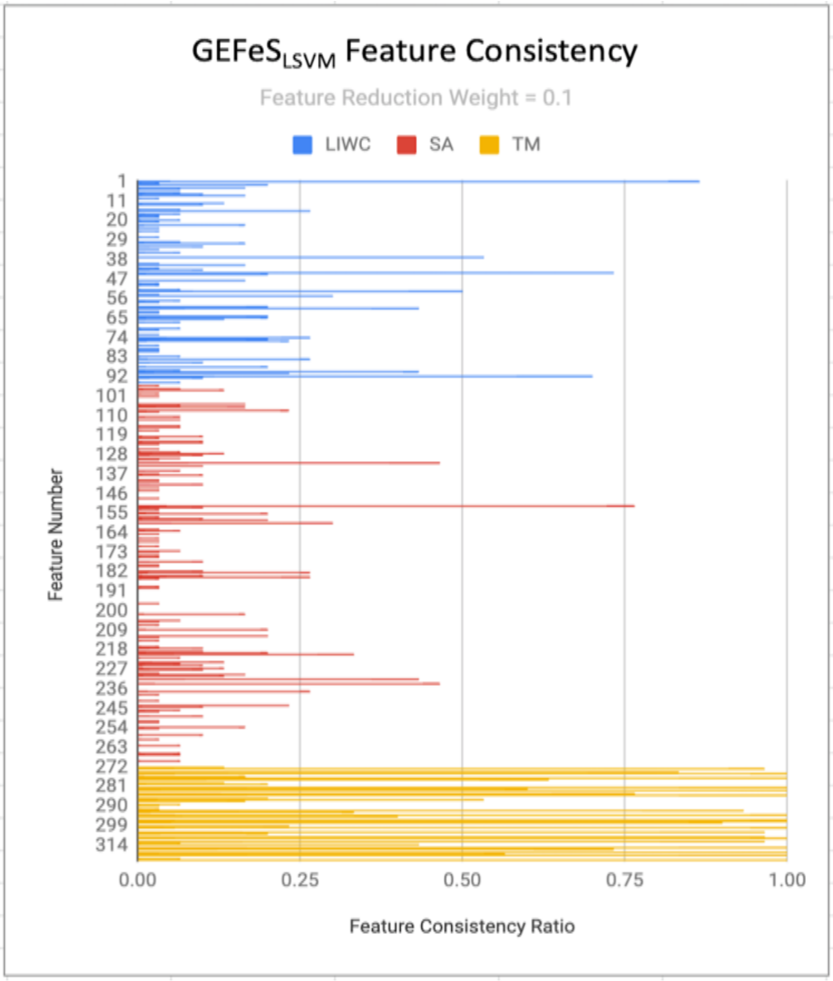


Figure 4.1. GEFeS$_{LSVM}$ Feature Consistency, Feature Reduction Weight = 0.1.

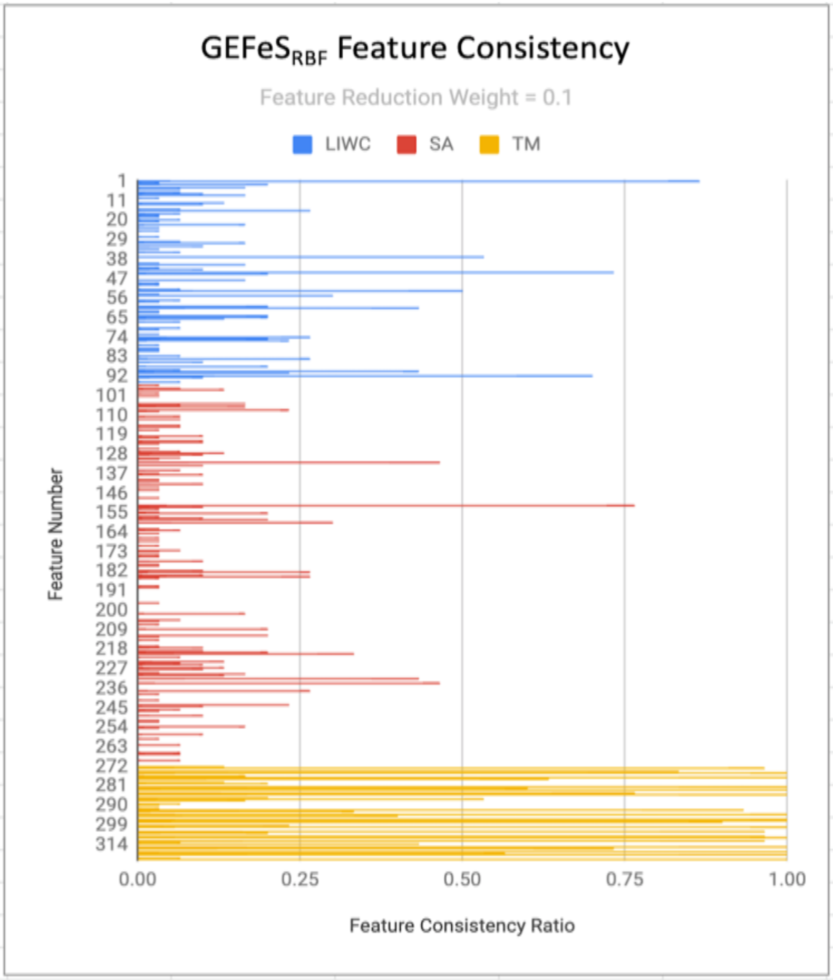


Figure 4.2. GEFeS$_{RBF}$ Feature Consistency, Feature Reduction Weight = 0.1.

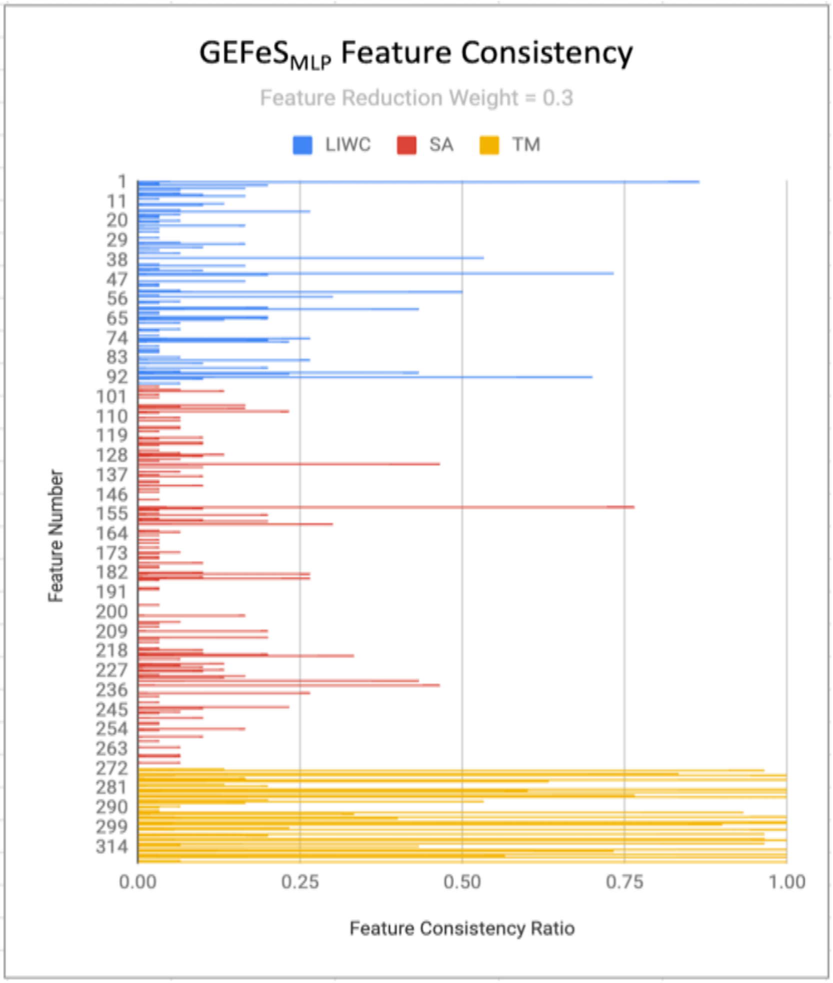


Figure 4.3. GEFeS$_{MLP}$ Feature Consistency, Feature Reduction Weight = 0.3.

One might incorrectly anticipate that optimizing purely for accuracy (i.e., $\omega = 0.0$) would yield the best observed accuracy across all feature weights, but surprisingly, this is seldom the case. Instead, we see a general trend where best accuracy is achieved when ω is between 0.1 to 0.9. This attests to the benefit of feature reduction beyond computational advantages. On the other hand, $\omega = 1.0$, which increases the bias towards feature reduction, intuitively diminishes accuracy.

Also, it is interesting to note that GEFeS was very successful at significantly reducing the number of features, while still maintaining high accuracy across all AIdS configurations. ASO was moderately successful at reducing the number of features, whereas with the other four algorithms, one can observe there is some influence of the feature weight on the number of features, but not nearly to the extent of GEFeS or ASO.

As indicated in Table 4.3, although we discovered one fewer equivalence class for the LSVM AIdS configuration, we can still recognize a general trend. We see that GEFeS is the dominant feature selection algorithm. ASO is consistently in the second equivalence classes, and PSO appears in each of the third equivalence classes. ABCO and GSO are in the lower equivalence classes and, as expected, RAND is the worst performer.

These results give us a foundation for understanding the effects of genetic-based and Swarm Intelligence-based feature selection with respect to several diverse feature sets, values of ω, and classification algorithms. We see from this chapter that feature selection is an effective way of improving AIdSs. We see that different feature sets perform differently, so in future chapters we can be more judicious about selecting feature sets. Also, having determined that the classification algorithm is not an interesting factor, the work in future chapters will use only LSVM.

Some attributes not explored in this chapter, but which are worthy of consideration are varying the dataset, including the number of authors, the number of samples and the sample length. This chapter focuses on discrete feature selection. It will be valuable to consider these same combinations of algorithms using feature weighting [8].

Chapter 5

The Good, the Bad and the Ugly of Using Genetic-Based Feature Selection for Author Identification

5.1 Motivation

Researchers continue to investigate approaches to refine author identification. As we saw in the previous chapter, one such approach uses genetic search to reduce the number of features, to improve accuracy, and to reduce computational costs. However, it is necessary to consider the exposures that result from these otherwise helpful improvements. In this chapter, we compare the effects of author identification adversarial attacks when using and not using genetic-based feature selection. We compare these results across four feature sets including Linguistic Inquiry & Word Count, Topic Modeling, Stylometry as well as a multimodal hybrid of all three.

5.2 The Genetic-Based Feature Selection for Author Identification Experiment

The intent of this chapter is to investigate the interaction between author identification, genetic-based feature selection and various feature sets including LIWC, Topic Modeling, Stylometry, and a hybrid of the three. We preprocess the resulting features in the same manner as outlined in Chapter 4, except we substitute Stylometry for Sentiment Analysis. As in Chapter 4, this experiment uses a subset of the CASIS-1000 dataset [36, 37]. Recall that the CASIS-1000 dataset consists of blog posts from 1,000 authors. Each author has four samples for a total of 4,000 samples. In this experiment, we use a subset of 25 authors, with four samples per author, for a total of 100 samples. We refer to this dataset as the *CASIS-25* dataset [36, 37, 43]. In addition to the original dataset, we created 25 more adversarial samples, one per author. Each of these samples is a derivation of the fourth authors' samples. We use AuthorCAAT-V and AIM-IT [43] to create the adversarial versions of the samples. All other processing, including calculating fitness and reporting accuracy, follows the procedure outlined in Chapter 4.

We assess the AIdS accuracy across four feature sets with and without feature selection as well as with and without adversarial texts. The AIdS uses a Linear Support Vector Machine (LSVM) [14, 18, 19, 46, 47] classifier algorithm.

We perform two sets of similar experiments that differ only in the way we test for accuracy. Since the dataset consists of four writing samples per author, in both sets of experiments we train the AIdS on three of the samples. Then, for the first set of experiments, we test the accuracy on the fourth samples. We denote this approach with the notation: 75+(org = 25, adv = 25). This notation indicates that we train on 75 writing samples, test on 25 of the original writing samples, and then test again on the corresponding 25 adversarial writing samples of the 25 authors. In the second set of experiments, we test using all four writing samples. Then for comparison's sake, we swap the fourth samples for an adversarial version of the sample and test accuracy again as previously described. We denote this as 75+(org = 100, adv = 100). In this case, the notation indicates that we train on 75 of the writing samples, test on all 100 original writing samples, and then perform a second test on the 75 original writing samples combined with the 25 adversarial writing samples. Both sets of experiments calculate accuracy across all four feature sets (i.e., LIWC, Topic Modeling, Stylometry and a Hybrid of all three). All experiments also vary the feature reduction weight (ω), as in Chapter 4, using values of 0.0, 0.1, 0.3, 0.5, 0.7, 0.9 and 1.0. In addition, we test all feature sets with no feature selection as a baseline.

Following the same procedure outlined in Chapter 4, we run GEFeS 30 times and allow each run a total of 15,000 feature mask evaluations. For the baseline results, since no feature selection is involved, we only need to run the tests once.

To review the GEFeS configuration as described in Chapter 4, we configure the GEFeS hyper-parameters as described in [19]. We start with an initial population of 100 randomly generated individuals (i.e., feature masks) consisting of approximately 50% of the features selected. We then run 14,900 generations, where we create a single child for each generation. We create the child by selecting two parents each using binary tournament selection, which is to say that for each parent, we select two individuals form the population at random and use the best as the parent. Given the two parents, we use crossover with a 2% mutation rate to generate the child. Since this is a steady state GA, we eliminate the least fit individual from the population. Note that the algorithm must initially evaluate all members of the population as well as each child. Since we have 100 individuals in the population and 14,900 children, the algorithm requires a total of 15,000 evaluations.

5.3 Genetic-Based Feature Selection for Author Identification Results Analysis

This section reviews the results of the experiment to identify general trends and points of interest. See Appendix B for a detailed treatment of the results.

We introduce a new metric, *Use?*, in this chapter, which is a value indicating a feature set's relative susceptibility to adversarial attacks. We define this metric formally in Table 5.1.

Table 5.1. A Formal Definition of the *Use?* Value.

$$\boldsymbol{Use}? = \frac{\boldsymbol{\alpha_g^u - \alpha_b^u}}{\boldsymbol{\alpha_b^u}} + \frac{\boldsymbol{\alpha_g^a - \alpha_b^a}}{\boldsymbol{\alpha_b^a}}$$

Where:

$\boldsymbol{\alpha_g^u}$ is the GEFeS accuracy of original samples

$\boldsymbol{\alpha_b^u}$ is the baseline accuracy of original samples

$\boldsymbol{\alpha_g^a}$ is the GEFeS accuracy of adversarial samples

$\boldsymbol{\alpha_b^a}$ is the baseline accuracy of adversarial samples

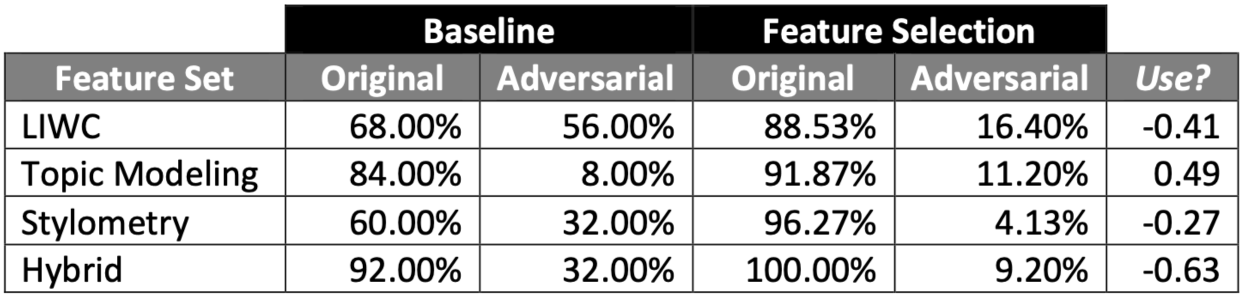
$\boldsymbol{Use}? \geq \boldsymbol{0}$: use feature selection

$\boldsymbol{Use}? < \boldsymbol{0}$: do not use feature selection

The value of *Use?* captures the relationship of the effectiveness of the adversarial attack without feature selection, compared to the same attack with feature selection. Positive *Use?* values favor the use of feature selection and negative values favor the avoidance of feature selection.

Table 5.2 is a representative summary of the results from this experiment. While there are many data points resulting from this experiment (see Appendix B), this table serves as a basis for the general intuition we glean from this experiment. In the table, the first column is a feature set label for each of the rows. The second and third columns show accuracies measured without feature selection (i.e., Baseline). The third and fourth columns show accuracies when using feature selection. Columns two and four, labeled *Original*, are measured accuracies without adversarial samples, and columns three and five are measured accuracies with adversarial samples. Column six is the calculated *Use?* metric value.

Table 5.2. Representative Summary Data (ω = 0.5, 75+(org = 25, adv = 25)).



	Baseline		Feature Selection		
Feature Set	**Original**	**Adversarial**	**Original**	**Adversarial**	***Use?***
LIWC	68.00%	56.00%	88.53%	16.40%	-0.41
Topic Modeling	84.00%	8.00%	91.87%	11.20%	0.49
Stylometry	60.00%	32.00%	96.27%	4.13%	-0.27
Hybrid	92.00%	32.00%	100.00%	9.20%	-0.63

If we first look at the baseline original accuracies, we might rank the feature sets based on effectiveness from an AIdS perspective as follows:

1. Hybrid – 92.00%
2. Topic Modeling – 84.00%
3. LIWC - 68.00%
4. Stylometry – 60.00%

Notice that the Hybrid feature set out-performs any of the individual feature sets (LIWC, Topic Modeling and Stylometry). Clearly, combining feature sets can yield synergy. Also, as we saw in Chapter 4, Topic Modeling outperforms the other non-hybrid feature sets. We hypothesize that this is because topic features are particularly well suited to the blogpost dataset.

Next, consider how the rankings change when we apply feature selection:

1. Hybrid – 100.00%
2. Stylometry – 96.27%
3. Topic Modeling – 91.87%
4. LIWC – 88.53%

We see that, when using feature selection, accuracies improve across all feature sets. We also see that the Hybrid feature set is the top performer. Based on these results alone, we might be convinced that all feature selection is good, and that the hybrid feature set is the most effective. But let us dig deeper into the data.

When we introduce adversarial samples and consider the accuracies, we see that the Hybrid feature set drops to the middle of the pack and LIWC dominates (in both the Baseline column and the Feature selection column). Based on these data, we might be persuaded to resort to using the LIWC feature set. However, if we use the LIWC feature set exclusively and there are no adversarial samples, we will have inadvertently reduced the AIdS effectiveness.

This is where the *Use?* metric comes into play. By comparing the effects of adversarial attacks on the baseline relative to adversarial attacks with feature selection, we can determine which feature set displays the most relative susceptibility. Using the *Use?* metric we can rank the feature sets from least to most susceptible as follows:

1. Topic Modeling
2. Stylometry
3. LIWC
4. Hybrid

In this case, Topic Modeling emerges as the most robust feature set.

One final point of interest; when we consider the results from the second set of experiments 75+(org = 100, adv = 100), see a slight movement in favor of using feature selection due to the dilution of the number of adversarial samples. However, these feature sets maintain roughly the same susceptibility rankings (see Appendix B for details). This indicates that benefits we see from feature selection, when applying a weaker adversarial attack, would need to be diluted even further before we would strongly consider using the Hybrid feature set.

5.4 Summary

To summarize this chapter, we might ask: what can we conclude? Unfortunately, the answer is: not enough!

This experiment in this chapter shows that, although feature selection can improve AIdS accuracy, in the face of an adversarial attack, in some cases, feature selection can be a disadvantage. However, this appears to be true only for some feature sets. We make three conclusions from this experiment:

1. The Good: Feature selection increases accuracy while decreasing the number of features needed.
2. The Bad: Our results show that, in some cases, feature selection can make AIdSs more susceptible to adversarial author identification attacks.
3. The Ugly: We do not yet have a clear understanding of the characteristics of this susceptibility.

Later chapters are devoted to characterizing, understanding, and predicting when susceptibility to adversarial author identification attacks will occur. The results laid out in this chapter form a foundation for warranted additional work. In the following chapters, we will consider varying the dataset size and constitution as well as feature selection mechanisms.

Chapter 6

Adversarial Authorship, Swarm Intelligence Feature Selection, and CAISIS-50 Dataset

6.1 Motivation

The research in the previous chapter is evidence of vulnerability when using genetic-based feature selection on the CASIS-25 dataset. This research piques the interest in understanding the parameters of adversarial authorship vulnerabilities to feature selection. In this chapter we consider additional feature selection approaches and increase the size of the dataset.

Specifically, in addition to Genetic & Evolutionary Feature Selection (GEFeS), we consider several Swarm Intelligence (SI) based feature selection algorithms, including Particle Swarm Intelligence (PSO), Artificial Bee Colony Optimization (ABCO), Ant System Optimization (ASO), and Glowworm Swarm Optimization (GSO). In addition, we employ a Random (RAND) feature selection algorithm as a baseline algorithm. We continue to use the same four feature sets from the previous chapter (i.e., LIWC, Topic Modeling, Stylometry, and Hybrid). We double the dataset size with the CASIS-50 dataset.

The intent of the work in this chapter is to consider the extent of the vulnerabilities when using these addition feature selection algorithms. We also want to investigate the effects of increasing the dataset size, specifically, increasing the number of authors.

6.2 Adversarial Authorship, Swarm Intelligence Feature Selection and CASIS-50 Dataset Experiment

As mentioned in the previous section, this experiment uses the same six feature selection algorithms GEFeS, PSO, ABCO, ASO, GSO and RAND) as discussed in Chapter 4, and the

CASIS-25 and CASIS-50 datasets. We configure the six algorithms as described in Section 4.2. Also, we continue to use a Linear Support Vector Machine (LSVM) for author classification as described in the previous chapter, including the fitness function described in Equation 4.1. Again, we vary the values of the feature reduction weighting factor (ω) using the same seven values consistent with previous chapters (i.e., 0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0). Further, the experiment runs 15,000 evaluations per feature selection algorithm per experiment measurement and uses 30 measurements.

For each of the measurements, we trained on all but 25 samples, and we calculated accuracy by testing with and without adversarial samples, both individually (i.e., on only the held out 25 samples) and across all samples. This means, for the CASIS-50 dataset, we trained on 175 samples (3 samples for the first 25 authors, plus four samples for the other 25 authors), and tested using the 25 samples from the first 25 authors, and all 200 samples, swapping in the adversarial samples as well. For the CASIS-25 dataset we trained on 75 samples and tested on 25 and 100 samples, respectively.

The experiment outlined in this chapter uses, in addition to the CASIS-25 dataset, the CASIS-50 dataset, which is a subset of the CASIS-1000 dataset [36, 37], and a superset of the CASIS-25 dataset. Like the CASIS-25 dataset, the CASIS-50 dataset consists of blog posts, with four posts per author. The CASIS-50 dataset contains all the samples of the CASIS-25 dataset as describe in the previous chapter, plus an additional 25 authors with four samples each for a total of 200 samples. In addition to the original 200 samples, we used 25 adversarial samples, one sample from each of the first 25 authors (these are the same adversarial samples used in the previous chapter). These samples were created by taking the fourth writing sample from each of the first 25 authors and applying Author-CAAT-V and AIM-IT [43]. For both the CASIS-25 and

CASIS-50 datasets, we use the preprocessing and feature extraction mechanisms as described in Chapter 4.

6.3 Adversarial Authorship, Swarm Intelligence Feature Selection and CASIS-50 Dataset Experiment Analysis

This section reviews the results of the experiment to identify general trends and points of interest. See Appendix C for a detailed treatment of the results.

This experiment yields results where the baseline accuracies are zero. As a result, we extend the definition of the *Use?* metric from what we defined in Table 5.1 to that shown in Table 6.1, which avoids zeros in the denominator.

Table 6.1. The Enhanced Definition of the *Use?* value.

$$\boldsymbol{Use}? = \begin{cases} \dfrac{\boldsymbol{\alpha_g^u - \alpha_b^u}}{\boldsymbol{\alpha_b^u}} + \dfrac{\boldsymbol{\alpha_g^a - \alpha_b^a}}{\boldsymbol{\alpha_b^a}}, & (\boldsymbol{\alpha_b^u \neq 0}) \wedge (\boldsymbol{\alpha_b^a \neq 0}) \\ \boldsymbol{\alpha_g^u - \alpha_b^u}, & (\boldsymbol{\alpha_b^u = 0}) \vee (\boldsymbol{\alpha_b^a = 0}) \end{cases}$$

Where:

$\boldsymbol{\alpha_g^u}$ is feature selection accuracy of original samples

$\boldsymbol{\alpha_b^u}$ is baseline accuracy of original samples

$\boldsymbol{\alpha_g^a}$ is feature selection accuracy of adversarial samples

$\boldsymbol{\alpha_b^a}$ is baseline accuracy of adversarial samples

$\boldsymbol{Use}? > \boldsymbol{0}$: use feature selection

$\boldsymbol{Use}? \leq \boldsymbol{0}$: do not use feature selection

Table 6.2 shows a representative summary of the results from this experiment. So as to summarize, Table 6.2 only shows results for three (GEFeS, ASO, PSO) of the six feature selection algorithms for *ω* = 0.5. The first column (Feature Set) identifies the feature set used to observe the results. The second column (Dataset) distinguishes between the CASIS-25 and the CASIS-50 datasets. The third column (Alg) identifies which of the three representative feature selection algorithms was used. The remaining columns show the results of the experiment for each feature set, dataset, and algorithm configuration. The columns labeled *Baseline* show results when not using feature selection, and the columns labeled *Feature Selection* show results when using feature selection. Under each of these labels are two columns that show results without adversarial samples (Original) and with adversarial samples (Adversarial). The last column indicates the *Use?* value as defined in Table 6.1.

When *Use?* values are negative, the configuration demonstrates a vulnerability to feature selection. An intuitive review of the *Use?* values in Table 6.2 reveals that nearly all the Topic Modeling configurations (with the single exception of PSO/CASIS-25/100) sport positive values. This observation is consistent with that made in the previous chapter, namely that Topic Modeling seems resilient to the feature selection vulnerability. Of course, the Topic Modeling/CASIS-50/25 *Use?* values are positive due to the complete author identification failure in the baseline using adversarial samples. Therefore, since feature selection allows some identification, these values are positive.

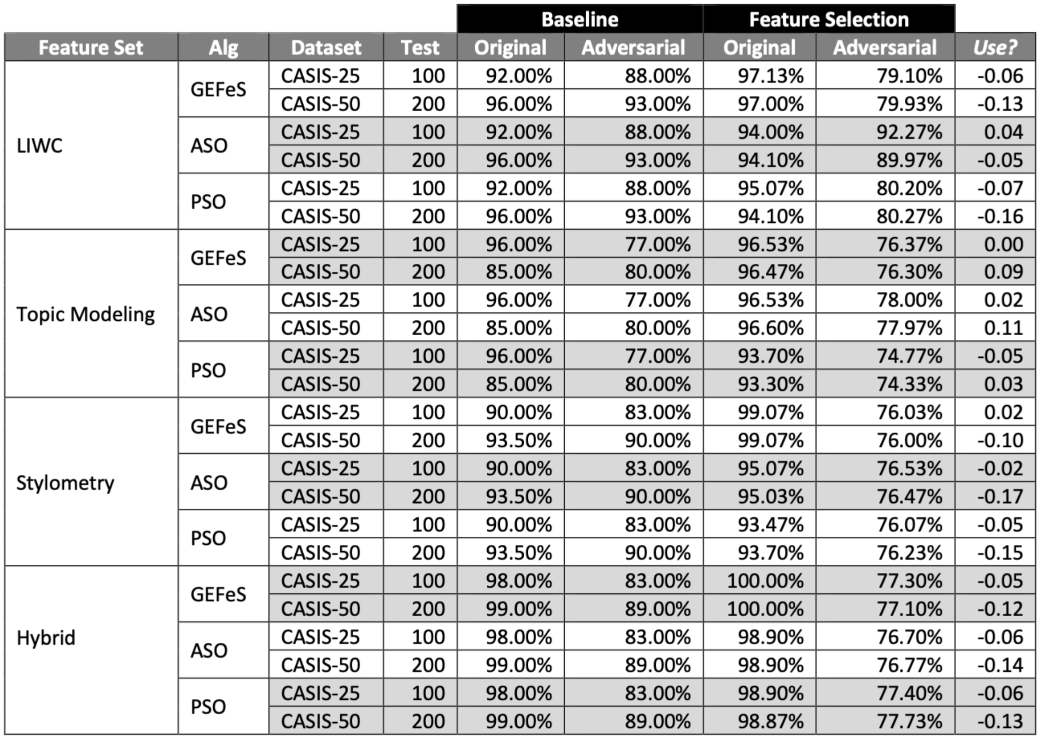
Looking beyond Topic Modeling in the results, we find a handful of positive *Use?* values associated with ASO. The frequency of these positive values is hardly enough to warrant referring to this as a trend, but it may indicate some tendency towards immunity from the feature selection vulnerability. The only other positive value is for the configuration of Stylometry/CASIS-25/GEFeS/100.

Table 6.2. Representative Summary Data ($\omega = 0.5$).

					Baseline		Feature Selection		
Feature Set		Dataset	Alg	Test	Original	Adversarial	Original	Adversarial	*Use?*
LIWC		CASIS-25	GEFeS	25	68.00%	56.00%	88.53%	16.40%	-0.41
				100	92.00%	88.00%	97.13%	79.10%	-0.06
			ASO	25	68.00%	56.00%	76.00%	61.07%	0.21
				100	92.00%	88.00%	94.00%	92.27%	0.04
			PSO	25	68.00%	56.00%	80.40%	20.93%	-0.44
				100	92.00%	88.00%	95.07%	80.20%	-0.07
		CASIS-50	GEFeS	25	68.00%	44.00%	88.00%	19.73%	-0.26
				200	96.00%	93.00%	97.00%	79.93%	-0.13
			ASO	25	68.00%	44.00%	76.40%	59.87%	0.48
				200	96.00%	93.00%	94.10%	89.97%	-0.05
			PSO	25	68.00%	44.00%	76.40%	21.07%	-0.40
				200	96.00%	93.00%	94.10%	80.27%	-0.16
Topic Modeling		CASIS-25	GEFeS	25	84.00%	8.00%	91.87%	11.20%	0.49
				100	96.00%	77.00%	96.53%	76.37%	0.00
			ASO	25	84.00%	8.00%	86.13%	12.00%	0.53
				100	96.00%	77.00%	96.53%	78.00%	0.02
			PSO	25	84.00%	8.00%	85.33%	9.60%	0.22
				100	96.00%	77.00%	93.70%	74.77%	-0.05
		CASIS-50	GEFeS	25	40.00%	0.00%	91.60%	10.93%	1.29
				200	85.00%	80.00%	96.47%	76.30%	0.09
			ASO	25	40.00%	0.00%	86.40%	11.87%	1.16
				200	85.00%	80.00%	96.60%	77.97%	0.11
			PSO	25	40.00%	0.00%	85.07%	9.20%	1.13
				200	85.00%	80.00%	93.30%	74.33%	0.03
Stylometry		CASIS-25	GEFeS	25	60.00%	32.00%	96.27%	4.13%	-0.27
				100	90.00%	83.00%	99.07%	76.03%	0.02
			ASO	25	60.00%	32.00%	80.27%	6.13%	-0.47
				100	90.00%	83.00%	95.07%	76.53%	-0.02
			PSO	25	60.00%	32.00%	73.87%	4.27%	-0.64
				100	90.00%	83.00%	93.47%	76.07%	-0.05
		CASIS-50	GEFeS	25	48.00%	20.00%	96.27%	4.00%	0.21
				200	93.50%	90.00%	99.07%	76.00%	-0.10
			ASO	25	48.00%	20.00%	81.13%	5.87%	0.04
				200	93.50%	90.00%	95.03%	76.47%	-0.17
			PSO	25	48.00%	20.00%	74.80%	4.93%	-0.20
				200	93.50%	90.00%	93.70%	76.23%	-0.15
Hybrid		CASIS-25	GEFeS	25	92.00%	32.00%	100.00%	9.20%	-0.63
				100	98.00%	83.00%	100.00%	77.30%	-0.05
			ASO	25	92.00%	32.00%	95.60%	6.8%	-0.75
				100	98.00%	83.00%	98.90%	76.70%	-0.06
			PSO	25	92.00%	32.00%	95.60%	9.60%	-0.66
				100	98.00%	83.00%	98.90%	77.40%	-0.06
		CASIS-50	GEFeS	25	92.00%	12.00%	100.00%	8.40%	-0.21
				200	99.00%	89.00%	100.00%	77.10%	-0.12
			ASO	25	92.00%	12.00%	95.60%	7.07%	-0.37
				200	99.00%	89.00%	98.90%	76.77%	-0.14
			PSO	25	92.00%	12.00%	95.47%	10.93%	-0.05
				200	99.00%	89.00%	98.87%	77.73%	-0.13

When we consider the effects increasing the number of authors by comparing similar configurations that vary only in the number of tests, we see that *Use?* values increase for all configurations with 25 tests. However, when testing using all samples (100 or 200), except for the Topic Modeling configurations, we see just the opposite where *Use?* values decrease. Table 6.3 makes it easier to compare *Use?* values as this table is limited to testing using all samples, and places similar Feature Set/Alg in adjacent rows. From Table 6.3, we discover that, with the exception of the Topic Modeling feature set, increasing the number of authors and testing all samples appears to increase the feature selection vulnerability.

Table 6.3. Representative Summary Data Arranged to Compare Dataset Size ($\omega = 0.5$, Test = 100 or Test = 200).



				Baseline		Feature Selection		
Feature Set	Alg	Dataset	Test	Original	Adversarial	Original	Adversarial	*Use?*
LIWC	GEFeS	CASIS-25	100	92.00%	88.00%	97.13%	79.10%	-0.06
		CASIS-50	200	96.00%	93.00%	97.00%	79.93%	-0.13
	ASO	CASIS-25	100	92.00%	88.00%	94.00%	92.27%	0.04
		CASIS-50	200	96.00%	93.00%	94.10%	89.97%	-0.05
	PSO	CASIS-25	100	92.00%	88.00%	95.07%	80.20%	-0.07
		CASIS-50	200	96.00%	93.00%	94.10%	80.27%	-0.16
Topic Modeling	GEFeS	CASIS-25	100	96.00%	77.00%	96.53%	76.37%	0.00
		CASIS-50	200	85.00%	80.00%	96.47%	76.30%	0.09
	ASO	CASIS-25	100	96.00%	77.00%	96.53%	78.00%	0.02
		CASIS-50	200	85.00%	80.00%	96.60%	77.97%	0.11
	PSO	CASIS-25	100	96.00%	77.00%	93.70%	74.77%	-0.05
		CASIS-50	200	85.00%	80.00%	93.30%	74.33%	0.03
Stylometry	GEFeS	CASIS-25	100	90.00%	83.00%	99.07%	76.03%	0.02
		CASIS-50	200	93.50%	90.00%	99.07%	76.00%	-0.10
	ASO	CASIS-25	100	90.00%	83.00%	95.07%	76.53%	-0.02
		CASIS-50	200	93.50%	90.00%	95.03%	76.47%	-0.17
	PSO	CASIS-25	100	90.00%	83.00%	93.47%	76.07%	-0.05
		CASIS-50	200	93.50%	90.00%	93.70%	76.23%	-0.15
Hybrid	GEFeS	CASIS-25	100	98.00%	83.00%	100.00%	77.30%	-0.05
		CASIS-50	200	99.00%	89.00%	100.00%	77.10%	-0.12
	ASO	CASIS-25	100	98.00%	83.00%	98.90%	76.70%	-0.06
		CASIS-50	200	99.00%	89.00%	98.90%	76.77%	-0.14
	PSO	CASIS-25	100	98.00%	83.00%	98.90%	77.40%	-0.06
		CASIS-50	200	99.00%	89.00%	98.87%	77.73%	-0.13

6.4 Summary

As the results of this chapter show, we see further evidence that feature selection renders AIdSs susceptible to adversarial attacks under certain situations. The important question is: what are the specific conditions that lead to susceptibility?

Based on the results of this chapter, it appears that adversarial susceptibility is generally independent of the feature selection algorithm. Also, some feature sets (specifically Topic Modeling) seem to perform better under attack than others. However, this appears to be less pronounced in the case of fewer authors (CASIS-25) when testing across all samples. With more authors (CASIS-50), we see that Topic Modeling performs well for both testing strategies.

Topic Modeling may perform well because of the adversarial techniques employed in the study. That is to say that techniques based on iterative translation and paraphrasing will not affect the topic of discourse. This may indicate that either Topic Modeling is an effective feature set against adversarial attacks (especially with a blog-post dataset where authors tend to discuss their topics of interest), or it may be necessary to develop effective topic-related adversarial techniques.

Chapter 7

An Analysis of the Effects of Genetic and Swarm Intelligence Based Feature Selection on Adversarial Author Identification Using Many Authors

7.1. Motivation

In previous chapters, we have assessed the effectiveness of genetic-based and SI-based feature selection for AIdS. Further, we have provided evidence of the vulnerability these feature selection methods expose to adversarial author Identification. In the previous chapter, besides introducing SI-based feature selection, we explored the effects of doubling the number of authors in the dataset from 25 to 50. In this chapter, we extend the research along this dimension by, again, doubling the number of authors from 50 to 100, giving us an additional datapoint in our understanding.

7.2. Genetic and Swarm Intelligence Based Feature Selection on Adversarial Author Identification Using Many Authors Experiment

This experiment uses precisely the same configuration as detailed in Section 6.2; except we introduce a third dataset. Recall that the experiment in Chapter 6 used the CASIS-25 and the CASIS-50 datasets. In this experiment, in addition to these two previous datasets, we introduce the CASIS-100 dataset, which is a superset of the CASIS-50 dataset, and a subset of the CASIS-1000 dataset (see Section 3.3). The CASIS-100 dataset consists of four samples for each of 100 authors. Since the CASIS-100 dataset is a superset of the CASIS-25 dataset, we can also continue to use the 25 adversarial samples from the CASIS-25 dataset. Following the training and testing process used in Chapter 6, we train on 375 of the 400 samples from the CASIS-100 dataset, and test on the remaining 25 samples as well as all 400 samples, with and without the adversarial samples.

To summarize the experiment configurations, we use six feature selection algorithms (GEFeS, ABCO, ASO, GSO, PSO, RAND), an LSVM AIdS, seven values of $\boldsymbol{\omega}$ (0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0), four feature sets (LIWC, Topic Modeling, Stylometry, Hybrid), three datasets (CASIS-25, CASIS-50, CASIS-100), with and without adversarial samples, with 30 runs of each configuration for a total of 30,240 observations.

7.3. Genetic and Swarm Intelligence Based Feature Selection on Adversarial Author Identification Using Many Authors Experiment Analysis

This section provides a summary of the results observed from this experiment. For a discussion of detailed results, see Appendix D.

We present the summary results using six charts. There is one chart for each dataset/test-method combination. The charts reflect the resulting *Use?* values from these experiments as defined in Table 6.1. Recall that the *Use?* value represents the relative improvement of accuracies due to feature selection, with and without adversarial samples. When the *Use?* value is positive, it indicates a preference for using feature selection, even when under adversarial attack. A value less-than or equal to zero indicates a preference to avoid feature selection when under adversarial attack. In the event that the accuracy is zero (either with or without feature selection), then *Use?* takes on the value of the difference attained by using feature selection (notice that this difference is bounded by zero on the lower end).

Each of Figures 7.1-7.6, are graphs with lines representing a combination of a feature selection algorithm and a feature set. Each graph considers a combination of a dataset (i.e., CASIS-

25, CASIS-50, CASIS100), and a testing/training approach (i.e., testing on only the adversarial samples versus all the samples), for a total of six graphs. We considered the top three feature selection algorithms (i.e., GEFeS, PSO, ASO), and four feature sets (i.e., LIWC, Topic Modeling, Stylometry, Hybrid), for a total of 12 curves per graph. The vertical axis of the graph represents the *Use?* values and contains values that range from negative values to positive values. The line mid-way up the vertical access labeled "0" represents the separation between using (above the line) feature selection, and not using (below the line) feature selection as calculated according to Table 6.1. The horizontal axis represents the various values of ω.

Figure 7.1 shows the *Use?* values for the CASIS-25 dataset where we trained on 75 samples and tested on only the 25 samples, swapping in the adversarial samples to calculate the *Use?* values as per Table 6.1. Notice that all feature selection algorithms may be used for the Topic Modeling feature set, and that ASO also works for LIWC. All other feature selection/feature set combinations demonstrate a vulnerability to adversarial attacks.

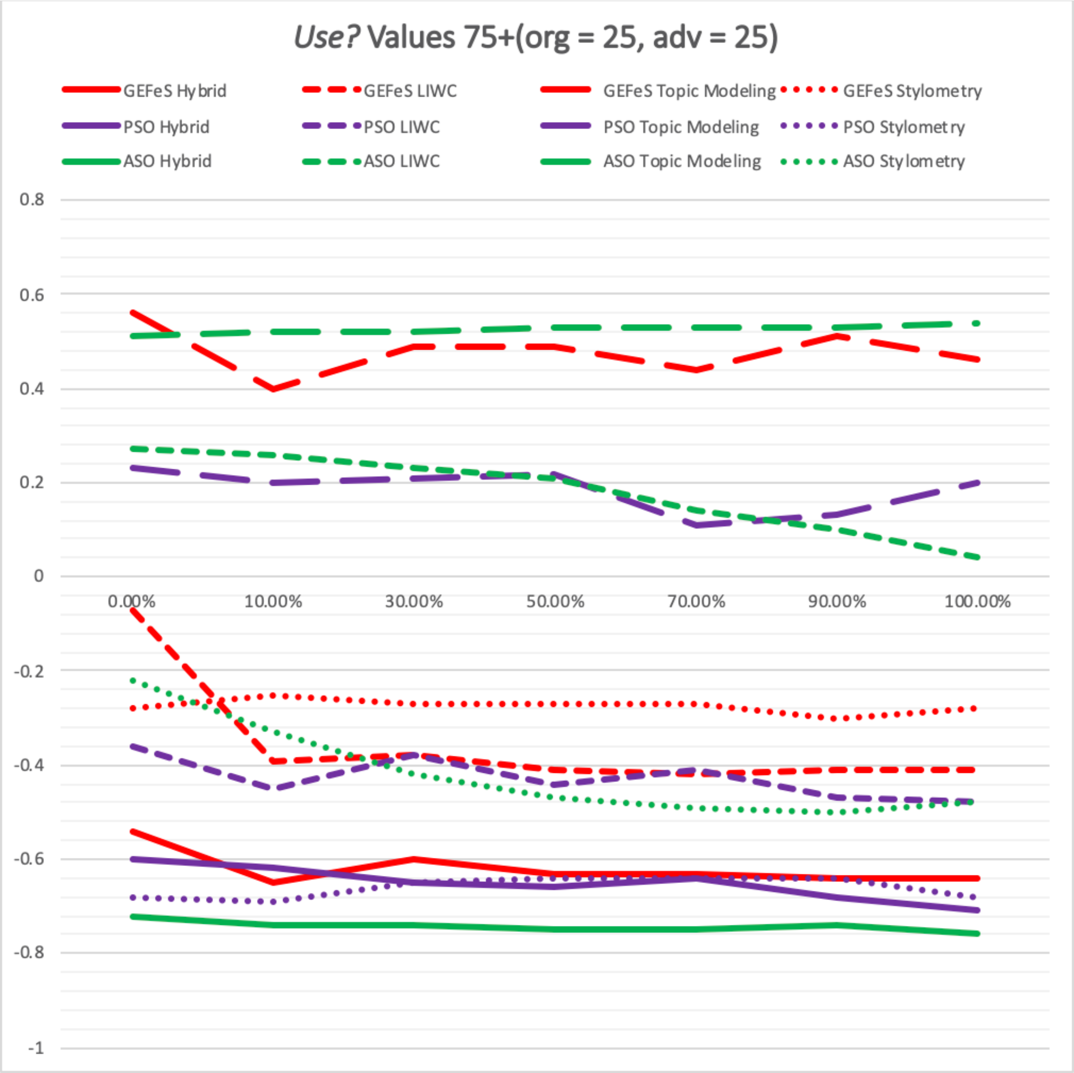


Figure 7.1. *Use?* Values CASIS-25, 75+(org = 25, adv = 25).

Figure 7.2 uses the same dataset, but tests using all samples (again swapping in the 25 adversarial samples to be able to calculate the *Use?* values according to Table 6.1). We see from this graph, that ASO continues to work well for LIWC and Topic Modeling feature sets, as does GEFeS with Stylometry. GEFeS, using Topic Modeling, is marginal (mostly close to zero), with one instance positive and two are negative. However, all other feature selection/feature set combinations appear to be vulnerable to adversarial attacks.

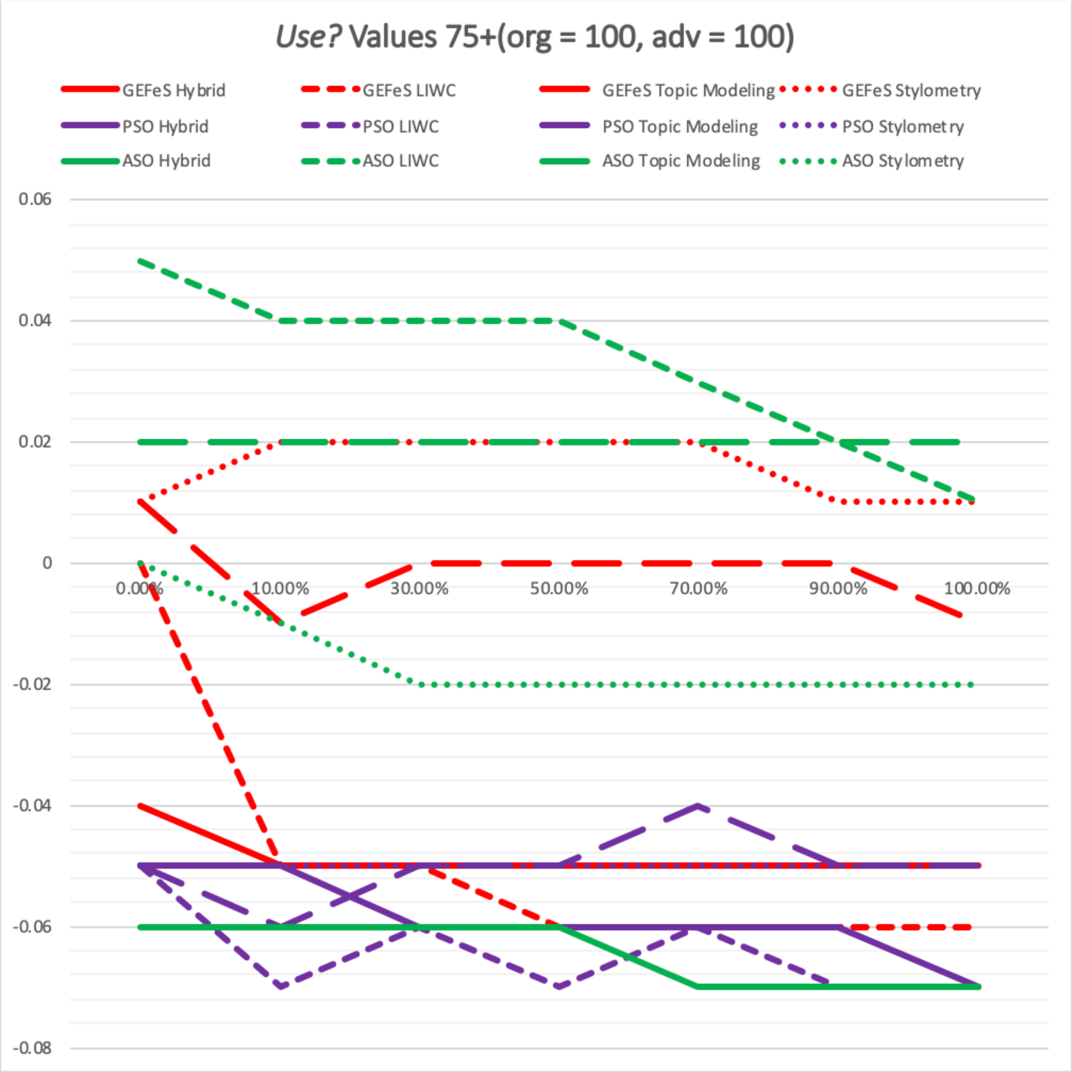


Figure 7.2. *Use?* Values CASIS-25, 75+(org = 100, adv = 100).

Figure 7.3 switches to the CASIS-50 dataset and employs the testing on the 25 samples (like Figures 7.1 and 7.5). Here, we see that Topic Modeling is resilient to adversarial attacks for all three feature selection algorithms. We also see that ASO is positive for the LIWC feature set and the Stylometry feature set, as is GEFeS also (for Stylometry). Other combinations appear to be mostly susceptible to adversarial attacks.

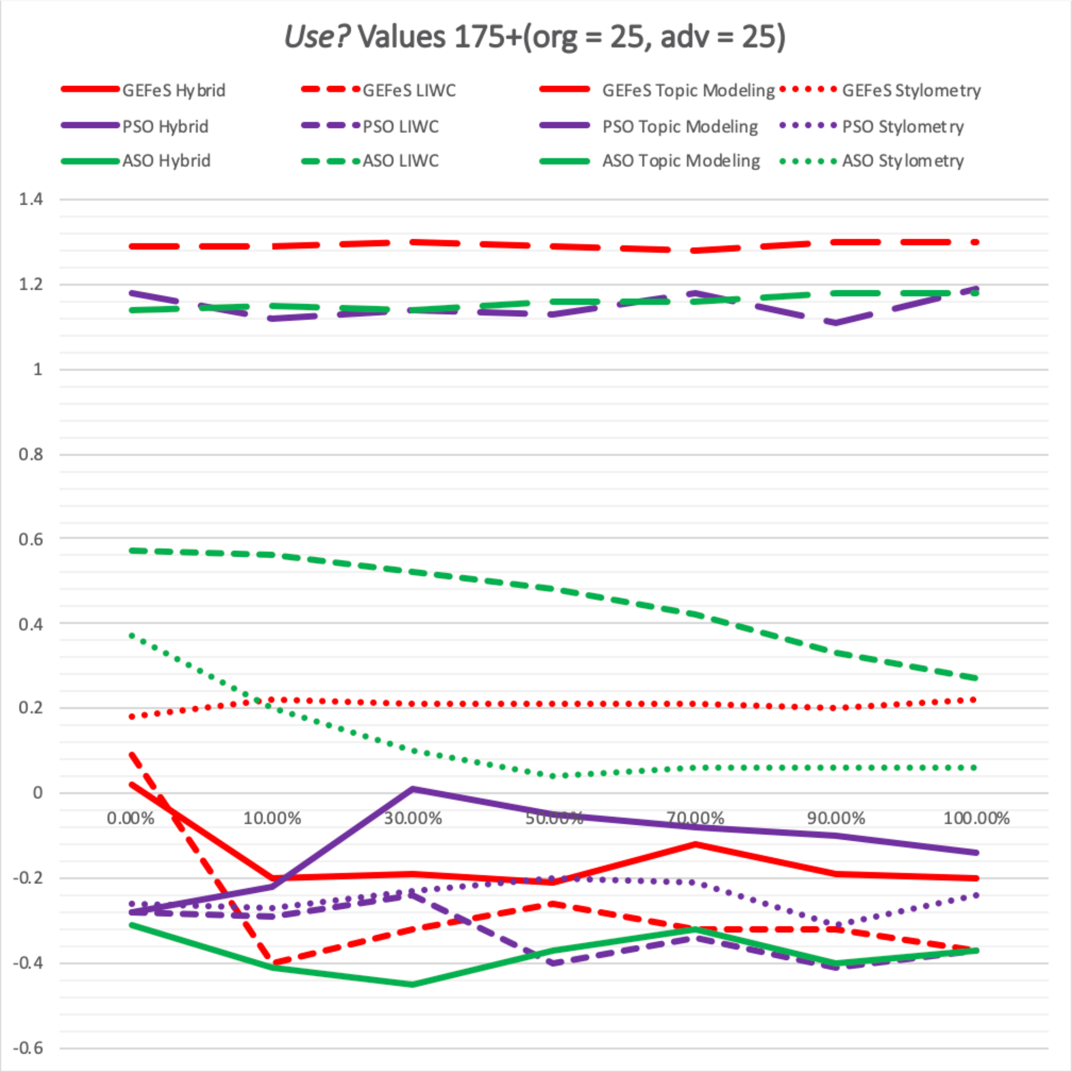


Figure 7.3. *Use?* Values CASIS-50, 175+(org = 25, adv = 25).

Figure 7.4 again uses the CASIS-50 dataset with testing across all samples (similar to Figures 7.2 and 7.6). In this graph we see that only Topic Modeling (but across all feature selection algorithms) resists adversarial attacks.

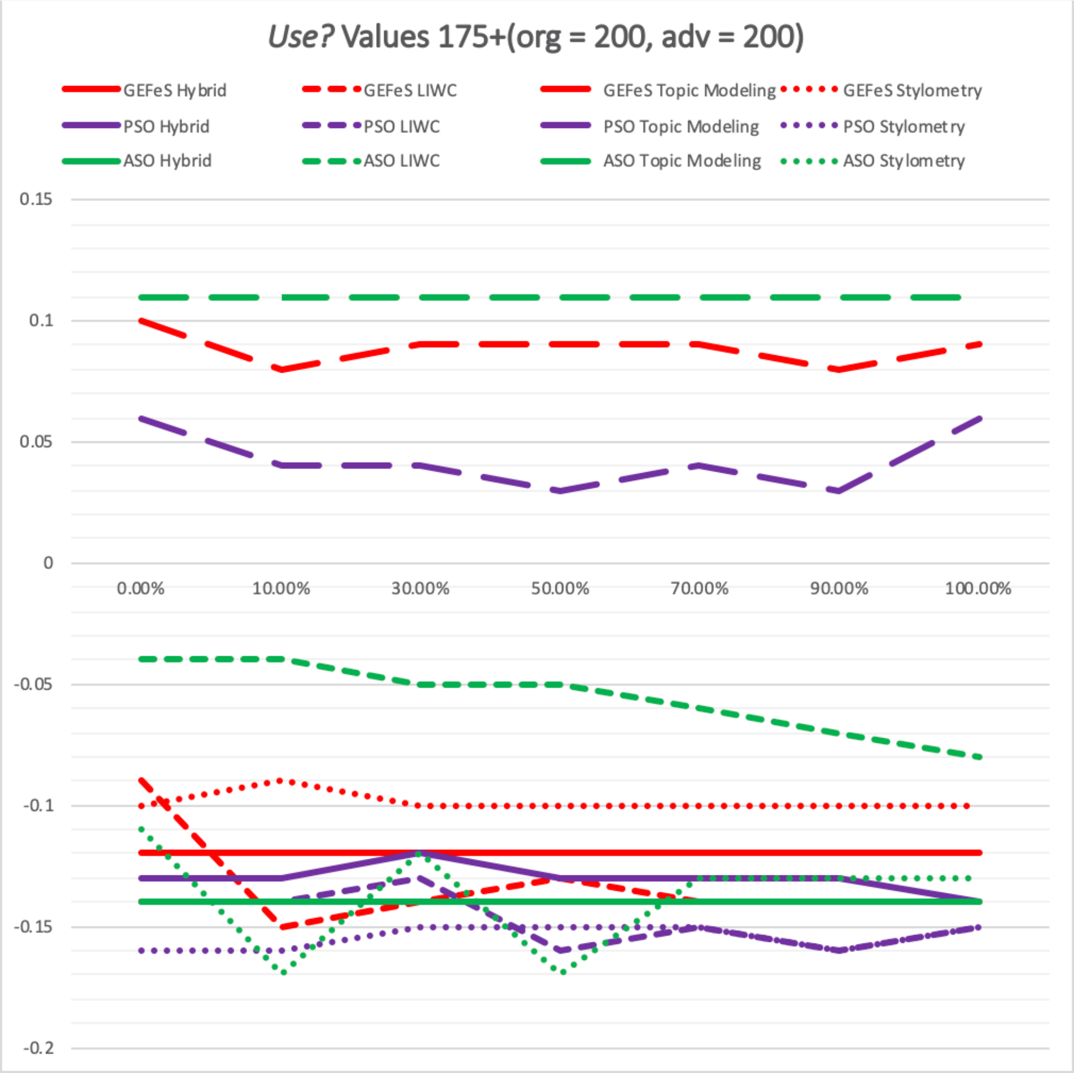


Figure 7.4. *Use?* Values CASIS-50, 175+(org = 200, adv = 200).

Figure 7.5 shows results using the CASIS-100 dataset and testing on only 25 samples. We see that Topic Modeling performs well, except for ASO which is marginal. However, interestingly we see that ASO and Stylometry perform well for lower values of ω, but as ω increases, ASO becomes susceptible to adversarial attacks. We also see that the combination of GEFeS and Stylometry is marginally positive.

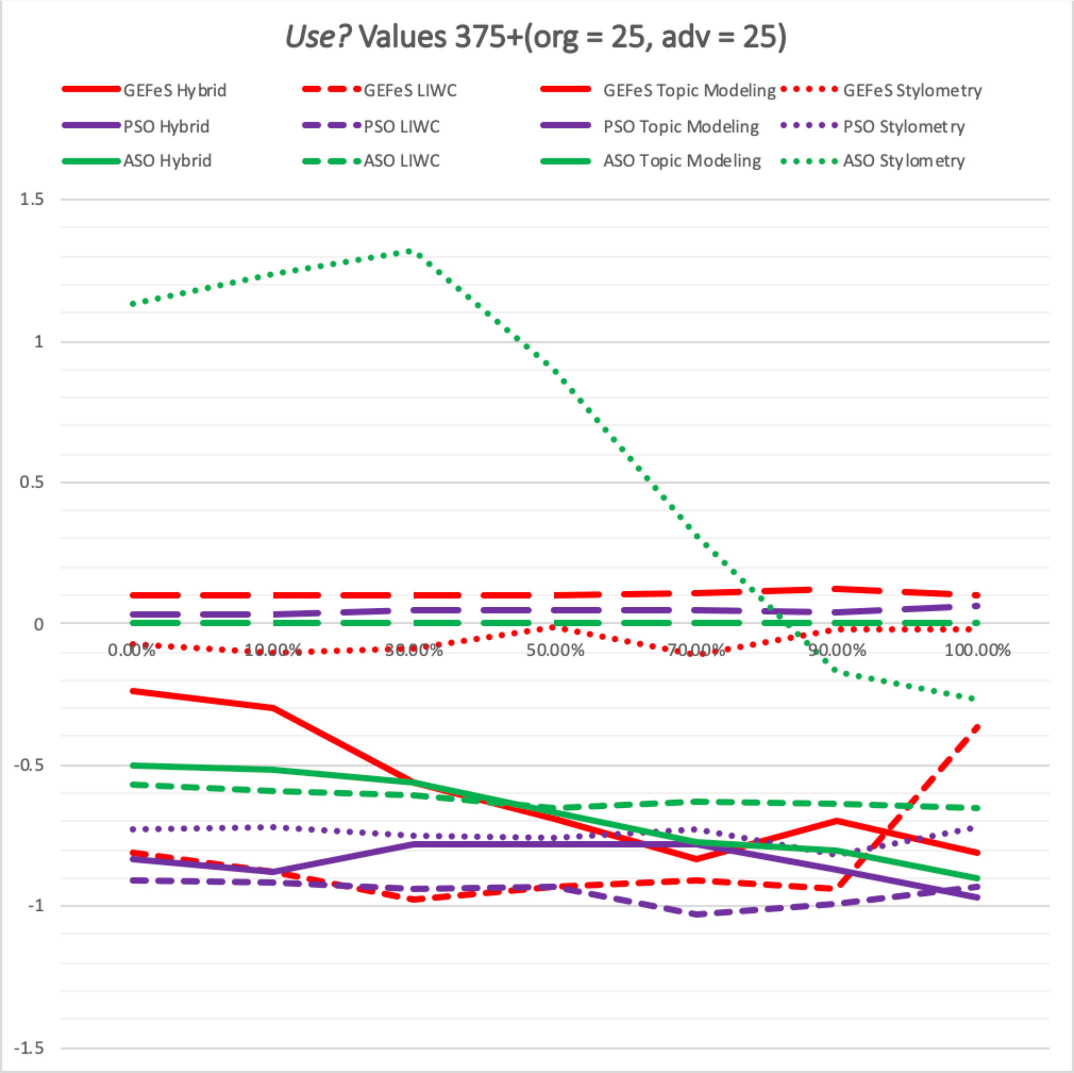


Figure 7.5. *Use?* Values CASIS-100, 375+(org = 25, adv = 25).

Figure 7.6 Uses the CASIS-100 dataset and tests on all samples. In this experiment, Stylometry is the only feature set that resists Adversarial attacks mostly only for GEFeS and ASO. All other combinations of feature selection and feature sets are marginal at best.

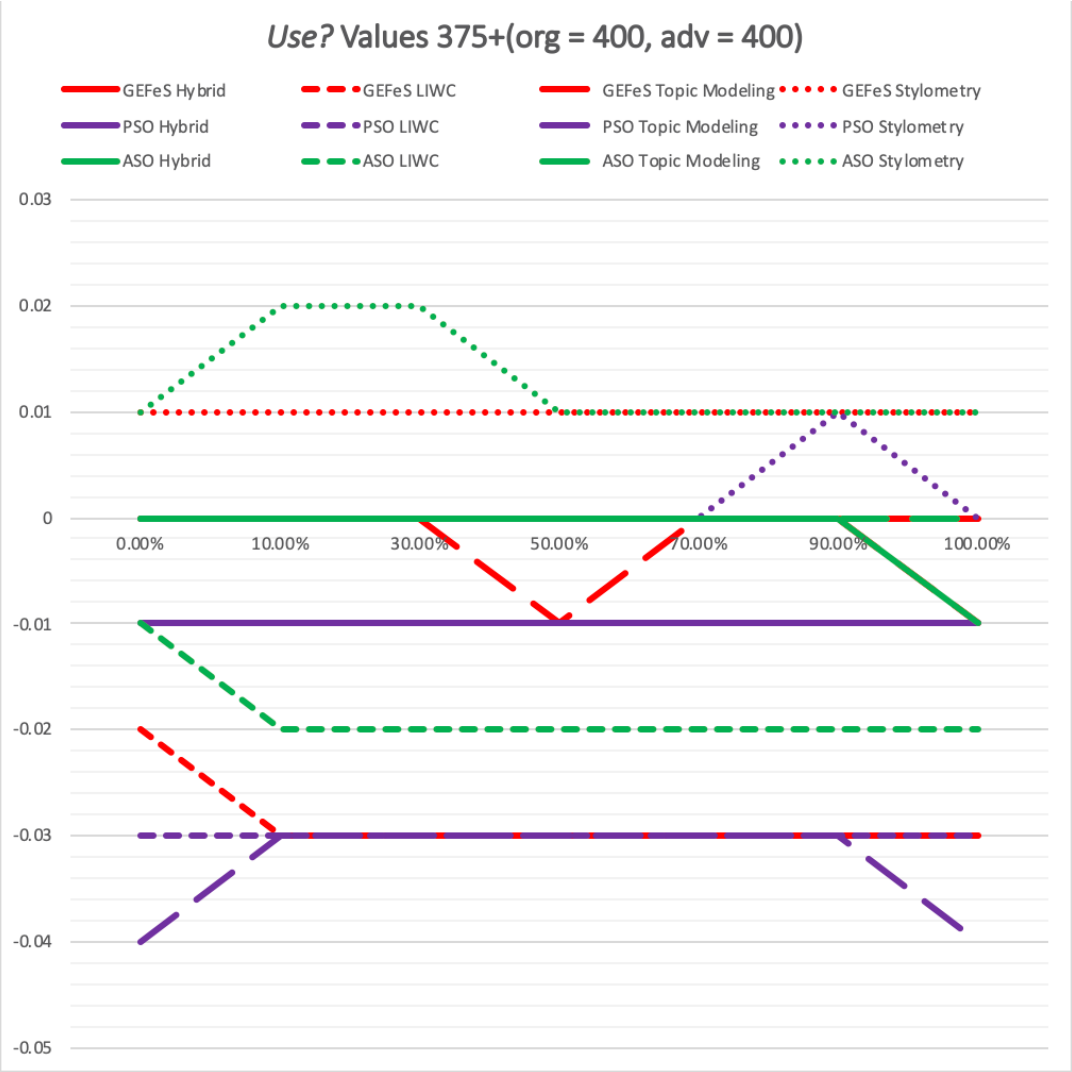


Figure 7.6. *Use?* Values CASIS-100, 375+(org = 400, adv = 400).

Finally, we include three representative feature consistency charts for each of the three datasets studied in this chapter (CASIS-25, CASIS-50, CASIS-100). All three charts represent results observed when using GEFeS for feature selection with $\omega = 0.5$. In these bar charts, the bars represent the ratio that each feature was selected across the 30 runs. The chart is divided into three sections by color, where the blue bars represent LIWC features, red bars represent Topic Modeling features, and green bars represent Stylometry features.

Figure 7.7 shows the feature consistency of the CASIS-25 dataset. This chart shows a dramatic difference between the feature sets, with the Topic Modeling feature set (red) having most of the consistency. The Stylometry feature set (green) is least consistent, and the LIWC feature set (blue) is moderately in between.

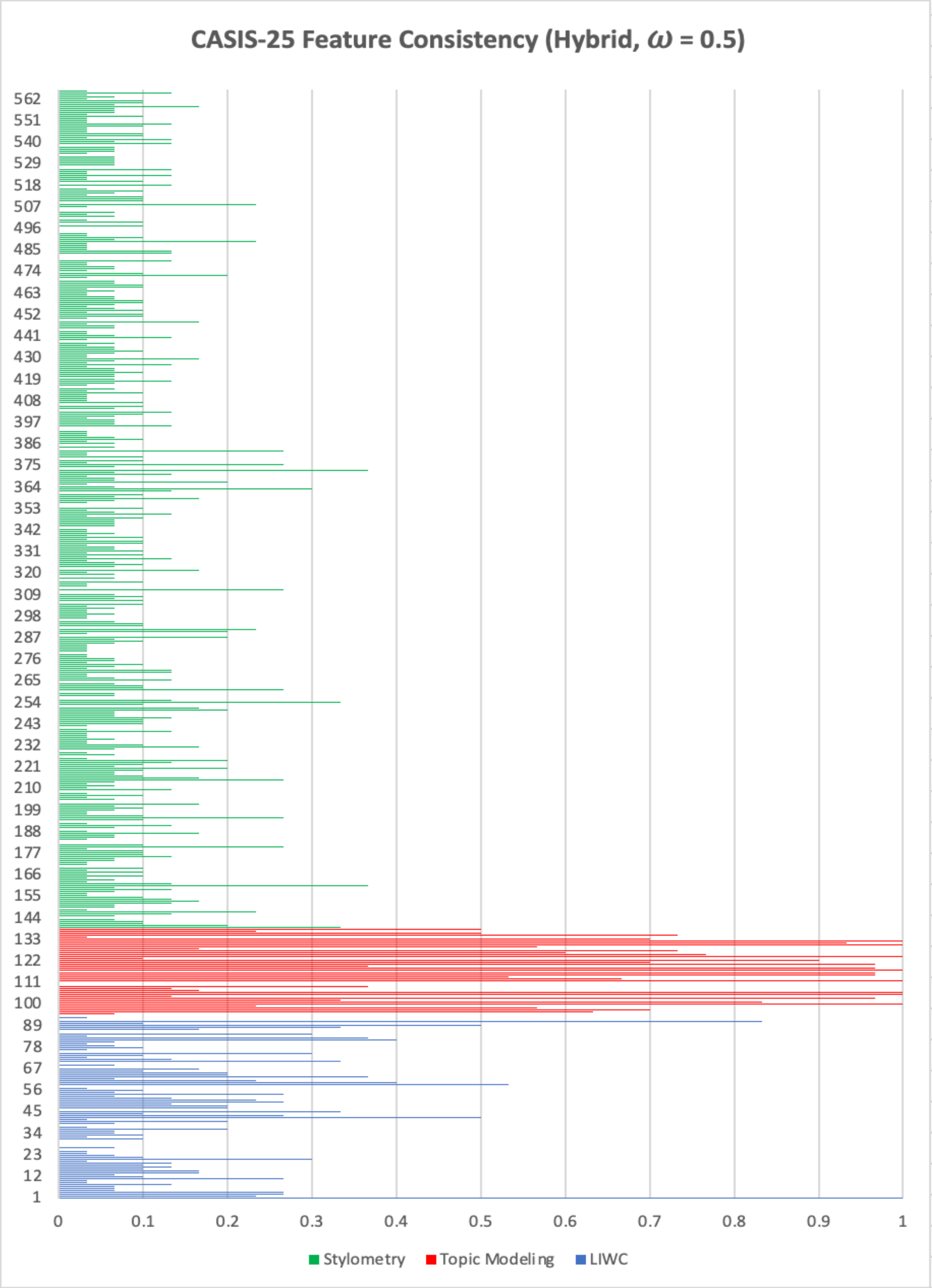


Figure 7.7. Feature Consistency for the CASIS-25 Hybrid Feature Set ($\omega = 0.5$).

Figure 7.8, which shows the feature consistency for the CASIS-50 dataset, looks similar to Figure 7.7 and maintains the same relative ranking of consistency between the three feature sets. However, as we increase the number of authors from 25 to 50, while the Topic Modeling feature set (red) is clearly still the dominantly most consistent feature set, we can see a bit less differentiation between the LIWC feature set (blue) and the Stylometry feature set (green).

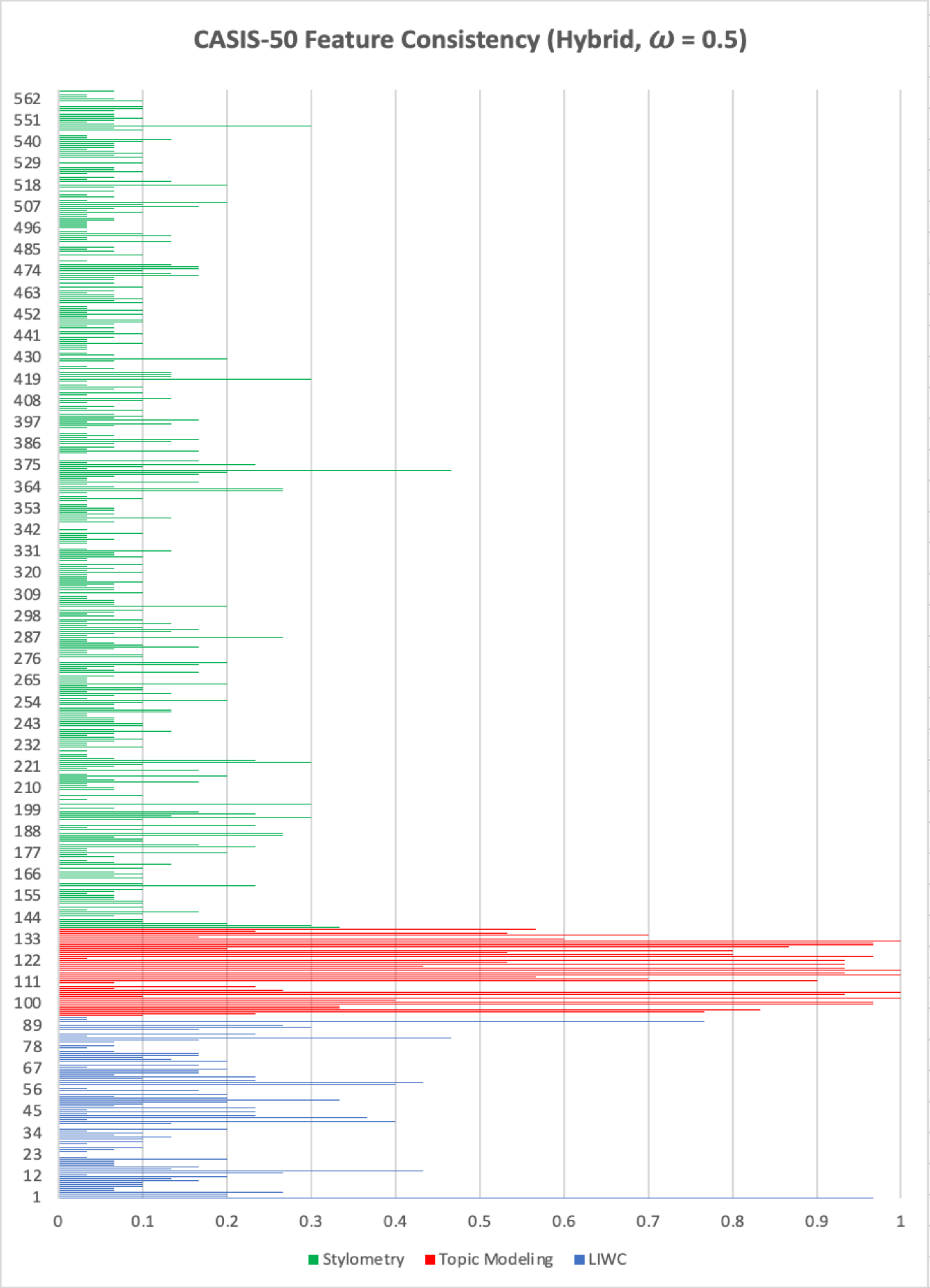


Figure 7.8. Feature Consistency for the CASIS-50 Hybrid Feature Set ($\omega = 0.5$).

Figure 7.9 shows feature consistency for the CASIS-100 dataset. With the increase in the number of authors, the Topic Modeling feature set (red) is still clearly most consistent. However, the differentiation between the LIWC feature set (blue) and the Stylometry feature set is much less dramatic. In fact, it is not intuitively obvious which feature set has the most consistently used features.

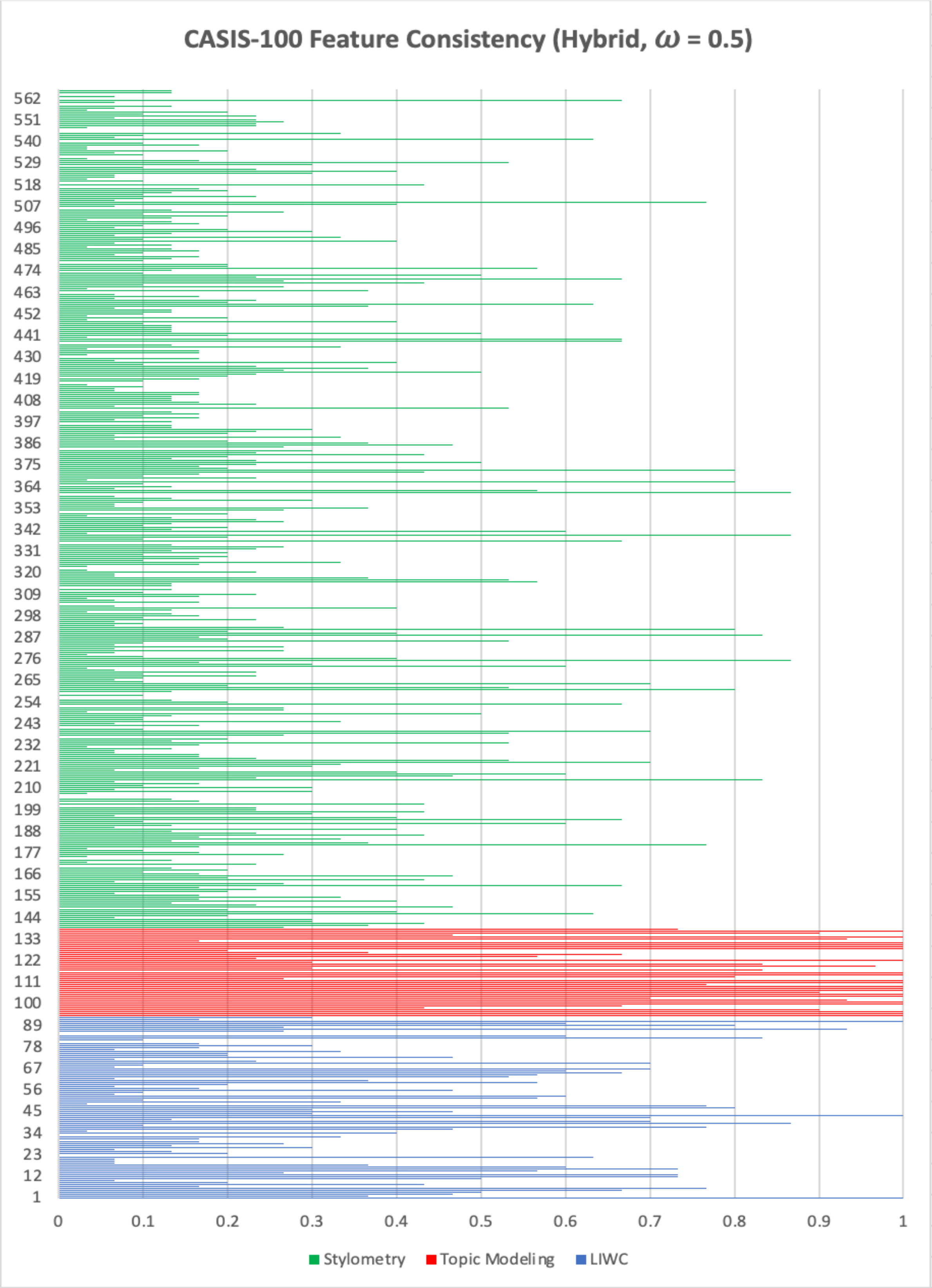


Figure 7.9. Feature Consistency for the CASIS-100 Hybrid Feature Set ($\omega = 0.5$).

7.4. Summary

In this chapter, we demonstrate that feature selection improves AIdS accuracy, but at a possible cost. We demonstrate that sometimes feature selection is susceptible to adversarial authorship attacks. Further, while it is not yet clear what all the factors are that affect feature selection susceptibility, a significant factor appears to be the feature set itself.

Future research needs to continue to explore and eventually characterize the susceptibility of feature selection to adversarial authorship attacks. We need to come to understand and predict feature selection susceptibilities. This includes exploring the parameters of datasets such as dataset size, number of authors and dataset constitution.

Chapter 8

A Study of the Harmful Effects of Genetic and Swarm Based Feature Selection with Respect to Adversarial Author Profiling

8.1. Motivation

The previous chapters detail experiments in genetic-based and SI-based feature selection algorithms and the effects of this feature selection on adversarial author identification. Specifically, chapter 4 demonstrated that genetic-based and SI-based feature selection can improve AIdS accuracy. Chapters 5, 6 and 7 give evidence that, while feature selection may improve accuracy for non-adversarial configurations, this same feature selection renders AIdSs vulnerable to adversarial attacks. All work done in chapters 4, 5, 6, and 7 has focused on a single dataset type, namely blog posts as contained in the CASIS-25, CASIS-50, and CASIS-100 datasets. This leads to a question of the possible influence of dataset type on the feature selection vulnerability.

The work in this chapter uses the same feature sets as well as the same feature selection algorithms but varies the dataset to use Twitter-based writing samples to shed further light on this susceptibility phenomena. The feature set used in this chapter's study has the added advantage that all samples have been labeled not only with an author identifier, but also an author type (male, female, or bot). The author type designation allows us to also explore the effects of genetic-based and SI-based feature selection on author profiling.

8.2. Genetic and Swarm Based Feature Selection with Respect to Adversarial Author Profiling Experiment

This chapter's experiment builds on previous chapters' experiments to investigate the interaction between author identification, various EC feature selection algorithms and feature sets, but extends the previous work by addressing a different dataset. Since this chapter builds on previous work, this experiment uses the same feature selection algorithms as chapters 6 and 7, namely LIWC, Topic Modeling, Stylometry and a hybrid of the previous three. Similarly, this work employs the same feature selection algorithms chapters 4, 6 and 7, namely GEFeS, PSO, ABCO, ASO, GSO, and RAND. The dataset for this work is a subset of the PAN19 dataset as described in Section 3.4. This data set consists of tweets contributed by humans (both male and female) and bots. For this research, we created a subset of the PAN19 dataset by selecting 25 authors (eight male, eight female and nine bots) we designate as the PAN19-25 dataset. We created five samples for each author by combining four tweets per sample. We combined the tweets to create large enough samples to give the AIdS a fair chance at functioning well. Since there are 25 authors with five samples each, we use a total of 125 classified samples. We then selected one of the five samples from each of the authors for adversarial processing, which generated another 25 adversarial samples. It may be noted that some of the tweet content (two of the 25 samples) did not lend itself well to adversarial processing, so the effects of the processing were limited on these two samples. This configuration allows us to perform five-fold cross-validation. Other than the changes resulting from introducing a new dataset, this experiment follows the same process used in chapters 6 and 7.

8.3 Genetic and Swarm Based Feature Selection with Respect to Adversarial Author Profiling Experiment Analysis

We calculate a derivative metric from the results referred to as *Use?* based on the work in Chapter 6. The idea of *Use?* is to have an indicator that tells if feature selection should be used or not, based on the effects of adversarial attacks. Table 8.1 gives a formal definition of *Use?*. Note that this definition of *Use?* varies slightly from that in Table 6.1 in that we have introduced ρ, which allows us to adjust the amount of concern we place on the adversarial effect. Note that for a value of $\rho = 0$, we are interested only in the improvement obtained from feature selection. Alternatively, when $\rho = 1$ we are concerned with the adversarial effects of feature selection. The explanation in Table 8.1 refers to feature selection accuracy versus baseline accuracy, and original samples versus adversarial samples. Feature selection accuracy refers to accuracy achieved using feature selection, and baseline accuracy refers to accuracy achieved using all features of the specified feature set. Original samples are samples without adversarial texts and adversarial samples include the adversarial texts.

Intuitively, the *Use?* indicator represents the relative change in accuracies (when using and not using feature selection) with and without adversarial samples. If the value of *Use?* is less than zero, feature selection creates a susceptibility to adversarial attack. But if *Use?* is greater than or equal to zero, the benefit of using feature selection outweighs a risk of susceptibility to adversarial attacks.

Table 8.1. The Definition of the *Use?* Value.

$$\boldsymbol{Use}? = \begin{cases} (\mathbf{1}-\boldsymbol{\rho})\dfrac{\boldsymbol{\alpha_g^u}-\boldsymbol{\alpha_b^u}}{\boldsymbol{\alpha_b^u}} + \boldsymbol{\rho}\dfrac{\boldsymbol{\alpha_g^a}-\boldsymbol{\alpha_b^a}}{\boldsymbol{\alpha_b^a}}, & (\boldsymbol{\alpha_b^u} \neq \mathbf{0}) \wedge (\boldsymbol{\alpha\,{}_b^a} \neq \mathbf{0}) \\ \boldsymbol{\alpha_g^u}-\boldsymbol{\alpha_b^u}, & (\boldsymbol{\alpha_b^u} = \mathbf{0}) \vee (\boldsymbol{\alpha_b^a} = \mathbf{0}) \end{cases}$$

Where:

$\boldsymbol{\rho}$ is adversarial risk weighting factor

$\boldsymbol{\alpha_g^u}$ is feature selection accuracy original samples

$\boldsymbol{\alpha_b^u}$ is baseline accuracy of original samples

$\boldsymbol{\alpha_g^a}$ is feature selection accuracy adversarial samples

$\boldsymbol{\alpha_b^a}$ is baseline accuracy of adversarial samples

$\boldsymbol{Use}? > \mathbf{0}$: use feature selection

$\boldsymbol{Use}? \leq \mathbf{0}$: do not use feature selection

To help make sense of these raw data, we have compiled two charts (figures 8.1 & 8.2) that map *Use?* values as a function of ω. Figure 8.1 represents results trained on 100 samples and tested on 25 (*100+(org = 25, adv = 25)*) evaluation configuration, and Figure 8.2 also trains on 100 samples, but tests on all 125 (*100+(org = 125, adv = 125)*). In these charts, the line color indicates the feature selection algorithms, and the line type indicates the feature set. Note that we have omitted GSO, ABCO, and RAND feature selection algorithms to make the charts more legible. See Appendix E for a treatment of the detailed results.

In Figure 8.1 notice that the Topic Modeling lines for each of the three feature selection algorithms remain in the positive space. Also, notice that the GEFeS Hybrid line is dominant, and the ASO Hybrid line is sometimes positive and sometimes negative. The remainder of the lines are distinctly negative.

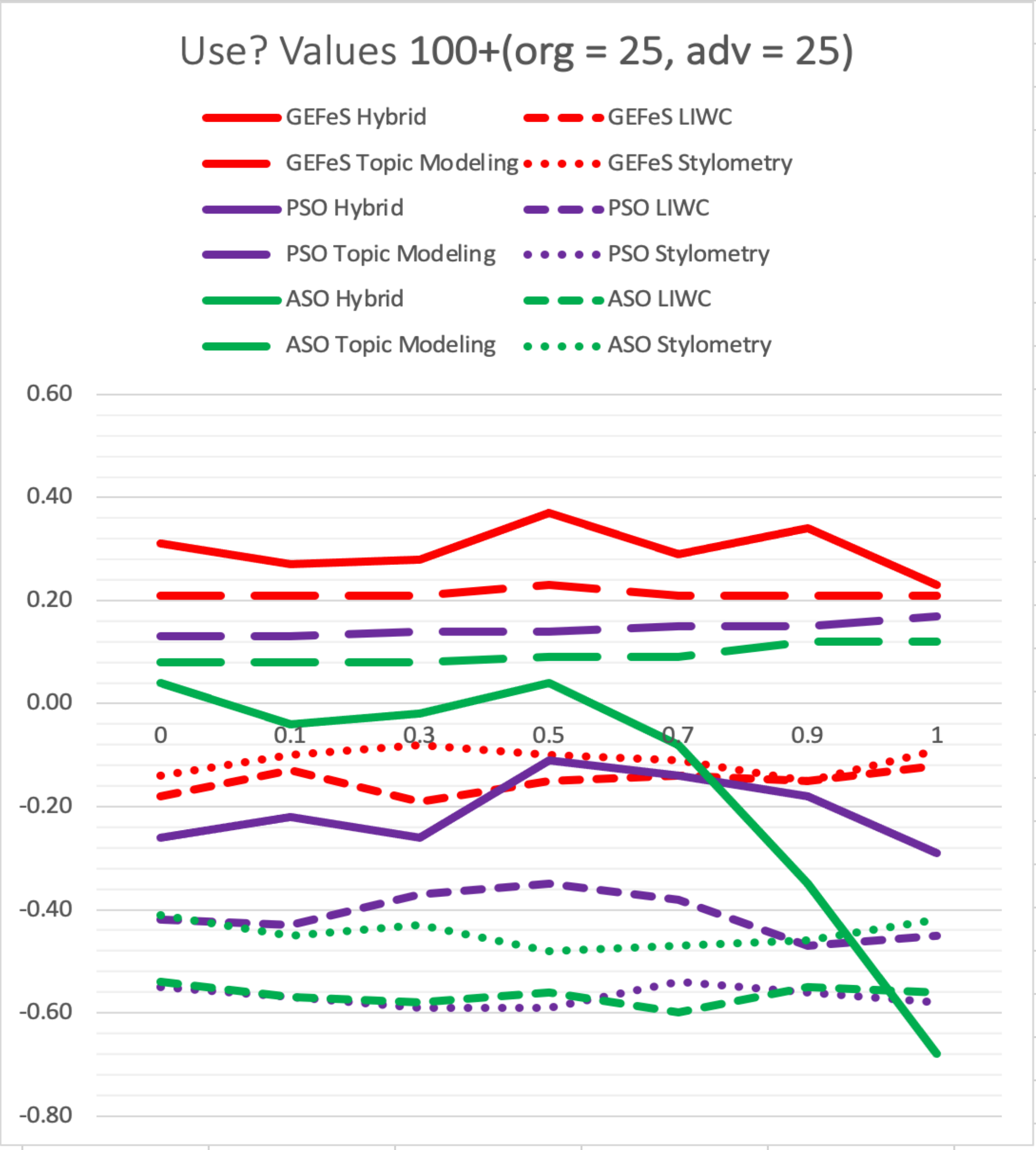


Figure 8.1. *Use?* Values as a Function of ω, PAN19-25, 100+(org = 25, adv = 25).

Figure 8.2 tells a very different story with the Hybrid lines remaining manly in the positive space. All other lines are negative, and Topic Modeling clearly is most negative.

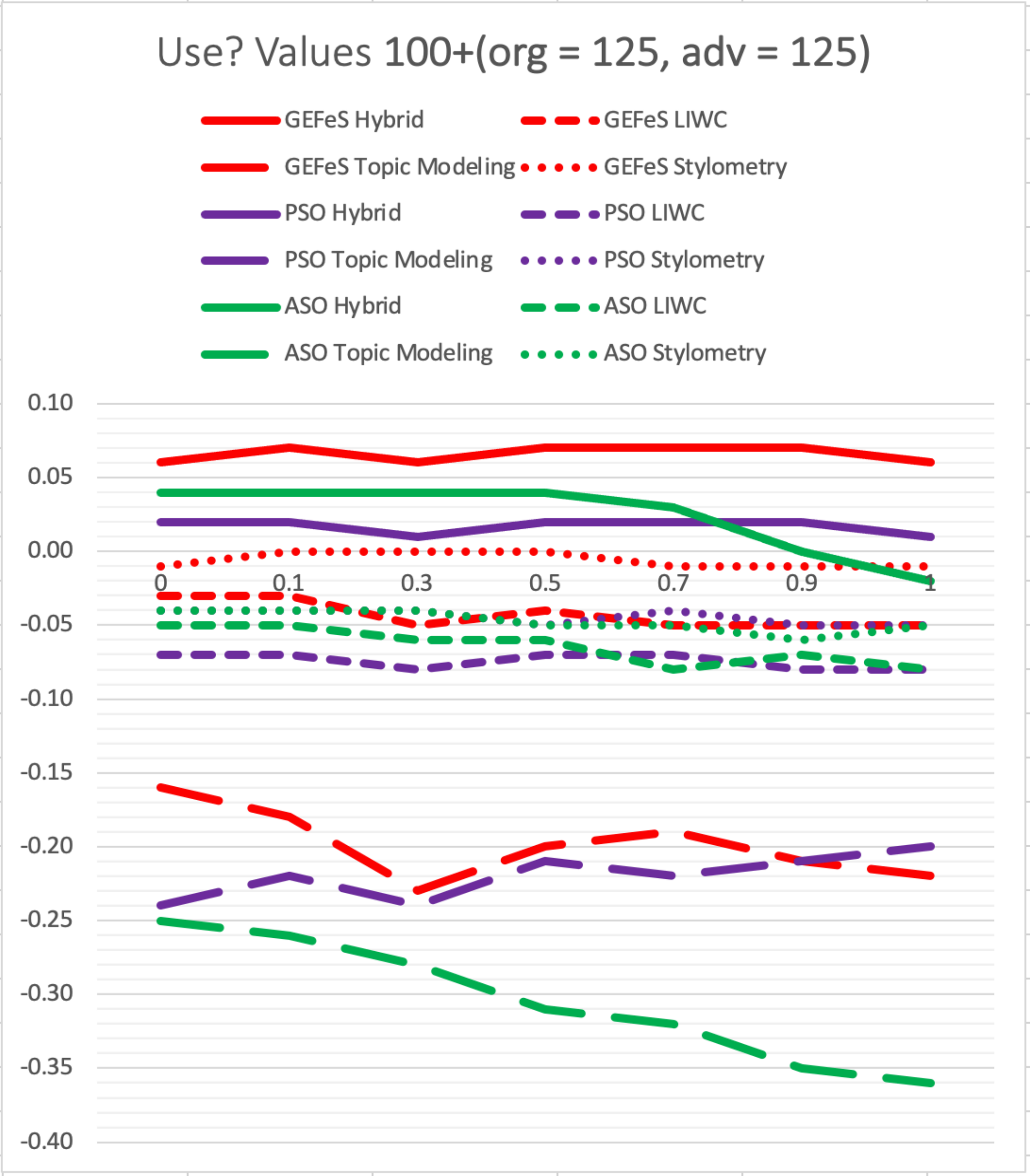


Figure 8.2. *Use?* Values as a Function of ω, PAN19-25, 100+(org = 125, adv = 125).

Consider the feature consistency as shown in Figure 8.3. This figure color codes the three feature sets: LIWC is blue, Topic Modeling is red and Stylometry is green. From this figure we see that Topic Modeling has four out of 46 (8.70%) of the most dominant (greater than 90% consistent) features, whereas LIWC has three out of 93 features (3.23%) and Stylometry only has 1 out of 427 features (0.002%). This is further evidence that Topic Modeling is a productive feature set for this dataset (both relatively and absolutely).

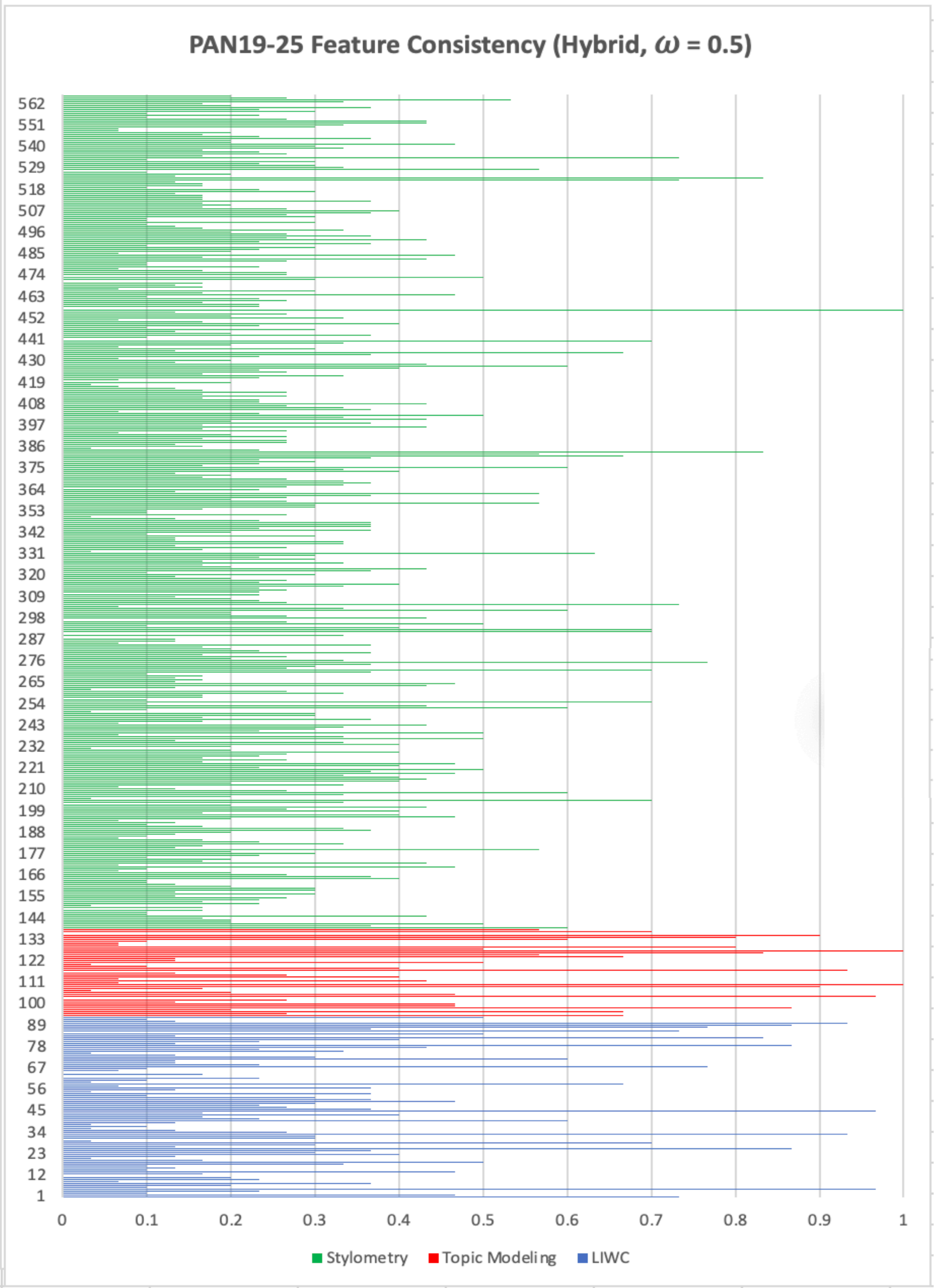

PAN19-25 Feature Consistency (Hybrid, ω = 0.5)
562
551
540
529
518
507
496
485
474
463
452
441
430
419
408
397
386
375
364
353
342
331
320
309
298
287
276
265
254
243
232
221
210
199
188
177
166
155
144
133
122
111
100
89
78
67
56
45
34
23
12
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
🟩 Stylometry
🟥 Topic Modeling
🟦 LIWC

Figure 8.3. Feature Consistency for the Hybrid Feature Set ($\omega = 0.5$).

The intent of this research is to try to understand characteristics of adversarial attacks on author identification using feature selection. As such, we make the following observations based on figures 8.1 and 8.2.

The first observation is that intuitively, the line types in the chart seem to cluster more than do the line colors. This is easiest to notice in Figure 8.2. This observation means that feature sets are a better indicator of susceptibility *for this dataset*, than is the feature selection algorithm.

A second observation is that although the feature set is likely the most significant factor in determining susceptibility, GEFeS almost always dominates the other feature selection algorithms. Although the results are mixed, more often than not, PSO dominates ASO. However, it is important to note that for some feature sets, GEFeS would still not be advised based on the *Use?* values.

In addition to the author identification work included herein, the PAN19-25 dataset is tagged with an attribute that identifies authors as bot, male or female. This affords an opportunity to consider the effects of feature selection on adversarial Author Profiling. To explore these effects, we created confusion matrices for the feature selection mechanism and feature set combination that performed best (i.e., had highest *Use?* values) for each of the seven values of ω. These matrices are included as tables 8.2 through 8.8 for the non-adversarial samples, and tables 8.9 through 8.15 with adversarial samples. The row and columns are both labeled 1-25, which is the author identification number. Author identification numbers 1-9 are bot samples, 10-17 are male samples and 18-25 are female samples. The rows of the matrix represent the correct identification and columns represent the predicted identification. Each matrix represents the cumulative results for the 30 runs, and the value in each cell represents the ratio of selection. Therefore, the sum of the values of each row equals 1.0 (unless there are sight rounding errors). A perfect system of prediction would have 1’s along the diagonal. We have highlighted the mistaken predictions with a color-code where the cells that represent incorrect predictions less than one third of the time are green, cells with incorrect predictions that occur between one-third and two-thirds are yellow, and cell with incorrect predictions two-thirds or more are red. The blue diagonal cells contain the ratio of time the system correctly identified the author.

The cell color-coding yields an intuitive feel for the effectiveness of both author identification and profiling. For example, a quick glance at Table 8.2 versus its adversarial counterpart, 8.9 shows much more color on Table 8.9, indicating that adversarial attacks are effective. We also see that as ω increases (from Table 8.2 through Table 8.8, or Table 8.9 through Table 8.15), which emphasizes reducing features, the color of the matrices increases, indicating that reducing the features does have a cost with respect to accuracy.

Intuitively, as shown in Table 8.2, in the absence of adversarial attack and with ω value of 0.0, we see that ASO is very effective at identifying the author. The limited missteps appear to occur most frequently when identifying bots, with a slight propensity for profiling bots as females. However, compare Table 8.2 with Table 8.6 (still ASO, but with $\omega = 0.7$). As ω increases, we see that incorrect profiling generally still predicts females (for bots and females), but incorrect bot predictions also increase, such that males are least frequently incorrectly profiled. Varying the feature selection mechanism and increasing ω (Table 8.8) does not appear to have a significant impact on this trend.

In Table 8.9, we see that the incorrect predictions appear to occur most frequently for males followed by bots and females. Also, it appears that bots and females are more difficult to predict than males. Compare this to Table 8.13 ($\omega = 0.7$) where females have fewer incorrect predictions than males or bots.

Finally, in Table 8.15 (PSO with $\omega = 1.0$), we see that bots appears to be more frequently predicted correctly than males or females. This shift appears to occur when transitioning from Table 8.13 to Table 8.14, which may be evidence that the shift is due to the feature selection mechanism (over a change in ω), since tables 8.9-8.13 use ASO, and tables 8.14 and 8.15 use PSO.

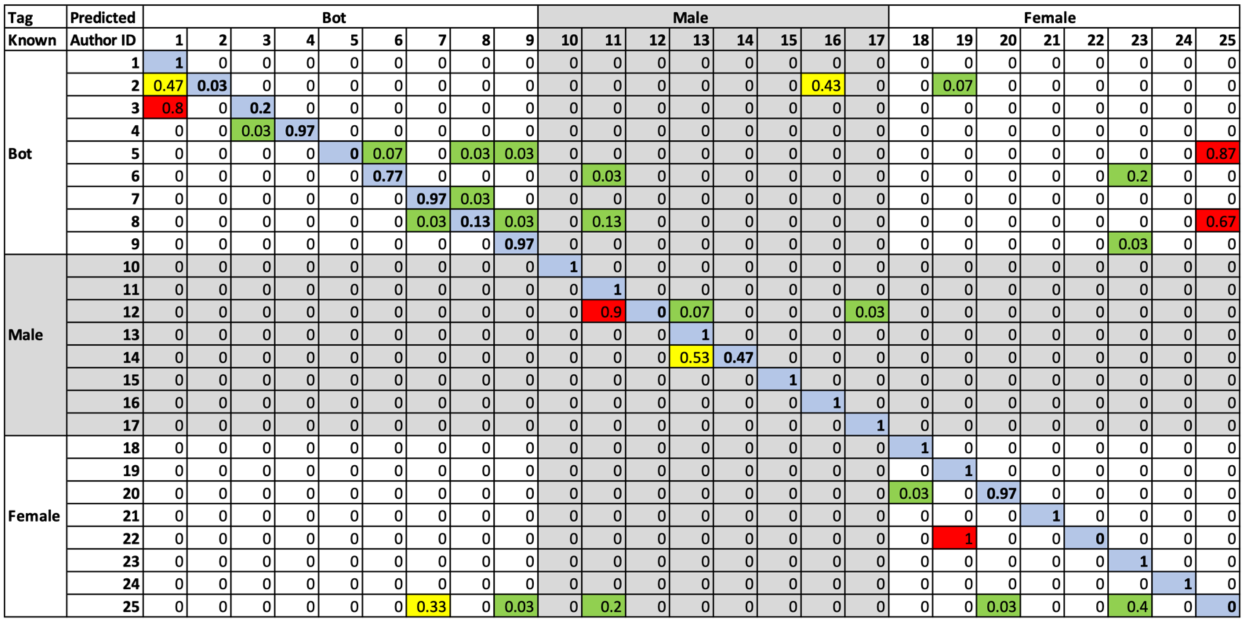
Table 8.2. Confusion Matrix for ASO, Hybrid, ω = 0.0.

Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.1	0.07	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0.33	0	0	0.17	0	0	0	0	0	0
	3	0.93	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0.03	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9
	6	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
	7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0.03	0.07	0	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.6
	9	0	0	0	0	0	0	0	0	0.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0.07
Male	10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0.7	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0.03	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.97	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	25	0	0	0	0	0	0	0.47	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.43	0	0

Table 8.3. Confusion Matrix for ASO, Hybrid, ω = 0.1.

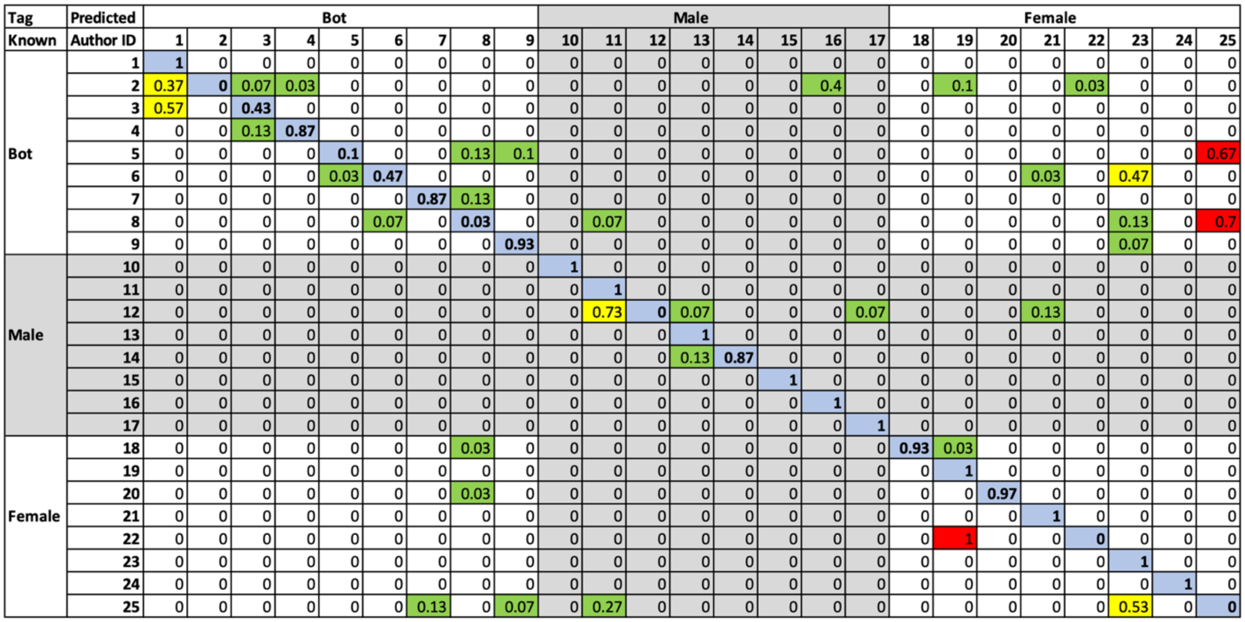
Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.3	0.07	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0
	3	0.9	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.93
	6	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0
	7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0.1	0.1	0	0.13	0	0	0	0	0	0	0	0	0	0.03	0	0.03	0	0.6
	9	0	0	0	0	0	0	0	0	0.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0
Male	10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0.97	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0.37	0.63	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0.03	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0.97	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	25	0	0	0	0	0	0	0.53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.47	0	0

Table 8.4. Confusion Matrix for ASO, Hybrid, ω = 0.3.



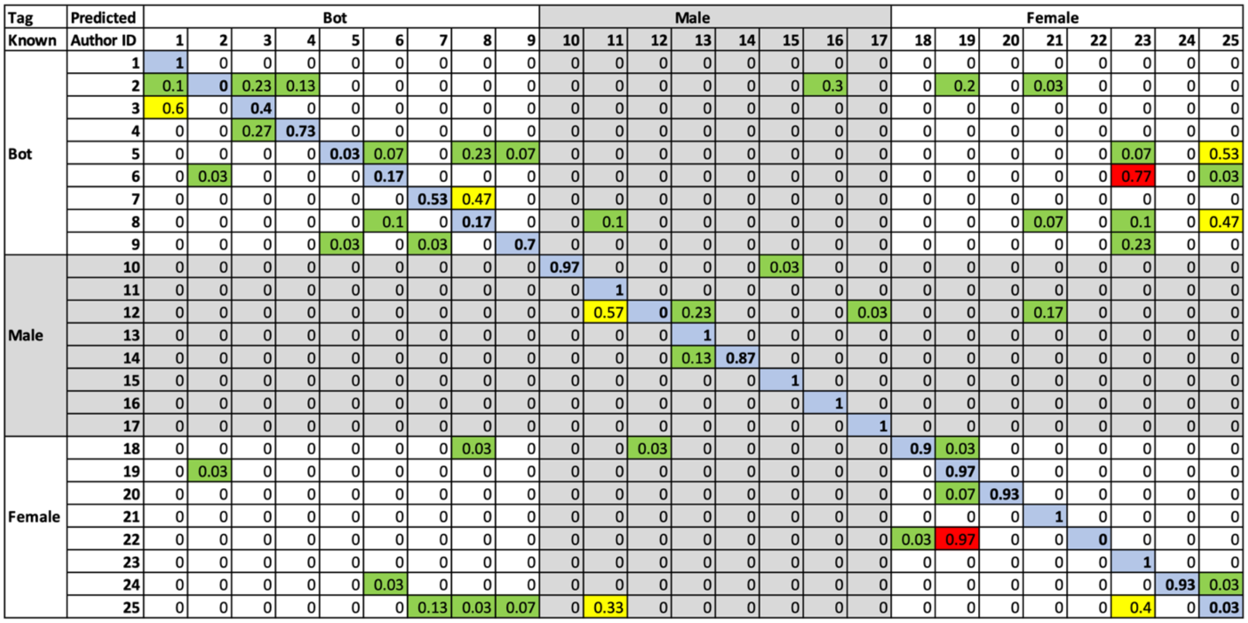
Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.47	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0.43	0	0	0.07	0	0	0	0	0	0
	3	0.8	0	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.03	0.97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0.07	0	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.87
	6	0	0	0	0	0	0.77	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0
	7	0	0	0	0	0	0	0.97	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0.03	0.13	0.03	0	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.67
	9	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0
Male	10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0.9	0	0.07	0	0	0	0.03	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0.53	0.47	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0.97	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	25	0	0	0	0	0	0	0.33	0	0.03	0	0.2	0	0	0	0	0	0	0	0	0.03	0	0	0.4	0	0

Table 8.5. Confusion Matrix for ASO, Hybrid, ω = 0.5.



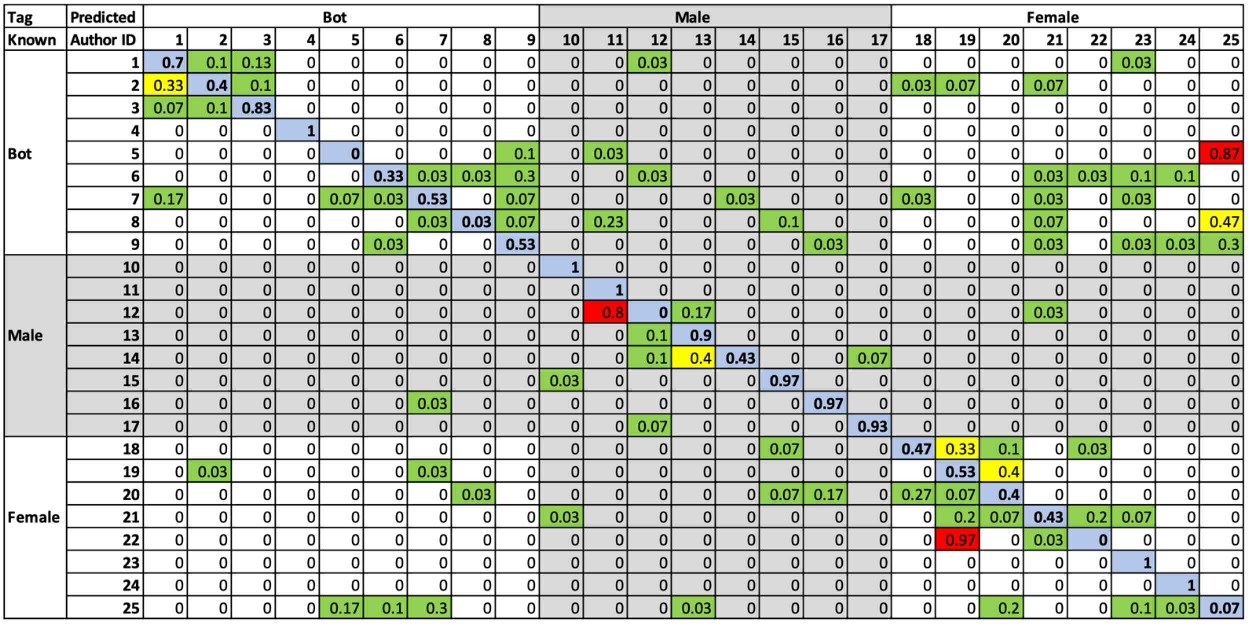
Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.37	0	0.07	0.03	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0.1	0	0	0.03	0	0	0
	3	0.57	0	0.43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.13	0.87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0.1	0	0	0.13	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.67
	6	0	0	0	0	0.03	0.47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0.47	0	0
	7	0	0	0	0	0	0	0.87	0.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0.07	0	0.03	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0	0.13	0	0.7
	9	0	0	0	0	0	0	0	0	0.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0
Male	10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0.73	0	0.07	0	0	0	0.07	0	0	0	0.13	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0.13	0.87	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0.93	0.03	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	25	0	0	0	0	0	0	0.13	0	0.07	0	0.27	0	0	0	0	0	0	0	0	0	0	0	0.53	0	0

Table 8.6. Confusion Matrix for ASO, Hybrid, ω = 0.7.



Tag	Predicted	Bot									Male								Female							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	1	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.1	**0**	0.23	0.13	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0.2	0	0.03	0	0	0	0
	3	0.6	0	**0.4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.27	**0.73**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	**0.03**	0.07	0	0.23	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0.53
	6	0	0.03	0	0	0	**0.17**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.77	0	0.03
	7	0	0	0	0	0	0	**0.53**	0.47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0.1	0	**0.17**	0	0	0.1	0	0	0	0	0	0	0	0	0	0.07	0	0.1	0	0.47
	9	0	0	0	0	0.03	0	0.03	0	**0.7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.23	0	0
Male	10	0	0	0	0	0	0	0	0	0	**0.97**	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0.57	**0**	0.23	0	0	0	0.03	0	0	0	0.17	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0.13	**0.87**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0.03	0	0	0	0.03	0	0	0	0	0	**0.9**	0.03	0	0	0	0	0	0
	19	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.97**	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	**0.93**	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.97	0	0	**0**	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
	24	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.93**	0.03
	25	0	0	0	0	0	0	0.13	0.03	0.07	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0.4	0	**0.03**

Table 8.7. Confusion Matrix for PSO, Hybrid, ω = 0.9.



Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	**0.7**	0.1	0.13	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0.03	0	0
	2	0.33	**0.4**	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.07	0	0.07	0	0	0	0
	3	0.07	0.1	**0.83**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	**0**	0	0	0	0.1	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0.87
	6	0	0	0	0	0	**0.33**	0.03	0.03	0.3	0	0	0.03	0	0	0	0	0	0	0	0	0.03	0.03	0.1	0.1	0
	7	0.17	0	0	0	0.07	0.03	**0.53**	0	0.07	0	0	0	0	0.03	0	0	0	0.03	0	0	0.03	0	0.03	0	0
	8	0	0	0	0	0	0	0.03	**0.03**	0.07	0	0.23	0	0	0	0.1	0	0	0	0	0	0.07	0	0	0	0.47
	9	0	0	0	0	0	0.03	0	0	**0.53**	0	0	0	0	0	0	0.03	0	0	0	0	0.03	0	0.03	0.03	0.3
Male	10	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0.8	**0**	0.17	0	0	0	0	0	0	0	0.03	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0.1	**0.9**	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0.1	0.4	**0.43**	0	0	0.07	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	**0.97**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	**0.97**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0	**0.93**	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	**0.47**	0.33	0.1	0	0.03	0	0	0
	19	0	0.03	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	**0.53**	0.4	0	0	0	0	0
	20	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0.07	0.17	0	0.27	0.07	**0.4**	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0.2	0.07	**0.43**	0.2	0.07	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0.03	**0**	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
	25	0	0	0	0	0.17	0.1	0.3	0	0	0	0	0	0.03	0	0	0	0	0	0	0.2	0	0	0.1	0.03	**0.07**

Table 8.8. Confusion Matrix for PSO, Hybrid, ω = 1.0.

Tag	Predicted	Bot									Male								Female							
Known	Author ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Bot	1	**0.73**	0.03	0.2	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.17	**0.23**	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0.33	0	0	0	0
	3	0.17	0	**0.83**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	**0**	0.13	0	0.03	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.77
	6	0	0	0	0	0.03	**0.3**	0.07	0.07	0.33	0	0.03	0	0	0	0	0	0	0	0.07	0	0	0	0	0.1	0
	7	0.07	0	0	0	0.03	0.03	**0.57**	0	0.07	0.1	0	0	0	0	0	0	0	0.07	0	0	0.03	0	0	0	0.03
	8	0	0	0	0	0	0	0.03	**0**	0.03	0	0.2	0	0	0	0	0	0	0	0	0	0.03	0	0.1	0	0.6
	9	0	0	0	0	0.17	0.03	0	0	**0.53**	0	0	0	0.03	0	0	0	0	0	0	0	0.03	0	0	0.03	0.17
Male	10	0	0	0	0	0	0	0	0	0	**0.9**	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0.03	0.63	**0.03**	0.17	0	0	0	0.03	0	0	0	0.1	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0.13	**0.87**	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0.13	0.3	**0.57**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	**0.93**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0	**0.97**	0	0	0	0	0	0	0	0
Female	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	**0.53**	0.33	0.03	0	0.07	0	0	0
	19	0	0.03	0	0	0	0.03	0	0	0	0	0	0	0	0	0.03	0	0	0.03	**0.43**	0.43	0	0	0	0	0
	20	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0.1	0.07	0	0.3	0.1	**0.4**	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0.13	0	**0.63**	0.17	0	0	0
	22	0	0	0	0.03	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0.1	0.83	0	0	**0**	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0
	25	0	0	0	0	0.17	0.03	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0.47	0	0	0.07	0	**0.07**

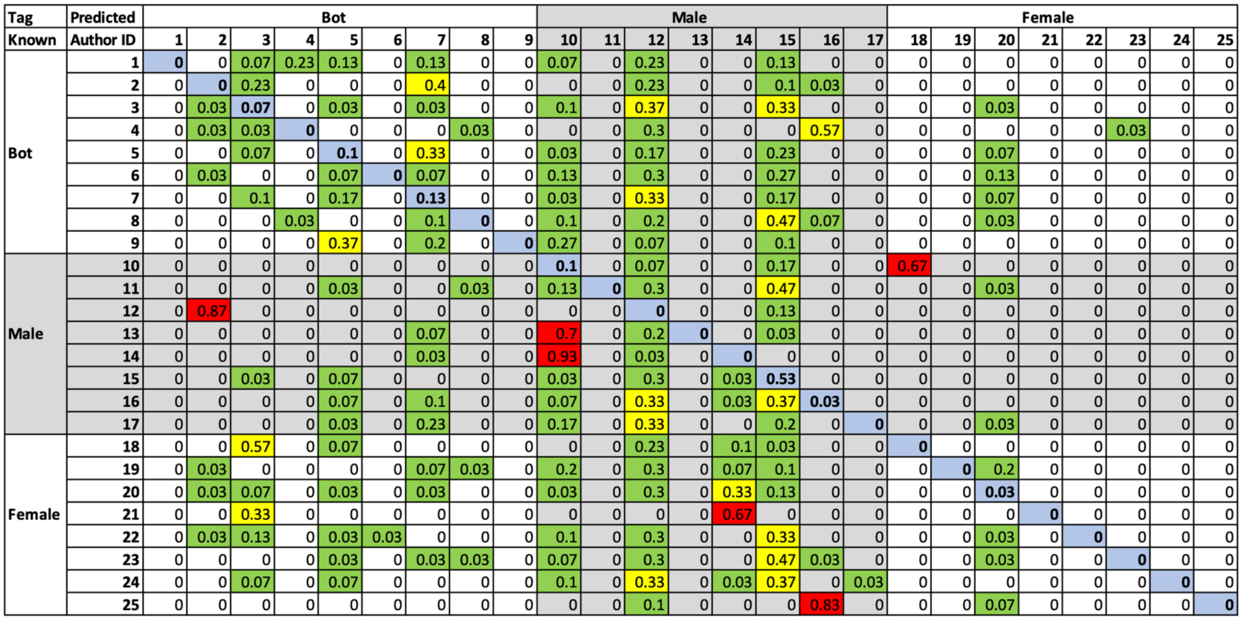
Table 8.9. Confusion Matrix for ASO with Adversarial Samples, Hybrid, $\omega = 0.0$.

Tag	**Predicted**	**Bot**									**Male**								**Female**							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0	0	0.23	0.03	0	0	0	0	0.03	0	0.63	0	0	0.07	0	0	0	0	0	0	0	0	0	0
	2	0	**0**	0.43	0	0	0	0.33	0	0	0	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0.1	**0.03**	0	0	0	0	0	0	0.03	0	0.7	0	0	0.13	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.07	**0**	0	0	0	0	0	0	0	0.5	0	0	0	0.43	0	0	0	0	0	0	0	0	0
	5	0	0	0.87	0	**0**	0	0.1	0	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0.03	0.03	0	0	**0**	0.07	0	0.2	0.07	0	0.37	0	0	0.07	0	0	0	0	0.17	0	0	0	0	0
	7	0	0	0.5	0	0.2	0	**0.07**	0	0	0	0	0.23	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0.07	0	0	0.2	**0**	0	0.1	0	0.23	0	0	0.17	0.2	0	0	0	0	0	0	0.03	0	0
	9	0	0	0	0	0.63	0	0.03	0	**0**	0.23	0	0.07	0	0	0	0	0	0	0	0	0	0	0.03	0	0
Male	**10**	0	0.03	0	0	0	0	0	0	0	**0.6**	0	0.03	0	0	0.1	0	0	0.2	0.03	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0.1	**0**	0.67	0	0	0.2	0.03	0	0	0	0	0	0	0	0	0
	12	0	1	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0.53	0	0.47	**0**	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	1	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0.53	0	0.07	**0.4**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0.03	0	0.73	0	0.07	0.07	**0.1**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0.03	0.03	0	0.73	0	0	0.07	0.07	**0.03**	0	0	0	0	0	0	0.03	0
Female	**18**	0	0	0.6	0	0	0	0	0	0	0	0	0.27	0	0	0.03	0	0	**0**	0	0	0	0.07	0	0	0.03
	19	0	0	0	0	0	0	0	0	0	0.13	0	0.47	0	0.03	0.03	0	0	0	**0**	0.33	0	0	0	0	0
	20	0	0	0.1	0	0.03	0	0	0	0	0	0	0.77	0	0.03	0.07	0	0	0	0	**0**	0	0	0	0	0
	21	0	0	0.53	0	0	0	0	0	0	0	0	0.03	0	0.4	0	0	0	0	0	0	**0**	0.03	0	0	0
	22	0	0	0.2	0	0.03	0	0	0	0	0.07	0	0.6	0	0	0.1	0	0	0	0	0	0	**0**	0	0	0
	23	0	0	0	0	0	0	0.03	0	0	0.07	0	0.7	0	0	0.13	0.07	0	0	0	0	0	0	**0**	0	0
	24	0	0	0.03	0	0	0	0	0	0	0.03	0	0.67	0	0	0.1	0	0.13	0	0	0	0	0	0	**0.03**	0
	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.97	0	0	0	0.03	0	0	0	0	**0**

Table 8.10. Confusion Matrix for ASO with Adversarial Samples, Hybrid, ω = 1.0.

Tag	**Predicted**	**Bot**									**Male**								**Female**							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0	0	0.03	0.07	0	0.03	0	0	0	0	0.77	0	0	0.07	0	0	0.03	0	0	0	0	0	0	0
	2	0	**0**	0.43	0	0	0	0.27	0	0	0	0	0.27	0	0	0.03	0	0	0	0	0	0	0	0	0	0
	3	0	0.03	**0.03**	0	0	0	0	0	0	0	0	0.77	0	0	0.17	0	0	0	0	0	0	0	0	0	0
	4	0	0	0.03	**0**	0	0	0	0	0	0	0	0.7	0	0	0	0.27	0	0	0	0	0	0	0	0	0
	5	0	0	0.57	0	**0.07**	0	0.1	0	0	0	0	0.17	0	0	0.07	0	0	0	0	0.03	0	0	0	0	0
	6	0	0	0.03	0	0.03	**0**	0	0	0.07	0.03	0	0.63	0	0	0.1	0	0	0	0	0.1	0	0	0	0	0
	7	0	0	0.2	0	0.23	0	**0**	0	0	0	0	0.53	0	0	0.03	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0.03	0	0	0.1	**0**	0	0.17	0	0.37	0	0	0.2	0.1	0	0	0	0	0	0	0.03	0	0
	9	0	0	0	0	0.57	0	0.03	0	**0**	0.2	0	0.17	0	0	0	0.03	0	0	0	0	0	0	0	0	0
Male	**10**	0	0	0	0	0	0	0	0	0	**0.37**	0	0.03	0	0	0	0	0	0.6	0	0	0	0	0	0	0
	11	0	0	0	0	0.03	0	0	0	0	0.03	**0**	0.73	0	0	0.2	0	0	0	0	0	0	0	0	0	0
	12	0	1	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0.37	0	0.6	**0**	0	0.03	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0.87	0	0.13	0	**0**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0.03	0	0	0	0	0	0	0	0	0.73	0	0	**0.23**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0.03	0	0.03	0	0	0	0	0	0	0.77	0	0.03	0.1	**0.03**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0.03	0	0	0	0	0.07	0	0.8	0	0	0.03	0.07	**0**	0	0	0	0	0	0	0	0
Female	**18**	0	0	0.37	0	0	0	0	0	0	0	0	0.6	0	0	0.03	0	0	**0**	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0.13	0	0.77	0	0	0.03	0	0	0	**0**	0.07	0	0	0	0	0
	20	0	0	0.07	0	0	0	0	0	0	0	0	0.73	0	0.07	0.1	0	0	0	0	**0.03**	0	0	0	0	0
	21	0	0	0.27	0	0	0	0	0	0	0	0	0	0	0.73	0	0	0	0	0	0	**0**	0	0	0	0
	22	0	0	0.07	0	0	0	0	0	0	0	0	0.77	0	0	0.13	0	0	0	0	0.03	0	**0**	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0.17	0	0	0	0	0.03	0	0	**0**	0	0
	24	0	0	0.03	0	0.03	0	0	0	0	0	0	0.73	0	0	0.13	0	0.07	0	0	0	0	0	0	**0**	0
	25	0	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0.93	0	0	0	0.03	0	0	0	0	**0**

Table 8.11. Confusion Matrix for ASO with Adversarial Samples, Hybrid, $\omega = 0.3$.



Tag	**Predicted**	**Bot**									**Male**								**Female**							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0	0.07	0.23	0.13	0	0.13	0	0	0.07	0	0.23	0	0	0.13	0	0	0	0	0	0	0	0	0	0
	2	0	**0**	0.23	0	0	0	0.4	0	0	0	0	0.23	0	0	0.1	0.03	0	0	0	0	0	0	0	0	0
	3	0	0.03	**0.07**	0	0.03	0	0.03	0	0	0.1	0	0.37	0	0	0.33	0	0	0	0	0.03	0	0	0	0	0
	4	0	0.03	0.03	**0**	0	0	0	0.03	0	0	0	0.3	0	0	0	0.57	0	0	0	0	0	0	0.03	0	0
	5	0	0	0.07	0	**0.1**	0	0.33	0	0	0.03	0	0.17	0	0	0.23	0	0	0	0	0.07	0	0	0	0	0
	6	0	0.03	0	0	0.07	**0**	0.07	0	0	0.13	0	0.3	0	0	0.27	0	0	0	0	0.13	0	0	0	0	0
	7	0	0	0.1	0	0.17	0	**0.13**	0	0	0.03	0	0.33	0	0	0.17	0	0	0	0	0.07	0	0	0	0	0
	8	0	0	0	0.03	0	0	0.1	**0**	0	0.1	0	0.2	0	0	0.47	0.07	0	0	0	0.03	0	0	0	0	0
	9	0	0	0	0	0.37	0	0.2	0	**0**	0.27	0	0.07	0	0	0.1	0	0	0	0	0	0	0	0	0	0
Male	**10**	0	0	0	0	0	0	0	0	0	**0.1**	0	0.07	0	0	0.17	0	0	0.67	0	0	0	0	0	0	0
	11	0	0	0	0	0.03	0	0	0.03	0	0.13	**0**	0.3	0	0	0.47	0	0	0	0	0.03	0	0	0	0	0
	12	0	0.87	0	0	0	0	0	0	0	0	0	**0**	0	0	0.13	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	0.07	0	0	0.7	0	0.2	**0**	0	0.03	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0.03	0	0	0.93	0	0.03	0	**0**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0.03	0	0.07	0	0	0	0	0.03	0	0.3	0	0.03	**0.53**	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0.07	0	0.1	0	0	0.07	0	0.33	0	0.03	0.37	**0.03**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0.03	0	0.23	0	0	0.17	0	0.33	0	0	0.2	0	**0**	0	0	0.03	0	0	0	0	0
Female	**18**	0	0	0.57	0	0.07	0	0	0	0	0	0	0.23	0	0.1	0.03	0	0	**0**	0	0	0	0	0	0	0
	19	0	0.03	0	0	0	0	0.07	0.03	0	0.2	0	0.3	0	0.07	0.1	0	0	0	**0**	0.2	0	0	0	0	0
	20	0	0.03	0.07	0	0.03	0	0.03	0	0	0.03	0	0.3	0	0.33	0.13	0	0	0	0	**0.03**	0	0	0	0	0
	21	0	0	0.33	0	0	0	0	0	0	0	0	0	0	0.67	0	0	0	0	0	0	**0**	0	0	0	0
	22	0	0.03	0.13	0	0.03	0.03	0	0	0	0.1	0	0.3	0	0	0.33	0	0	0	0	0.03	0	**0**	0	0	0
	23	0	0	0	0	0.03	0	0.03	0.03	0	0.07	0	0.3	0	0	0.47	0.03	0	0	0	0.03	0	0	**0**	0	0
	24	0	0	0.07	0	0.07	0	0	0	0	0.1	0	0.33	0	0.03	0.37	0	0.03	0	0	0	0	0	0	**0**	0
	25	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0.83	0	0	0	0.07	0	0	0	0	**0**

Table 8.12. Confusion Matrix for ASO with Adversarial Samples, Hybrid, ω = 0.5.

Tag	Predicted	Bot									Male								Female							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0	0.23	0.07	0.07	0	0.27	0	0.03	0	0	0.07	0	0	0	0	0	0.27	0	0	0	0	0	0	0
	2	0	**0**	0.13	0	0	0	0.4	0	0	0	0	0.03	0	0	0.07	0.3	0.03	0.03	0	0	0	0	0	0	0
	3	0	0.07	**0.33**	0	0.03	0	0.1	0	0.03	0	0	0.07	0	0	0.07	0.17	0	0.13	0	0	0	0	0	0	0
	4	0	0	0	**0**	0	0	0	0	0.03	0	0	0.03	0	0	0	0.93	0	0	0	0	0	0	0	0	0
	5	0	0	0.23	0	**0.17**	0	0.33	0	0	0	0	0.03	0	0	0.17	0.07	0	0	0	0	0	0	0	0	0
	6	0	0	0.03	0	0.1	**0.03**	0.07	0	0.1	0.03	0	0.07	0	0	0.13	0.4	0	0	0	0.03	0	0	0	0	0
	7	0	0	0.23	0	0.23	0.03	**0.07**	0	0	0	0	0.07	0	0	0.2	0.13	0	0	0	0.03	0	0	0	0	0
	8	0	0	0	0	0	0.03	0	**0**	0.03	0	0	0.03	0	0.07	0.4	0.43	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0.13	0	0.23	0	**0.03**	0.03	0	0.07	0	0	0	0.5	0	0	0	0	0	0	0	0	0
Male	**10**	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	11	0	0	0	0	0.07	0.03	0	0	0.03	0	**0**	0.07	0	0	0.43	0.33	0	0.03	0	0	0	0	0	0	0
	12	0	0.6	0	0	0	0	0	0	0	0	0	**0**	0	0	0.33	0.07	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0.07	0	0.33	0	0.03	0.3	0	0.07	**0**	0	0.17	0	0	0.03	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0.4	0	0	0.57	0	0.03	0	**0**	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0.1	0	0.03	0	0.03	0	0	0.07	0	0.03	**0.63**	0.1	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0.1	0.03	0.03	0	0.03	0.03	0	0.07	0	0.03	0.27	**0.4**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0.03	0	0.07	0	0.03	0	0	0.07	0	0	0.2	0.6	**0**	0	0	0	0	0	0	0	0
Female	**18**	0	0	0.5	0	0.1	0.03	0	0	0	0	0	0.07	0	0.1	0	0.2	0	**0**	0	0	0	0	0	0	0
	19	0	0	0	0	0	0.03	0.03	0	0.03	0.03	0	0.07	0	0.5	0.1	0.1	0	0	**0**	0.1	0	0	0	0	0
	20	0	0	0.03	0	0.07	0.03	0	0	0	0	0	0.07	0	0.43	0.1	0.23	0	0	0	**0.03**	0	0	0	0	0
	21	0	0	0.37	0	0	0.07	0	0	0	0	0	0	0	0.57	0	0	0	0	0	0	**0**	0	0	0	0
	22	0	0	0.17	0	0.03	0.03	0	0	0.03	0	0	0.07	0	0.03	0.33	0.27	0	0.03	0	0	0	**0**	0	0	0
	23	0	0	0	0	0.07	0.03	0	0	0.03	0	0	0.07	0	0	0.43	0.37	0	0	0	0	0	0	**0**	0	0
	24	0	0	0.27	0	0.17	0.03	0	0	0.03	0	0	0.07	0	0	0.17	0.27	0	0	0	0	0	0	0	**0**	0
	25	0	0	0	0	0	0	0	0	0.03	0	0	0	0	0	0	0.97	0	0	0	0	0	0	0	0	**0**

Table 8.13. Confusion Matrix for ASO with Adversarial Samples, Hybrid, ω = 0.7.

Tag	**Predicted**	**Bot**									**Male**								**Female**							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0	0.1	0.3	0.07	0	0.27	0	0	0	0	0.07	0	0	0	0.17	0	0.03	0	0	0	0	0	0	0
	2	0	**0**	0.07	0.03	0	0	0.07	0	0	0	0	0.1	0	0	0.03	0.5	0.17	0	0	0	0	0	0	0.03	0
	3	0	0.03	**0.2**	0.03	0.07	0.03	0.03	0	0	0.03	0	0.1	0	0	0.03	0.37	0	0.07	0	0	0	0	0	0	0
	4	0	0	0.07	**0**	0.03	0	0	0	0	0	0	0.1	0	0	0.03	0.73	0	0	0.03	0	0	0	0	0	0
	5	0	0	0.13	0.03	**0.3**	0.1	0.07	0	0	0	0	0	0	0	0.07	0.27	0	0	0.03	0	0	0	0	0	0
	6	0	0	0.03	0.03	0.2	**0.1**	0.07	0	0	0	0	0.03	0	0	0.03	0.4	0	0	0	0.1	0	0	0	0	0
	7	0	0	0.07	0.03	0.2	0.23	**0.03**	0	0	0	0	0.03	0	0	0.07	0.3	0	0	0	0.03	0	0	0	0	0
	8	0	0	0	0.03	0.03	0.1	0	**0**	0	0	0	0.03	0	0	0.13	0.63	0	0	0.03	0	0	0	0	0	0
	9	0	0	0	0.03	0.37	0	0.13	0	**0**	0	0	0	0	0	0.03	0.43	0	0	0	0	0	0	0	0	0
Male	**10**	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0.9	0.1	0	0	0	0	0	0
	11	0	0	0	0	0.13	0.17	0.03	0	0	0.03	**0**	0.03	0	0	0.13	0.47	0	0	0	0	0	0	0	0	0
	12	0	0.7	0	0	0	0.03	0.03	0	0	0	0	**0**	0	0.03	0.2	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0.13	0	0.3	0	0	0.27	0	0	**0**	0	0.1	0.2	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0.03	0	0.5	0	0	0.43	0	0	0	**0**	0	0.03	0	0	0	0	0	0	0	0	0
	15	0	0	0.07	0.03	0.27	0.13	0.03	0	0	0	0	0.03	0	0.03	**0.2**	0.2	0	0	0	0	0	0	0	0	0
	16	0	0	0.03	0	0.17	0.17	0.07	0	0	0.03	0	0.03	0	0.03	0.03	**0.43**	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0.07	0.03	0.1	0	0	0.03	0	0.07	0	0.03	0	0.63	**0.03**	0	0	0	0	0	0	0	0
Female	**18**	0	0	0.27	0.03	0.27	0.07	0	0	0	0	0	0	0	0.03	0	0.33	0	**0**	0	0	0	0	0	0	0
	19	0	0	0	0.03	0.07	0.3	0	0	0	0.03	0	0.03	0	0.2	0.03	0.27	0	0	**0**	0.03	0	0	0	0	0
	20	0	0	0.03	0	0.13	0.2	0.03	0	0	0	0	0	0	0.1	0.07	0.43	0	0	0	**0**	0	0	0	0	0
	21	0	0	0.33	0	0	0.07	0	0	0	0	0	0	0	0.2	0	0.13	0	0	0	0	**0**	0.27	0	0	0
	22	0	0	0.07	0.03	0.1	0.3	0	0	0	0	0	0.03	0	0.03	0.07	0.37	0	0	0	0	0	**0**	0	0	0
	23	0	0	0	0	0.1	0.13	0.07	0	0	0.03	0	0.03	0	0	0.13	0.5	0	0	0	0	0	0	**0**	0	0
	24	0	0	0.13	0.03	0.3	0.1	0	0	0	0	0	0.1	0	0	0	0.3	0	0	0	0	0	0	0	**0.03**	0
	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**0**

Table 8.14. Confusion Matrix for PSO with Adversarial Samples, Hybrid, ω = 0.9.

Tag	Predicted	Bot									Male								Female							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0**	0.1	0.03	0.27	0	0	0	0.03	0	0.07	0	0.13	0	0.1	0.07	0.03	0	0.03	0	0	0.03	0	0.1	0	0
	2	0	**0.03**	0.23	0	0	0	0.07	0	0	0.17	0	0.03	0	0.1	0	0.07	0.07	0.1	0	0.03	0.03	0	0.07	0	0
	3	0	0.3	**0.13**	0	0	0	0	0.03	0	0.13	0	0.03	0	0.03	0.03	0.03	0.13	0	0	0.03	0.03	0.03	0.03	0	0
	4	0	0.03	0.03	**0**	0	0	0.03	0	0	0.1	0	0.13	0	0	0.1	0.1	0.03	0	0.13	0	0.13	0.03	0.1	0.03	0
	5	0	0	0.13	0	**0**	0	0.03	0	0	0.13	0.03	0	0	0.03	0.07	0.13	0.07	0.03	0	0.07	0	0.03	0.23	0	0
	6	0	0.27	0.03	0	0	**0**	0.1	0.03	0	0.2	0	0.03	0	0.03	0.03	0.03	0	0	0	0.1	0	0.03	0.1	0	0
	7	0	0	0.07	0	0	0	**0.03**	0	0	0.1	0	0.2	0.03	0.1	0.07	0.07	0	0	0	0.1	0.1	0.03	0.1	0	0
	8	0	0.1	0	0	0	0	0.03	**0**	0	0.17	0	0	0	0.03	0.1	0.17	0	0	0	0	0	0.03	0.37	0	0
	9	0	0	0	0	0	0	0	0	**0**	0.33	0.03	0	0.03	0.03	0	0.37	0.07	0	0	0	0	0	0.13	0	0
Male	**10**	0	0	0	0	0.07	0	0	0	0	**0.2**	0	0	0.03	0.03	0.07	0.1	0	0.1	0.27	0	0.03	0.03	0.07	0	0
	11	0	0.07	0	0	0	0	0	0	0	0.23	**0**	0	0.03	0.03	0.1	0.17	0	0	0	0	0.03	0.03	0.3	0	0
	12	0	0.37	0.03	0	0	0.03	0	0	0	0.07	0	**0.03**	0	0.03	0.27	0.03	0	0	0	0	0	0.03	0.1	0	0
	13	0	0	0	0	0	0	0	0.03	0	0.57	0	0.03	**0**	0.03	0	0	0.03	0	0	0	0	0	0.3	0	0
	14	0.03	0	0	0	0	0	0.03	0.03	0	0.73	0	0.03	0	**0.03**	0	0.03	0	0	0	0	0	0	0.07	0	0
	15	0	0.03	0.03	0	0.03	0.07	0	0	0	0.2	0	0.03	0	0.07	**0.27**	0.07	0	0.03	0	0	0	0.03	0.13	0	0
	16	0	0	0	0	0	0.03	0.07	0	0	0.17	0	0.07	0	0.03	0.03	**0.17**	0	0	0	0	0.2	0.07	0.17	0	0
	17	0	0	0	0	0.03	0	0	0.03	0	0.13	0	0.2	0	0.07	0.07	0.1	**0.1**	0.03	0	0.07	0.07	0	0.1	0	0
Female	**18**	0	0.03	0.27	0	0	0	0	0	0	0.07	0	0.03	0	0.1	0.13	0.17	0	**0**	0	0	0	0.03	0.17	0	0
	19	0.13	0	0	0	0	0	0	0	0	0.33	0	0	0.07	0.13	0.03	0	0	0	**0**	0.07	0	0.03	0.2	0	0
	20	0	0	0.07	0	0	0	0.07	0	0	0.2	0	0.07	0.03	0.13	0.03	0.07	0	0.03	0	**0.03**	0	0.07	0.2	0	0
	21	0	0	0.1	0	0	0	0.07	0	0	0.17	0	0.03	0	0.17	0.03	0.2	0.03	0	0	0.07	**0**	0.03	0.1	0	0
	22	0.07	0	0.07	0	0	0.03	0	0.07	0	0.17	0	0.03	0.03	0.03	0.1	0.17	0	0.03	0	0	0	**0**	0.2	0	0
	23	0	0.03	0	0	0	0	0	0.07	0	0.1	0	0.1	0	0.03	0.03	0.2	0	0	0	0.3	0.03	0	**0.1**	0	0
	24	0	0	0.03	0	0	0.03	0	0	0.03	0.17	0	0.07	0	0.03	0.1	0.03	0.37	0.03	0	0	0	0.03	0.07	**0**	0
	25	0	0	0	0	0	0	0.07	0.03	0	0.13	0.03	0	0	0	0.03	0.57	0	0	0	0	0	0	0.13	0	**0**

Table 8.15. Confusion Matrix for PSO with Adversarial Samples, Hybrid, $\omega = 1.0$.

Tag	Predicted	Bot									Male								Female							
Known	**Author ID**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
Bot	**1**	**0.03**	0.03	0	0.2	0	0	0.03	0	0	0.03	0.03	0.3	0	0	0.03	0.03	0.03	0.03	0.03	0.03	0	0	0.13	0	0
	2	0.03	**0**	0.13	0	0.03	0	0.1	0	0	0.2	0.03	0.2	0	0	0.03	0	0.07	0.03	0	0	0.03	0	0.1	0	0
	3	0.03	0.33	**0.03**	0	0	0	0	0	0	0.17	0.03	0.17	0	0	0	0	0.07	0	0	0	0.07	0	0.1	0	0
	4	0.03	0.07	0	**0**	0	0	0	0.03	0	0.13	0	0.3	0	0	0	0.07	0.07	0.03	0.13	0	0	0	0.13	0	0
	5	0	0	0.07	0	**0**	0.03	0	0.03	0	0.17	0	0.2	0	0	0.03	0.03	0.03	0.03	0.07	0.03	0	0	0.27	0	0
	6	0	0.27	0	0	0	**0**	0	0	0	0.1	0	0.2	0	0	0.03	0.1	0.03	0.03	0	0.13	0	0	0.1	0	0
	7	0	0	0.1	0	0.03	0	**0**	0	0	0.1	0	0.37	0	0	0.03	0	0.07	0.03	0.03	0.07	0	0.03	0.13	0	0
	8	0	0	0	0	0	0	0	**0**	0	0.17	0	0.13	0	0	0.03	0.17	0.03	0.03	0	0	0	0	0.43	0	0
	9	0	0	0	0	0.03	0	0	0	**0.03**	0.43	0	0.13	0	0	0	0.13	0.07	0	0	0	0	0	0.17	0	0
Male	**10**	0	0	0	0	0.1	0	0	0	0	**0.13**	0	0.1	0	0	0	0	0.03	0.13	0.43	0	0	0	0.07	0	0
	11	0.03	0.03	0	0	0.03	0	0	0.03	0	0.23	**0.03**	0.17	0	0	0.13	0.07	0.03	0	0	0	0	0	0.2	0	0
	12	0	0.4	0	0	0	0	0	0.03	0	0.1	0.03	**0.17**	0	0	0.13	0	0.03	0	0	0	0	0	0.1	0	0
	13	0	0	0	0	0	0	0	0.03	0	0.5	0	0.2	**0**	0	0.03	0.07	0.03	0	0	0	0	0	0.13	0	0
	14	0	0	0	0	0	0	0	0	0	0.63	0	0.2	0	**0**	0.03	0.03	0.03	0	0.03	0	0	0	0.03	0	0
	15	0	0	0	0	0.03	0	0	0.03	0	0.3	0	0.2	0	0.03	**0.17**	0.07	0.07	0.03	0	0	0	0	0.07	0	0
	16	0	0	0	0	0.03	0	0.03	0	0	0.13	0.03	0.23	0	0	0.03	**0.13**	0.03	0.03	0.03	0	0.1	0.03	0.13	0	0
	17	0	0	0	0	0.03	0	0	0.03	0	0.17	0	0.4	0	0	0.07	0	**0.13**	0.03	0	0.03	0	0	0.07	0.03	0
Female	**18**	0	0	0.17	0	0.03	0	0	0	0	0.17	0	0.13	0	0.03	0	0.1	0.07	**0.03**	0.03	0	0	0.1	0.13	0	0
	19	0	0	0	0	0	0	0	0.03	0	0.27	0	0.23	0	0.07	0.03	0.03	0.07	0.03	**0**	0.07	0	0	0.17	0	0
	20	0	0	0	0	0.1	0	0	0.07	0	0.2	0.03	0.2	0	0.07	0	0.03	0.07	0.03	0.03	**0.03**	0	0	0.13	0	0
	21	0	0	0.1	0	0	0.03	0	0	0	0.1	0.03	0.07	0	0.17	0.07	0.23	0.03	0.03	0	0	**0**	0.03	0.1	0	0
	22	0	0.03	0	0	0	0	0.03	0.13	0	0.1	0.03	0.23	0.03	0.03	0	0.03	0.03	0.03	0	0	0	**0**	0.27	0	0
	23	0	0	0	0	0	0	0	0	0	0.2	0	0.2	0	0	0.07	0.17	0.07	0.03	0	0.13	0	0	**0.13**	0	0
	24	0	0	0	0	0	0	0	0	0	0.23	0	0.13	0	0	0.03	0	0.47	0.03	0	0	0	0	0.07	**0.03**	0
	25	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0.03	0.6	0.03	0	0	0	0	0	0.13	0	**0**

8.4 Summary

In this chapter, we show that feature selection improves the accuracy of author identification. However, we also show evidence that feature selection renders author identification systems susceptible to adversarial attacks. Further, with respect to these two previous assertions, this work is consistent with the findings of chapters 6 and 7, but on different datasets. Also consistent with chapter 7, the feature set appears to be a dominate factor in predicting this phenomenon.

This work is far from definitive! While the evidence mounts as to the causes and effects of adversarial attacks on AIdSs using feature selection, still it is not precisely clear how to predict the vulnerability. We call for future research to investigate additional feature sets and datasets to help illuminate these issues.

Chapter 9

Conclusion & Future Directions

The arms race between author identification and adversarial authorship continues to escalate. This work examined an aspect of this arms race, namely the impact of EC-based feature selection on adversarial authorship. Specifically, this research first compared various EC-based feature selection algorithms, including GEFeS, PSO, ASO, ABCO and GSO. The results indicate that while all these feature selection algorithms perform well, GEFeS maintains a slight edge. Next, this research investigated the susceptibility of EC-based feature selection to adversarial authorship attacks, which verified that for at least some feature sets and feature selection combinations, feature selection presents a vulnerability. To attempt to better understand this vulnerability, this research slightly increased the number of authors and adjusted the feature sets, which continued to demonstrate the vulnerability. The research then further explored this issue by increasing the number of authors again. According to the results, as the number of authors increase, author identification generally becomes less accurate. Finally, this research explored the effects of varying the dataset from a blog dataset to a twitter dataset. This dataset change shows evidence that the feature set and dataset match-up is important, as some feature sets appear to work better for certain datasets.

To recap the significant events in the author arms race, authors created anonymous text. Then, identification researchers extracted stylometric features and learned to match them to authors. Anonymous authors retaliated by finding ways to obfuscate these stylometric features.

Identification researchers responded by exploring feature sets beyond stylometry, to which anonymous authors looked for workarounds. As the number of features used by researchers has grown, it has become important to perform feature selection, not only to reduce computational costs, but also to improve identification accuracy. Finally, adversarial attacks exploit feature selection.

The top-of-mind question for researchers in this area must be how will identification researchers respond to the current escalation? With companies, such as OpenAI, creating massively deep neural networks such as GPT-3 [60], perhaps the next step will be a matter of overpowering computational horsepower for both author identification and author obfuscation. While such massive computation may mean researchers no longer need to consider the computational impact of adding more features to a problem, this does not address the demonstrated fact that feature selection can improve accuracy. However, as shown in this work, feature selection also creates a vulnerability *as* it improves accuracy. So, massive computation does not make this research moot. In fact, it may become even more important as researchers explore the vulnerabilities of these massive artificial intelligence systems. Certainly, one of the impacts of these massive systems is that the table-stakes for research are escalating along with the arms race.

While this research has made great strides in demonstrating and beginning to explain the vulnerability resulting from feature selection, we cannot claim that this research is comprehensive! Based on this research, it does not appear likely that there will be a silver-bullet feature set that is immune to adversarial attacks. However, it may be possible to further understand the interplay between datasets and feature sets, with the intent being to create metametrics that will predict the effects of this interplay. More results from additional datasets, feature sets and possible feature set selection algorithms are warranted.

REFERENCES

[1] Liu, Y., Passino, K., Swarm Intelligence: Literature Overview, Department of electrical engineering, the Ohio State University. http://www.eleceng.ohio-state.edu/~passino/swarms.pdf

[2] Zhu, Y., Tang, X. (2010). Overview of swarm intelligence, *2010 International Conference on Computer Application and System Modeling*, 9, 400-403. https://doi.org/10.1109/ICCASM.2010.5623005

[3] Krause, J., Graeme, D., Krause, S. (2010). Swarm Intelligence in Animals and Humans", *Trends in Ecology and Evolution*, 25(1), 28-34. https://doi.org/10.1016/j.tree.2009.06.016

[4] Zhao, Y., Zobel, J. (2007). Searching With Style: Authorship Attribution in Classic Literature, *Proceedings of the Thirtieth Australasian Conference on Computer Science*, 62, 59-68. https://dl.acm.org/doi/10.5555/1273749.1273757

[5] Brocardo, M., Traore, I., Saad, S., Woungang, I. (2013). Authorship verification for short messages using stylometry, *2013 International Conference on Computer, Information and Telecommunication Systems*, 1-6. https://doi.org/10.1109/CITS.2013.6705711

[6] Keselj, V., Peng, F., Cercone, N., Thomas, C. (2003). N-Gram-Based Author Profiles for Authorship Attribution, *Proceedings of the Conference Pacific Association for Computational Linguistics*, 255–264. http://www.cs.dal.ca/~vlado/papers/pacling03.pdf

[7] Boyd R. (2018). Mental profile mapping: A psychological single-candidate authorship attribution method. *PloS one*, *13*(7), e0200588. https://doi.org/10.1371/journal.pone.0200588

[8] Alford, A., Popplewell, K., Dozier, G., Bryant, K., Kelly, J., Adams, J., Abegaz, T., Shelton, J. (2011). GEFeWS: A Hybrid Genetic-Based Feature Weighting and Selection Algorithm for Multi-Biometric Recognition, *Proceedings of The 22nd Midwest Artificial Intelligence and Cognitive Science Conference 2011*, 710. 86-90. https://www.researchgate.net/publication/220833248_GEFeWS_A_Hybrid_Genetic-Based_Feature_Weighting_and_Selection_Algorithm_for_Multi-Biometric_Recognition

[9] Sari, Y., Stevenson, M., Vlachos, A. (2018). Topic or style? exploring the most useful features for authorship attribution, *Proceedings of the 27th International Conference on Computational Linguistics*, 343–353. https://aclanthology.org/C18-1029

[10] Stamatatos, E. (2009). A survey of modern authorship attribution methods, *Journal of the American Society for Information Science and Technology*, 60, 538-556. https://doi.org/10.1002/asi.21001

[11] Srinivas, M., Patnaik, L. (1994). Genetic algorithms: a survey, *Computer*, 27(6), 17-26. https://doi.org/10.1109/2.294849

[12] Wang, L., and Geng, X., (2009). *Behavioral Biometrics for Human Identification: Intelligent Applications*, Hershey, New York: Medical Information Science Reference.

[13] Nguyen, B., Xue, B., Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining, *Swarm and Evolutionary Computating*, 54, 100663. https://doi.org/10.1016/j.swevo.2020.100663

[14] Gaston, J., Narayanan, M., Dozier, G., Cothran, D., Arms-Chavez, C., Rossi, M., King, M., Xu, J. (2018). Authorship Attribution vs. Adversarial Authorship from a LIWC and Sentiment Analysis Perspective, *2018 IEEE Symposium Series on Computational Intelligence*, 920-927. https://doi.org/10.1109/SSCI.2018.8628769

[15] Reynolds, C. (1987). Flocks, herds and schools: a distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4), 25-34. https://doi.org/10.1145/37402.37406

[16] Clerc, M. (1999). The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. *Proceedings of the 1999 Congress on Evolutionary Computation*, 3, 1951-1957. https://doi.org/10.1109/CEC.1999.785513

[17] Parpinelli, R., Lopes, H. (2011). New inspirations in swarm intelligence: A survey, International Journal of Bio-Inspired Computation, 3(1), 1-16. https://doi.org/10.1504/IJBIC.2011.038700

[18] Narayanan, M., Gaston, J., Dozier, G., Cothran, L., Arms-Chavez, C., Rossi, M., King, M., Bryant, K. (2018). Adversarial Authorship, Sentiment Analysis, and the AuthorWeb Zoo, *2018 IEEE Symposium Series on Computational Intelligence*, 928-932. https://doi.org/10.1109/SSCI.2018.8628806

[19] Gaston, J., Narayanan, M. Dozier, G., Cothran, D., Arms-Chavez, C., Rossi, M., King, M., Xu, J. (2018). Authorship Attribution via Evolutionary Hybridization of Sentiment Analysis, LIWC, and Topic Modeling Features, *2018 IEEE Symposium Series on Computational Intelligence*, 933-940. https://doi.org/10.1109/SSCI.2018.8628647

[20] Pennebaker, J., Boyd, R., Jordan, K., Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin. https://doi.org/10.15781/T29G6Z

[21] Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S. (2005). OpinionFinder: A System for Subjectivity Analysis, *'05 Proceedings of HLT/EMNLP on Interactive Demonstrations*, 34-35. https://doi.org/10.3115/1225733.1225751

[22] Wiebe, J. Wilson, T., and Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language, *Language Resourses and Evaluation*, 39, 165-210. https://doi.org/10.1007/s10579-005-7880-9

[23] Riloff, E., Wiebe, J., Wilson, T., (2003). Learning Subjective Nouns Using Extraction Pattern Bootstrapping, *Seventh Conference on Natural Language Learning* 4, 25-32. https://dl.acm.org/doi/10.3115/1119176.1119180

[24] Riloff, E., Wiebe, J. (2003). Learning Extraction Patterns for Subjective Expressions, *Conference on Empirical Methods in Natural Language Processing*, 105-112. https://aclanthology.org/W03-1014

[25] Wilson, T., Wiebe, J., Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis, *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing*, 347-354. https://aclanthology.org/H05-1044

[26] Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. https://doi.org/10.1145/2133806.2133826

[27] McCallum, A. (2002). MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu. 2002. http://www.cs.umass.edu/~mccallum/mallet

[28] Xue, B., Zhang, M., Browne, W., Yao, X. (2016). A Survey on Evolutionary Computation Approaches to Feature Selection, *IEEE Transactions on Evolutionary Computation*, 20(4), 606-626. https://doi.org/10.1109/TEVC.2015.2504420

[29] Beni, G., Wang, J. (1993). Swarm Intelligence in Cellular Robotic Systems. *Proceedings NATO Advanced Workshop on Robots and Biological Systems,* 703–712. https://doi.org/10.1007/978-3-642-58069-7_38

[30] Dozier, G., Purrington, K., Popplewell, K., Shelton, J., Bryant, K., Adams, J., Woodard, D., Miller, P. (2011). GEFeS: Genetic & Evolutionary Feature Selection for Periocular Biometric Recognition, *Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, 4(3), 220-245. https://doi.org/10.1504/IJBM.2012.047642

[31] Kennedy, J. Eberhart, R. (1995). Particle swarm optimization, *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942-1948. https://ieeexplore.ieee.org/document/488968

[32] Carlisle, A. Dozier, G. (2001). An off-the-shelf PSO, *Proceedings of the Workshop on Particle Swarm Optimization*. https://www.researchgate.net/publication/216300408_An_off-the-shelf_PSO

[33] Karaboga, D. (2005). An Idea Based on Honey Bee Swarm for Numerical Optimization, Technical Report - TR06, *Technical Report, Erciyes University*. http://lia.disi.unibo.it/Courses/SistInt/articoli/bee-colony1.pdf

[34] Dorigo, M., Maniezzo, V., Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 26(1), 29-41. https://doi.org/10.1109/3477.484436

[35] Krishnanand, K., Ghose, D. (2005). Detection of multiple source locations using a glowworm metaphor with applications to collective robotics, *Proceedings 2005 IEEE Swarm Intelligence Symposium*, 84-91. https://doi.org/10.1109/SIS.2005.1501606

[36] Faust, C., Dozier, G., Xu, J., King, M. (2017). Adversarial Authorship, Interactive Evolutionary Hill-Climbing, and AuthorCAAT-III, *2017 IEEE Symposium Series on*

Computational Intelligence, pp. 1-8. https://www.eng.auburn.edu/~doziegv/dozier_Auburn_05232019.pdf

[37] Mack, N., Bowers, J., Williams, H., Dozier, G., Shelton, J. (2015). The Best Way to a Strong Defense is a Strong Offense: Mitigating Deanonymization Attacks via Iterative Language Translation, *International Journal of Machine Learning and Computing*, 5, 409-413. https://doi.org/10.7763/IJMLC.2015.V5.543

[38] Baltrušaitis, T., Ahuja, C., Morency, L. (2019). Multimodal Machine Learning: A Survey and Taxonomy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443. https://doi.org/10.1109/TPAMI.2018.2798607

[39] Snoek, C., Worring, M., Smeulders, A. (2005). Early versus late fusion in semantic video analysis, *Proceedings of the 13th annual ACM international conference on Multimedia*, 399-402. https://doi.org/10.1145/1101149.1101236

[40] Brennan, M., Afroz, S., and Greenstadt, R. (2011). Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity, *ACM Transactions on Information and System Security*, 15(3). https://doi.org/10.1145/2382448.2382450

[41] Brennan, M., Greenstadt, R. (2009). Practical attacks against authorship recognition techniques, *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence*. https://aaai.org/ocs/index.php/IAAI/IAAI09/paper/view/257

[42] Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., Woodard, D. (2017). Surveying Stylometry Techniques and Applications. *ACM Computuing Surveys*, 50(6). https://doi.org/10.1145/3132039

[43] Allred, J., Packer, S., Dozier, G., Aykent, S., Richardson, A., King, M. (2020). Towards a Human-AI Hybrid for Adversarial Authorship, *2020 SoutheastCon,* 1-8. https://doi.org/10.1109/SoutheastCon44009.2020.9249682

[44] Xue, B., Zhang, M., Browne, W. (2015). A Comprehensive Comparison on Evolutionary Feature Selection Approaches to Classification, *International Journal of Computational Intelligence and* Applications, 14(2), 1550008. https://doi.org/10.1142/S146902681550008X

[45] Zhang, Y., Gong, D., Cheng, J. (2017). Multi-Objective Particle Swarm Optimization Approach for Cost-Based Feature Selection in Classification, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 14(1), 64-75. https://doi.org/10.1109/TCBB.2015.2476796.

[46] Halladay, S. Dozier, G. (2020). A Comparison of Genetic & Swarm Intelligence-Based Feature Selection Algorithms for Author Identification, *2020 IEEE Symposium Series on Computational Intelligence*, 1731-1738. https://doi.org/10.1109/SSCI47803.2020.9308343

[47] Richardson, A., Dozier, G., King, M, Chapman, R. (2020). 'Uh-oh Spaghetti-oh': When Successful Genetic and Evolutionary Feature Selection Makes You More Susceptible to Adversarial Authorship Attacks, *2020 IEEE International Conference on Systems, Man, and Cybernetics*, 567-571, https://doi.org/10.1109/SMC42975.2020.9283352

[48] Mendenhall, T. (1887). The Characteristic Curves of Composition, *Science*, ns-9(214S), 237–246. https://doi.org/10.1126/science.ns-9.214S.237

[49] Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques, *Literary and Linguistic Computing*, 22(3), 251-270. http://dx.doi.org/10.1093/llc/fqm020

[50] Wanner, L., Soler, J. (2017). On the Relevance of Syntactic and Discourse Features for Author Profiling and Identification, *Proceedings of the 15th Conference of the European*

Chapter of the Association for Computational Linguistics, 2. https://aclanthology.org/E17-2108

[51] Hernández-Castañeda, Á., Calvo, H. (2017). Author Verification Using a Semantic Space Model, *Computacion y Sistemas*, 21, 167-179. https://doi.org/10.13053/CyS-21-2-2732

[52] Schlapbach, A., Kilchherr, V., Bunke, H. (2005). Improving writer identification by means of feature selection and extraction, *Eighth International Conference on Document Analysis and Recognition*, 1, 131-135. https://doi.org/10.1109/ICDAR.2005.139

[53] Chandrashekar, G., Sahin F. (2014). A Survey on feature selection methods, *Computors & Electrical Engeering*, 40(1), 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[54] Back, T., Hammel, U., Schwefel, H. (1997). Evolutionary computation: comments on the history and current state, *IEEE Transactions on Evolutionary Computation*, 1(1), 3-17. https://doi.org/10.1109/4235.585888

[55] Mavrovouniotis, M., Li, C., Yang, S. (2017). A survey of swarm intelligence for dynamic optimization: algorithms and applications, *Swarm Evolutionary Computation*, 33, 1-17. http://dx.doi.org/10.1016/j.swevo.2016.12.005

[56] de la Iglesia, B. (2013). Evolutionary computation for feature selection in classification problems. *WIREs Data Mining Knowledge Discovery*, 3(6) 381-407. https://doi.org/10.1002/widm.1106

[57] Rangel F., Celli F., Rosso P., Potthast M., Stein B., Daelemans W. (2015). Overview of the 3rd Author Profiling Task at PAN 2015. *CLEF*, 1391. http://www.sensei-conversation.eu/wp-content/uploads/2015/09/15-pan@clef.pdf

[58] Rostami, M., Beramand, K., Forouzandeh, S. (2020). Review of swarm intelligence-based feature selection methods, *Engineering Applications of Artificial Intelligence*, 100, 104210. https://doi.org/10.1016/j.engappai.2021.104210

[59] Halladay, S. Dozier, G. (2021) The Good, the Bad and the Ugly of Using Genetic-Based Feature Selection, *Submitted to IEEE SoutheastCon*.

[60] Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, *27*(1), 113-118. https://doi.org/10.1017/S1351324920000601

[61] Singhi, S., Lui, H. (2006). Feature subset selection bias for classification learning, *Proceedings of the 23rd international conference on machine learning*, 849-856. https://doi.org/10.1145/1143844.1143951

[62] Sapkota, U., Bethard, S., Montes, M., Solorio, T. (2015). Not All Character N-grams Are Created Equal in Authorship Attribution. *Proceedings of the 2015 Conference of the North Amaerican Chapter of the Association for Computer Linguistics: Human Language Technologies*, pages 93-102. https://doi.org/10.3115/v1/N15-1010

[63] Zheng, R., Li, J., Chen, H., Huang, Z. (2006). A Framework for Authorship Identification of online messages: Writing style features and classification techniques. *Journal of the American Society of Information Science and Technology, 57*(3), 378-393. https://doi.org/10.1002/asi.20316

[64] Zhao, Y., Zobel, J. (2005). Effective and scalable authorship attribution using function words. *Proceedings of the 2nd Asia Information Retrieval Symposium*. 174-189. https://doi.org/10.1007/11562382_14

[65] Koppel, M., Schler, J., Bonchek-Dokow, E. (2007). Measuring differentiability: Unmasking pseudonymous authors. *Journal of Machine Learning Research, 8(45)*, 1261–1276. https://jmlr.org/papers/volume8/koppel07a/koppel07a.pdf

[66] Hearst, M., Dumais, S., Osuna, E., Platt, J., Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications, 13*(4), 18-28. https://doi.org/10.1109/5254.708428.

[67] Orr, M. (1996). Introduction to radial basis function networks. https://faculty.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf

[68] Fogel, D. (1993). Applying evolutionary programming to selected traveling salesman problems. *Cybernetics and Systems*, 24(1), 27-36. https://doi.org/10.1080/01969729308961697

[69] Wehrens, R., Buydens, L. (1998). Evolutionary optimization: A tutorial. *TrAC Trends in Analytical Chemistry, 17*, (4), 193-203. https://doi.org/10.1016/S0165-9936(98)00011-9

[70] Brezočnik, L., Iztok F., Podgorelec, V. (2018). Swarm intelligence algorithms for feature selection: a review. *Applied Sciences, 8*(9), 1521. https://doi.org/10.3390/app8091521

[71] Ding, S., Liu, X. (2009). Evolutionary computing optimization for parameter determination and feature selection of support vector machines. *2009 International Conference on Computational Intelligence and Software Engineering*, 1-5, https://doi.org/10.1109/CISE.2009.5366095

[72] Rostami, M., Berahmand, K., Nasiri, E., Forouzandeh, S. (2021). Review of swarm intelligence-based feature selection methods. *Engineering Applications of Artificial Intelligence, 100*, 104210.

[73] Burger J., Henderson J., Kim G., Zarella G. (2011). Discriminating Gender on Twitter. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 1301-1309. https://aclanthology.org/D11-1120.pdf

[74] Schler, J., Koppel, M., Argamon, S., Pennebaker, J. (2006). Effects of Age and Gender on Blogging. *Compuational Approaches to Analyzing Weblogs: Papers from the 2006 AAAI Spring Symposium*. 199-205. https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-039.pdf

[75] Mukherjee, A., Liu, B., (2010). Improving gender classification of blog authors. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 207-217. https://aclanthology.org/D10-1021.pdf

[76] Rosso, P., Rangel, F., Farías, I.H., Cagnina, L., Zaghouani, W., Charfi, A. A survey on author profiling, deception, and irony detection for the Arabic language. *Lang Linguist Compass*. 2018; 12:e12275. https://doi.org/10.1111/lnc3.12275

Appendix A

Feature Selection Algorithm Comparison Detailed Results

This following is a detailed explanation of the results from the experiments described in Chapter 4 and in [46]. The following is extracted directly from the results section of [46].

While we train the algorithms according to the fitness function presented in Equation 4.1, we evaluate the algorithms based on accuracy alone. Tables A.1 - A.6 show results for each feature selection algorithm (i.e., ABCO, ASO, GSO, PSO, GEFeS and RAND). Each table shows accuracy and feature reduction performance of feature selection algorithms for each AIdS. The first column, labeled "Hybrid" indicates the feature selection algorithm with the AIdS (subscripted) hybrid. The second column, "ω", is the feature reduction weight as presented in Equation 4.1. The third column, "% Accuracy", lists the accuracy achieved for the combination of feature selection algorithm, AIdS, and feature reduction weight (ω). Each cell shows the best accuracy achieved, as well as the mean (shown in parenthesis) across the 30 samples. The fourth column, "% Features", indicates the percentage of the 314 features that were used. In the fourth column, each cell shows the percentage of features used for the corresponding best accuracy (*not to be confused with the best feature reduction*), as well as the mean number of features used for all 30 samples. *Note that because, in the fourth column, the best feature reduction corresponds to the sample with the best accuracy, it is possible for the best feature reduction to have a value higher than the mean feature reduction value.*

For each category of feature selection algorithm, AIdS and feature reduction weight (ω), the tables highlight, in bold font, the best accuracy, mean accuracy, feature reduction, and mean feature reduction values. Note that when there was a tie for best accuracy, the mean accuracy was used to break the tie.

Table A.1 presents the results of the ABCO feature selection algorithm. For LSVM, the best and best mean accuracy (98.00, 96.07) was achieved when $\omega = 0.1$. The best feature reduction (46.50) and best mean feature reduction (44.02) was achieved with $\omega = 1.0$. For RBF, the best accuracy (97.00) was achieved with $\omega = 0.3$, and the best mean accuracy (94.43) was achieved with $\omega = 0.1$. The best feature reduction (44.27) and best mean feature reduction (44.05) was achieved at $\omega = 1.0$ and $\omega = 0.9$ respectively. The MLP AIdS achieved best accuracy (96.00) with $\omega = 0.7$ and best mean accuracy (89.40) with $\omega = 0.3$. The best feature reduction (44.27) was achieved with $\omega = 0.9$ and the best mean feature reduction (45.14) was achieved using $\omega = 1.0$.

Table A.2 presents the results for ASO feature selection. The best accuracy and best mean accuracy for LSVM (99.00, 97.40) and RBF (98.00, 96.00) was achieved when $\omega = 0.3$. For MLP, the best accuracy (100.0) occurred where $\omega = 0.9$ and the best mean (96.43) occurred where $\omega = 0.7$. For LSVM, RBF and MLP, the best feature reduction (20.06, 20.38, 22.29) and best mean feature reduction (20.59, 19.45, 22.85) occurred where $\omega = 1.0$.

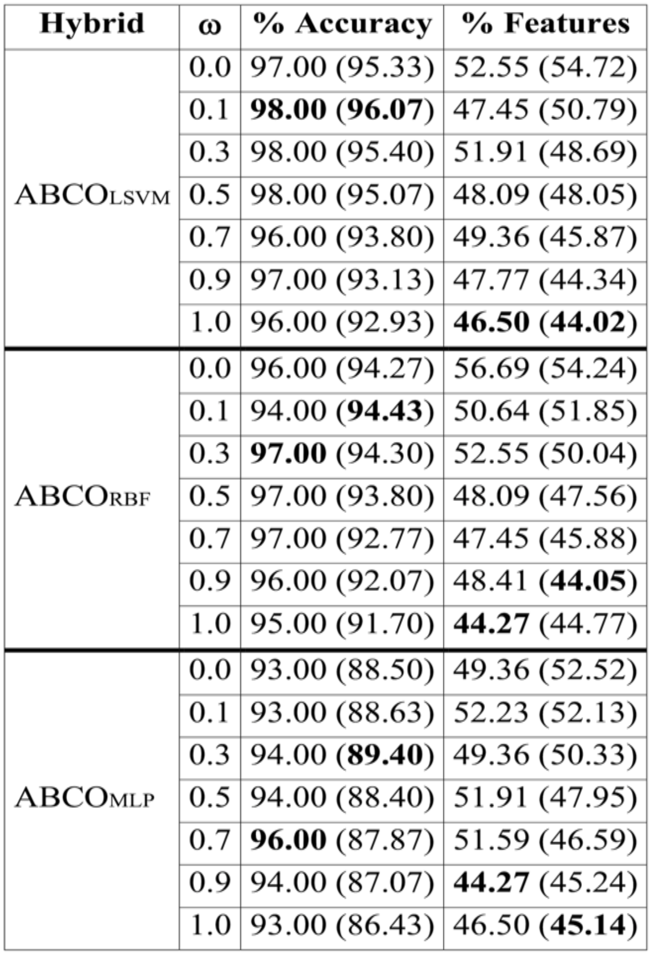
The results for GSO are in Table A.3. The best accuracy (99.00) and best mean accuracy (93.87) for LSVM occurred where $\omega = 0.0$, whereas the best feature reduction (46.18) and the best mean feature reduction (45.08) occurred using $\omega = 1.0$. For RBF, the best accuracy (96.00) occurred using $\omega = 0.3$ and the best mean accuracy (92.30) occurred using $\omega = 0.0$. The best feature reduction (41.72) and best mean feature reduction (46.13) was observed using $\omega = 0.9$. For MLP, the best accuracy (96.00) was observed using $\omega = 0.9$, the best mean accuracy (90.17) using $\omega = 0.1$, the best feature reduction (43.31) using $\omega = 0.3$, and the best mean feature reduction (44.66) using $\omega = 1.0$.

Table A.4 presents the results for PSO. For LSVM, the best accuracy (100.00) occurred with $\omega = 0.3$, the best mean accuracy (95.83) with $\omega = 0.1$, the best feature reduction (45.22) using $\omega = 0.1$, and the best mean feature reduction (43.13) using $\omega = 0.9$. For RBF, the best accuracy (98.00) occurred with $\omega = 0.3$, the best mean accuracy (94.70) with $\omega = 0.1$, the best feature reduction (44.27) using $\omega = 0.7$, and the best mean feature reduction (43.78) using $\omega = 1.0$. For MLP, the best accuracy (96.00) occurred with $\omega = 0.5$, the best mean accuracy (90.17) with $\omega = 0.3$, the best feature reduction (38.45) using $\omega = 0.9$, and the best mean feature reduction (43.94) using $\omega = 1.0$.

Table A.5 presents the results of the RAND algorithm. For LSVM, the best accuracy (87.00) and best mean accuracy (72.73) occurred with $\omega = 0.9$, the best feature reduction (45.54) using $\omega = 0.9$, and the best mean feature reduction (45.76) using $\omega = 1.0$. For RBF, the best accuracy (85.00) occurred with $\omega = 0.3$, the best mean accuracy (69.03) with $\omega = 0.9$, the best feature reduction (44.90) using $\omega = 0.7$, and the best mean feature reduction (46.68) using $\omega = 0.9$. For MLP, the best accuracy (82.00) occurred with $\omega = 0.1$, the best mean accuracy (69.90) with $\omega = 0.9$, the best feature reduction (43.95) using $\omega = 0.5$, and the best mean feature reduction (46.17) using $\omega = 1.0$.

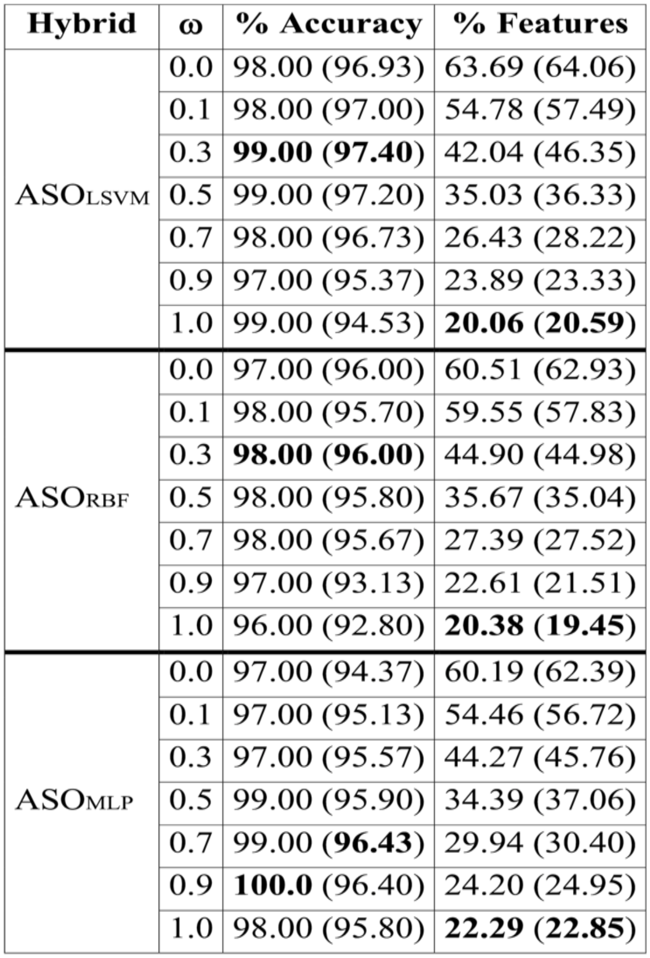
Table A.6 presents the GEFeS results. The best accuracy and best mean accuracy occurred at $\omega = 0.1$ for LSVM (100.0, 99.97) and RBF (100.0, 99.83), and $\omega = 0.3$ for MLP (100.0, 98.37). The best feature reduction and best mean feature reduction occurred at $\omega = 1.0$ for LSVM (12.74, 13.47) and RBF (14.97, 14.32). For MLP, the best feature reduction (15.29) occurred using $\omega = 0.9$, and best mean feature reduction (14.14) using $\omega = 1.0$.

Table A.1. ABCO Algorithm Accuracy and Feature Reduction Performance.



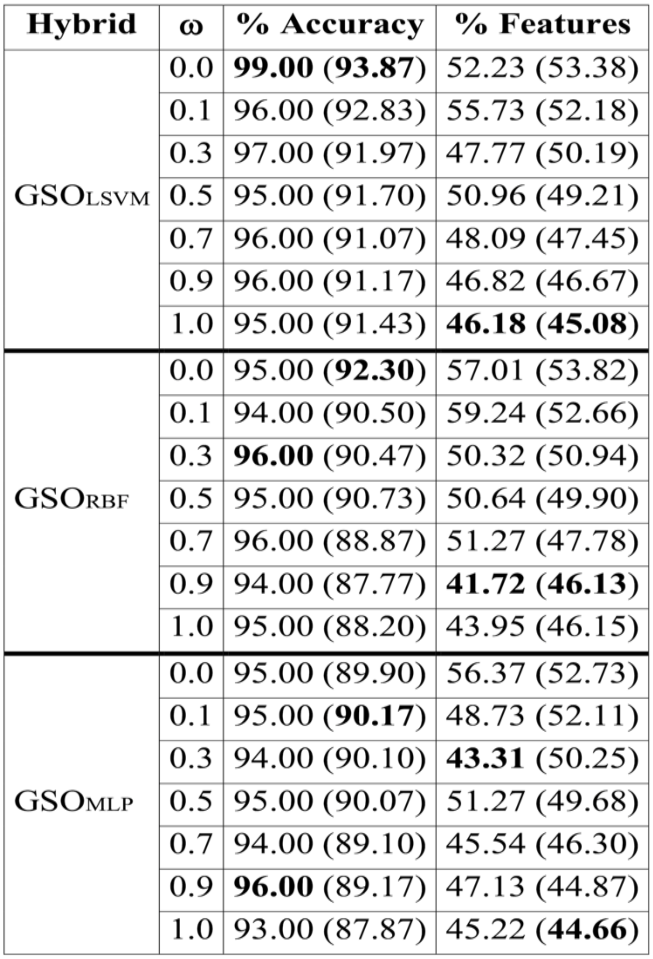
Hybrid	ω	% Accuracy	% Features
ABCO$_{LSVM}$	0.0	97.00 (95.33)	52.55 (54.72)
	0.1	**98.00** (**96.07**)	47.45 (50.79)
	0.3	98.00 (95.40)	51.91 (48.69)
	0.5	98.00 (95.07)	48.09 (48.05)
	0.7	96.00 (93.80)	49.36 (45.87)
	0.9	97.00 (93.13)	47.77 (44.34)
	1.0	96.00 (92.93)	**46.50** (**44.02**)
ABCO$_{RBF}$	0.0	96.00 (94.27)	56.69 (54.24)
	0.1	94.00 (**94.43**)	50.64 (51.85)
	0.3	**97.00** (94.30)	52.55 (50.04)
	0.5	97.00 (93.80)	48.09 (47.56)
	0.7	97.00 (92.77)	47.45 (45.88)
	0.9	96.00 (92.07)	48.41 (**44.05**)
	1.0	95.00 (91.70)	**44.27** (44.77)
ABCO$_{MLP}$	0.0	93.00 (88.50)	49.36 (52.52)
	0.1	93.00 (88.63)	52.23 (52.13)
	0.3	94.00 (**89.40**)	49.36 (50.33)
	0.5	94.00 (88.40)	51.91 (47.95)
	0.7	**96.00** (87.87)	51.59 (46.59)
	0.9	94.00 (87.07)	**44.27** (45.24)
	1.0	93.00 (86.43)	46.50 (**45.14**)

Table A.2. ASO Algorithm Accuracy and Feature Reduction Performance.



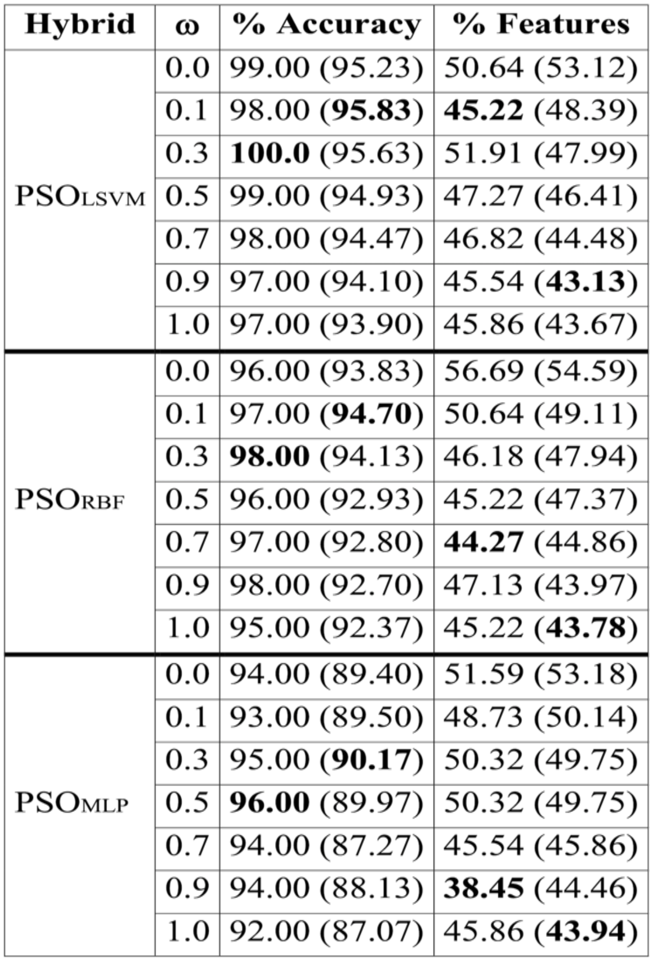
Hybrid	**ω**	**% Accuracy**	**% Features**
ASO$_{LSVM}$	0.0	98.00 (96.93)	63.69 (64.06)
	0.1	98.00 (97.00)	54.78 (57.49)
	0.3	**99.00** (**97.40**)	42.04 (46.35)
	0.5	99.00 (97.20)	35.03 (36.33)
	0.7	98.00 (96.73)	26.43 (28.22)
	0.9	97.00 (95.37)	23.89 (23.33)
	1.0	99.00 (94.53)	**20.06** (**20.59**)
ASO$_{RBF}$	0.0	97.00 (96.00)	60.51 (62.93)
	0.1	98.00 (95.70)	59.55 (57.83)
	0.3	**98.00** (**96.00**)	44.90 (44.98)
	0.5	98.00 (95.80)	35.67 (35.04)
	0.7	98.00 (95.67)	27.39 (27.52)
	0.9	97.00 (93.13)	22.61 (21.51)
	1.0	96.00 (92.80)	**20.38** (**19.45**)
ASO$_{MLP}$	0.0	97.00 (94.37)	60.19 (62.39)
	0.1	97.00 (95.13)	54.46 (56.72)
	0.3	97.00 (95.57)	44.27 (45.76)
	0.5	99.00 (95.90)	34.39 (37.06)
	0.7	99.00 (**96.43**)	29.94 (30.40)
	0.9	**100.0** (96.40)	24.20 (24.95)
	1.0	98.00 (95.80)	**22.29** (**22.85**)

Table A.3. GSO Algorithm Accuracy and Feature Reduction Performance.



Hybrid	ω	% Accuracy	% Features
GSOLSVM	0.0	**99.00** (**93.87**)	52.23 (53.38)
	0.1	96.00 (92.83)	55.73 (52.18)
	0.3	97.00 (91.97)	47.77 (50.19)
	0.5	95.00 (91.70)	50.96 (49.21)
	0.7	96.00 (91.07)	48.09 (47.45)
	0.9	96.00 (91.17)	46.82 (46.67)
	1.0	95.00 (91.43)	**46.18** (**45.08**)
GSORBF	0.0	95.00 (**92.30**)	57.01 (53.82)
	0.1	94.00 (90.50)	59.24 (52.66)
	0.3	**96.00** (90.47)	50.32 (50.94)
	0.5	95.00 (90.73)	50.64 (49.90)
	0.7	96.00 (88.87)	51.27 (47.78)
	0.9	94.00 (87.77)	**41.72** (**46.13**)
	1.0	95.00 (88.20)	43.95 (46.15)
GSOMLP	0.0	95.00 (89.90)	56.37 (52.73)
	0.1	95.00 (**90.17**)	48.73 (52.11)
	0.3	94.00 (90.10)	**43.31** (50.25)
	0.5	95.00 (90.07)	51.27 (49.68)
	0.7	94.00 (89.10)	45.54 (46.30)
	0.9	**96.00** (89.17)	47.13 (44.87)
	1.0	93.00 (87.87)	45.22 (**44.66**)

Table A.4. PSO Algorithm Accuracy and Feature Reduction Performance.

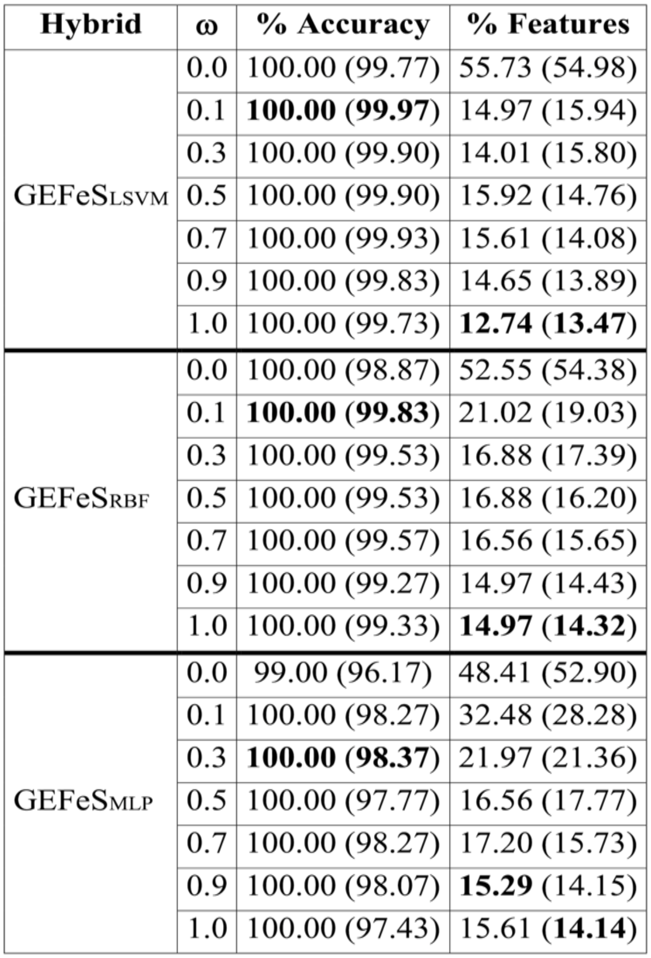


Hybrid	**ω**	**% Accuracy**	**% Features**
PSOLSVM	0.0	99.00 (95.23)	50.64 (53.12)
	0.1	98.00 (**95.83**)	**45.22** (48.39)
	0.3	**100.0** (95.63)	51.91 (47.99)
	0.5	99.00 (94.93)	47.27 (46.41)
	0.7	98.00 (94.47)	46.82 (44.48)
	0.9	97.00 (94.10)	45.54 (**43.13**)
	1.0	97.00 (93.90)	45.86 (43.67)
PSORBF	0.0	96.00 (93.83)	56.69 (54.59)
	0.1	97.00 (**94.70**)	50.64 (49.11)
	0.3	**98.00** (94.13)	46.18 (47.94)
	0.5	96.00 (92.93)	45.22 (47.37)
	0.7	97.00 (92.80)	**44.27** (44.86)
	0.9	98.00 (92.70)	47.13 (43.97)
	1.0	95.00 (92.37)	45.22 (**43.78**)
PSOMLP	0.0	94.00 (89.40)	51.59 (53.18)
	0.1	93.00 (89.50)	48.73 (50.14)
	0.3	95.00 (**90.17**)	50.32 (49.75)
	0.5	**96.00** (89.97)	50.32 (49.75)
	0.7	94.00 (87.27)	45.54 (45.86)
	0.9	94.00 (88.13)	**38.45** (44.46)
	1.0	92.00 (87.07)	45.86 (**43.94**)

Table A.5. RAND Algorithm Accuracy and Feature Reduction Performance.

Hybrid	ω	% Accuracy	% Features
RAND$_{LSVM}$	0.0	85.00 (69.13)	54.78 (52.74)
	0.1	84.00 (71.07)	51.91 (52.13)
	0.3	85.00 (69.33)	51.27 (50.69)
	0.5	84.00 (70.13)	52.23 (49.11)
	0.7	83.00 (70.37)	47.13 (47.89)
	0.9	**87.00 (72.73)**	**45.54** (46.57)
	1.0	84.00 (72.10)	49.68 (**45.76**)
RAND$_{RBF}$	0.0	78.00 (69.00)	54.14 (52.64)
	0.1	81.00 (67.77)	51.27 (53.03)
	0.3	**85.00** (66.53)	53.82 (51.00)
	0.5	78.00 (65.77)	48.73 (51.07)
	0.7	81.00 (66.80)	**44.90** (48.51)
	0.9	78.00 (**69.03**)	45.22 (**46.68**)
	1.0	78.00 (65.97)	45.54 (47.31)
RAND$_{MLP}$	0.0	79.00 (68.03)	52.87 (52.81)
	0.1	**82.00** (68.53)	50.64 (52.49)
	0.3	81.00 (68.80)	52.23 (51.52)
	0.5	81.00 (67.33)	**43.95** (47.62)
	0.7	81.00 (68.50)	49.36 (48.21)
	0.9	80.00 (**69.90**)	45.54 (46.38)
	1.0	79.00 (67.07)	47.77 (**46.17**)

Table A.6. GEFeS Algorithm Accuracy and Feature Reduction Performance.



Hybrid	ω	% Accuracy	% Features
GEFeS$_{LSVM}$	0.0	100.00 (99.77)	55.73 (54.98)
	0.1	**100.00** (**99.97**)	14.97 (15.94)
	0.3	100.00 (99.90)	14.01 (15.80)
	0.5	100.00 (99.90)	15.92 (14.76)
	0.7	100.00 (99.93)	15.61 (14.08)
	0.9	100.00 (99.83)	14.65 (13.89)
	1.0	100.00 (99.73)	**12.74** (**13.47**)
GEFeS$_{RBF}$	0.0	100.00 (98.87)	52.55 (54.38)
	0.1	**100.00** (**99.83**)	21.02 (19.03)
	0.3	100.00 (99.53)	16.88 (17.39)
	0.5	100.00 (99.53)	16.88 (16.20)
	0.7	100.00 (99.57)	16.56 (15.65)
	0.9	100.00 (99.27)	14.97 (14.43)
	1.0	100.00 (99.33)	**14.97** (**14.32**)
GEFeS$_{MLP}$	0.0	99.00 (96.17)	48.41 (52.90)
	0.1	100.00 (98.27)	32.48 (28.28)
	0.3	**100.00** (**98.37**)	21.97 (21.36)
	0.5	100.00 (97.77)	16.56 (17.77)
	0.7	100.00 (98.27)	17.20 (15.73)
	0.9	100.00 (98.07)	**15.29** (14.15)
	1.0	100.00 (97.43)	15.61 (**14.14**)

Appendix B

Genetic-Based Feature Selection for Author Identification Detailed Results

This following is a detailed explanation of the results from the experiments described in Chapter 5.

Although we train the algorithm according to the fitness function presented in Equation 4.1, we report only the accuracy, not the fitness value. Tables B.2 – B.5 show the results for the first set of experiments, which uses each of the feature sets and calculates accuracy only on the fourth sample (including swapping the adversarial sample). Tables B.6 – B.9 show the results for the second set of experiments, which also uses each of the feature sets, but calculates the accuracy using all four samples (also including swapping the adversarial sample).

In each of these tables, the first column, ω, represents the feature reduction weight as presented in Equation 4.1. The next pair of columns, labeled *Baseline*, shows the results achieved with no feature mask. The first of the baseline columns, labeled *Original*, shows the accuracy using the original author samples, whereas the second column, labeled *Adversarial*, shows the accuracy when introducing the adversarial samples. Note that since there is no feature mask, the feature reduction weight has no effect on the accuracy. Therefore, the values of these columns are denormalized for the sake of readability and are the same for all rows of a given feature set.

The next pair of columns, labeled *GEFeS* shows accuracy using GEFeS feature selection. The first of the pair of columns, labeled *Original*, shows accuracy using original author samples (as opposed to the adversarial samples modified to disguise the author), and the second column of the pair, labeled *Adversarial*, shows the accuracy achieved when swapping in the disguised adversarial samples.

The final column, labeled *Use?*, is a value that indicates the AIdS's susceptibility to adversarial authorship attacks using feature selection for that feature set given the specific feature reduction value (ω). Formally, the expression for this value is as defined in Table 5.1.

The value of *Use?* captures the relationship of the effectiveness of the adversarial attack without feature selection, compared to the same attack with feature selection. Positive *Use?* values favor the use of feature selection and negative values favor the avoidance of feature selection.

Table B.1 shows the results for the experiments using only the LIWC feature set, where the AIdS trained on the first three samples and tested on the fourth original and adversarial sample texts. One can see that the baseline (i.e., no feature selection) accuracy for the non-adversarial, or original, samples is 68.00%, whereas the adversarial attack lowers the baseline accuracy to 56.00%. As previously mentioned, ω has little effect on accuracy, so the measured accuracies remain similar – as they do for these two columns across all eight tables. The fourth column shows accuracies for the non-adversarial, or original, writing samples using GEFeS feature selection ranging from 87.73% (ω = 1.0) to 89.07% (ω = 0.3). The fifth column shows that with adversarial samples and GEFeS feature selection, the accuracy drops to a range of 15.73% (ω = 0.7) to 35.07% (ω = 0.0). The last column indicates that the most favorable accuracy drop (from the AIdS perspective) occurs with ω = 0.0, yielding a *Use?* value, as defined in Table 5.1, of -0.07. The least favorable drop occurs with ω = 0.7, yielding a *Use?* value of -0.42.

In Table B.2, which is based on experiments using the Topic Modeling feature set, the baseline accuracies are 84.00% and 8.00% respectively for the original and adversarial samples. Using GEFeS, the accuracies improve to ranges of 91.33% (ω = 0.1) to 92.00% (ω = 0.1) for original samples, and 10.53% ($\omega = 0.1$) to 11.73% ($\omega = 0.0$) with adversarial samples. Unlike Table B.1, all *Use?* values (see Table 5.1) in Table B.2 are positive, ranging from 0.40 ($\omega = 0.1$) to 0.56 ($\omega = 0.0$), indicating that feature selection is desirable for all values of ω using Topic Modeling.

The results in Table B.3 reflect experiments using the Stylometry feature set. The baseline accuracies are 60.00% for the original samples and 32.00% for the adversarial samples. Of the three independent feature sets, Stylometry yields the highest range of original GEFeS feature selection accuracies ranging from 95.33% ($\omega = 0.0$) to 97.47% ($\omega = 0.1$). Stylometry also yields the lowest range of adversarial GEFeS accuracies ranging from 4.00% ($\omega = 0.3$, 0.7, 0.9 & 1.0) to 4.27% ($\omega = 0.0$). Despite this most significant swing among the three independent feature sets when using feature selection, the *Use?* indicators, as defined in Table 5.1, are moderately negative due to the relative drop in baseline accuracies.

Table B.4 shows the results of the hybrid feature set, which is a combination of the previous three feature sets. Synergistically, this feature set yields the highest non-adversarial, or original, accuracies for both the baseline and GEFeS, at 92.00% and a range of 99.73% ($\omega = 0.0$) to 100.00% ($\omega = 0.1 – 0.7$) respectively. The adversarial baseline is 32.00%. The drop in the GEFeS adversarial accuracy is moderate ranging from 8.27 ($\omega = 0.1$) to 12.13% ($\omega = 0.0$). The notably good performance of the baseline compared with the lackluster performance of feature selection explains why the *Use?* values, as defined in Table 5.1, are the most disparaging, ranging from -0.54 ($\omega = 0.0$) to -0.65 ($\omega = 0.1$).

Table B.5 revisits the LIWC feature set with baseline accuracies of 92.00% and 88.00% for original and adversarial samples respectively. The GEFeS accuracies range from 96.93% ($\omega = 1.0$) to 97.27% ($\omega = 0.3$) and 79.00% ($\omega = 0.9$) to 83.77% ($\omega = 0.0$) for original and adversarial samples.

The *Use?* values, as defined in Table 5.1, are slightly negative ranging from 0.00 (ω = 0.0) to -0.06 (ω = 0.5 - 1.0).

Table B.6 uses the Topic Modeling feature set. This table shows baseline accuracies of 96.00% for non-adversarial, or original, samples and 77.00% for adversarial samples. When we employ feature selection, we see similar numbers with a range of 96.23% (ω = 0.1 & 1.0) to 96.87% (ω = 0.0), and a range of 75.10% (ω = 1.0) to 76.80 (ω = 0.0), for original and adversarial respectively. The *Use?* value, as defined in Table 5.1, remains relatively stable near 0.00, fluctuating only by -0.01 (ω = 0.1 & 1.0).

Table B.7 shows the results for the Stylometry feature set. The baseline accuracies are 90.00% for original samples and 83.00% for adversarial samples. Using GEFeS, the non-adversarial sample accuracies range from 98.83% (ω = 0.0) to 99.87 (ω = 1.0), and for adversarial samples, the accuracies remain relatively close to 76.00%, ranging from 76.00% (ω = 0.3, 0.7, 0.9 & 1.0) to 76.07% (ω = 0.0). The *Use?* values, as defined in Table 5.1, are slightly positive with values of 0.01 (ω = 0.0, 0.9 & 1.0) and 0.02 (ω = 0.1 - 0.7), which is slightly better than the results for Topic Modeling shown in Table B.6.

Table B.8 is the Hybrid feature set, which consists of a combination of the three previous feature sets, with baseline accuracies of 98.00% and 83.00% for non-adversarial, or original, and adversarial samples. Feature selection yields an original accuracy near or at 100.00%, with the lowest accuracy being 99.93% (ω = 0.0). The adversarial samples have an accuracy ranging from 77.07% (ω = 0.1), to 78.03% (ω = 0.0). The *Use?* values, as defined in Table 5.1, are all negative with a value of -0.05 with the exception of ω = 0.0, which has a value of -0.04.

If we compare the baseline accuracies of tables B.1 - B.4 for original samples, we would likely rank the usefulness of the feature sets, from an AIdS perspective, from best to worst as:

1. Hybrid
2. Topic Modeling
3. LIWC
4. Stylometry

It is interesting that the Hybrid feature set outperforms any of its individual constituent feature sets. This demonstrates the synergy of feature set combinations. It is also interesting to note the high ranking of Topic Modeling among the base feature sets. We hypothesize that this is because the samples are blog posts and that bloggers tend to focus on specific topics.

If we now compare the relative rankings of the feature sets when using feature selection, we see a ranking of:

1. Hybrid
2. Stylometry
3. Topic Modeling
4. LIWC

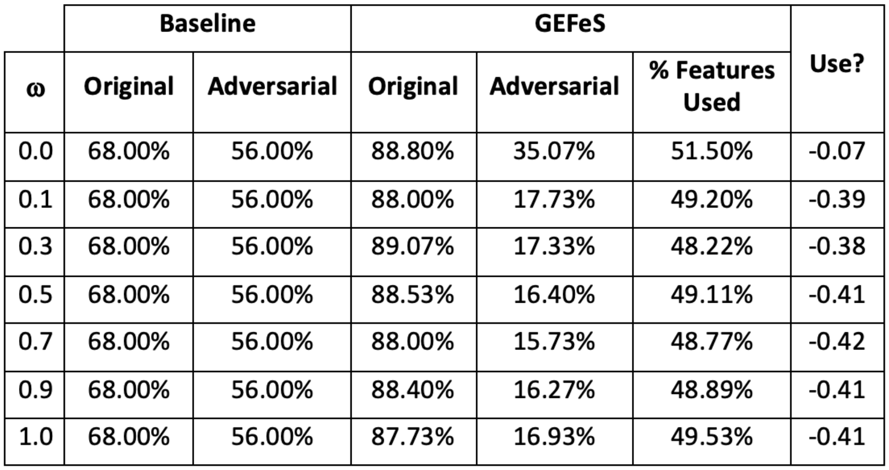
Once again, Hybrid ranks as the top feature set. With no analysis of the susceptibility of adversarial attacks, we would happily prefer the Hybrid feature set. But the story is not quite that simple.

If we consider the susceptibility of each of the feature sets, using tables B.1 - B.4 we can rank them from least susceptible to most susceptible as:

1. Topic Modeling
2. Stylometry
3. LIWC
4. Hybrid

Recall that tables B.5 - B.8 are the results with accuracy testing using all four samples. As a result, we see a slight skew in favor of using feature selection due to the dilution of the number of adversarial samples. However, these feature sets maintain roughly the same susceptibility ranking based on tables B.5 - B.8. This indicates that benefits we see from feature selection, when applying a weaker adversarial attack, would need to be diluted even further before we would strongly consider using the Hybrid feature set.

Table B.1. A Comparison of Adversarial Author Identification with and without Feature Selection Using the LIWC Feature Set - 93 Features, 75+(org = 25, adv = 25).



	Baseline		**GEFeS**			**Use?**
ω	**Original**	**Adversarial**	**Original**	**Adversarial**	**% Features Used**	
0.0	68.00%	56.00%	88.80%	35.07%	51.50%	-0.07
0.1	68.00%	56.00%	88.00%	17.73%	49.20%	-0.39
0.3	68.00%	56.00%	89.07%	17.33%	48.22%	-0.38
0.5	68.00%	56.00%	88.53%	16.40%	49.11%	-0.41
0.7	68.00%	56.00%	88.00%	15.73%	48.77%	-0.42
0.9	68.00%	56.00%	88.40%	16.27%	48.89%	-0.41
1.0	68.00%	56.00%	87.73%	16.93%	49.53%	-0.41

Table B.2. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Topic Modeling Feature Set - 45 Features, 75+(org = 25, adv = 25).

	Baseline		GEFeS			Use?
ω	Original	Adversarial	Original	Adversarial	% Features Used	
0.0	84.00%	8.00%	92.00%	11.73%	52.20%	0.56
0.1	84.00%	8.00%	91.33%	10.53%	51.57%	0.40
0.3	84.00%	8.00%	91.47%	11.20%	50.64%	0.49
0.5	84.00%	8.00%	91.87%	11.20%	51.47%	0.49
0.7	84.00%	8.00%	91.47%	10.80%	51.12%	0.44
0.9	84.00%	8.00%	91.47%	11.33%	51.54%	0.51
1.0	84.00%	8.00%	91.47%	10.93%	51.54%	0.46

Table B.3. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Stylometry Feature Set - 428 Features, 75+(org = 25, adv = 25).

	Baseline		GEFeS			Use?
ω	Original	Adversarial	Original	Adversarial	% Features Used	
0.0	60.00%	32.00%	95.33%	4.27%	49.34%	-0.28
0.1	60.00%	32.00%	97.47%	4.13%	39.06%	-0.25
0.3	60.00%	32.00%	96.00%	4.00%	36.64%	-0.27
0.5	60.00%	32.00%	96.27%	4.13%	34.38%	-0.27
0.7	60.00%	32.00%	96.40%	4.00%	32.06%	-0.27
0.9	60.00%	32.00%	94.67%	4.00%	29.87%	-0.30
1.0	60.00%	32.00%	95.47%	4.00%	28.96%	-0.28

Table B.4. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Hybrid Feature Set -566 Features, 75+(org = 25, adv = 25).

	Baseline		GEFeS			Use?
ω	**Original**	**Adversarial**	**Original**	**Adversarial**	**% Features Used**	
0.0	92.00%	32.00%	99.73%	12.13%	52.33%	-0.54
0.1	92.00%	32.00%	100.00%	8.27%	13.77%	-0.65
0.3	92.00%	32.00%	100.00%	10.13%	12.89%	-0.60
0.5	92.00%	32.00%	100.00%	9.2%	12.77%	-0.63
0.7	92.00%	32.00%	100.00%	9.07%	12.40%	-0.63
0.9	92.00%	32.00%	99.87%	8.67%	11.67%	-0.64
1.0	92.00%	32.00%	99.87%	8.67%	11.20%	-0.64

Table B.5. A Comparison of Adversarial Author Identification with and without Feature Selection Using the LIWC Feature Set - 93 Features, 75+(org = 100, adv = 100).

	Baseline		GEFeS			Use?
ω	**Original**	**Adversarial**	**Original**	**Adversarial**	**% Features Used**	
0.0	92.00%	88.00%	97.20%	83.77%	51.50%	0.00
0.1	92.00%	88.00%	97.00%	79.43%	49.20%	-0.05
0.3	92.00%	88.00%	97.27%	79.33%	48.22%	-0.05
0.5	92.00%	88.00%	97.13%	79.10%	49.11%	-0.06
0.7	92.00%	88.00%	97.00%	78.93%	48.77%	-0.06
0.9	92.00%	88.00%	97.03%	79.00%	48.89%	-0.06
1.0	92.00%	88.00%	96.93%	79.23%	49.53%	-0.06

Table B.6. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Topic Modeling Feature Set - 45 Features, 75+(org = 100, adv = 100).

	Baseline		GEFeS			Use?
ω	Original	Adversarial	Original	Adversarial	% Features Used	
0.0	96.00%	77.00%	96.87%	76.80%	52.20%	0.01
0.1	96.00%	77.00%	96.23%	76.03%	51.57%	-0.01
0.3	96.00%	77.00%	96.50%	76.43%	50.64%	0.00
0.5	96.00%	77.00%	96.53%	76.37%	51.47%	0.00
0.7	96.00%	77.00%	96.50%	76.33%	51.12%	0.00
0.9	96.00%	77.00%	96.63%	76.60%	51.54%	0.00
1.0	96.00%	77.00%	96.23%	75.10%	51.54%	-0.01

Table B.7. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Stylometry Feature Set - 428 Features, 75+(org = 100, adv = 100).

	Baseline		GEFeS			Use?
ω	Original	Adversarial	Original	Adversarial	% Features Used	
0.0	90.00%	83.00%	98.83%	76.07%	49.34%	0.01
0.1	90.00%	83.00%	99.37%	76.03%	39.06%	0.02
0.3	90.00%	83.00%	99.00%	76.00%	36.64%	0.02
0.5	90.00%	83.00%	99.07%	76.03%	34.38%	0.02
0.7	90.00%	83.00%	99.10%	76.00%	32.06%	0.02
0.9	90.00%	83.00%	99.67%	76.00%	29.87%	0.01
1.0	90.00%	83.00%	99.87%	76.00%	28.96%	0.01

Table B.8. A Comparison of Adversarial Author Identification with and without Feature Selection Using the Hybrid Feature Set - 566 Features, 75+(org = 100, adv = 100).

	Baseline		GEFeS			Use?
ω	Original	Adversarial	Original	Adversarial	% Features Used	
0.0	98.00%	83.00%	99.93%	78.03%	52.33%	-0.04
0.1	98.00%	83.00%	100.00%	77.07%	13.77%	-0.05
0.3	98.00%	83.00%	100.00%	77.53%	12.89%	-0.05
0.5	98.00%	83.00%	100.00%	77.30%	12.77%	-0.05
0.7	98.00%	83.00%	100.00%	77.27%	12.40%	-0.05
0.9	98.00%	83.00%	99.97%	77.17%	11.67%	-0.05
1.0	98.00%	83.00%	99.97%	77.17%	11.20%	-0.05

Appendix C

Adversarial Authorship, Swarm Intelligence Feature Selection, and CAISIS-50 Dataset Detailed Results

This following is a detailed explanation of the results from the experiments described in Chapter 6.

We summarize the results of these experiments in tables C.3 – C.18, included in this chapter. Each table represents the results for a specific dataset (CASIS-25 or CASIS-50) and a specific training/testing strategy. For CASIS-25 the possible testing/training strategies include training on 75 samples and testing on 25 samples (labeled as *75+(org = 25, adv = 25)*) or training on 75 samples and testing on all 100 samples (labeled as *75+(org = 100, adv = 100)*). For the CASIS-50 dataset, the training/testing strategies included training on 175 samples and testing on 25 samples (labeled as 1*75+(org = 25, adv = 25)*) or training on 175 samples and testing using all 200 samples (labeled as 1*75+(org = 200, adv = 200)*). We choose to test on only 25 samples for the two testing strategies because we have 25 adversarial samples, which allows us to compare original and adversarial sample results.

Each table consists of eight columns. The first column, labeled *ω*, is the feature reduction weighting factor as described in Equation 4.1. Notice that each table has seven super-rows for each of the seven weighting factor values (i.e., 0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0). The second column, labeled *FS Alg*, lists the feature selection algorithm used for the measurements of the row. Note that each feature selection algorithm appears in a row corresponding to each of the seven feature reduction weighting values. So, with seven weighting values and six feature selection algorithms, each table has 42 rows of measurements. The third and fourth columns have a super-label of *Baseline*. These columns represent author classification accuracy without feature selection. The third column, labeled *Orig*, indicates author classification accuracy without adversarial samples. The fourth column, labeled *Adv*, indicates the classification accuracy when adversarial samples are swapped in. Since no feature selection is used for the accuracies reported in these two columns, the accuracies for each column are the same for all rows. The fifth, sixth and seventh columns have the super-label *With Feature Selection*, indicating that these columns show values measured when using the feature selection algorithms. The fifth and sixth columns labeled *Orig* and *Adv* respectively, represent the accuracies measured without and with adversarial samples. The seventh column, labeled *% Features Used*, indicates the percent of features used in the most fit mask. Column eight is labeled *Use?*. This column represents a value calculated using the previous columns as shown in Table 6.1. The *Use?* value is an indicator of when it is useful to use feature selection, based on the relative degradation due to feature selection. *Use?* values greater that zero indicate it is favorable to use feature selection, whereas values less than or equal to zero indicate not using feature selection.

Tables C.2 – C.9, report results for the CASIS-25 dataset, and tables C.10 – C.17 report results for the CASIS-50 dataset. Tables C.2 – C.5, train on 75 samples and test using 25 samples (75+(org = 25, adv = 25)). Tables C.6 – C.9, also train on 75 samples, but test using all 100 samples (75+(org = 100, adv = 100)). Similarly, tables C.10 – C.13, train on 175 samples and test using 25 samples (175+(org = 25, adv = 25)). Whereas tables C.14 – C.17, train on 175 samples but test on all 200 samples (175+(org = 200, adv = 200)). Each of these sets of four tables use a different feature selection algorithm. Table C.1 explains the relationships between the tables to help navigate the results.

Table C.1. Results Tables Relationships.

	CASIS-25		**CASIS-50**	
	75 + (org = 25, adv = 25)	**75 + (org = 100, adv = 100)**	**175 + (org = 25, adv = 25)**	**175 + (org = 200, adv = 200)**
LIWC	Table C.2	Table C.6	Table C.10	Table C.14
Topic Modeling	Table C.3	Table C.7	Table C.11	Table C.15
Stylometry	Table C.4	Table C.8	Table C.12	Table C.16
Hybrid	Table C.5	Table C.9	Table C.13	Table C.17

Table C.2 shows measurements taken using the LIWC dataset and 75+(org = 25, adv = 25). Without feature selection (i.e., baseline), we see baseline accuracies of 68.00% without adversarial samples (i.e., original samples), and 56.00% when introducing adversarial samples. When we employ feature selection but not adversarial samples, we see accuracies ranging from 72.27% with ω of 1.0 using RAND feature selection, to 89.07% with ω of 0.3 using GEFeS. By introducing adversarial samples, we see the range drop to between 15.73% ($\omega = 0.7$, GEFeS) and 64.53% ($\omega = 0.0$, ASO). These accuracies yield negative *Use?* values ranging from -0.07 ($\omega = 0.0$, GEFeS) to -0.59 ($\omega = 0.9$, RAND).

Table C.3 reflects the measurements using the Topic Modeling feature set and 75+(org = 25, adv = 25). We see a large drop in the baseline accuracies, going from 84.00% to 8.00%. When we introduce feature selection, we see a general improvement in accuracies, ranging from 77.47% ($\omega = 0.9$, RAND) to 92.00% ($\omega = 0.0$, GEFeS) without adversarial samples, and a range of 8.67% ($\omega = 0.7$, PSO) to 12.00% (several occurrences). These improved accuracies yield positive *Use?* values ranging from 0.02 ($\omega = 0.0$ & 0.9, RAND) to 0.56 ($\omega = 0.0$, GEFeS).

The results of Table C.4 used the Stylometry feature set and 75+(org = 25, adv = 25). We see moderate baseline accuracies of 60.00% for original samples and 32.00% when we introduce adversarial samples. While we see a significant improvement in accuracies with feature selection for original samples ranging from 64.40% (ω = 0.7, RAND) to 97.47% (ω = 0.1, GEFeS), we also see a drop in accuracies when we introduce adversarial samples ranging from 4.00% to an outlying 14.80% (ω = 0.0, ASO). These badly impacted accuracies, due to feature selection, unsurprisingly yield firmly negative *Use?* values ranging from -0.22 (ω = 0.0, ASO) to -0.80 (ω = 0.7 & 1.0, RAND).

In Table C.5, we see results using the Hybrid feature set and 75+(org = 25, adv = 25). The baseline original accuracy is a stellar 92.00%, and when introducing adversarial samples, the accuracy drops to 32.00%. Feature selection helps the original sample accuracies slightly to a range of 90.00% (ω = 1.0, RAND) to 100.00% (several values of ω, GEFeS). But the introduction of adversarial samples damages the accuracies to a range of 6.80% (ω = 0.5, ASO) to 12.93% (ω = 0.1, GSO). As we would expect, the *Use?* values reflect the adversarial vulnerability with negative values ranging from -0.54 (ω = 0.0, GEFeS) to -0.76 (ω = 1.0, ASO).

With Table C.6, we return to the LIWC dataset, this time testing on all samples 75+(org = 100, adv = 100). Since we are testing on all samples, we see the baseline accuracies are higher at 92.00% for original samples and 88.00% with the adversarial samples. When we employ feature selection, the change in accuracies is subtle ranging from 93.07% (ω = 0.1 & 0.9, RAND) to 97.27% (ω = 0.3, GEFeS) for original samples, and 78.93 (ω = 0.7, GEFeS) to 91.13% (ω = 0.0, ASO). The *Use?* values are marginal, with ASO showing some slightly positive values near 0.04, and the other algorithms having slightly negative values ranging from 0.00 to -0.09.

Table C.7 uses the Topic Modeling dataset, which performed relatively well under adversarial attack (see Table 6.4), but we see a different story when we change the testing strategy to 75+(org = 100, adv = 100). We see the baseline accuracies improve to 96.00% for the original samples and a whopping 77.00% with the adversarial samples. However, in this case, feature selection often reduces accuracies with original samples ranging from 90.73% (ω = 0.9, RAND) to 96.87% (ω = 0.0, GEFeS), and with adversarial samples, a range of 47.70 (ω = 0.9, RAND) to 85.04% (ω = 0.0, ASO). Not surprisingly, the *Use?* values are mostly slightly negative ranging from 0.02 to -0.10.

The measurements of Table C.8 reflect using the Stylometry dataset and 75+(org = 100, adv = 100). The baseline accuracies are 90.00% for original samples and 83.00% for adversarial samples. Feature selection renders the original sample accuracies slightly better, ranging from 91.10% (ω = 0.7, RAND) to 99.87% (ω = 1.0, GEFeS). But the adversarial sample accuracies suffer, ranging from 76.00% to 78.70% (ω = 0.0, ASO). Therefore, the *Use?* values are mostly slightly negative ranging from 0.02 to -0.07.

Table C.9 is the final table using the CASIS-25 dataset. This table reflects measurements using the Hybrid dataset 75+(org = 100, adv = 100). The baseline accuracies are 98.00% and 83.00% for original and adversarial samples, respectively. While there is not a lot of room for improvement in original accuracies, feature selection manages some improvement, ranging from 97.50 (ω = 1.0, RAND) to 100.00%. However, we see a general reduction in accuracy when using feature selection with adversarial samples ranging from 76.70% (ω = 0.5, ASO) to 78.23% (ω = 0.1, GSO). As expected, the *Use?* values are negative ranging from -0.04 (ω = 0.0, GEFeS), to -0.07.

Table C.10 is the first of the CASIS-50 dataset tables and uses the LIWC dataset 175+(org = 25, adv = 25). Like Table 6.3, we do not have the advantage of testing with previously seen samples, so the baseline accuracies drop back to 68.00% for original samples and 44.00% for adversarial samples. Feature selection improves original sample accuracies with a range of 70.80% (ω = 0.5, RAND) to 89.97% (ω = 0.0, GEFeS). But we see a general drop in accuracies with adversarial samples ranging from 13.33% (ω = 0.1, GEFeS) to 64.13% for the outlier ASO algorithm (ω = 0.0, ASO). As a result, the *Use?* values are positive for ASO ranging from 0.27 to 0.57, while the other feature selection algorithms' *Use?* values are usually negative.

Table C.11 uses Topic Modeling 175+(org = 25, adv = 25). The original sample baseline accuracy is 40%, but the adversarial samples reduce the accuracy to complete failure (0.00%). Feature selection significantly improves accuracies for original samples, and also offers some improvement for adversarial samples. The baseline accuracy of 0.00% causes the conditional clause of Table 6.1 to fire, and since the adversarial samples have some accuracy ranging from 9.20% to 12.00%, the *Use?* values are all positive, because some accuracy with feature selection is better than no accuracy.

In Table C.12, the measurements reflect using Stylometry 175+(org = 25, adv = 25). The baseline accuracies without and with adversarial samples are 48.00% and 20.00%, respectively. With feature selection, we see improvement in original sample accuracies, ranging from 63.73% (ω = 0.0, RAND) to 96.93% (ω = 0.1, GEFeS). But we see reduced accuracies for feature selection with the introduction of adversarial samples ranging from 4.00% to 14.27% (ω = 0.0, ASO). Because GEFeS and ASO performed better than the other algorithms when under adversarial attack with feature selection, their *Use?* values were positive (0.04 to 0.37), while the remainder of the *Use?* values were negative (-0.20 to -0.47).

Table C.13 reflects measurements using the Hybrid feature set and 175+(org = 25, adv = 25). The baseline accuracies were 92.00% and 12.00% for original and adversarial samples. Feature selection generally improved accuracies for original samples with a range of 89.07% (ω = 0.9, RAND) to 100.00% (GEFeS). The results of feature selection with adversarial samples were mixed with some minor improvements and mostly slight degradation ranging from 6.00% (ω = 0.3, ASO) to 12.67% (ω = 0.7, ABCO). Correspondingly, the *Use?* values are mixed, but mostly negative ranging from -0.41 (ω = 0.1, ASO) to 0.08 (ω = 0.0, GSO).

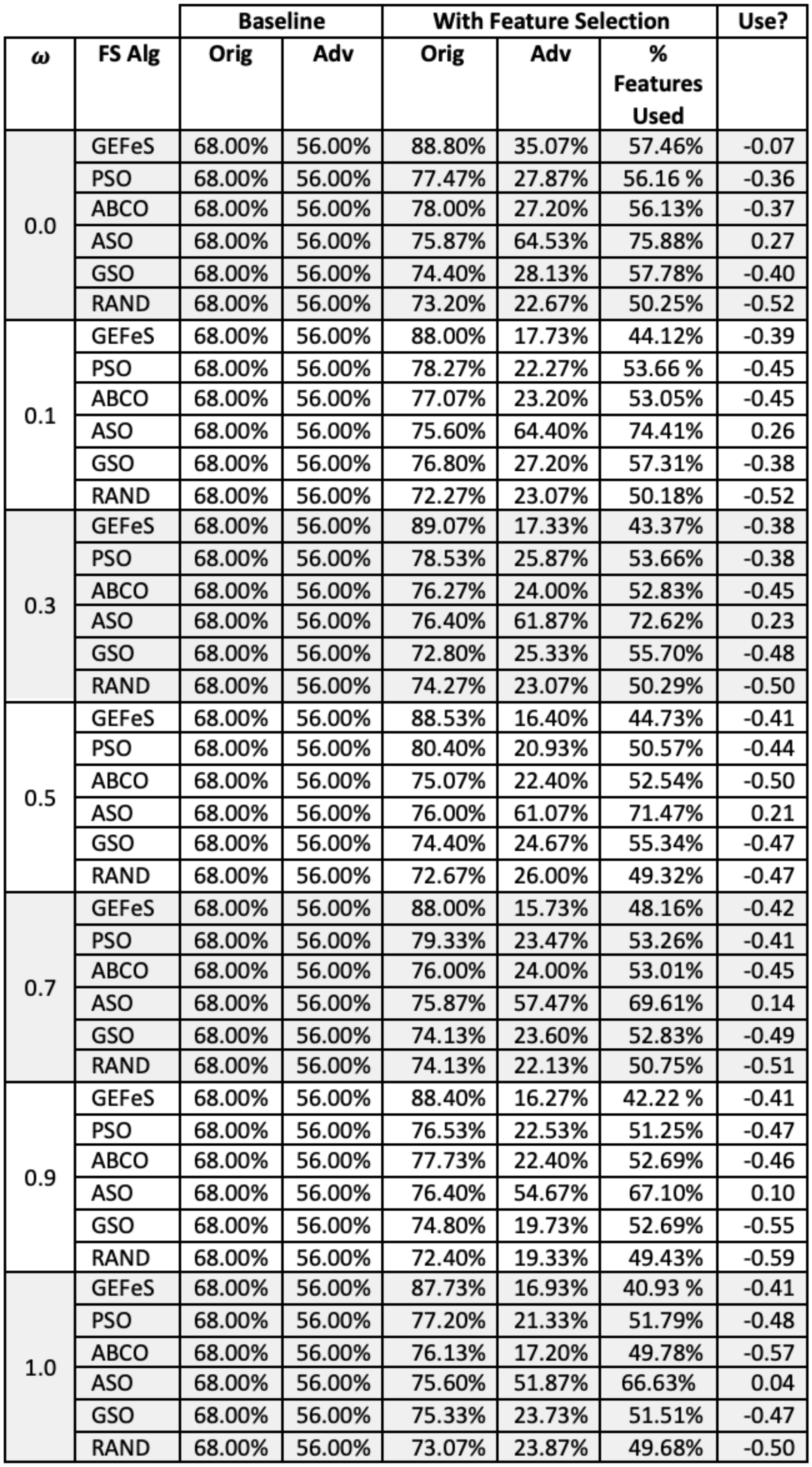
Table C.14 returns one last time to the LIWC feature set but uses 175+(org = 200, adv = 200). The baseline accuracies are 96.00% and 93.00% for original and adversarial samples. Feature selection was mostly unhelpful with original sample accuracies ranging from 92.67 (ω = 0.5, RAND) to 97.47% (ω = 0.0, GEFeS). As a result, the *Use?* values are consistently negative ranging from -0.04 (ω = 0.1, ASO) to -0.18 (ω = 0.7, RAND).

Table C.15 uses Topic Modeling 175+(org = 200, adv = 200). The baseline accuracies are 85.00% and 80.00% for original and adversarial samples. Feature selection demonstrates an improvement when using original samples ranging from 90.73% (ω = 0.7, RAND) to 96.83 (ω = 1.0, ASO), and shows only slight degradation when introducing adversarial samples ranging from 70.00% (ω = 0.5, GEFeS) to 85.33% (ω = 0.1, ASO). The resulting *Use?* values are mostly positive ranging from -0.02 (ω = 0.7, RAND) to 0.11 (ASO).

In Table C.16, we see results using Stylometry 175+(org = 200, adv = 200) with baseline accuracies of 93.50% and 90.00% for original and adversarial samples. Feature selection does not introduce much improvement for original sample accuracies, which range from 91.17% (ω = 1.0, RAND) to 99.23% (ω = 0.1, GEFeS). Adversarial samples performed even worse with a range of 76.00% to 78.57 (ω = 0.0, ASO). The resulting *Use?* values are negative ranging from -0.09 (ω = 0.1, GEFeS) to -0.18 (RAND).

Finally, Table C.17, shows measurements for the Hybrid feature set 175+(org = 200, adv = 200). Baseline accuracies are high at 99.00% for original samples and 89.00% for adversarial samples. Since there is not much headroom for the original samples, feature selection accuracies are only slightly better ranging from 97.27% (ω = 0.9, RAND) to 100.00% (GEFeS). The adversarial sample accuracies show worse performance ranging from 76.50% (ω = 0.3, ASO) to 78.17% (ω = 0.7, ABCO). Therefore, the *Use?* values are negative clustering very close to -0.13.

Table C.2. CASIS-25, LIWC Feature Set - 93 Features, 75+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	68.00%	56.00%	88.80%	35.07%	57.46%	-0.07
	PSO	68.00%	56.00%	77.47%	27.87%	56.16 %	-0.36
	ABCO	68.00%	56.00%	78.00%	27.20%	56.13%	-0.37
	ASO	68.00%	56.00%	75.87%	64.53%	75.88%	0.27
	GSO	68.00%	56.00%	74.40%	28.13%	57.78%	-0.40
	RAND	68.00%	56.00%	73.20%	22.67%	50.25%	-0.52
0.1	GEFeS	68.00%	56.00%	88.00%	17.73%	44.12%	-0.39
	PSO	68.00%	56.00%	78.27%	22.27%	53.66 %	-0.45
	ABCO	68.00%	56.00%	77.07%	23.20%	53.05%	-0.45
	ASO	68.00%	56.00%	75.60%	64.40%	74.41%	0.26
	GSO	68.00%	56.00%	76.80%	27.20%	57.31%	-0.38
	RAND	68.00%	56.00%	72.27%	23.07%	50.18%	-0.52
0.3	GEFeS	68.00%	56.00%	89.07%	17.33%	43.37%	-0.38
	PSO	68.00%	56.00%	78.53%	25.87%	53.66%	-0.38
	ABCO	68.00%	56.00%	76.27%	24.00%	52.83%	-0.45
	ASO	68.00%	56.00%	76.40%	61.87%	72.62%	0.23
	GSO	68.00%	56.00%	72.80%	25.33%	55.70%	-0.48
	RAND	68.00%	56.00%	74.27%	23.07%	50.29%	-0.50
0.5	GEFeS	68.00%	56.00%	88.53%	16.40%	44.73%	-0.41
	PSO	68.00%	56.00%	80.40%	20.93%	50.57%	-0.44
	ABCO	68.00%	56.00%	75.07%	22.40%	52.54%	-0.50
	ASO	68.00%	56.00%	76.00%	61.07%	71.47%	0.21
	GSO	68.00%	56.00%	74.40%	24.67%	55.34%	-0.47
	RAND	68.00%	56.00%	72.67%	26.00%	49.32%	-0.47
0.7	GEFeS	68.00%	56.00%	88.00%	15.73%	48.16%	-0.42
	PSO	68.00%	56.00%	79.33%	23.47%	53.26%	-0.41
	ABCO	68.00%	56.00%	76.00%	24.00%	53.01%	-0.45
	ASO	68.00%	56.00%	75.87%	57.47%	69.61%	0.14
	GSO	68.00%	56.00%	74.13%	23.60%	52.83%	-0.49
	RAND	68.00%	56.00%	74.13%	22.13%	50.75%	-0.51
0.9	GEFeS	68.00%	56.00%	88.40%	16.27%	42.22 %	-0.41
	PSO	68.00%	56.00%	76.53%	22.53%	51.25%	-0.47
	ABCO	68.00%	56.00%	77.73%	22.40%	52.69%	-0.46
	ASO	68.00%	56.00%	76.40%	54.67%	67.10%	0.10
	GSO	68.00%	56.00%	74.80%	19.73%	52.69%	-0.55
	RAND	68.00%	56.00%	72.40%	19.33%	49.43%	-0.59
1.0	GEFeS	68.00%	56.00%	87.73%	16.93%	40.93 %	-0.41
	PSO	68.00%	56.00%	77.20%	21.33%	51.79%	-0.48
	ABCO	68.00%	56.00%	76.13%	17.20%	49.78%	-0.57
	ASO	68.00%	56.00%	75.60%	51.87%	66.63%	0.04
	GSO	68.00%	56.00%	75.33%	23.73%	51.51%	-0.47
	RAND	68.00%	56.00%	73.07%	23.87%	49.68%	-0.50

Table C.3. CASIS-25, Topic Modeling Feature Set - 45 Features, 75+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	84.00%	8.00%	92.00%	11.73%	77.11%	0.56
	PSO	84.00%	8.00%	85.20%	9.73%	73.19%	0.23
	ABCO	84.00%	8.00%	87.07%	10.00%	75.56%	0.29
	ASO	84.00%	8.00%	85.87%	11.87%	85.04%	0.51
	GSO	84.00%	8.00%	85.20%	9.20%	74.74%	0.16
	RAND	84.00%	8.00%	78.53%	8.67%	50.67%	0.02
0.1	GEFeS	84.00%	8.00%	91.33%	10.53%	70.74%	0.40
	PSO	84.00%	8.00%	87.07%	9.33%	70.22%	0.20
	ABCO	84.00%	8.00%	88.13%	9.73%	71.26%	0.27
	ASO	84.00%	8.00%	85.47%	12.00%	84.81%	0.52
	GSO	84.00%	8.00%	84.13%	9.33%	71.70%	0.17
	RAND	84.00%	8.00%	78.53%	8.93%	50.44%	0.05
0.3	GEFeS	84.00%	8.00%	91.47%	11.20%	69.41%	0.49
	PSO	84.00%	8.00%	87.60%	9.33%	71.85%	0.21
	ABCO	84.00%	8.00%	88.27%	9.73%	71.48%	0.27
	ASO	84.00%	8.00%	85.73%	12.00%	84.22%	0.52
	GSO	84.00%	8.00%	84.80%	9.20%	72.22%	0.16
	RAND	84.00%	8.00%	78.80%	8.80%	48.74%	0.04
0.5	GEFeS	84.00%	8.00%	91.87%	11.20%	70.00%	0.49
	PSO	84.00%	8.00%	85.33%	9.60%	70.22%	0.22
	ABCO	84.00%	8.00%	87.87%	9.07%	71.70%	0.18
	ASO	84.00%	8.00%	86.13%	12.00%	84.07%	0.53
	GSO	84.00%	8.00%	85.20%	9.20%	72.89%	0.16
	RAND	84.00%	8.00%	78.67%	9.20%	49.41%	0.09
0.7	GEFeS	84.00%	8.00%	91.47%	10.80%	70.15%	0.44
	PSO	84.00%	8.00%	86.40%	8.67%	70.07%	0.11
	ABCO	84.00%	8.00%	87.87%	9.20%	74.00%	0.20
	ASO	84.00%	8.00%	86.67%	12.00%	83.19%	0.53
	GSO	84.00%	8.00%	84.00%	9.47%	72.37%	0.18
	RAND	84.00%	8.00%	81.33%	9.20%	49.85%	0.12
0.9	GEFeS	84.00%	8.00%	91.47%	11.33%	69.85%	0.51
	PSO	84.00%	8.00%	85.47%	8.93%	68.89%	0.13
	ABCO	84.00%	8.00%	87.73%	9.60%	72.89%	0.24
	ASO	84.00%	8.00%	86.93%	12.00%	83.04%	0.53
	GSO	84.00%	8.00%	85.87%	9.33%	71.93%	0.19
	RAND	84.00%	8.00%	77.47%	8.80%	47.70%	0.02
1.0	GEFeS	84.00%	8.00%	91.47%	10.93%	70.15%	0.46
	PSO	84.00%	8.00%	86.40%	9.33%	71.41%	0.20
	ABCO	84.00%	8.00%	87.47%	9.47%	71.26%	0.22
	ASO	84.00%	8.00%	87.20%	12.00%	82.67%	0.54
	GSO	84.00%	8.00%	85.33%	9.60%	72.89%	0.22
	RAND	84.00%	8.00%	78.93%	8.93%	48.07%	0.06

Table C.4. CASIS-25, Stylometry Feature Set - 428 Features, 75+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	60.00%	32.00%	95.33%	4.27%	48.77%	-0.28
	PSO	60.00%	32.00%	71.47%	4.27%	51.25%	-0.68
	ABCO	60.00%	32.00%	70.27%	4.00%	51.31%	-0.70
	ASO	60.00%	32.00%	79.33%	14.80%	57.31%	-0.22
	GSO	60.00%	32.00%	69.87%	4.13%	51.85%	-0.71
	RAND	60.00%	32.00%	65.87%	4.67%	49.99%	-0.76
0.1	GEFeS	60.00%	32.00%	97.47%	4.13%	34.21%	-0.25
	PSO	60.00%	32.00%	71.07%	4.13%	48.36%	-0.69
	ABCO	60.00%	32.00%	71.07%	4.13%	49.51%	-0.69
	ASO	60.00%	32.00%	80.13%	10.67%	54.05%	-0.33
	GSO	60.00%	32.00%	71.73%	4.40%	51.09%	-0.67
	RAND	60.00%	32.00%	68.93%	4.00%	50.32%	-0.73
0.3	GEFeS	60.00%	32.00%	96.00%	4.00%	32.38%	-0.27
	PSO	60.00%	32.00%	73.33%	4.13%	48.27%	-0.65
	ABCO	60.00%	32.00%	70.67%	4.40%	48.89%	-0.68
	ASO	60.00%	32.00%	79.33%	8.40%	47.47%	-0.42
	GSO	60.00%	32.00%	68.40%	4.27%	49.14%	-0.73
	RAND	60.00%	32.00%	65.73%	4.13%	50.38%	-0.78
0.5	GEFeS	60.00%	32.00%	96.27%	4.13%	29.43%	-0.27
	PSO	60.00%	32.00%	73.87%	4.27%	47.31%	-0.64
	ABCO	60.00%	32.00%	71.47%	4.53%	48.96%	-0.67
	ASO	60.00%	32.00%	80.27%	6.13%	41.64%	-0.47
	GSO	60.00%	32.00%	71.20%	4.00%	49.00%	-0.69
	RAND	60.00%	32.00%	65.87%	4.00%	49.91%	-0.78
0.7	GEFeS	60.00%	32.00%	96.40%	4.00%	26.10%	-0.27
	PSO	60.00%	32.00%	74.00%	4.00%	47.18%	-0.64
	ABCO	60.00%	32.00%	71.20%	4.00%	47.06%	-0.69
	ASO	60.00%	32.00%	81.87%	4.67%	36.51%	-0.49
	GSO	60.00%	32.00%	69.20%	4.00%	48.70%	-0.72
	RAND	60.00%	32.00%	64.40%	4.13%	49.60%	-0.80
0.9	GEFeS	60.00%	32.00%	94.67%	4.00%	23.29%	-0.30
	PSO	60.00%	32.00%	73.73%	4.13%	45.69%	-0.64
	ABCO	60.00%	32.00%	72.53%	4.00%	46.90%	-0.67
	ASO	60.00%	32.00%	82.67%	4.00%	32.76%	-0.50
	GSO	60.00%	32.00%	68.80%	4.00%	46.88%	-0.73
	RAND	60.00%	32.00%	65.73%	4.40%	50.79%	-0.77
1.0	GEFeS	60.00%	32.00%	95.47%	4.00%	21.54%	-0.28
	PSO	60.00%	32.00%	71.20%	4.13%	45.33%	-0.68
	ABCO	60.00%	32.00%	69.47%	4.27%	45.80%	-0.71
	ASO	60.00%	32.00%	83.60%	4.00%	30.97%	-0.48
	GSO	60.00%	32.00%	68.87%	4.27%	46.14%	-0.72
	RAND	60.00%	32.00%	64.53%	4.00%	50.10%	-0.80

Table C.5. CASIS-25, Hybrid Feature Set – 566 Features, 75+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	92.00%	32.00%	99.73%	12.13%	52.09%	-0.54
	PSO	92.00%	32.00%	96.27%	11.20%	51.21%	-0.60
	ABCO	92.00%	32.00%	94.67%	10.93%	51.64%	-0.63
	ASO	92.00%	32.00%	96.13%	7.47%	57.04%	-0.72
	GSO	92.00%	32.00%	95.07%	11.87%	51.48%	-0.60
	RAND	92.00%	32.00%	92.80%	11.60%	50.37%	-0.63
0.1	GEFeS	92.00%	32.00%	100.00%	8.27%	13.47%	-0.65
	PSO	92.00%	32.00%	96.53%	10.53%	47.83%	-0.62
	ABCO	92.00%	32.00%	95.20%	10.93%	49.58%	-0.62
	ASO	92.00%	32.00%	96.13%	6.93%	51.01%	-0.74
	GSO	92.00%	32.00%	94.13%	12.93%	50.21%	-0.57
	RAND	92.00%	32.00%	94.13%	11.73%	49.77%	-0.61
0.3	GEFeS	92.00%	32.00%	100.00%	10.13%	13.12 %	-0.60
	PSO	92.00%	32.00%	96.00%	9.73%	46.77%	-0.65
	ABCO	92.00%	32.00%	95.47%	10.27%	47.89%	-0.64
	ASO	92.00%	32.00%	96.13%	6.93%	39.08%	-0.74
	GSO	92.00%	32.00%	92.80%	12.13%	49.55%	-0.61
	RAND	92.00%	32.00%	93.33%	10.13%	50.22%	-0.67
0.5	GEFeS	92.00%	32.00%	100.00%	9.20%	12.74%	-0.63
	PSO	92.00%	32.00%	95.60%	9.60%	45.47%	-0.66
	ABCO	92.00%	32.00%	95.73%	9.47%	47.35%	-0.66
	ASO	92.00%	32.00%	95.60%	6.80%	29.86%	-0.75
	GSO	92.00%	32.00%	93.60%	9.60%	48.10%	-0.68
	RAND	92.00%	32.00%	92.00%	11.60%	50.24%	-0.64
0.7	GEFeS	92.00%	32.00%	100.00%	9.07%	12.40%	-0.63
	PSO	92.00%	32.00%	96.27%	10.00%	44.69%	-0.64
	ABCO	92.00%	32.00%	94.80%	10.40%	45.94%	-0.64
	ASO	92.00%	32.00%	94.93%	7.07%	22.82%	-0.75
	GSO	92.00%	32.00%	94.13%	8.93%	47.39%	-0.70
	RAND	92.00%	32.00%	90.67%	10.00%	50.17%	-0.70
0.9	GEFeS	92.00%	32.00%	99.87%	8.67%	11.81%	-0.64
	PSO	92.00%	32.00%	93.07%	9.73%	43.82%	-0.68
	ABCO	92.00%	32.00%	94.13%	10.80%	45.27%	-0.64
	ASO	92.00%	32.00%	93.73%	7.73%	17.92%	-0.74
	GSO	92.00%	32.00%	93.20%	10.53%	45.76%	-0.66
	RAND	92.00%	32.00%	90.67%	9.73%	49.67%	-0.71
1.0	GEFeS	92.00%	32.00%	99.87%	8.67%	11.28%	-0.64
	PSO	92.00%	32.00%	93.73%	8.53%	43.42%	-0.71
	ABCO	92.00%	32.00%	93.33%	12.27%	44.59%	-0.60
	ASO	92.00%	32.00%	93.73%	7.07%	16.49%	-0.76
	GSO	92.00%	32.00%	92.00%	12.00%	45.55%	-0.63
	RAND	92.00%	32.00%	90.00%	12.27%	49.91%	-0.64

Table C.6. CASIS-25, LIWC Feature Set - 93 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	92.00%	88.00%	97.20%	83.77%	57.46%	0.00
	PSO	92.00%	88.00%	94.37%	81.97%	56.16 %	-0.05
	ABCO	92.00%	88.00%	94.47%	81.77%	56.13%	-0.05
	ASO	92.00%	88.00%	93.97%	91.13%	75.88%	0.05
	GSO	92.00%	88.00%	93.53%	81.97%	57.78%	-0.06
	RAND	92.00%	88.00%	93.27%	80.63%	50.25%	-0.08
0.1	GEFeS	92.00%	88.00%	97.00%	79.43%	44.12%	-0.05
	PSO	92.00%	88.00%	94.50%	80.50%	53.66 %	-0.07
	ABCO	92.00%	88.00%	94.27%	80.80%	53.05%	-0.07
	ASO	92.00%	88.00%	93.90%	91.10%	74.41%	0.04
	GSO	92.00%	88.00%	94.20%	81.80%	57.31%	-0.06
	RAND	92.00%	88.00%	93.07%	80.83%	50.18%	-0.08
0.3	GEFeS	92.00%	88.00%	97.27%	79.33%	43.37%	-0.05
	PSO	92.00%	88.00%	94.63%	81.47%	53.66%	-0.06
	ABCO	92.00%	88.00%	94.00%	80.90%	52.83%	-0.07
	ASO	92.00%	88.00%	94.10%	90.47%	72.62%	0.04
	GSO	92.00%	88.00%	93.20%	81.33%	55.70%	-0.07
	RAND	92.00%	88.00%	93.50%	80.70%	50.29%	-0.08
0.5	GEFeS	92.00%	88.00%	97.13%	79.10%	44.73%	-0.06
	PSO	92.00%	88.00%	95.07%	80.20%	50.57%	-0.07
	ABCO	92.00%	88.00%	93.77%	80.60%	52.54%	-0.08
	ASO	92.00%	88.00%	94.00%	90.27%	71.47%	0.04
	GSO	92.00%	88.00%	95.53%	81.10%	55.34%	-0.07
	RAND	92.00%	88.00%	93.13%	81.47%	49.32%	-0.07
0.7	GEFeS	92.00%	88.00%	97.00%	78.93%	48.16%	-0.06
	PSO	92.00%	88.00%	94.80%	80.83%	53.26%	-0.06
	ABCO	92.00%	88.00%	93.93%	80.93%	53.01%	-0.07
	ASO	92.00%	88.00%	93.97%	89.37%	69.61%	0.03
	GSO	92.00%	88.00%	93.50%	80.87%	52.83%	-0.08
	RAND	92.00%	88.00%	93.40%	80.40%	50.75%	-0.08
0.9	GEFeS	92.00%	88.00%	97.03%	79.00%	42.22 %	-0.06
	PSO	92.00%	88.00%	94.13%	80.63%	51.25%	-0.07
	ABCO	92.00%	88.00%	94.40%	80.57%	52.69%	-0.07
	ASO	92.00%	88.00%	94.10%	88.67%	67.10%	0.02
	GSO	92.00%	88.00%	93.67%	79.90%	52.69%	-0.08
	RAND	92.00%	88.00%	93.07%	79.80%	49.43%	-0.09
1.0	GEFeS	92.00%	88.00%	96.93%	79.23%	40.93 %	-0.06
	PSO	92.00%	88.00%	94.30%	80.33%	51.79%	-0.07
	ABCO	92.00%	88.00%	94.03%	79.30%	49.78%	-0.09
	ASO	92.00%	88.00%	93.90%	87.97%	66.63%	0.01
	GSO	92.00%	88.00%	93.80%	80.90%	51.51%	-0.07
	RAND	92.00%	88.00%	93.27%	80.97%	49.68%	-0.08

Table C.7. CASIS-25, Topic Modeling Feature Set - 45 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	96.00%	77.00%	96.87%	76.80%	77.11%	0.01
	PSO	96.00%	77.00%	93.73%	74.87%	73.19%	-0.05
	ABCO	96.00%	77.00%	95.03%	75.77%	75.56%	-0.03
	ASO	96.00%	77.00%	96.47%	77.97%	85.04%	0.02
	GSO	96.00%	77.00%	94.30%	75.30%	74.74%	-0.04
	RAND	96.00%	77.00%	91.33%	73.78%	50.67%	-0.09
0.1	GEFeS	96.00%	77.00%	96.23%	76.03%	70.74%	-0.01
	PSO	96.00%	77.00%	93.73%	74.30%	70.22%	-0.06
	ABCO	96.00%	77.00%	94.93%	75.33%	71.26%	-0.03
	ASO	96.00%	77.00%	96.37%	78.00%	84.81%	0.02
	GSO	96.00%	77.00%	93.67%	74.97%	71.70%	-0.05
	RAND	96.00%	77.00%	91.07%	73.67%	50.44%	-0.09
0.3	GEFeS	96.00%	77.00%	96.50%	76.43%	69.41%	0.00
	PSO	96.00%	77.00%	94.33%	74.77%	71.85%	-0.05
	ABCO	96.00%	77.00%	95.00%	75.37%	71.48%	-0.03
	ASO	96.00%	77.00%	96.43%	78.00%	84.22%	0.02
	GSO	96.00%	77.00%	93.53%	74.63%	72.22%	-0.06
	RAND	96.00%	77.00%	91.20%	73.70%	48.74%	-0.09
0.5	GEFeS	96.00%	77.00%	96.53%	76.37%	70.00%	0.00
	PSO	96.00%	77.00%	93.70%	74.77%	70.22%	-0.05
	ABCO	96.00%	77.00%	95.13%	75.43%	71.70%	-0.03
	ASO	96.00%	77.00%	96.53%	78.00%	84.07%	0.02
	GSO	96.00%	77.00%	93.70%	74.70%	72.89%	-0.05
	RAND	96.00%	77.00%	90.87%	73.50%	49.41%	-0.10
0.7	GEFeS	96.00%	77.00%	96.50%	76.33%	70.15%	0.00
	PSO	96.00%	77.00%	94.33%	74.90%	70.07%	-0.04
	ABCO	96.00%	77.00%	94.90%	75.23%	74.00%	-0.03
	ASO	96.00%	77.00%	96.67%	78.00%	83.19%	0.02
	GSO	96.00%	77.00%	93.90%	75.27%	72.37%	-0.04
	RAND	96.00%	77.00%	92.57%	74.53%	49.85%	-0.07
0.9	GEFeS	96.00%	77.00%	96.63%	76.60%	69.85%	0.00
	PSO	96.00%	77.00%	94.00%	74.87%	68.89%	-0.05
	ABCO	96.00%	77.00%	95.43%	75.90%	72.89%	-0.02
	ASO	96.00%	77.00%	96.73%	78.00%	83.04%	0.02
	GSO	96.00%	77.00%	94.37%	75.23%	71.93%	-0.04
	RAND	96.00%	77.00%	90.73%	73.57%	47.70%	-0.10
1.0	GEFeS	96.00%	77.00%	96.23%	75.10%	70.15%	-0.01
	PSO	96.00%	77.00%	94.07%	74.80%	71.41%	-0.05
	ABCO	96.00%	77.00%	94.97%	75.47%	71.26%	-0.03
	ASO	96.00%	77.00%	96.80%	78.00%	82.67%	0.02
	GSO	96.00%	77.00%	94.03%	75.10%	72.89%	-0.05
	RAND	96.00%	77.00%	91.07%	73.57%	48.07%	-0.10

Table C.8. CASIS-25, Stylometry Feature Set - 428 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	90.00%	83.00%	98.83%	76.07%	48.77%	0.01
	PSO	90.00%	83.00%	92.87%	76.07%	51.25%	-0.05
	ABCO	90.00%	83.00%	92.57%	76.00%	51.31%	-0.06
	ASO	90.00%	83.00%	94.83%	78.70%	57.31%	0.00
	GSO	90.00%	83.00%	92.47%	76.03%	51.85%	-0.06
	RAND	90.00%	83.00%	91.47%	76.17%	49.99%	-0.07
0.1	GEFeS	90.00%	83.00%	99.37%	76.03%	34.21%	0.02
	PSO	90.00%	83.00%	92.77%	76.03%	48.36%	-0.05
	ABCO	90.00%	83.00%	92.77%	76.03%	49.51%	-0.05
	ASO	90.00%	83.00%	95.03%	77.67%	54.05%	-0.01
	GSO	90.00%	83.00%	92.93%	76.10%	51.09%	-0.05
	RAND	90.00%	83.00%	92.23%	76.00%	50.32%	-0.06
0.3	GEFeS	90.00%	83.00%	99.00%	76.00%	32.38%	0.02
	PSO	90.00%	83.00%	93.33%	76.03%	48.27%	-0.05
	ABCO	90.00%	83.00%	92.67%	76.10%	48.89%	-0.05
	ASO	90.00%	83.00%	94.83%	77.10%	47.47%	-0.02
	GSO	90.00%	83.00%	92.10%	76.07%	49.14%	-0.06
	RAND	90.00%	83.00%	91.43%	76.03%	50.38%	-0.07
0.5	GEFeS	90.00%	83.00%	99.07%	76.03%	29.43%	0.02
	PSO	90.00%	83.00%	93.47%	76.07%	47.31%	-0.05
	ABCO	90.00%	83.00%	92.87%	76.13%	48.96%	-0.05
	ASO	90.00%	83.00%	95.07%	76.53%	41.64%	-0.02
	GSO	90.00%	83.00%	92.80%	76.00%	49.00%	-0.05
	RAND	90.00%	83.00%	91.47%	76.00%	49.91%	-0.07
0.7	GEFeS	90.00%	83.00%	99.10%	76.00%	26.10%	0.02
	PSO	90.00%	83.00%	93.50%	76.00%	47.18%	-0.05
	ABCO	90.00%	83.00%	92.80%	76.00%	47.06%	-0.05
	ASO	90.00%	83.00%	95.47%	76.17%	36.51%	-0.02
	GSO	90.00%	83.00%	92.30%	76.00%	48.70%	-0.06
	RAND	90.00%	83.00%	91.10%	76.03%	49.60%	-0.07
0.9	GEFeS	90.00%	83.00%	99.67%	76.00%	23.29%	0.01
	PSO	90.00%	83.00%	93.43%	76.03%	45.69%	-0.05
	ABCO	90.00%	83.00%	93.13%	76.00%	46.90%	-0.05
	ASO	90.00%	83.00%	95.67%	76.00%	32.76%	-0.02
	GSO	90.00%	83.00%	92.20%	76.00%	46.88%	-0.06
	RAND	90.00%	83.00%	91.43%	76.10%	50.79%	-0.07
1.0	GEFeS	90.00%	83.00%	99.87%	76.00%	21.54%	0.01
	PSO	90.00%	83.00%	92.80%	76.03%	45.33%	-0.05
	ABCO	90.00%	83.00%	92.37%	76.07%	45.80%	-0.06
	ASO	90.00%	83.00%	95.90%	76.00%	30.97%	-0.02
	GSO	90.00%	83.00%	92.17%	76.07%	46.14%	-0.06
	RAND	90.00%	83.00%	91.13%	76.00%	50.10%	-0.07

Table C.9. CASIS-25, Hybrid Feature Set - 566 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	98.00%	83.00%	99.93%	78.03%	52.09%	-0.04
	PSO	98.00%	83.00%	99.07%	77.80%	51.21%	-0.05
	ABCO	98.00%	83.00%	98.67%	77.73%	51.64%	-0.06
	ASO	98.00%	83.00%	99.03%	76.87%	57.04%	-0.06
	GSO	98.00%	83.00%	98.77%	77.97%	51.48%	-0.05
	RAND	98.00%	83.00%	98.20%	77.90%	50.37%	-0.06
0.1	GEFeS	98.00%	83.00%	100.00%	77.07%	13.47%	-0.05
	PSO	98.00%	83.00%	99.13%	77.63%	47.83%	-0.05
	ABCO	98.00%	83.00%	98.80%	77.73%	49.58%	-0.06
	ASO	98.00%	83.00%	99.03%	76.73%	51.01%	-0.06
	GSO	98.00%	83.00%	98.53%	78.23%	50.21%	-0.06
	RAND	98.00%	83.00%	98.53%	77.93%	49.77%	-0.06
0.3	GEFeS	98.00%	83.00%	100.00%	77.53%	13.12 %	-0.05
	PSO	98.00%	83.00%	99.00%	77.43%	46.77%	-0.06
	ABCO	98.00%	83.00%	98.87%	77.57%	47.89%	-0.06
	ASO	98.00%	83.00%	99.03%	76.73%	39.08%	-0.06
	GSO	98.00%	83.00%	98.20%	78.03%	49.55%	-0.06
	RAND	98.00%	83.00%	98.33%	77.53%	50.22%	-0.06
0.5	GEFeS	98.00%	83.00%	100.00%	77.30%	12.74%	-0.05
	PSO	98.00%	83.00%	98.90%	77.40%	45.47%	-0.06
	ABCO	98.00%	83.00%	98.93%	77.37%	47.35%	-0.06
	ASO	98.00%	83.00%	98.90%	76.70%	29.86%	-0.06
	GSO	98.00%	83.00%	98.40%	77.40%	48.10%	-0.06
	RAND	98.00%	83.00%	98.00%	77.90%	50.24%	-0.07
0.7	GEFeS	98.00%	83.00%	100.00%	77.27%	12.40%	-0.05
	PSO	98.00%	83.00%	99.07%	77.50%	44.69%	-0.06
	ABCO	98.00%	83.00%	98.70%	77.60%	45.94%	-0.06
	ASO	98.00%	83.00%	98.73%	76.77%	22.82%	-0.07
	GSO	98.00%	83.00%	98.53%	77.23%	47.39%	-0.06
	RAND	98.00%	83.00%	97.67%	77.50%	50.17%	-0.07
0.9	GEFeS	98.00%	83.00%	99.97%	77.17%	11.81%	-0.05
	PSO	98.00%	83.00%	98.27%	77.43%	43.82%	-0.06
	ABCO	98.00%	83.00%	98.53%	77.70%	45.27%	-0.06
	ASO	98.00%	83.00%	98.43%	76.93%	17.92%	-0.07
	GSO	98.00%	83.00%	98.30%	77.63%	45.76%	-0.06
	RAND	98.00%	83.00%	97.67%	77.43%	49.67%	-0.07
1.0	GEFeS	98.00%	83.00%	99.97%	77.17%	11.28%	-0.05
	PSO	98.00%	83.00%	98.43%	77.13%	43.42%	-0.07
	ABCO	98.00%	83.00%	98.33%	78.07%	44.59%	-0.06
	ASO	98.00%	83.00%	98.43%	76.77%	16.49%	-0.07
	GSO	98.00%	83.00%	98.00%	78.00%	45.55%	-0.06
	RAND	98.00%	83.00%	97.50%	78.07%	49.91%	-0.06

Table C.10. CASIS-50, LIWC Feature Set - 93 Features, 175+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	68.00%	44.00%	89.97%	34.00%	57.67%	0.09
	PSO	68.00%	44.00%	78.53%	25.07%	55.66%	-0.28
	ABCO	68.00%	44.00%	75.20%	25.60%	57.74%	-0.31
	ASO	68.00%	44.00%	75.73%	64.13%	75.91%	0.57
	GSO	68.00%	44.00%	73.47%	25.47%	57.85%	-0.34
	RAND	68.00%	44.00%	72.00%	23.60%	48.57%	-0.40
0.1	GEFeS	68.00%	44.00%	88.40%	13.33%	43.19%	-0.40
	PSO	68.00%	44.00%	78.00%	24.93%	52.76%	-0.29
	ABCO	68.00%	44.00%	78.00%	22.80%	52.62%	-0.33
	ASO	68.00%	44.00%	76.00%	63.47%	74.87%	0.56
	GSO	68.00%	44.00%	75.87%	26.53%	56.88%	-0.28
	RAND	68.00%	44.00%	74.53%	22.67%	49.68%	-0.39
0.3	GEFeS	68.00%	44.00%	89.07%	16.13%	45.05%	-0.32
	PSO	68.00%	44.00%	80.13%	25.60%	51.72%	-0.24
	ABCO	68.00%	44.00%	77.60%	23.47%	54.44%	-0.33
	ASO	68.00%	44.00%	76.13%	61.60%	72.80%	0.52
	GSO	68.00%	44.00%	75.73%	23.07%	54.91%	-0.36
	RAND	68.00%	44.00%	72.67%	22.93%	50.11%	-0.41
0.5	GEFeS	68.00%	44.00%	88.00%	19.73%	42.97%	-0.26
	PSO	68.00%	44.00%	76.40%	21.07%	53.80%	-0.40
	ABCO	68.00%	44.00%	77.73%	18.67%	53.84%	-0.43
	ASO	68.00%	44.00%	76.40%	59.87%	71.43%	0.48
	GSO	68.00%	44.00%	75.33%	23.20%	53.66%	-0.36
	RAND	68.00%	44.00%	70.80%	19.20%	49.43%	-0.52
0.7	GEFeS	68.00%	44.00%	89.47%	16.00%	42.44%	-0.32
	PSO	68.00%	44.00%	77.87%	22.53%	51.47%	-0.34
	ABCO	68.00%	44.00%	77.47%	19.20%	53.76%	-0.42
	ASO	68.00%	44.00%	76.67%	56.67%	69.39%	0.42
	GSO	68.00%	44.00%	73.47%	23.13%	53.91%	-0.42
	RAND	68.00%	44.00%	71.33%	21.73%	50.32%	-0.46
0.9	GEFeS	68.00%	44.00%	88.00%	16.80%	42.01%	-0.32
	PSO	68.00%	44.00%	77.07%	20.13%	51.22%	-0.41
	ABCO	68.00%	44.00%	77.20%	20.13%	51.40%	-0.41
	ASO	68.00%	44.00%	75.33%	53.60%	67.60%	0.33
	GSO	68.00%	44.00%	74.27%	25.87%	53.12%	-0.32
	RAND	68.00%	44.00%	73.47%	23.20%	50.47%	-0.39
1.0	GEFeS	68.00%	44.00%	87.87%	14.93%	41.86%	-0.37
	PSO	68.00%	44.00%	77.20%	21.60%	51.72%	-0.37
	ABCO	68.00%	44.00%	78.40%	22.53%	51.72%	-0.33
	ASO	68.00%	44.00%	75.69%	50.80%	66.52%	0.27
	GSO	68.00%	44.00%	75.33%	23.33%	54.12%	-0.36
	RAND	68.00%	44.00%	71.47%	19.07%	50.90%	-0.52

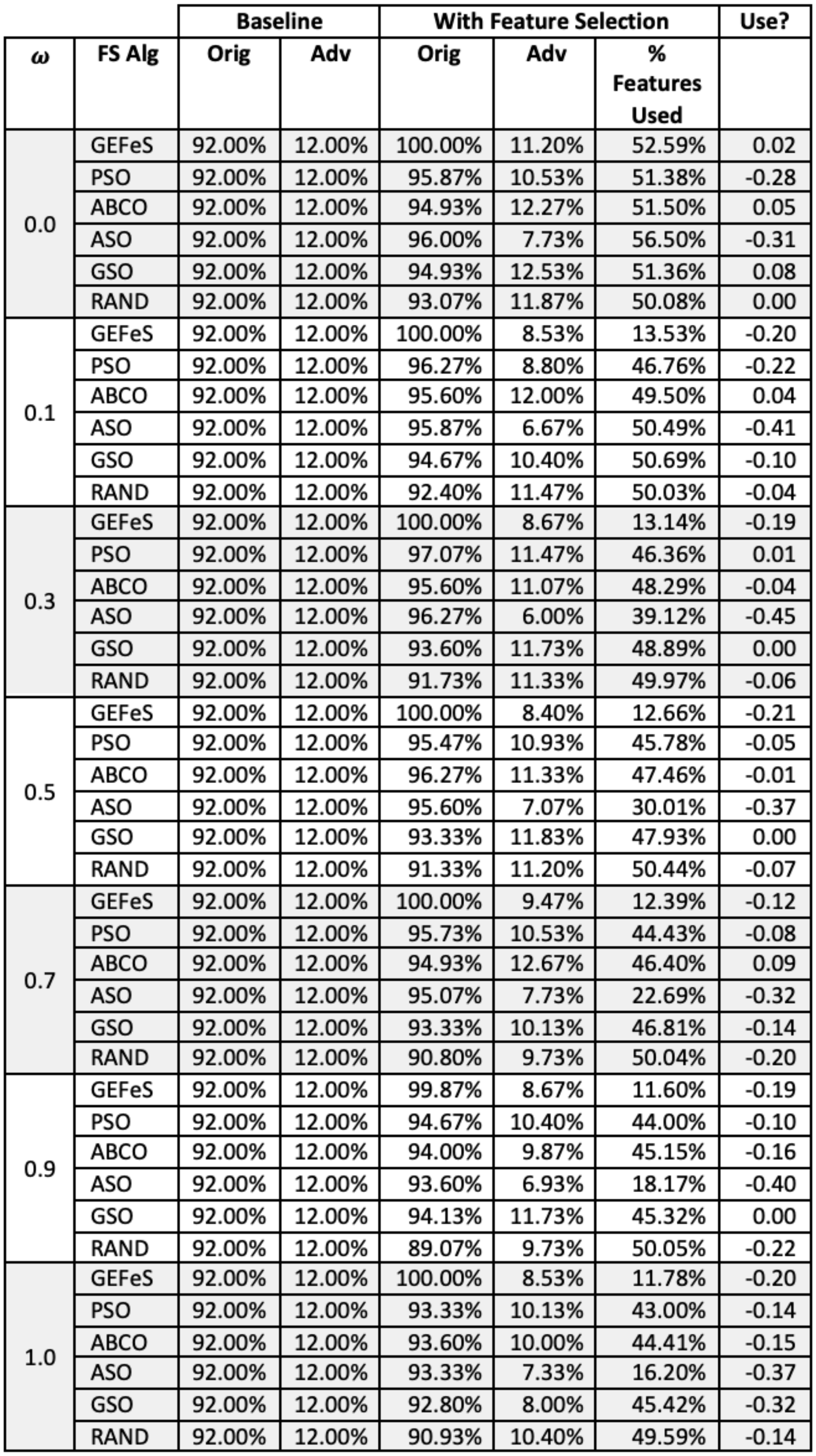
Table C.11. CASIS-50, Topic Modeling Feature Set - 45 Features, 175+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	40.00%	0.00%	91.47%	10.80%	76.96%	1.29
	PSO	40.00%	0.00%	87.07%	10.27%	73.26%	1.18
	ABCO	40.00%	0.00%	87.20%	9.60%	74.89%	1.18
	ASO	40.00%	0.00%	85.73%	11.73%	84.96%	1.14
	GSO	40.00%	0.00%	84.67%	9.20%	72.59%	1.12
	RAND	40.00%	0.00%	79.60%	9.47%	74.89%	0.99
0.1	GEFeS	40.00%	0.00%	91.47%	10.67%	70.15%	1.29
	PSO	40.00%	0.00%	84.93%	9.07%	70.52%	1.12
	ABCO	40.00%	0.00%	87.20%	9.60%	73.26%	1.18
	ASO	40.00%	0.00%	85.87%	12.00%	85.33%	1.15
	GSO	40.00%	0.00%	84.80%	9.87%	73.78%	1.12
	RAND	40.00%	0.00%	79.33%	9.73%	73.26%	0.98
0.3	GEFeS	40.00%	0.00%	91.87%	10.67%	71.04%	1.30
	PSO	40.00%	0.00%	85.60%	10.00%	70.74%	1.14
	ABCO	40.00%	0.00%	88.47%	9.87%	72.52%	1.19
	ASO	40.00%	0.00%	85.73%	12.00%	84.44%	1.14
	GSO	40.00%	0.00%	86.00%	9.33%	75.70%	1.15
	RAND	40.00%	0.00%	79.33%	9.20%	72.52%	0.98
0.5	GEFeS	40.00%	0.00%	91.60%	10.93%	70.00%	1.29
	PSO	40.00%	0.00%	85.07%	9.20%	70.44%	1.13
	ABCO	40.00%	0.00%	86.80%	10.00%	72.44%	1.17
	ASO	40.00%	0.00%	86.40%	11.87%	83.33%	1.16
	GSO	40.00%	0.00%	83.87%	9.60%	73.04%	1.10
	RAND	40.00%	0.00%	80.13%	9.33%	72.44%	1.00
0.7	GEFeS	40.00%	0.00%	91.20%	11.60%	70.30%	1.28
	PSO	40.00%	0.00%	87.20%	9.33%	70.89%	1.18
	ABCO	40.00%	0.00%	88.00%	9.87%	71.70%	1.20
	ASO	40.00%	0.00%	86.40%	12.00%	83.19%	1.16
	GSO	40.00%	0.00%	85.33%	8.93%	71.19%	1.13
	RAND	40.00%	0.00%	78.80%	9.20%	71.70%	0.97
0.9	GEFeS	40.00%	0.00%	92.13%	10.53%	70.30%	1.30
	PSO	40.00%	0.00%	84.53%	9.20%	71.85%	1.11
	ABCO	40.00%	0.00%	87.20%	9.47%	72.30%	1.18
	ASO	40.00%	0.00%	87.07%	12.00%	82.81%	1.18
	GSO	40.00%	0.00%	86.13%	9.60%	71.48%	1.15
	RAND	40.00%	0.00%	79.87%	10.00%	72.30%	1.00
1.0	GEFeS	40.00%	0.00%	92.13%	10.67%	69.85%	1.30
	PSO	40.00%	0.00%	87.73%	9.60%	69.19%	1.19
	ABCO	40.00%	0.00%	88.27%	9.73%	72.15%	1.21
	ASO	40.00%	0.00%	87.33%	12.00%	83.11%	1.18
	GSO	40.00%	0.00%	83.87%	10.13%	72.81%	1.10
	RAND	40.00%	0.00%	78.93%	9.33%	72.15%	0.97

Table C.12. CASIS-50, Stylometry Feature Set - 428 Features, 175+(org = 25, adv = 25).

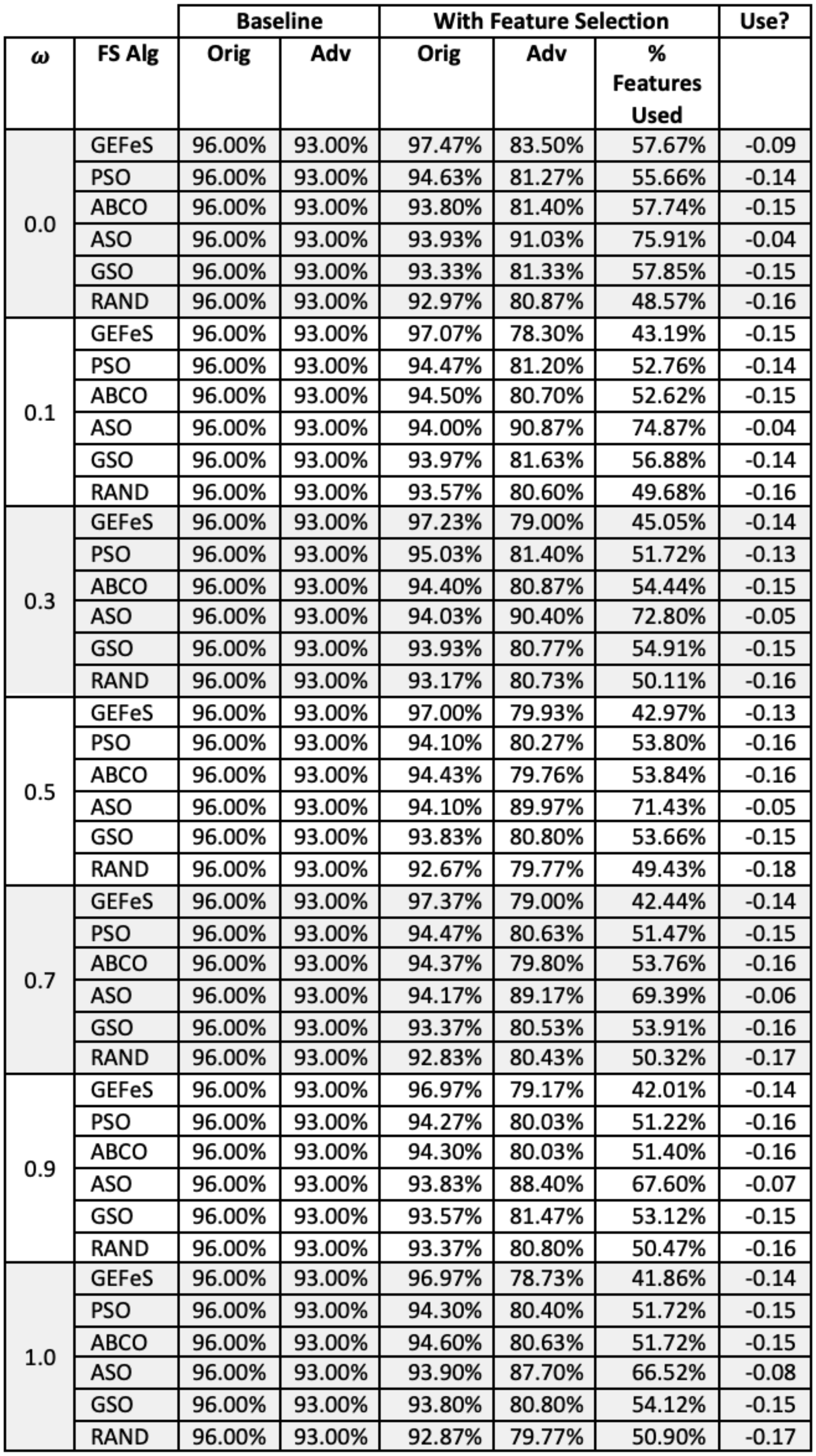
		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	48.00%	20.00%	94.80%	4.13%	49.55%	0.18
	PSO	48.00%	20.00%	72.93%	4.40%	50.79%	-0.26
	ABCO	48.00%	20.00%	69.87%	4.27%	51.21%	-0.33
	ASO	48.00%	20.00%	79.33%	14.27%	57.25%	0.37
	GSO	48.00%	20.00%	72.93%	4.13%	51.51%	-0.27
	RAND	48.00%	20.00%	63.73%	4.00%	50.07%	-0.47
0.1	GEFeS	48.00%	20.00%	96.93%	4.00%	34.35%	0.22
	PSO	48.00%	20.00%	72.93%	4.13%	48.47%	-0.27
	ABCO	48.00%	20.00%	71.73%	4.20%	49.09%	-0.29
	ASO	48.00%	20.00%	79.33%	10.93%	54.05%	0.20
	GSO	48.00%	20.00%	69.60%	4.40%	51.27%	-0.33
	RAND	48.00%	20.00%	66.93%	4.40%	50.40%	-0.39
0.3	GEFeS	48.00%	20.00%	96.27%	4.00%	32.54%	0.21
	PSO	48.00%	20.00%	75.47%	4.00%	48.12%	-0.23
	ABCO	48.00%	20.00%	69.87%	4.27%	49.31%	-0.33
	ASO	48.00%	20.00%	79.07%	9.07%	47.10%	0.10
	GSO	48.00%	20.00%	71.07%	4.27%	49.20%	-0.31
	RAND	48.00%	20.00%	68.80%	4.00%	49.70%	-0.37
0.5	GEFeS	48.00%	20.00%	96.27%	4.00%	29.88%	0.21
	PSO	48.00%	20.00%	74.80%	4.93%	47.74%	-0.20
	ABCO	48.00%	20.00%	70.93%	4.27%	48.15%	-0.31
	ASO	48.00%	20.00%	81.13%	5.87%	41.29%	0.04
	GSO	48.00%	20.00%	68.67%	4.27%	48.36%	-0.36
	RAND	48.00%	20.00%	65.07%	4.80%	49.82%	-0.40
0.7	GEFeS	48.00%	20.00%	96.00%	4.27%	26.00%	0.21
	PSO	48.00%	20.00%	75.47%	4.40%	46.97%	-0.21
	ABCO	48.00%	20.00%	72.53%	4.27%	47.98%	-0.28
	ASO	48.00%	20.00%	81.20%	4.93%	36.21%	0.06
	GSO	48.00%	20.00%	71.07%	4.13%	47.62%	-0.31
	RAND	48.00%	20.00%	66.53%	4.40%	50.19%	-0.39
0.9	GEFeS	48.00%	20.00%	96.73%	4.13%	23.16%	0.20
	PSO	48.00%	20.00%	71.73%	4.00%	46.50%	-0.31
	ABCO	48.00%	20.00%	70.00%	4.00%	46.44%	-0.34
	ASO	48.00%	20.00%	83.47%	4.00%	32.22%	0.06
	GSO	48.00%	20.00%	68.80%	4.27%	47.30%	-0.35
	RAND	48.00%	20.00%	66.20%	4.40%	50.97%	-0.42
1.0	GEFeS	48.00%	20.00%	96.40%	4.13%	21.81%	0.22
	PSO	48.00%	20.00%	74.40%	4.13%	45.56%	-0.24
	ABCO	48.00%	20.00%	69.47%	4.00%	46.16%	-0.35
	ASO	48.00%	20.00%	83.60%	4.00%	30.50%	0.06
	GSO	48.00%	20.00%	68.40%	4.27%	47.23%	-0.36
	RAND	48.00%	20.00%	64.67%	4.27%	49.78%	-0.44

Table C.13. CASIS-50, Hybrid Feature Set - 566 Features, 175+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	92.00%	12.00%	100.00%	11.20%	52.59%	0.02
	PSO	92.00%	12.00%	95.87%	10.53%	51.38%	-0.28
	ABCO	92.00%	12.00%	94.93%	12.27%	51.50%	0.05
	ASO	92.00%	12.00%	96.00%	7.73%	56.50%	-0.31
	GSO	92.00%	12.00%	94.93%	12.53%	51.36%	0.08
	RAND	92.00%	12.00%	93.07%	11.87%	50.08%	0.00
0.1	GEFeS	92.00%	12.00%	100.00%	8.53%	13.53%	-0.20
	PSO	92.00%	12.00%	96.27%	8.80%	46.76%	-0.22
	ABCO	92.00%	12.00%	95.60%	12.00%	49.50%	0.04
	ASO	92.00%	12.00%	95.87%	6.67%	50.49%	-0.41
	GSO	92.00%	12.00%	94.67%	10.40%	50.69%	-0.10
	RAND	92.00%	12.00%	92.40%	11.47%	50.03%	-0.04
0.3	GEFeS	92.00%	12.00%	100.00%	8.67%	13.14%	-0.19
	PSO	92.00%	12.00%	97.07%	11.47%	46.36%	0.01
	ABCO	92.00%	12.00%	95.60%	11.07%	48.29%	-0.04
	ASO	92.00%	12.00%	96.27%	6.00%	39.12%	-0.45
	GSO	92.00%	12.00%	93.60%	11.73%	48.89%	0.00
	RAND	92.00%	12.00%	91.73%	11.33%	49.97%	-0.06
0.5	GEFeS	92.00%	12.00%	100.00%	8.40%	12.66%	-0.21
	PSO	92.00%	12.00%	95.47%	10.93%	45.78%	-0.05
	ABCO	92.00%	12.00%	96.27%	11.33%	47.46%	-0.01
	ASO	92.00%	12.00%	95.60%	7.07%	30.01%	-0.37
	GSO	92.00%	12.00%	93.33%	11.83%	47.93%	0.00
	RAND	92.00%	12.00%	91.33%	11.20%	50.44%	-0.07
0.7	GEFeS	92.00%	12.00%	100.00%	9.47%	12.39%	-0.12
	PSO	92.00%	12.00%	95.73%	10.53%	44.43%	-0.08
	ABCO	92.00%	12.00%	94.93%	12.67%	46.40%	0.09
	ASO	92.00%	12.00%	95.07%	7.73%	22.69%	-0.32
	GSO	92.00%	12.00%	93.33%	10.13%	46.81%	-0.14
	RAND	92.00%	12.00%	90.80%	9.73%	50.04%	-0.20
0.9	GEFeS	92.00%	12.00%	99.87%	8.67%	11.60%	-0.19
	PSO	92.00%	12.00%	94.67%	10.40%	44.00%	-0.10
	ABCO	92.00%	12.00%	94.00%	9.87%	45.15%	-0.16
	ASO	92.00%	12.00%	93.60%	6.93%	18.17%	-0.40
	GSO	92.00%	12.00%	94.13%	11.73%	45.32%	0.00
	RAND	92.00%	12.00%	89.07%	9.73%	50.05%	-0.22
1.0	GEFeS	92.00%	12.00%	100.00%	8.53%	11.78%	-0.20
	PSO	92.00%	12.00%	93.33%	10.13%	43.00%	-0.14
	ABCO	92.00%	12.00%	93.60%	10.00%	44.41%	-0.15
	ASO	92.00%	12.00%	93.33%	7.33%	16.20%	-0.37
	GSO	92.00%	12.00%	92.80%	8.00%	45.42%	-0.32
	RAND	92.00%	12.00%	90.93%	10.40%	49.59%	-0.14

Table C.14. CASIS-50, LIWC Feature Set - 93 Features, 175+(org = 200, adv = 200).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	96.00%	93.00%	97.47%	83.50%	57.67%	-0.09
	PSO	96.00%	93.00%	94.63%	81.27%	55.66%	-0.14
	ABCO	96.00%	93.00%	93.80%	81.40%	57.74%	-0.15
	ASO	96.00%	93.00%	93.93%	91.03%	75.91%	-0.04
	GSO	96.00%	93.00%	93.33%	81.33%	57.85%	-0.15
	RAND	96.00%	93.00%	92.97%	80.87%	48.57%	-0.16
0.1	GEFeS	96.00%	93.00%	97.07%	78.30%	43.19%	-0.15
	PSO	96.00%	93.00%	94.47%	81.20%	52.76%	-0.14
	ABCO	96.00%	93.00%	94.50%	80.70%	52.62%	-0.15
	ASO	96.00%	93.00%	94.00%	90.87%	74.87%	-0.04
	GSO	96.00%	93.00%	93.97%	81.63%	56.88%	-0.14
	RAND	96.00%	93.00%	93.57%	80.60%	49.68%	-0.16
0.3	GEFeS	96.00%	93.00%	97.23%	79.00%	45.05%	-0.14
	PSO	96.00%	93.00%	95.03%	81.40%	51.72%	-0.13
	ABCO	96.00%	93.00%	94.40%	80.87%	54.44%	-0.15
	ASO	96.00%	93.00%	94.03%	90.40%	72.80%	-0.05
	GSO	96.00%	93.00%	93.93%	80.77%	54.91%	-0.15
	RAND	96.00%	93.00%	93.17%	80.73%	50.11%	-0.16
0.5	GEFeS	96.00%	93.00%	97.00%	79.93%	42.97%	-0.13
	PSO	96.00%	93.00%	94.10%	80.27%	53.80%	-0.16
	ABCO	96.00%	93.00%	94.43%	79.76%	53.84%	-0.16
	ASO	96.00%	93.00%	94.10%	89.97%	71.43%	-0.05
	GSO	96.00%	93.00%	93.83%	80.80%	53.66%	-0.15
	RAND	96.00%	93.00%	92.67%	79.77%	49.43%	-0.18
0.7	GEFeS	96.00%	93.00%	97.37%	79.00%	42.44%	-0.14
	PSO	96.00%	93.00%	94.47%	80.63%	51.47%	-0.15
	ABCO	96.00%	93.00%	94.37%	79.80%	53.76%	-0.16
	ASO	96.00%	93.00%	94.17%	89.17%	69.39%	-0.06
	GSO	96.00%	93.00%	93.37%	80.53%	53.91%	-0.16
	RAND	96.00%	93.00%	92.83%	80.43%	50.32%	-0.17
0.9	GEFeS	96.00%	93.00%	96.97%	79.17%	42.01%	-0.14
	PSO	96.00%	93.00%	94.27%	80.03%	51.22%	-0.16
	ABCO	96.00%	93.00%	94.30%	80.03%	51.40%	-0.16
	ASO	96.00%	93.00%	93.83%	88.40%	67.60%	-0.07
	GSO	96.00%	93.00%	93.57%	81.47%	53.12%	-0.15
	RAND	96.00%	93.00%	93.37%	80.80%	50.47%	-0.16
1.0	GEFeS	96.00%	93.00%	96.97%	78.73%	41.86%	-0.14
	PSO	96.00%	93.00%	94.30%	80.40%	51.72%	-0.15
	ABCO	96.00%	93.00%	94.60%	80.63%	51.72%	-0.15
	ASO	96.00%	93.00%	93.90%	87.70%	66.52%	-0.08
	GSO	96.00%	93.00%	93.80%	80.80%	54.12%	-0.15
	RAND	96.00%	93.00%	92.87%	79.77%	50.90%	-0.17

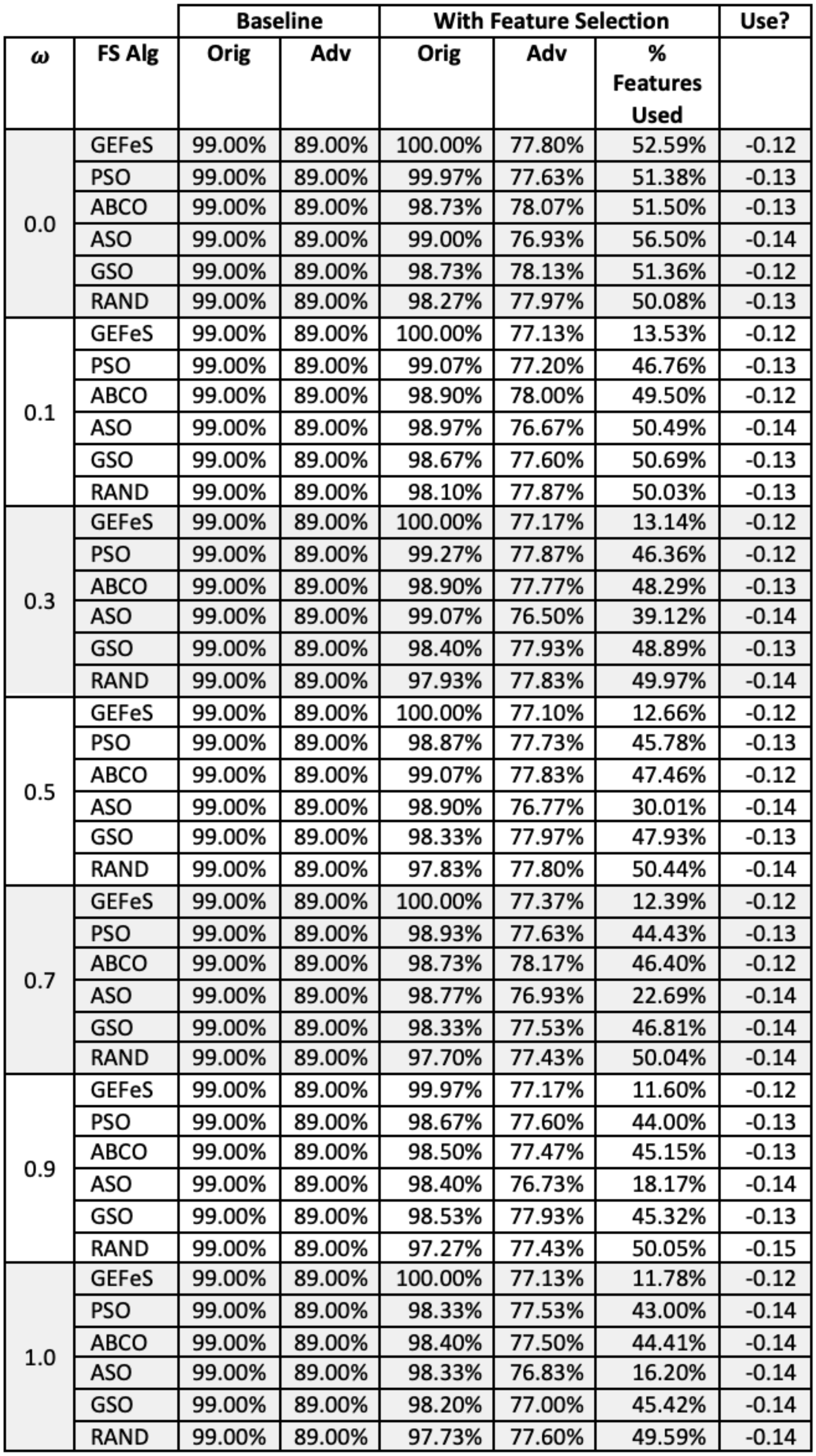
Table C.15. CASIS-50, Topic Modeling Feature Set - 45 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	85.00%	80.00%	96.90%	76.73%	76.96%	0.10
	PSO	85.00%	80.00%	94.87%	75.67%	73.26%	0.06
	ABCO	85.00%	80.00%	95.53%	76.13%	74.89%	0.08
	ASO	85.00%	80.00%	96.43%	77.93%	84.96%	0.11
	GSO	85.00%	80.00%	94.37%	75.50%	72.59%	0.05
	RAND	85.00%	80.00%	92.03%	74.50%	74.89%	0.01
0.1	GEFeS	85.00%	80.00%	96.10%	75.90%	70.15%	0.08
	PSO	85.00%	80.00%	93.67%	74.70%	70.52%	0.04
	ABCO	85.00%	80.00%	94.93%	75.53%	73.26%	0.06
	ASO	85.00%	80.00%	96.47%	78.00%	85.33%	0.11
	GSO	85.00%	80.00%	94.03%	75.30%	73.78%	0.05
	RAND	85.00%	80.00%	91.37%	73.97%	73.26%	0.00
0.3	GEFeS	85.00%	80.00%	96.53%	76.23%	71.04%	0.09
	PSO	85.00%	80.00%	93.77%	74.87%	70.74%	0.04
	ABCO	85.00%	80.00%	95.00%	75.60%	72.52%	0.06
	ASO	85.00%	80.00%	96.43%	78.00%	84.44%	0.11
	GSO	85.00%	80.00%	94.53%	75.%37	75.70%	0.05
	RAND	85.00%	80.00%	91.13%	73.60%	72.52%	-0.01
0.5	GEFeS	85.00%	80.00%	96.47%	76.30%	70.00%	0.09
	PSO	85.00%	80.00%	93.30%	74.33%	70.44%	0.03
	ABCO	85.00%	80.00%	94.73%	75.53%	72.44%	0.06
	ASO	85.00%	80.00%	96.60%	77.97%	83.33%	0.11
	GSO	85.00%	80.00%	93.50%	74.93%	73.04%	0.04
	RAND	85.00%	80.00%	91.40%	73.70%	72.44%	0.00
0.7	GEFeS	85.00%	80.00%	96.37%	76.47%	70.30%	0.09
	PSO	85.00%	80.00%	93.93%	74.47%	70.89%	0.04
	ABCO	85.00%	80.00%	95.33%	75.80%	71.70%	0.07
	ASO	85.00%	80.00%	96.60%	78.00%	83.19%	0.11
	GSO	85.00%	80.00%	94.03%	74.93%	71.19%	0.04
	RAND	85.00%	80.00%	90.63%	73.23%	71.70%	-0.02
0.9	GEFeS	85.00%	80.00%	96.30%	75.90%	70.30%	0.08
	PSO	85.00%	80.00%	93.23%	74.40%	71.85%	0.03
	ABCO	85.00%	80.00%	94.63%	75.20%	72.30%	0.05
	ASO	85.00%	80.00%	96.77%	78.00%	82.81%	0.11
	GSO	85.00%	80.00%	94.63%	75.50%	71.48%	0.06
	RAND	85.00%	80.00%	91.47%	74.00%	72.30%	0.00
1.0	GEFeS	85.00%	80.00%	96.57%	76.20%	69.85%	0.09
	PSO	85.00%	80.00%	94.97%	75.43%	69.19%	0.06
	ABCO	85.00%	80.00%	95.30%	75.67%	72.15%	0.07
	ASO	85.00%	80.00%	96.83%	78.00%	83.11%	0.11
	GSO	85.00%	80.00%	93.33%	74.90%	72.81%	0.03
	RAND	85.00%	80.00%	91.17%	73.77%	72.15%	-0.01

Table C.16. CASIS-50, Stylometry Feature Set - 428 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	93.50%	90.00%	98.70%	76.03%	49.55%	-0.10
	PSO	93.50%	90.00%	93.23%	76.10%	50.79%	-0.16
	ABCO	93.50%	90.00%	92.47%	76.07%	51.21%	-0.17
	ASO	93.50%	90.00%	94.83%	78.57%	57.25%	-0.11
	GSO	93.50%	90.00%	93.23%	76.03%	51.51%	-0.16
	RAND	93.50%	90.00%	90.93%	76.00%	50.07%	-0.18
0.1	GEFeS	93.50%	90.00%	99.23%	76.00%	34.35%	-0.09
	PSO	93.50%	90.00%	93.23%	76.03%	48.47%	-0.16
	ABCO	93.50%	90.00%	92.93%	76.10%	49.09%	-0.16
	ASO	93.50%	90.00%	94.83%	77.73%	54.05%	-0.17
	GSO	93.50%	90.00%	92.40%	76.10%	51.27%	-0.17
	RAND	93.50%	90.00%	91.73%	76.10%	50.40%	-0.17
0.3	GEFeS	93.50%	90.00%	99.07%	76.00%	32.54%	-0.10
	PSO	93.50%	90.00%	93.87%	76.00%	48.12%	-0.15
	ABCO	93.50%	90.00%	92.47%	76.07%	49.31%	-0.13
	ASO	93.50%	90.00%	94.77%	77.27%	47.10%	-0.12
	GSO	93.50%	90.00%	92.77%	76.07%	49.20%	-0.16
	RAND	93.50%	90.00%	92.20%	76.00%	49.70%	-0.17
0.5	GEFeS	93.50%	90.00%	99.07%	76.00%	29.88%	-0.10
	PSO	93.50%	90.00%	93.70%	76.23%	47.74%	-0.15
	ABCO	93.50%	90.00%	92.73%	76.07%	48.15%	-0.13
	ASO	93.50%	90.00%	95.03%	76.47%	41.29%	-0.17
	GSO	93.50%	90.00%	92.17%	76.07%	48.36%	-0.17
	RAND	93.50%	90.00%	91.27%	76.20%	49.82%	-0.18
0.7	GEFeS	93.50%	90.00%	99.00%	76.07%	26.00%	-0.10
	PSO	93.50%	90.00%	93.87%	76.10%	46.97%	-0.15
	ABCO	93.50%	90.00%	93.13%	76.07%	47.98%	-0.16
	ASO	93.50%	90.00%	95.30%	76.23%	36.21%	-0.13
	GSO	93.50%	90.00%	92.77%	76.03%	47.62%	-0.16
	RAND	93.50%	90.00%	91.63%	76.10%	50.19%	-0.17
0.9	GEFeS	93.50%	90.00%	98.93%	76.03%	23.16%	-0.10
	PSO	93.50%	90.00%	92.93%	76.00%	46.50%	-0.16
	ABCO	93.50%	90.00%	92.50%	76.00%	46.44%	-0.17
	ASO	93.50%	90.00%	95.87%	76.00%	32.22%	-0.13
	GSO	93.50%	90.00%	92.20%	76.07%	47.30%	-0.17
	RAND	93.50%	90.00%	91.30%	76.10%	50.97%	-0.18
1.0	GEFeS	93.50%	90.00%	99.10%	76.03%	21.81%	-0.10
	PSO	93.50%	90.00%	93.60%	76.03%	45.56%	-0.15
	ABCO	93.50%	90.00%	92.37%	76.00%	46.16%	-0.17
	ASO	93.50%	90.00%	95.90%	76.00%	30.50%	-0.13
	GSO	93.50%	90.00%	92.10%	76.07%	47.23%	-0.17
	RAND	93.50%	90.00%	91.17%	76.07%	49.78%	-0.18

Table C.17. CASIS-50, Hybrid Feature Set - 566 Features, 175+(org = 200, adv = 200).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	99.00%	89.00%	100.00%	77.80%	52.59%	-0.12
	PSO	99.00%	89.00%	99.97%	77.63%	51.38%	-0.13
	ABCO	99.00%	89.00%	98.73%	78.07%	51.50%	-0.13
	ASO	99.00%	89.00%	99.00%	76.93%	56.50%	-0.14
	GSO	99.00%	89.00%	98.73%	78.13%	51.36%	-0.12
	RAND	99.00%	89.00%	98.27%	77.97%	50.08%	-0.13
0.1	GEFeS	99.00%	89.00%	100.00%	77.13%	13.53%	-0.12
	PSO	99.00%	89.00%	99.07%	77.20%	46.76%	-0.13
	ABCO	99.00%	89.00%	98.90%	78.00%	49.50%	-0.12
	ASO	99.00%	89.00%	98.97%	76.67%	50.49%	-0.14
	GSO	99.00%	89.00%	98.67%	77.60%	50.69%	-0.13
	RAND	99.00%	89.00%	98.10%	77.87%	50.03%	-0.13
0.3	GEFeS	99.00%	89.00%	100.00%	77.17%	13.14%	-0.12
	PSO	99.00%	89.00%	99.27%	77.87%	46.36%	-0.12
	ABCO	99.00%	89.00%	98.90%	77.77%	48.29%	-0.13
	ASO	99.00%	89.00%	99.07%	76.50%	39.12%	-0.14
	GSO	99.00%	89.00%	98.40%	77.93%	48.89%	-0.13
	RAND	99.00%	89.00%	97.93%	77.83%	49.97%	-0.14
0.5	GEFeS	99.00%	89.00%	100.00%	77.10%	12.66%	-0.12
	PSO	99.00%	89.00%	98.87%	77.73%	45.78%	-0.13
	ABCO	99.00%	89.00%	99.07%	77.83%	47.46%	-0.12
	ASO	99.00%	89.00%	98.90%	76.77%	30.01%	-0.14
	GSO	99.00%	89.00%	98.33%	77.97%	47.93%	-0.13
	RAND	99.00%	89.00%	97.83%	77.80%	50.44%	-0.14
0.7	GEFeS	99.00%	89.00%	100.00%	77.37%	12.39%	-0.12
	PSO	99.00%	89.00%	98.93%	77.63%	44.43%	-0.13
	ABCO	99.00%	89.00%	98.73%	78.17%	46.40%	-0.12
	ASO	99.00%	89.00%	98.77%	76.93%	22.69%	-0.14
	GSO	99.00%	89.00%	98.33%	77.53%	46.81%	-0.14
	RAND	99.00%	89.00%	97.70%	77.43%	50.04%	-0.14
0.9	GEFeS	99.00%	89.00%	99.97%	77.17%	11.60%	-0.12
	PSO	99.00%	89.00%	98.67%	77.60%	44.00%	-0.13
	ABCO	99.00%	89.00%	98.50%	77.47%	45.15%	-0.13
	ASO	99.00%	89.00%	98.40%	76.73%	18.17%	-0.14
	GSO	99.00%	89.00%	98.53%	77.93%	45.32%	-0.13
	RAND	99.00%	89.00%	97.27%	77.43%	50.05%	-0.15
1.0	GEFeS	99.00%	89.00%	100.00%	77.13%	11.78%	-0.12
	PSO	99.00%	89.00%	98.33%	77.53%	43.00%	-0.14
	ABCO	99.00%	89.00%	98.40%	77.50%	44.41%	-0.14
	ASO	99.00%	89.00%	98.33%	76.83%	16.20%	-0.14
	GSO	99.00%	89.00%	98.20%	77.00%	45.42%	-0.14
	RAND	99.00%	89.00%	97.73%	77.60%	49.59%	-0.14

Appendix D

An Analysis of the Effects of Genetic and Swarm Intelligence Feature Selection on Adversarial Author Identification Using Many Authors Detailed Results

We generated the results in this appendix using six feature selection algorithms (i.e., GEFeS, PSO, ABCO, ASO and GSO) and followed the process outlined Chapter 7. Tables D.1-D.24, show the results of these experiments. In the table title, we indicate the dataset (CASIS-25, CASIS-50, or CASIS-100), the feature set (LIWC, Topic Modeling, Stylometry or Hybrid), and the training/testing configuration, where *75+(org = 25, adv = 25)* indicates we trained on 75 samples and tested on 25 original samples and 25 adversarial samples, *75+(org = 100, adv = 100)* indicates we trained on 75 samples and tested on all 100 samples (75 original samples, plus 25 samples, either original or adversarial), 1*75+(org = 25, adv = 25)* indicates we trained on 175 samples and tested on 25 original samples and 25 adversarial samples, and 1*75+(org = 200, adv = 200)* indicates we trained on 175 samples and tested using those 175 samples plus either the 25 original samples or the 25 adversarial samples. Similarly, 3*75+(org = 25, adv = 25)* means we trained on 375 samples and tested with 25 samples (adversarial and non-adversarial), whereas 3*75+(org = 400, adv = 400)* denotes training on 375 samples, but testing with all 400 samples, exchanging the adversarial samples as previously explained. Since there are three datasets, four feature sets and two testing approaches, we end up with 24 (3 x 4 x 2) tables of results.

Tables D.1-D.4, show the results of experiments for the CASIS-25 dataset using the *75+(org = 25, adv = 25)*, training/testing approach, where the individual tables vary the feature set. Tables 7.6-7.9, also use the CASIS-25 dataset, but use the *75+(org = 100, adv = 100)* training/testing approach. Similarly, tables D.9-D.12, use the CASIS-50 dataset and the 1*75+(org = 25, adv = 25)* training/testing approach. Whereas tables D.13-D.16, use the CASIS-50 dataset and the 1*75+(org = 200, adv = 200)* training/testing approach. Finally, tables D.17-D.20, use the CASIS-100 dataset using the 3*75+(org = 25, adv = 25)* training/testing approach, and tables D.21-D.24, also use the CASIS-100 dataset with the 3*75+(org = 400, adv = 400)* training/testing approach.

Tables D.1-D.24 are laid out as follows. The first column, labeled ω, is the feature reduction weighting parameter as described in Equation 4.1. The second column, labeled *FS Alg*, identifies the feature selection algorithm for each of the values of ω. The remainder of the columns are divided into three groups.

The first group of columns, labeled *Baseline*, are columns with data generated using no feature selection. Therefore, the values in each of the rows are the same since these values are affected neither by feature reduction nor feature selection. Notice that this first group has two columns labeled *Orig*, meaning original text, and *Adv*, meaning adversarial text. These values are the average accuracies across the 30 runs without and with the adversarial texts.

The second group of columns, labeled *With Feature Selection*, contains three columns representing results used when employing feature selection. The first two columns of the group, labeled *Orig* and *Adv*, correspond to accuracies without and with adversarial samples. The third column, labeled *% Features Used*, shows the average percent of features used across the 30 experimental evaluations.

The third group of columns, consisting of a single column labeled *Use?*, is a value indicating the preference to use, or not use, feature selection. Table 6.1 gives the definition for this value.

Table D.1 gives the measurements generated using the LIWC feature set and the CASIS-25 dataset training on three of the samples and testing on the fourth (i.e., *75+(org = 25, adv = 25)*). The baseline accuracies, without feature selection, are 68.00% with original samples and 56.00% when introducing adversarial samples. As previously mentioned, ω has little effect on accuracy, so the measured accuracies remain similar, as they do for these two columns across all 24 tables. The fifth column shows accuracies for the non-adversarial, or original, writing samples using the six feature selection algorithms ranging from 72.27% ($\omega = 0.1$ using RAND) to 89.07% ($\omega = 0.3$ using GEFeS). The sixth column shows that with adversarial samples and feature selection, the accuracy drops to a range of 15.73% ($\omega = 0.7$ using GEFeS) to 64.53% ($\omega = 0.0$ using ASO). The last column indicates that the most favorable accuracy drop (from the AIdS perspective) occurs with $\omega = 1.0$ using ASO, yielding a *Use?* value, as defined in Table 1, of 0.04. The least favorable drop occurs with $\omega = 0.9$ using RAND, yielding a *Use?* value of -0.59.

The method for generating the results for Table D.2, was similar to the method for Table D.1, except Table D.2 uses the Topic Modeling feature set. The baseline accuracies for original samples are 84.00%, and 8.00% when introducing adversarial samples. Using feature selection, the accuracies improve to ranges of 77.47% ($\omega = 0.9$ using RAND) to 92.00% ($\omega = 0.0$ using GEFeS) for original samples, and 8.67% ($\omega = 0.0$ using RAND and $\omega = 0.7$ using PSO) to 12.00% (using ASO for $\omega > 0.0$) with adversarial samples. Unlike Table D.1, all *Use?* values (see Table 6.1) in Table D.2 are positive, ranging from 0.02 (using RAND with $\omega = 0.0$ or $\omega = 0.9$) to 0.56 ($\omega = 0.0$ using GEFeS), indicating that feature selection is desirable for all values of ω using Topic Modeling.

Table D.3 is like Table D.1 and Table D.2 but uses the Stylometry feature set. In Table D.3, the baseline accuracy for original samples is 60.00%, and 32.00% with adversarial samples. When we compare the accuracies of original samples among the three independent feature sets using feature selection, we see that Stylometry has the broadest range of values, ranging from 64.40% (ω = 0.7 using RAND) to 97.47% (ω = 0.1 using GEFeS). We also see from Table D.3 that the adversarial accuracies when using feature selection are generally lower than the other three independent feature sets, ranging from 4.00% with several configurations of ω across all feature selection algorithms, to 14.80% (ω = 0.0 using ASO). With this significant swing when using feature selection, the *Use?* indicators, as defined in Table 6.1, are firmly negative, ranging from -0.22 (ω = 0.0 using ASO) to -0.80 (ω = 0.7 and 1.0 using RAND).

Table D.4 shows the results of the CASIS-25 dataset using *75+(org = 25, adv = 25)* with the hybrid feature set, which is a combination of the previous three feature sets. In comparison to the three independent feature sets, the Hybrid feature set sports the highest accuracies without adversarial samples (for both baseline and with feature selection, at 92.00% and a range of 90.00% (ω = 1.0 using RAND) to 100.00% (ω = 0.1, 0.3, 0.5 & 0.7 using GEFeS) respectively. However, when using adversarial samples, the baseline drops to 32.00% and when using feature selection drops to a range of 6.80% (ω = 0.5 using ASO) to 12.93% (ω = 0.1 using GSO). This massive swing in accuracies when introducing adversarial samples results in extremely negative *Use?* values, as defined in Table 6.1, ranging from -0.54 (ω = 0.0 using GEFeS) to -0.75 (ω = 0.5 & 0.7 using ASO).

Table D.5 starts over with the LIWC feature set and the CASIS-25 dataset, but this time using *75+(org = 100, adv = 100)*. Notice that this configuration is similar to that used in Table D.1 but differs in that this configuration tests all 100 samples (with and without the adversarial texts). The baseline accuracies are 92.00% and 88.00% for original and adversarial samples respectively. The feature selection accuracies range from 93.07% (ω = 0.1 using RAND, and ω = 0.9 using

RAND or GEFeS) to 97.27% (ω = 0.3 using GEFeS) and 78.93% (ω = 0.7 using GEFeS) to 91.13% (ω = 0.0 using ASO) for original and adversarial samples. The *Use?* values, as defined in Table 6.1, are mostly negative ranging from 0.05 (ω = 0.0 using ASO) to -0.09 (ω = 0.9 using RAND and ω = 1.0 using ABCO).

Table D.6 continues using the CASIS-25 dataset and *75+(org = 100, adv = 100)*, but employs the Topic Modeling feature set. The baseline accuracies in Table D.6, are 96.00% for original samples and 77.00% for adversarial samples. In Table D.6 the accuracies are not significantly influenced by the introduction of adversarial samples, with accuracies ranging from 90.73% (ω = 0.9 using RAND) to 96.87% (ω = 0.0 using GEFeS), and a range of 73.50 % (ω = 0.5 using RAND) to 78.00 (using ASO for all values except ω = 0.0), for original and adversarial respectively. We note that the *Use?* values do not show much variation, with only a minor movement from -0.10 (ω = 1.0 using RAND) to 0.02 (using ASO for all values of ω).

In Table D.7 we see the results again for the Stylometry feature set for CASIS-25, but this time testing with all samples (*75+(org = 100, adv = 100)).* The baseline accuracies are 90.00% and 83.00% for original samples and adversarial samples respectively. Using feature selection, the non-adversarial sample accuracies range from 91.10% (ω = 0.7 using RAND) to 99.87 (ω = 1.0 using GEFeS), and for adversarial samples, the accuracies remain relatively close to 76.00%, ranging from several occurrences of 76.00% to 77.67% (ω = 0.1 using ASO). The *Use?* value, as defined in Table 6.1, hovers near 0.0 with values of 0.02 (for various algorithms and values of ω), to -0.07 when using RAND for most values of ω.

Table D.8 CASIS-25 using *75+(org = 100, adv = 100)*, is the Hybrid feature set with baseline accuracies of 98.00% and 83.00% for non-adversarial, or original, and adversarial samples. Feature selection yields an original accuracy near or at 100.00% (GEFeS with ω = 0.1, 0.3, 0.5 &

0.7), with the lowest accuracy being 97.50% (ω = 1.0 using RAND). The adversarial samples have an accuracy ranging from 76.70% (ω = 0.5 using ASO), to 78.23% (ω = 0.1 using GSO). The *Use?* values, as defined in Table 6.1, are all slightly negative with values between -0.04 and -0.07.

Starting with Table D.9 we switch to using the CASIS-50 dataset, with tables D.9-D.12 using 1*75+(org = 25, adv = 25)*, which is to say we trained using 175 samples and tested on 25 with and without the adversarial samples. Tables D.13-D.16 use 1*75+(org = 200, adv = 200)*, which also trained on 175 samples, but tested using all 200 samples.

Table D.9 uses the LIWC feature set with accuracies for baseline (i.e., no feature selection) being 68.00% for original samples and 44.00% when using adversarial samples. Feature selection yields accuracies on original samples ranging from 71.33% (using RAND with ω = 0.7) to 89.97 (using GEFeS with ω = 0.0). Testing with adversarial samples, the accuracies drop to the range of 13.33% (ω = 0.1 using GEFeS) to 64.13% (ω = 0.0 using ASO). The *Use?* values have a broad spread of values ranging from a high of 0.57 (ω = 0.0 using ASO) to -0.42 ω = 0.7 using ABCO or GSO).

Table D.10 uses the Topic Modeling feature set. The baseline accuracy values are 40.00% for original samples and 0.00% for the adversarial samples. This poor performance demonstrates the adverse effect of increasing the number of authors when using topic modeling. Also, note that the complete failure to classify correctly a single baseline adversarial sample causes the *Use?* conditional (see Table 6.1) to fire resulting in unusually higher (positive) values. Notice that with feature selection, the accuracy increases to a range of 78.80% (ω = 0.7 using RAND) to 92.13% (ω = 0.9 & 1.0 using GEFeS) for original samples. Accuracies for adversarial samples range from 8.93% (ω = 0.7 using GSO) to 12.00% (ω = 0.1, 0.3, 0.7, 0.9 & 1.0 using ASO). The *Use?* conditional captures this case where, in the face of an adversarial attack, complete failure occurs without feature selection, but some success is realized with feature selection. Therefore, the *Use?* values are positive, ranging from 0.97 to 1.30.

The results of Table D.11, reflect using the Stylometry feature set with baseline accuracies of 48.00% for original samples, and 20.00% for adversarial samples. Feature selection raises the accuracies of the original samples to a range of 63.73% (ω = 0.0 using RAND) to 96.93 (ω = 0.1 using GEFeS). However, when we applied feature selection to the adversarial samples, we saw the accuracies drop to a range of 4.00% (for several values of ω and several algorithms) to 14.27% (ω = 0.0 using ASO). Generally, the *Use?* values are often negative, with a low-end of -0.47 (ω = 0.0 using RAND). However, because the baseline values are so low, we see some *Use?* values are positive, with a high-end of 0.37 (ω = 0.0 using ASO).

Table D.12 reflects the values using the Hybrid feature set. The baseline accuracies are 92.00% for original samples, and 12.00% for adversarial samples. Using feature selection on the original samples, we see a range of accuracies from 89.07% (ω = 0.9 using RAND) to 100.00% (for all values of ω except 0.9 using GEFeS). Accuracies using feature selection on adversarial samples range from 6.00% (ω = 0.3 using ASO) to 12.67 (ω = 0.7 using ABCO). The *Use?* values are mixed, ranging from a high of 0.09 (ω = 0.7 using ABCO) to a low of -0.45 (ω = 0.3 using ASO).

As previously noted, in tables D.13-D.16, we switch to using 1*75+(org = 200, adv = 200)*, still using the CASIS-50 dataset. The data in Table D.14, reflects using the LIWC feature set. The baseline accuracies are 96.00% for original samples and 93.00% for adversarial samples. We see a large jump in the baseline accuracies, compared to Table D.9, due to the change in testing using many of the samples also used in training. This same effect carries over to the use of feature selection with accuracies of original samples ranging from 92.67% (ω = 0.5 using RAND) to

97.47% (ω = 0.0 using GEFeS), and for adversarial samples, a range of 78.30% (ω = 0.1 using GEFeS) to 91.03% (ω = 0.0 using ASO). Because both the baseline and feature selection accuracies are high, the *Use?* values are slightly negative, ranging from a high of -0.04 (ω = 0.1 using ASO) to a low of -0.18 (ω = 0.5 using RAND).

The results shown in Table D.14, are based on the Topic Modeling feature set with baseline accuracies of 85.00% for original samples and 80.00% when swapping in the adversarial samples. The feature selection accuracies for original samples range from 90.63% (ω = 0.7 using RAND) to 96.90% (ω = 0.0 using GEFeS), and when swapping in the adversarial samples the accuracies range from 73.23 % (ω = 0.7 using RAND) to 78.00% (for several values of ω using ASO). The *Use?* values are generally positive, except for a few RAND values of -0.01 and -0.02 (ω = 0.3, 0.7 & 1.0). The high-end of the *Use?* range is 0.11 for ASO (all values of ω).

Table D.15 shows results when using the Stylometry feature set with baseline accuracies of 93.50% for original samples and 90.00% when including adversarial samples. Feature selection accuracies range from 90.93% (ω = 0.0 using RAND) to 99.23% (ω = 0.1 using GEFeS) for original samples, and for adversarial samples a range of 76.00% (for most of the feature selection algorithms at various settings for ω) to 78.57% (ω = 0.0 using ASO). The *Use?* values are all negative and range from -0.09 (ω = 0.1 using GEFeS) to -0.18 (ω = 0.0, 0.5, 0.9 & 1.0 using RAND).

The results of Table D.16 consider the Hybrid dataset with baseline values of 99.00% for original values and 89.00% when swapping in adversarial samples. Feature selection gives a range of accuracies on original samples of 97.27% (ω = 0.9 using RAND) to 100.00% (ω = 0.0-0.7 & 1.0 using GEFeS). Feature selection using the adversarial samples drops the accuracy range to 76.50% (ω = 0.3 using ASO) to 78.17% (ω = 0.7 using ABCO). Once again, the *Use?* values are negative clumping tightly between -0.12 and -0.15.

As previously mentioned, the results in tables D.17-D.24, are based on experiments using the CASIS-100 dataset. Table D.17 is like Table D.1 and Table D.9. Table D.17 trains on all but adversarial samples and tests with and without the adversarial samples using the LIWC feature set. We see the baseline original and adversarial accuracies are 52% and 44% respectively. When using feature selection, the non-adversarial accuracies range from 47.47% (ω = 0.9 using RAND) to 56.67% (ω = 0.1 using ASO). Adversarial accuracies range from 1.47% (ω = 0.3 using GEFeS) to 14.00% (ω = 0.1 using ASO). The *Use?* values are strongly negative ranging from -0.57 (ω = 0.0 using ASO) to -1.04 (ω = 0.9 using ABCO).

Table D.18, like tables D.2 and D.10, uses the Topic Modeling feature set. Baseline accuracies are low (0.00% to 32.00%). Feature selection improves accuracies only slightly ranging from 32.00% to 43.87% (ω = 0.9 using GEFeS) for non-adversarial samples, and 0.00% to 0.27% (ω = 1.0 using RAND). Because the baseline adversarial accuracies are zero, the conditional is used to calculate the *Use?* values, which mostly are only slightly positive (the only exception being a value of -0.02 ω = 0.1 using RAND), ranging from 0.00 to 0.12 (ω = 0.9 using GEFeS).

Table D.19 is similar to tables D.3 and D.11 using the Stylometry feature set. The baseline accuracies are 28.00% for original samples and 12.00% for adversarial samples. Using feature selection improves only the original sample accuracies, which range from just below 30.00% (ω = 0.5 using RAND), to 55.73% (ω = 0.5 using GEFeS). Adversarial accuracies with feature selection are abysmal for most of the feature selection algorithms ranging roughly between 0.00% - 1.00%. But ASO sometimes demonstrates an improvement over the baseline (especially for low values of ω) ranging from 0.00% (ω = 1.0) to 19.20% (ω = 0.1 and 0.3). This causes the *Use?* values to be mostly negative, except for lower values of ω for ASO, which are strongly positive (0.31 to 1.32).

Table D.20 uses the Hybrid feature set and training/testing configurations similar to tables D.4 and D.12. The baseline accuracy for original samples is promising at 72.00%. However, introducing adversarial samples drops the accuracy to 8.00%. Interestingly, feature selection is not always helpful with values that range from 65.07% (ω = 0.0 using RAND) to 83.73% (ω = 0.9 using GEFeS). Accuracies using adversarial samples causes a significant drop ranging from 0.27% (ω = 0.7 using GEFeS) to just over 5% (ω = 0.0 using GEFeS). This significant degradation forces the *Use?* values to be firmly negative ranging from -0.24 (ω = 0.0 using GEFeS) to -0.97 (ω = 1.0 using RAND).

Tables D.21-D.24 revisit the CASIS-100 dataset with the same feature sets (i.e., LIWC, Topic Modeling, Stylometry and Hybrid), but test using all samples. Table D.21 (compare with tables D.5 and D.13), shows results using the LIWC feature set. Baseline accuracies are high (96.75%) with only a slight drop due to adversarial samples (96.25). Feature selection doesn't have a significant effect on original accuracies with a range of 96.22% (ω = 1.0 using GSO) to 97.17% (ω = 0.0 using ASO). The *Use?* values are only slightly negative, ranging from -0.01 to -0.04.

Table D.22 uses the Topic Modeling feature set (like tables D.6 and D.14) with baseline accuracies for both original and adversarial samples at 88.75% and 86.75%. Feature selection generally is not very helpful with accuracies ranging from about low 87% (notice this is a drop from the baseline) to the high 88% for original samples and the high 84% to the mid 86% for adversarial samples. Notice that feature selection does not improve accuracies much, but also does not hinder adversarial accuracies either. As a result, *Use?* values are only slightly negative ranging from 0.00 to -0.04.

Table D.23, like tables D.7 and D.15, uses the Stylometric feature set with baseline accuracies of 95.50% and 94.50% (original and adversarial). Feature selection has a slightly positive effect on original samples causing the accuracies to range from 95.58% (ω = 0.5 using RAND) to 97.23% (ω = 0.5 using GEFeS). Accuracies using feature selection and adversarial samples generally remain high, ranging from 93.75% (only slightly worse than the baseline) to 94.95% (using ASO), which causes the *Use?* values mostly to be slightly positive, ranging from a few occurrences of -0.01 to 0.02.

Finally, Table D.24, uses the Hybrid feature set (like tables D.8 and D.16), with baseline accuracies of 98.25% and 94.25% (original and adversarial respectively). In these results we see that feature selection causes perhaps only a slight improvement in original samples with ranges of accuracies from 97.80% (ω = 0.0 using RAND) to 98.94% (ω = 0.5 using GEFeS). These ranges yield *Use?* values that hover around 0.00, ranging from -0.01 to 0.01.

As the intent of this work is to begin to shed some light on the susceptibility of feature selection on adversarial authorship attacks, it is helpful to make some general observations about these data.

The first observation has to do with the effects of increasing the number of authors in the dataset. For example, if we compare the baseline values in tables D.4, D.12, and D.20, where all tables use Hybrid features and test on 25 samples, we see that the increased number of authors has little impact on original accuracies (92.00%, 92.00% and 72.00), but some impact on adversarial accuracies (32.00%, 12.00% and 8.00%). Compare this to tables D.2, D.10, and D.18 which use the Topic Modeling feature set, where the same accuracies change more drastically (84.00%, 40.00% and 32.00% for original samples, and 8.00%, 0.00%, and 0.00% for adversarial samples). From these two examples, we can draw the conclusion that different feature sets react differently (at least when no feature selection is used) as the number of authors grow, both in terms of original samples and adversarial samples.

A second observation is that, for baseline original samples across all datasets, the Hybrid feature set dominates author identification accuracies (92.00%, 98.00%, 92.00%, 98.00%, 72.00% and 98.25%), whereas the LIWC feature set dominates (56.00%, 88.00%, 44.00%, 93.00%, 44.00%, 96.25%) when using adversarial samples. This indicates that, even without feature selection, feature sets react differently to adversarial attacks.

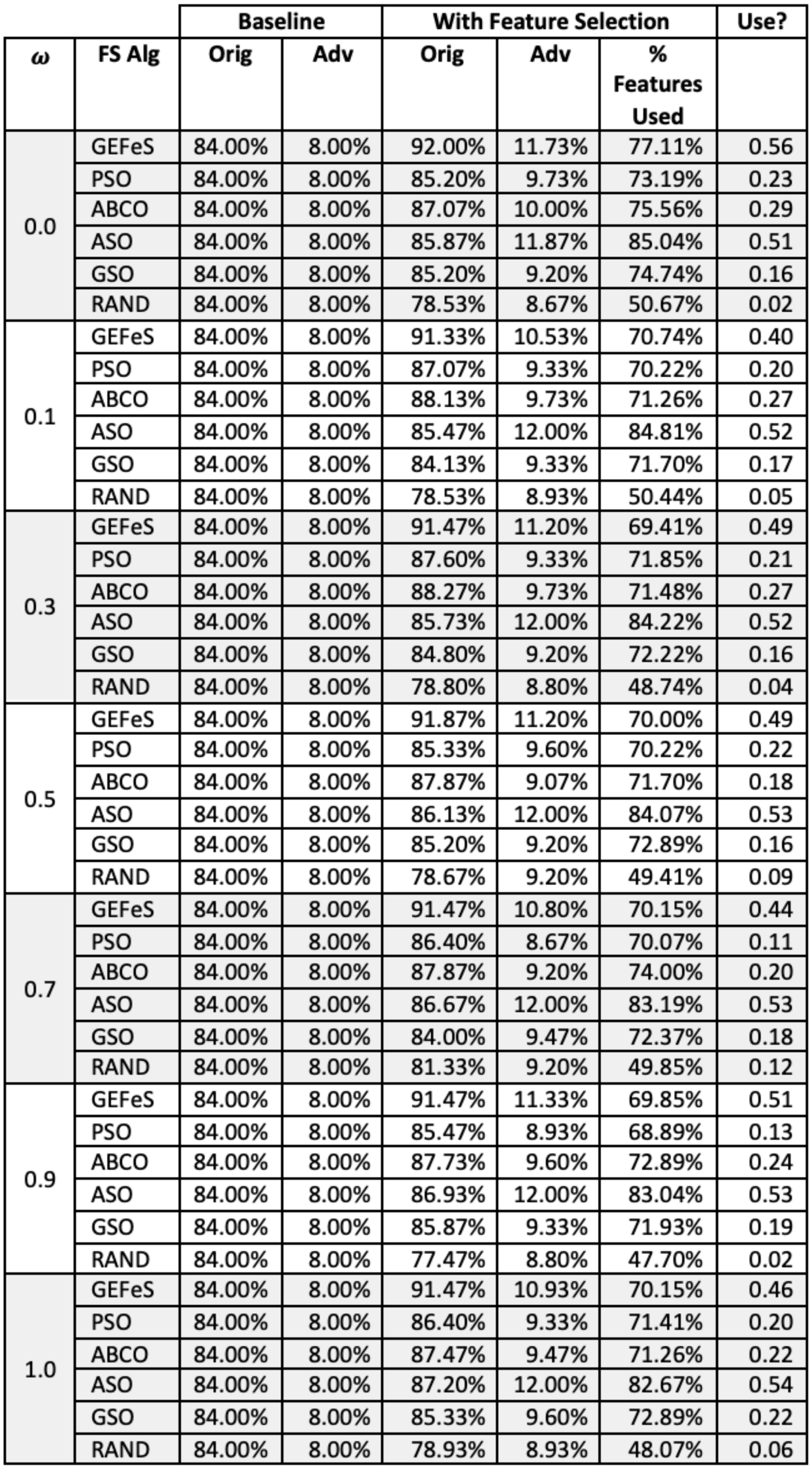
A third observation is a result of comparing *Use?* values across tables. We see that, for the LIWC, Stylometry and Hybrid feature sets, *Use?* values are generally neutral or negative. However, the Topic Modeling feature set has *Use?* values that are neutral or positive. From this observation we may draw the conclusion that, when using feature selection, feature sets reactive *relatively* differently, to adversarial attacks.

The fourth observation is that for all tables, except ASO in Table D.9, the *Use?* values appear to be clustered within the table (as opposed to the values of other tables), even across the various feature selection algorithms (including the RAND algorithm!), and across values of ω. This observation may indicate that the feature set is perhaps the dominate factor in determining adversarial author identification susceptibility.

Table D.1. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, LIWC Feature Set - 93 Features, 75+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	68.00%	56.00%	88.80%	35.07%	57.46%	-0.07
	PSO	68.00%	56.00%	77.47%	27.87%	56.16 %	-0.36
	ABCO	68.00%	56.00%	78.00%	27.205	56.13%	-0.37
	ASO	68.00%	56.00%	75.87%	64.53%	75.88%	0.27
	GSO	68.00%	56.00%	74.40%	28.13%	57.78%	-0.40
	RAND	68.00%	56.00%	73.20%	22.67%	50.25%	-0.52
0.1	GEFeS	68.00%	56.00%	88.00%	17.73%	44.12%	-0.39
	PSO	68.00%	56.00%	78.27%	22.27%	53.66 %	-0.45
	ABCO	68.00%	56.00%	77.07%	23.20%	53.05%	-0.45
	ASO	68.00%	56.00%	75.60%	64.40%	74.41%	0.26
	GSO	68.00%	56.00%	76.80%	27.20%	57.31%	-0.38
	RAND	68.00%	56.00%	72.27%	23.07%	50.18%	-0.52
0.3	GEFeS	68.00%	56.00%	89.07%	17.33%	43.37%	-0.38
	PSO	68.00%	56.00%	78.53%	25.87%	53.66%	-0.38
	ABCO	68.00%	56.00%	76.27%	24.00%	52.83%	-0.45
	ASO	68.00%	56.00%	76.40%	61.87%	72.62%	0.23
	GSO	68.00%	56.00%	72.80%	25.33%	55.70%	-0.48
	RAND	68.00%	56.00%	74.27%	23.07%	50.29%	-0.50
0.5	GEFeS	68.00%	56.00%	88.53%	16.40%	44.73%	-0.41
	PSO	68.00%	56.00%	80.40%	20.93%	50.57%	-0.44
	ABCO	68.00%	56.00%	75.07%	22.40%	52.54%	-0.50
	ASO	68.00%	56.00%	76.00%	61.07%	71.47%	0.21
	GSO	68.00%	56.00%	74.40%	24.67%	55.34%	-0.47
	RAND	68.00%	56.00%	72.67%	26.00%	49.32%	-0.47
0.7	GEFeS	68.00%	56.00%	88.00%	15.73%	48.16%	-0.42
	PSO	68.00%	56.00%	79.33%	23.47%	53.26%	-0.41
	ABCO	68.00%	56.00%	76.00%	24.00%	53.01%	-0.45
	ASO	68.00%	56.00%	75.87%	57.47%	69.61%	0.14
	GSO	68.00%	56.00%	74.13%	23.60%	52.83%	-0.49
	RAND	68.00%	56.00%	74.13%	22.13%	50.75%	-0.51
0.9	GEFeS	68.00%	56.00%	88.40%	16.27%	42.22 %	-0.41
	PSO	68.00%	56.00%	76.53%	22.53%	51.25%	-0.47
	ABCO	68.00%	56.00%	77.73%	22.40%	52.69%	-0.46
	ASO	68.00%	56.00%	76.40%	54.67%	67.10%	0.10
	GSO	68.00%	56.00%	74.80%	19.73%	52.69%	-0.55
	RAND	68.00%	56.00%	72.40%	19.33%	49.43%	-0.59
1.0	GEFeS	68.00%	56.00%	87.73%	16.93%	40.93 %	-0.41
	PSO	68.00%	56.00%	77.20%	21.33%	51.79%	-0.48
	ABCO	68.00%	56.00%	76.13%	17.20%	49.78%	-0.57
	ASO	68.00%	56.00%	75.60%	51.87%	66.63%	0.04
	GSO	68.00%	56.00%	75.33%	23.73%	51.51%	-0.47
	RAND	68.00%	56.00%	73.07%	23.87%	49.68%	-0.50

Table D.2. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Topic Modeling Feature Set - 45 Features, 75+(org = 25, adv = 25).

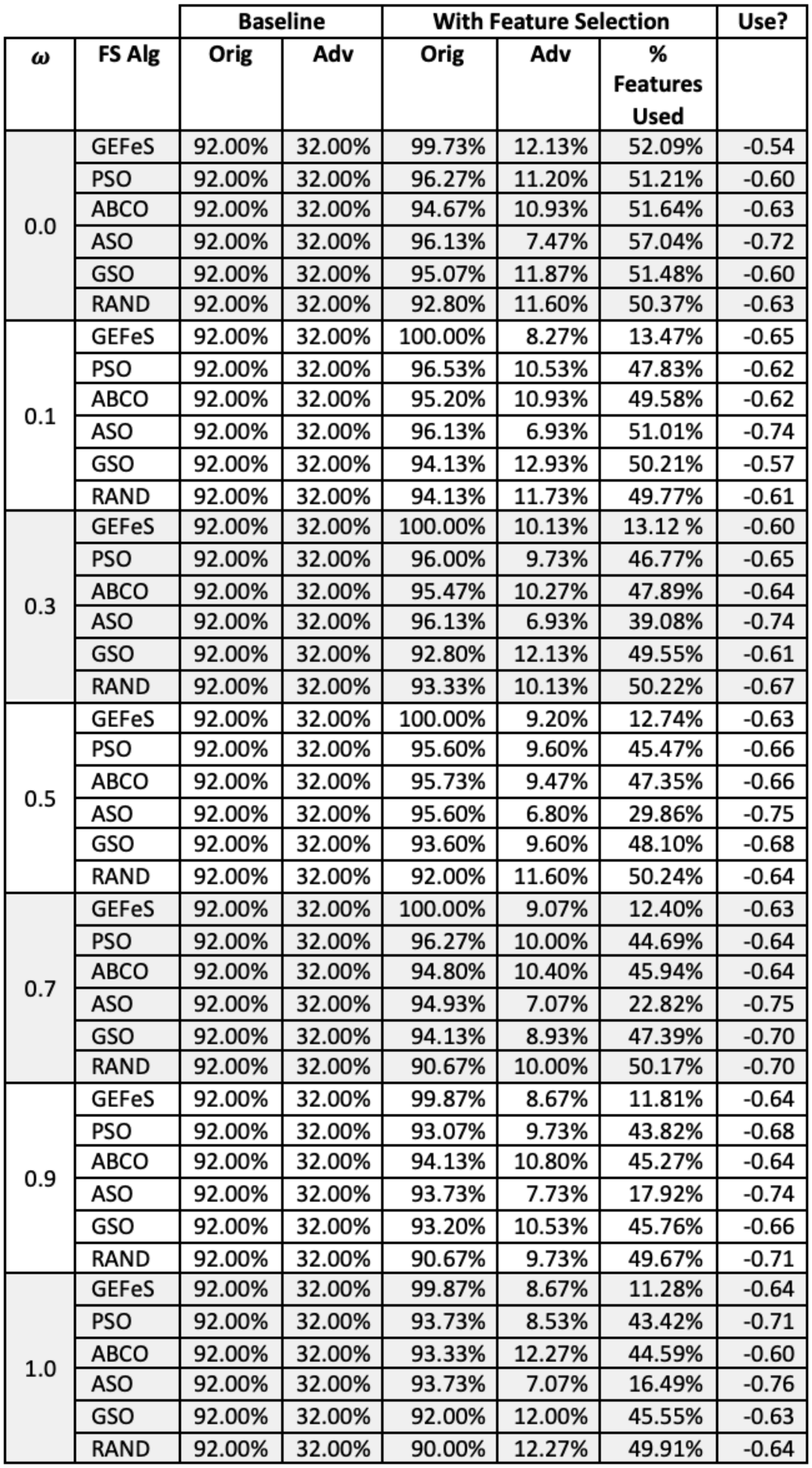


		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	84.00%	8.00%	92.00%	11.73%	77.11%	0.56
	PSO	84.00%	8.00%	85.20%	9.73%	73.19%	0.23
	ABCO	84.00%	8.00%	87.07%	10.00%	75.56%	0.29
	ASO	84.00%	8.00%	85.87%	11.87%	85.04%	0.51
	GSO	84.00%	8.00%	85.20%	9.20%	74.74%	0.16
	RAND	84.00%	8.00%	78.53%	8.67%	50.67%	0.02
0.1	GEFeS	84.00%	8.00%	91.33%	10.53%	70.74%	0.40
	PSO	84.00%	8.00%	87.07%	9.33%	70.22%	0.20
	ABCO	84.00%	8.00%	88.13%	9.73%	71.26%	0.27
	ASO	84.00%	8.00%	85.47%	12.00%	84.81%	0.52
	GSO	84.00%	8.00%	84.13%	9.33%	71.70%	0.17
	RAND	84.00%	8.00%	78.53%	8.93%	50.44%	0.05
0.3	GEFeS	84.00%	8.00%	91.47%	11.20%	69.41%	0.49
	PSO	84.00%	8.00%	87.60%	9.33%	71.85%	0.21
	ABCO	84.00%	8.00%	88.27%	9.73%	71.48%	0.27
	ASO	84.00%	8.00%	85.73%	12.00%	84.22%	0.52
	GSO	84.00%	8.00%	84.80%	9.20%	72.22%	0.16
	RAND	84.00%	8.00%	78.80%	8.80%	48.74%	0.04
0.5	GEFeS	84.00%	8.00%	91.87%	11.20%	70.00%	0.49
	PSO	84.00%	8.00%	85.33%	9.60%	70.22%	0.22
	ABCO	84.00%	8.00%	87.87%	9.07%	71.70%	0.18
	ASO	84.00%	8.00%	86.13%	12.00%	84.07%	0.53
	GSO	84.00%	8.00%	85.20%	9.20%	72.89%	0.16
	RAND	84.00%	8.00%	78.67%	9.20%	49.41%	0.09
0.7	GEFeS	84.00%	8.00%	91.47%	10.80%	70.15%	0.44
	PSO	84.00%	8.00%	86.40%	8.67%	70.07%	0.11
	ABCO	84.00%	8.00%	87.87%	9.20%	74.00%	0.20
	ASO	84.00%	8.00%	86.67%	12.00%	83.19%	0.53
	GSO	84.00%	8.00%	84.00%	9.47%	72.37%	0.18
	RAND	84.00%	8.00%	81.33%	9.20%	49.85%	0.12
0.9	GEFeS	84.00%	8.00%	91.47%	11.33%	69.85%	0.51
	PSO	84.00%	8.00%	85.47%	8.93%	68.89%	0.13
	ABCO	84.00%	8.00%	87.73%	9.60%	72.89%	0.24
	ASO	84.00%	8.00%	86.93%	12.00%	83.04%	0.53
	GSO	84.00%	8.00%	85.87%	9.33%	71.93%	0.19
	RAND	84.00%	8.00%	77.47%	8.80%	47.70%	0.02
1.0	GEFeS	84.00%	8.00%	91.47%	10.93%	70.15%	0.46
	PSO	84.00%	8.00%	86.40%	9.33%	71.41%	0.20
	ABCO	84.00%	8.00%	87.47%	9.47%	71.26%	0.22
	ASO	84.00%	8.00%	87.20%	12.00%	82.67%	0.54
	GSO	84.00%	8.00%	85.33%	9.60%	72.89%	0.22
	RAND	84.00%	8.00%	78.93%	8.93%	48.07%	0.06

Table D.3. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Stylometry Feature Set - 428 Features, 75+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	60.00%	32.00%	95.33%	4.27%	48.77%	-0.28
	PSO	60.00%	32.00%	71.47%	4.27%	51.25%	-0.68
	ABCO	60.00%	32.00%	70.27%	4.00%	51.31%	-0.70
	ASO	60.00%	32.00%	79.33%	14.80%	57.31%	-0.22
	GSO	60.00%	32.00%	69.87%	4.13%	51.85%	-0.71
	RAND	60.00%	32.00%	65.87%	4.67%	49.99%	-0.76
0.1	GEFeS	60.00%	32.00%	97.47%	4.13%	34.21%	-0.25
	PSO	60.00%	32.00%	71.07%	4.13%	48.36%	-0.69
	ABCO	60.00%	32.00%	71.07%	4.13%	49.51%	-0.69
	ASO	60.00%	32.00%	80.13%	10.67%	54.05%	-0.33
	GSO	60.00%	32.00%	71.73%	4.40%	51.09%	-0.67
	RAND	60.00%	32.00%	68.93%	4.00%	50.32%	-0.73
0.3	GEFeS	60.00%	32.00%	96.00%	4.00%	32.38%	-0.27
	PSO	60.00%	32.00%	73.33%	4.13%	48.27%	-0.65
	ABCO	60.00%	32.00%	70.67%	4.40%	48.89%	-0.68
	ASO	60.00%	32.00%	79.33%	8.40%	47.47%	-0.42
	GSO	60.00%	32.00%	68.40%	4.27%	49.14%	-0.73
	RAND	60.00%	32.00%	65.73%	4.13%	50.38%	-0.78
0.5	GEFeS	60.00%	32.00%	96.27%	4.13%	29.43%	-0.27
	PSO	60.00%	32.00%	73.87%	4.27%	47.31%	-0.64
	ABCO	60.00%	32.00%	71.47%	4.53%	48.96%	-0.67
	ASO	60.00%	32.00%	80.27%	6.13%	41.64%	-0.47
	GSO	60.00%	32.00%	71.20%	4.00%	49.00%	-0.69
	RAND	60.00%	32.00%	65.87%	4.00%	49.91%	-0.78
0.7	GEFeS	60.00%	32.00%	96.40%	4.00%	26.10%	-0.27
	PSO	60.00%	32.00%	74.00%	4.00%	47.18%	-0.64
	ABCO	60.00%	32.00%	71.20%	4.00%	47.06%	-0.69
	ASO	60.00%	32.00%	81.87%	4.67%	36.51%	-0.49
	GSO	60.00%	32.00%	69.20%	4.00%	48.70%	-0.72
	RAND	60.00%	32.00%	64.40%	4.13%	49.60%	-0.80
0.9	GEFeS	60.00%	32.00%	94.67%	4.00%	23.29%	-0.30
	PSO	60.00%	32.00%	73.73%	4.13%	45.69%	-0.64
	ABCO	60.00%	32.00%	72.53%	4.00%	46.90%	-0.67
	ASO	60.00%	32.00%	82.67%	4.00%	32.76%	-0.50
	GSO	60.00%	32.00%	68.80%	4.00%	46.88%	-0.73
	RAND	60.00%	32.00%	65.73%	4.40%	50.79%	-0.77
1.0	GEFeS	60.00%	32.00%	95.47%	4.00%	21.54%	-0.28
	PSO	60.00%	32.00%	71.20%	4.13%	45.33%	-0.68
	ABCO	60.00%	32.00%	69.47%	4.27%	45.80%	-0.71
	ASO	60.00%	32.00%	83.60%	4.00%	30.97%	-0.48
	GSO	60.00%	32.00%	68.87%	4.27%	46.14%	-0.72
	RAND	60.00%	32.00%	64.53%	4.00%	50.10%	-0.80

Table D.4. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Hybrid Feature Set -566 Features, 75+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	92.00%	32.00%	99.73%	12.13%	52.09%	-0.54
	PSO	92.00%	32.00%	96.27%	11.20%	51.21%	-0.60
	ABCO	92.00%	32.00%	94.67%	10.93%	51.64%	-0.63
	ASO	92.00%	32.00%	96.13%	7.47%	57.04%	-0.72
	GSO	92.00%	32.00%	95.07%	11.87%	51.48%	-0.60
	RAND	92.00%	32.00%	92.80%	11.60%	50.37%	-0.63
0.1	GEFeS	92.00%	32.00%	100.00%	8.27%	13.47%	-0.65
	PSO	92.00%	32.00%	96.53%	10.53%	47.83%	-0.62
	ABCO	92.00%	32.00%	95.20%	10.93%	49.58%	-0.62
	ASO	92.00%	32.00%	96.13%	6.93%	51.01%	-0.74
	GSO	92.00%	32.00%	94.13%	12.93%	50.21%	-0.57
	RAND	92.00%	32.00%	94.13%	11.73%	49.77%	-0.61
0.3	GEFeS	92.00%	32.00%	100.00%	10.13%	13.12 %	-0.60
	PSO	92.00%	32.00%	96.00%	9.73%	46.77%	-0.65
	ABCO	92.00%	32.00%	95.47%	10.27%	47.89%	-0.64
	ASO	92.00%	32.00%	96.13%	6.93%	39.08%	-0.74
	GSO	92.00%	32.00%	92.80%	12.13%	49.55%	-0.61
	RAND	92.00%	32.00%	93.33%	10.13%	50.22%	-0.67
0.5	GEFeS	92.00%	32.00%	100.00%	9.20%	12.74%	-0.63
	PSO	92.00%	32.00%	95.60%	9.60%	45.47%	-0.66
	ABCO	92.00%	32.00%	95.73%	9.47%	47.35%	-0.66
	ASO	92.00%	32.00%	95.60%	6.80%	29.86%	-0.75
	GSO	92.00%	32.00%	93.60%	9.60%	48.10%	-0.68
	RAND	92.00%	32.00%	92.00%	11.60%	50.24%	-0.64
0.7	GEFeS	92.00%	32.00%	100.00%	9.07%	12.40%	-0.63
	PSO	92.00%	32.00%	96.27%	10.00%	44.69%	-0.64
	ABCO	92.00%	32.00%	94.80%	10.40%	45.94%	-0.64
	ASO	92.00%	32.00%	94.93%	7.07%	22.82%	-0.75
	GSO	92.00%	32.00%	94.13%	8.93%	47.39%	-0.70
	RAND	92.00%	32.00%	90.67%	10.00%	50.17%	-0.70
0.9	GEFeS	92.00%	32.00%	99.87%	8.67%	11.81%	-0.64
	PSO	92.00%	32.00%	93.07%	9.73%	43.82%	-0.68
	ABCO	92.00%	32.00%	94.13%	10.80%	45.27%	-0.64
	ASO	92.00%	32.00%	93.73%	7.73%	17.92%	-0.74
	GSO	92.00%	32.00%	93.20%	10.53%	45.76%	-0.66
	RAND	92.00%	32.00%	90.67%	9.73%	49.67%	-0.71
1.0	GEFeS	92.00%	32.00%	99.87%	8.67%	11.28%	-0.64
	PSO	92.00%	32.00%	93.73%	8.53%	43.42%	-0.71
	ABCO	92.00%	32.00%	93.33%	12.27%	44.59%	-0.60
	ASO	92.00%	32.00%	93.73%	7.07%	16.49%	-0.76
	GSO	92.00%	32.00%	92.00%	12.00%	45.55%	-0.63
	RAND	92.00%	32.00%	90.00%	12.27%	49.91%	-0.64

Table D.5. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, LIWC Feature Set - 93 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	92.00%	88.00%	97.20%	83.77%	57.46%	0.00
	PSO	92.00%	88.00%	94.37%	81.97%	56.16 %	-0.05
	ABCO	92.00%	88.00%	94.47%	81.77%	56.13%	-0.05
	ASO	92.00%	88.00%	93.97%	91.13%	75.88%	0.05
	GSO	92.00%	88.00%	93.53%	81.97%	57.78%	-0.06
	RAND	92.00%	88.00%	93.27%	80.63%	50.25%	-0.08
0.1	GEFeS	92.00%	88.00%	97.00%	79.43%	44.12%	-0.05
	PSO	92.00%	88.00%	94.50%	80.50%	53.66 %	-0.07
	ABCO	92.00%	88.00%	94.27%	80.80%	53.05%	-0.07
	ASO	92.00%	88.00%	93.90%	91.10%	74.41%	0.04
	GSO	92.00%	88.00%	94.20%	81.80%	57.31%	-0.06
	RAND	92.00%	88.00%	93.07%	80.83%	50.18%	-0.08
0.3	GEFeS	92.00%	88.00%	97.27%	79.33%	43.37%	-0.05
	PSO	92.00%	88.00%	94.63%	81.47%	53.66%	-0.06
	ABCO	92.00%	88.00%	94.00%	80.90%	52.83%	-0.07
	ASO	92.00%	88.00%	94.10%	90.47%	72.62%	0.04
	GSO	92.00%	88.00%	93.20%	81.33%	55.70%	-0.07
	RAND	92.00%	88.00%	93.50%	80.70%	50.29%	-0.08
0.5	GEFeS	92.00%	88.00%	97.13%	79.10%	44.73%	-0.06
	PSO	92.00%	88.00%	95.07%	80.20%	50.57%	-0.07
	ABCO	92.00%	88.00%	93.77%	80.60%	52.54%	-0.08
	ASO	92.00%	88.00%	94.00%	90.27%	71.47%	0.04
	GSO	92.00%	88.00%	95.53%	81.10%	55.34%	-0.07
	RAND	92.00%	88.00%	93.13%	81.47%	49.32%	-0.07
0.7	GEFeS	92.00%	88.00%	97.00%	78.93%	48.16%	-0.06
	PSO	92.00%	88.00%	94.80%	80.83%	53.26%	-0.06
	ABCO	92.00%	88.00%	93.93%	80.93%	53.01%	-0.07
	ASO	92.00%	88.00%	93.97%	89.37%	69.61%	0.03
	GSO	92.00%	88.00%	93.50%	80.87%	52.83%	-0.08
	RAND	92.00%	88.00%	93.40%	80.40%	50.75%	-0.08
0.9	GEFeS	92.00%	88.00%	97.03%	79.00%	42.22 %	-0.06
	PSO	92.00%	88.00%	94.13%	80.63%	51.25%	-0.07
	ABCO	92.00%	88.00%	94.40%	80.57%	52.69%	-0.07
	ASO	92.00%	88.00%	94.10%	88.67%	67.10%	0.02
	GSO	92.00%	88.00%	93.67%	79.90%	52.69%	-0.08
	RAND	92.00%	88.00%	93.07%	79.80%	49.43%	-0.09
1.0	GEFeS	92.00%	88.00%	96.93%	79.23%	40.93 %	-0.06
	PSO	92.00%	88.00%	94.30%	80.33%	51.79%	-0.07
	ABCO	92.00%	88.00%	94.03%	79.30%	49.78%	-0.09
	ASO	92.00%	88.00%	93.90%	87.97%	66.63%	0.01
	GSO	92.00%	88.00%	93.80%	80.90%	51.51%	-0.07
	RAND	92.00%	88.00%	93.27%	80.97%	49.68%	-0.08

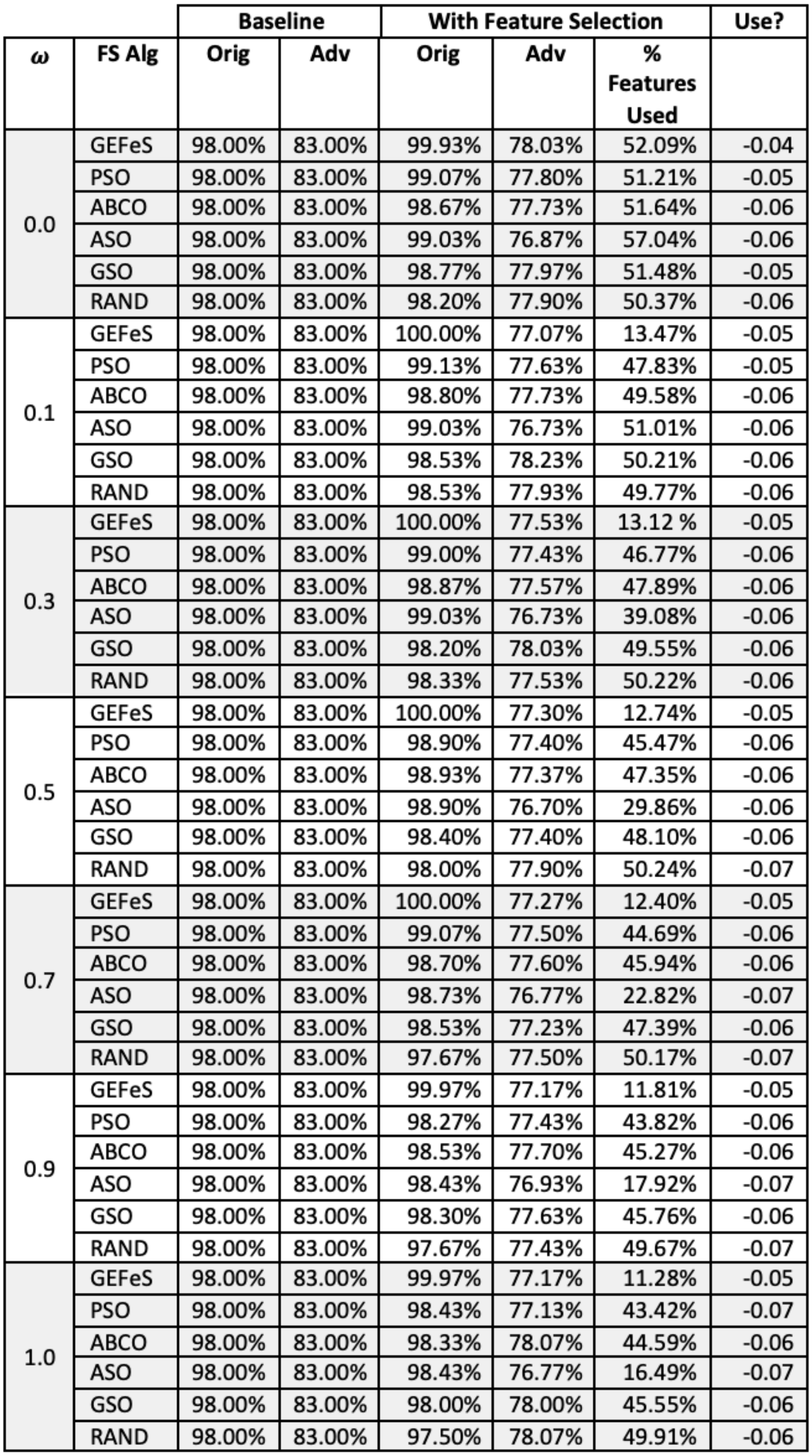
Table D.6. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Topic Modeling Feature Set - 45 Features, 75+(org = 100, adv = 100).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	96.00%	77.00%	96.87%	76.80%	77.11%	0.01
	PSO	96.00%	77.00%	93.73%	74.87%	73.19%	-0.05
	ABCO	96.00%	77.00%	95.03%	75.77%	75.56%	-0.03
	ASO	96.00%	77.00%	96.47%	77.97%	85.04%	0.02
	GSO	96.00%	77.00%	94.30%	75.30%	74.74%	-0.04
	RAND	96.00%	77.00%	91.33%	73.78%	50.67%	-0.09
0.1	GEFeS	96.00%	77.00%	96.23%	76.03%	70.74%	-0.01
	PSO	96.00%	77.00%	93.73%	74.30%	70.22%	-0.06
	ABCO	96.00%	77.00%	94.93%	75.33%	71.26%	-0.03
	ASO	96.00%	77.00%	96.37%	78.00%	84.81%	0.02
	GSO	96.00%	77.00%	93.67%	74.97%	71.70%	-0.05
	RAND	96.00%	77.00%	91.07%	73.67%	50.44%	-0.09
0.3	GEFeS	96.00%	77.00%	96.50%	76.43%	69.41%	0.00
	PSO	96.00%	77.00%	94.33%	74.77%	71.85%	-0.05
	ABCO	96.00%	77.00%	95.00%	75.37%	71.48%	-0.03
	ASO	96.00%	77.00%	96.43%	78.00%	84.22%	0.02
	GSO	96.00%	77.00%	93.53%	74.63%	72.22%	-0.06
	RAND	96.00%	77.00%	91.20%	73.70%	48.74%	-0.09
0.5	GEFeS	96.00%	77.00%	96.53%	76.37%	70.00%	0.00
	PSO	96.00%	77.00%	93.70%	74.77%	70.22%	-0.05
	ABCO	96.00%	77.00%	95.13%	75.43%	71.70%	-0.03
	ASO	96.00%	77.00%	96.53%	78.00%	84.07%	0.02
	GSO	96.00%	77.00%	93.70%	74.70%	72.89%	-0.05
	RAND	96.00%	77.00%	90.87%	73.50%	49.41%	-0.10
0.7	GEFeS	96.00%	77.00%	96.50%	76.33%	70.15%	0.00
	PSO	96.00%	77.00%	94.33%	74.90%	70.07%	-0.04
	ABCO	96.00%	77.00%	94.90%	75.23%	74.00%	-0.03
	ASO	96.00%	77.00%	96.67%	78.00%	83.19%	0.02
	GSO	96.00%	77.00%	93.90%	75.27%	72.37%	-0.04
	RAND	96.00%	77.00%	92.57%	74.53%	49.85%	-0.07
0.9	GEFeS	96.00%	77.00%	96.63%	76.60%	69.85%	0.00
	PSO	96.00%	77.00%	94.00%	74.87%	68.89%	-0.05
	ABCO	96.00%	77.00%	95.43%	75.90%	72.89%	-0.02
	ASO	96.00%	77.00%	96.73%	78.00%	83.04%	0.02
	GSO	96.00%	77.00%	94.37%	75.23%	71.93%	-0.04
	RAND	96.00%	77.00%	90.73%	73.57%	47.70%	-0.10
1.0	GEFeS	96.00%	77.00%	96.23%	75.10%	70.15%	-0.01
	PSO	96.00%	77.00%	94.07%	74.80%	71.41%	-0.05
	ABCO	96.00%	77.00%	94.97%	75.47%	71.26%	-0.03
	ASO	96.00%	77.00%	96.80%	78.00%	82.67%	0.02
	GSO	96.00%	77.00%	94.03%	75.10%	72.89%	-0.05
	RAND	96.00%	77.00%	91.07%	73.57%	48.07%	-0.10

Table D.7. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Stylometry Feature Set - 428 Features, 75+(org = 100, adv = 100).

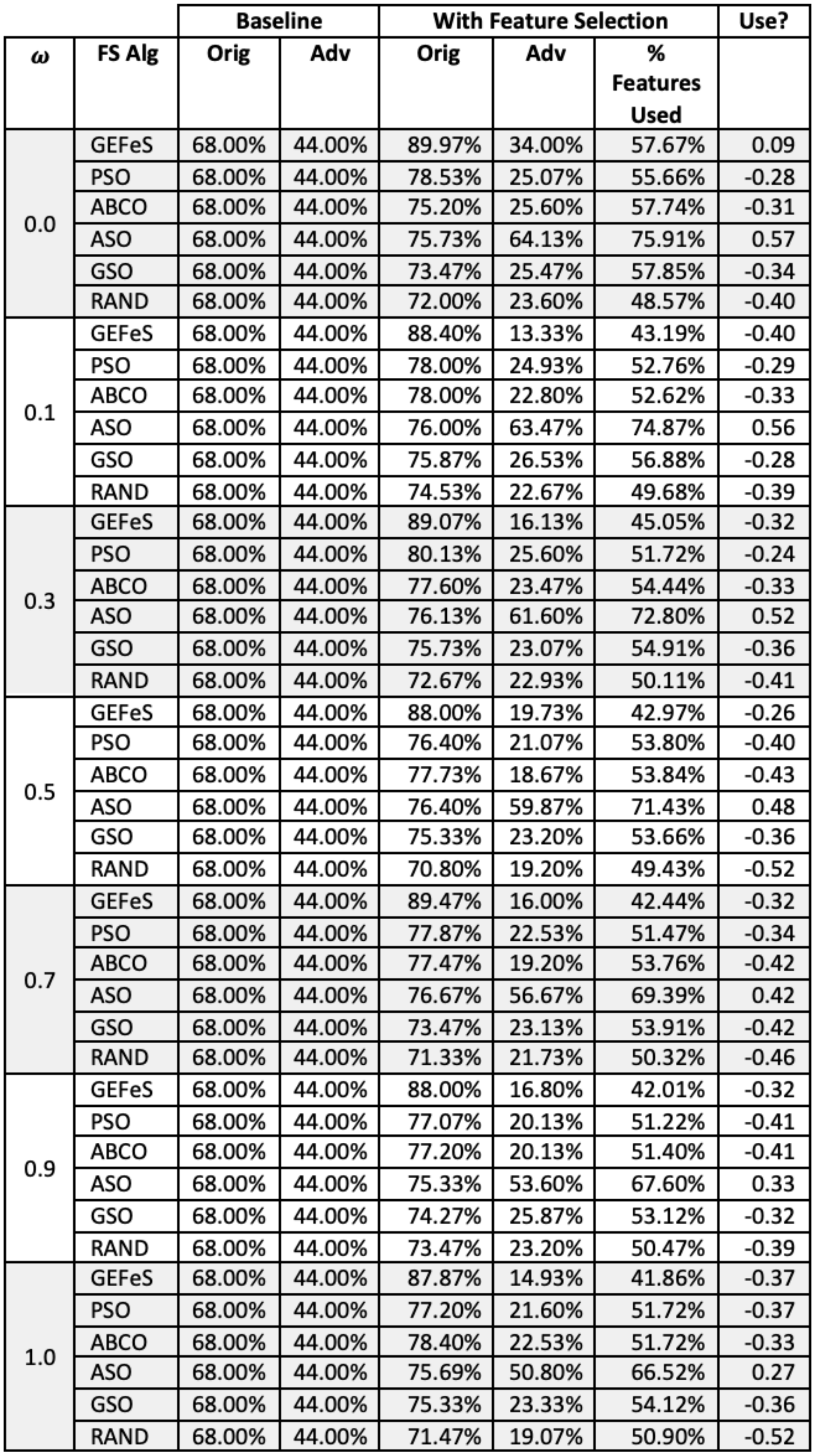
		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	90.00%	83.00%	98.83%	76.07%	48.77%	0.01
	PSO	90.00%	83.00%	92.87%	76.07%	51.25%	-0.05
	ABCO	90.00%	83.00%	92.57%	76.00%	51.31%	-0.06
	ASO	90.00%	83.00%	94.83%	78.70%	57.31%	0.00
	GSO	90.00%	83.00%	92.47%	76.03%	51.85%	-0.06
	RAND	90.00%	83.00%	91.47%	76.17%	49.99%	-0.07
0.1	GEFeS	90.00%	83.00%	99.37%	76.03%	34.21%	0.02
	PSO	90.00%	83.00%	92.77%	76.03%	48.36%	-0.05
	ABCO	90.00%	83.00%	92.77%	76.03%	49.51%	-0.05
	ASO	90.00%	83.00%	95.03%	77.67%	54.05%	-0.01
	GSO	90.00%	83.00%	92.93%	76.10%	51.09%	-0.05
	RAND	90.00%	83.00%	92.23%	76.00%	50.32%	-0.06
0.3	GEFeS	90.00%	83.00%	99.00%	76.00%	32.38%	0.02
	PSO	90.00%	83.00%	93.33%	76.03%	48.27%	-0.05
	ABCO	90.00%	83.00%	92.67%	76.10%	48.89%	-0.05
	ASO	90.00%	83.00%	94.83%	77.10%	47.47%	-0.02
	GSO	90.00%	83.00%	92.10%	76.07%	49.14%	-0.06
	RAND	90.00%	83.00%	91.43%	76.03%	50.38%	-0.07
0.5	GEFeS	90.00%	83.00%	99.07%	76.03%	29.43%	0.02
	PSO	90.00%	83.00%	93.47%	76.07%	47.31%	-0.05
	ABCO	90.00%	83.00%	92.87%	76.13%	48.96%	-0.05
	ASO	90.00%	83.00%	95.07%	76.53%	41.64%	-0.02
	GSO	90.00%	83.00%	92.80%	76.00%	49.00%	-0.05
	RAND	90.00%	83.00%	91.47%	76.00%	49.91%	-0.07
0.7	GEFeS	90.00%	83.00%	99.10%	76.00%	26.10%	0.02
	PSO	90.00%	83.00%	93.50%	76.00%	47.18%	-0.05
	ABCO	90.00%	83.00%	92.80%	76.00%	47.06%	-0.05
	ASO	90.00%	83.00%	95.47%	76.17%	36.51%	-0.02
	GSO	90.00%	83.00%	92.30%	76.00%	48.70%	-0.06
	RAND	90.00%	83.00%	91.10%	76.03%	49.60%	-0.07
0.9	GEFeS	90.00%	83.00%	99.67%	76.00%	23.29%	0.01
	PSO	90.00%	83.00%	93.43%	76.03%	45.69%	-0.05
	ABCO	90.00%	83.00%	93.13%	76.00%	46.90%	-0.05
	ASO	90.00%	83.00%	95.67%	76.00%	32.76%	-0.02
	GSO	90.00%	83.00%	92.20%	76.00%	46.88%	-0.06
	RAND	90.00%	83.00%	91.43%	76.10%	50.79%	-0.07
1.0	GEFeS	90.00%	83.00%	99.87%	76.00%	21.54%	0.01
	PSO	90.00%	83.00%	92.80%	76.03%	45.33%	-0.05
	ABCO	90.00%	83.00%	92.37%	76.07%	45.80%	-0.06
	ASO	90.00%	83.00%	95.90%	76.00%	30.97%	-0.02
	GSO	90.00%	83.00%	92.17%	76.07%	46.14%	-0.06
	RAND	90.00%	83.00%	91.13%	76.00%	50.10%	-0.07

Table D.8. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-25, Hybrid Feature Set - 566 Features, 75+(org = 100, adv = 100).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	98.00%	83.00%	99.93%	78.03%	52.09%	-0.04
	PSO	98.00%	83.00%	99.07%	77.80%	51.21%	-0.05
	ABCO	98.00%	83.00%	98.67%	77.73%	51.64%	-0.06
	ASO	98.00%	83.00%	99.03%	76.87%	57.04%	-0.06
	GSO	98.00%	83.00%	98.77%	77.97%	51.48%	-0.05
	RAND	98.00%	83.00%	98.20%	77.90%	50.37%	-0.06
0.1	GEFeS	98.00%	83.00%	100.00%	77.07%	13.47%	-0.05
	PSO	98.00%	83.00%	99.13%	77.63%	47.83%	-0.05
	ABCO	98.00%	83.00%	98.80%	77.73%	49.58%	-0.06
	ASO	98.00%	83.00%	99.03%	76.73%	51.01%	-0.06
	GSO	98.00%	83.00%	98.53%	78.23%	50.21%	-0.06
	RAND	98.00%	83.00%	98.53%	77.93%	49.77%	-0.06
0.3	GEFeS	98.00%	83.00%	100.00%	77.53%	13.12 %	-0.05
	PSO	98.00%	83.00%	99.00%	77.43%	46.77%	-0.06
	ABCO	98.00%	83.00%	98.87%	77.57%	47.89%	-0.06
	ASO	98.00%	83.00%	99.03%	76.73%	39.08%	-0.06
	GSO	98.00%	83.00%	98.20%	78.03%	49.55%	-0.06
	RAND	98.00%	83.00%	98.33%	77.53%	50.22%	-0.06
0.5	GEFeS	98.00%	83.00%	100.00%	77.30%	12.74%	-0.05
	PSO	98.00%	83.00%	98.90%	77.40%	45.47%	-0.06
	ABCO	98.00%	83.00%	98.93%	77.37%	47.35%	-0.06
	ASO	98.00%	83.00%	98.90%	76.70%	29.86%	-0.06
	GSO	98.00%	83.00%	98.40%	77.40%	48.10%	-0.06
	RAND	98.00%	83.00%	98.00%	77.90%	50.24%	-0.07
0.7	GEFeS	98.00%	83.00%	100.00%	77.27%	12.40%	-0.05
	PSO	98.00%	83.00%	99.07%	77.50%	44.69%	-0.06
	ABCO	98.00%	83.00%	98.70%	77.60%	45.94%	-0.06
	ASO	98.00%	83.00%	98.73%	76.77%	22.82%	-0.07
	GSO	98.00%	83.00%	98.53%	77.23%	47.39%	-0.06
	RAND	98.00%	83.00%	97.67%	77.50%	50.17%	-0.07
0.9	GEFeS	98.00%	83.00%	99.97%	77.17%	11.81%	-0.05
	PSO	98.00%	83.00%	98.27%	77.43%	43.82%	-0.06
	ABCO	98.00%	83.00%	98.53%	77.70%	45.27%	-0.06
	ASO	98.00%	83.00%	98.43%	76.93%	17.92%	-0.07
	GSO	98.00%	83.00%	98.30%	77.63%	45.76%	-0.06
	RAND	98.00%	83.00%	97.67%	77.43%	49.67%	-0.07
1.0	GEFeS	98.00%	83.00%	99.97%	77.17%	11.28%	-0.05
	PSO	98.00%	83.00%	98.43%	77.13%	43.42%	-0.07
	ABCO	98.00%	83.00%	98.33%	78.07%	44.59%	-0.06
	ASO	98.00%	83.00%	98.43%	76.77%	16.49%	-0.07
	GSO	98.00%	83.00%	98.00%	78.00%	45.55%	-0.06
	RAND	98.00%	83.00%	97.50%	78.07%	49.91%	-0.06

Table D.9. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, LIWC Feature Set - 93 Features, 175+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	68.00%	44.00%	89.97%	34.00%	57.67%	0.09
	PSO	68.00%	44.00%	78.53%	25.07%	55.66%	-0.28
	ABCO	68.00%	44.00%	75.20%	25.60%	57.74%	-0.31
	ASO	68.00%	44.00%	75.73%	64.13%	75.91%	0.57
	GSO	68.00%	44.00%	73.47%	25.47%	57.85%	-0.34
	RAND	68.00%	44.00%	72.00%	23.60%	48.57%	-0.40
0.1	GEFeS	68.00%	44.00%	88.40%	13.33%	43.19%	-0.40
	PSO	68.00%	44.00%	78.00%	24.93%	52.76%	-0.29
	ABCO	68.00%	44.00%	78.00%	22.80%	52.62%	-0.33
	ASO	68.00%	44.00%	76.00%	63.47%	74.87%	0.56
	GSO	68.00%	44.00%	75.87%	26.53%	56.88%	-0.28
	RAND	68.00%	44.00%	74.53%	22.67%	49.68%	-0.39
0.3	GEFeS	68.00%	44.00%	89.07%	16.13%	45.05%	-0.32
	PSO	68.00%	44.00%	80.13%	25.60%	51.72%	-0.24
	ABCO	68.00%	44.00%	77.60%	23.47%	54.44%	-0.33
	ASO	68.00%	44.00%	76.13%	61.60%	72.80%	0.52
	GSO	68.00%	44.00%	75.73%	23.07%	54.91%	-0.36
	RAND	68.00%	44.00%	72.67%	22.93%	50.11%	-0.41
0.5	GEFeS	68.00%	44.00%	88.00%	19.73%	42.97%	-0.26
	PSO	68.00%	44.00%	76.40%	21.07%	53.80%	-0.40
	ABCO	68.00%	44.00%	77.73%	18.67%	53.84%	-0.43
	ASO	68.00%	44.00%	76.40%	59.87%	71.43%	0.48
	GSO	68.00%	44.00%	75.33%	23.20%	53.66%	-0.36
	RAND	68.00%	44.00%	70.80%	19.20%	49.43%	-0.52
0.7	GEFeS	68.00%	44.00%	89.47%	16.00%	42.44%	-0.32
	PSO	68.00%	44.00%	77.87%	22.53%	51.47%	-0.34
	ABCO	68.00%	44.00%	77.47%	19.20%	53.76%	-0.42
	ASO	68.00%	44.00%	76.67%	56.67%	69.39%	0.42
	GSO	68.00%	44.00%	73.47%	23.13%	53.91%	-0.42
	RAND	68.00%	44.00%	71.33%	21.73%	50.32%	-0.46
0.9	GEFeS	68.00%	44.00%	88.00%	16.80%	42.01%	-0.32
	PSO	68.00%	44.00%	77.07%	20.13%	51.22%	-0.41
	ABCO	68.00%	44.00%	77.20%	20.13%	51.40%	-0.41
	ASO	68.00%	44.00%	75.33%	53.60%	67.60%	0.33
	GSO	68.00%	44.00%	74.27%	25.87%	53.12%	-0.32
	RAND	68.00%	44.00%	73.47%	23.20%	50.47%	-0.39
1.0	GEFeS	68.00%	44.00%	87.87%	14.93%	41.86%	-0.37
	PSO	68.00%	44.00%	77.20%	21.60%	51.72%	-0.37
	ABCO	68.00%	44.00%	78.40%	22.53%	51.72%	-0.33
	ASO	68.00%	44.00%	75.69%	50.80%	66.52%	0.27
	GSO	68.00%	44.00%	75.33%	23.33%	54.12%	-0.36
	RAND	68.00%	44.00%	71.47%	19.07%	50.90%	-0.52

Table D.10. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Topic Modeling Feature Set - 45 Features, 175+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	40.00%	0.00%	91.47%	10.80%	76.96%	1.29
	PSO	40.00%	0.00%	87.07%	10.27%	73.26%	1.18
	ABCO	40.00%	0.00%	87.20%	9.60%	74.89%	1.18
	ASO	40.00%	0.00%	85.73%	11.73%	84.96%	1.14
	GSO	40.00%	0.00%	84.67%	9.20%	72.59%	1.12
	RAND	40.00%	0.00%	79.60%	9.47%	74.89%	0.99
0.1	GEFeS	40.00%	0.00%	91.47%	10.67%	70.15%	1.29
	PSO	40.00%	0.00%	84.93%	9.07%	70.52%	1.12
	ABCO	40.00%	0.00%	87.20%	9.60%	73.26%	1.18
	ASO	40.00%	0.00%	85.87%	12.00%	85.33%	1.15
	GSO	40.00%	0.00%	84.80%	9.87%	73.78%	1.12
	RAND	40.00%	0.00%	79.33%	9.73%	73.26%	0.98
0.3	GEFeS	40.00%	0.00%	91.87%	10.67%	71.04%	1.30
	PSO	40.00%	0.00%	85.60%	10.00%	70.74%	1.14
	ABCO	40.00%	0.00%	88.47%	9.87%	72.52%	1.19
	ASO	40.00%	0.00%	85.73%	12.00%	84.44%	1.14
	GSO	40.00%	0.00%	86.00%	9.33%	75.70%	1.15
	RAND	40.00%	0.00%	79.33%	9.20%	72.52%	0.98
0.5	GEFeS	40.00%	0.00%	91.60%	10.93%	70.00%	1.29
	PSO	40.00%	0.00%	85.07%	9.20%	70.44%	1.13
	ABCO	40.00%	0.00%	86.80%	10.00%	72.44%	1.17
	ASO	40.00%	0.00%	86.40%	11.87%	83.33%	1.16
	GSO	40.00%	0.00%	83.87%	9.60%	73.04%	1.10
	RAND	40.00%	0.00%	80.13%	9.33%	72.44%	1.00
0.7	GEFeS	40.00%	0.00%	91.20%	11.60%	70.30%	1.28
	PSO	40.00%	0.00%	87.20%	9.33%	70.89%	1.18
	ABCO	40.00%	0.00%	88.00%	9.87%	71.70%	1.20
	ASO	40.00%	0.00%	86.40%	12.00%	83.19%	1.16
	GSO	40.00%	0.00%	85.33%	8.93%	71.19%	1.13
	RAND	40.00%	0.00%	78.80%	9.20%	71.70%	0.97
0.9	GEFeS	40.00%	0.00%	92.13%	10.53%	70.30%	1.30
	PSO	40.00%	0.00%	84.53%	9.20%	71.85%	1.11
	ABCO	40.00%	0.00%	87.20%	9.47%	72.30%	1.18
	ASO	40.00%	0.00%	87.07%	12.00%	82.81%	1.18
	GSO	40.00%	0.00%	86.13%	9.60%	71.48%	1.15
	RAND	40.00%	0.00%	79.87%	10.00%	72.30%	1.00
1.0	GEFeS	40.00%	0.00%	92.13%	10.67%	69.85%	1.30
	PSO	40.00%	0.00%	87.73%	9.60%	69.19%	1.19
	ABCO	40.00%	0.00%	88.27%	9.73%	72.15%	1.21
	ASO	40.00%	0.00%	87.33%	12.00%	83.11%	1.18
	GSO	40.00%	0.00%	83.87%	10.13%	72.81%	1.10
	RAND	40.00%	0.00%	78.93%	9.33%	72.15%	0.97

Table D.11. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Stylometry Feature Set - 428 Features, 175+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	48.00%	20.00%	94.80%	4.13%	49.55%	0.18
	PSO	48.00%	20.00%	72.93%	4.40%	50.79%	-0.26
	ABCO	48.00%	20.00%	69.87%	4.27%	51.21%	-0.33
	ASO	48.00%	20.00%	79.33%	14.27%	57.25%	0.37
	GSO	48.00%	20.00%	72.93%	4.13%	51.51%	-0.27
	RAND	48.00%	20.00%	63.73%	4.00%	50.07%	-0.47
0.1	GEFeS	48.00%	20.00%	96.93%	4.00%	34.35%	0.22
	PSO	48.00%	20.00%	72.93%	4.13%	48.47%	-0.27
	ABCO	48.00%	20.00%	71.73%	4.20%	49.09%	-0.29
	ASO	48.00%	20.00%	79.33%	10.93%	54.05%	0.20
	GSO	48.00%	20.00%	69.60%	4.40%	51.27%	-0.33
	RAND	48.00%	20.00%	66.93%	4.40%	50.40%	-0.39
0.3	GEFeS	48.00%	20.00%	96.27%	4.00%	32.54%	0.21
	PSO	48.00%	20.00%	75.47%	4.00%	48.12%	-0.23
	ABCO	48.00%	20.00%	69.87%	4.27%	49.31%	-0.33
	ASO	48.00%	20.00%	79.07%	9.07%	47.10%	0.10
	GSO	48.00%	20.00%	71.07%	4.27%	49.20%	-0.31
	RAND	48.00%	20.00%	68.80%	4.00%	49.70%	-0.37
0.5	GEFeS	48.00%	20.00%	96.27%	4.00%	29.88%	0.21
	PSO	48.00%	20.00%	74.80%	4.93%	47.74%	-0.20
	ABCO	48.00%	20.00%	70.93%	4.27%	48.15%	-0.31
	ASO	48.00%	20.00%	81.13%	5.87%	41.29%	0.04
	GSO	48.00%	20.00%	68.67%	4.27%	48.36%	-0.36
	RAND	48.00%	20.00%	65.07%	4.80%	49.82%	-0.40
0.7	GEFeS	48.00%	20.00%	96.00%	4.27%	26.00%	0.21
	PSO	48.00%	20.00%	75.47%	4.40%	46.97%	-0.21
	ABCO	48.00%	20.00%	72.53%	4.27%	47.98%	-0.28
	ASO	48.00%	20.00%	81.20%	4.93%	36.21%	0.06
	GSO	48.00%	20.00%	71.07%	4.13%	47.62%	-0.31
	RAND	48.00%	20.00%	66.53%	4.40%	50.19%	-0.39
0.9	GEFeS	48.00%	20.00%	96.73%	4.13%	23.16%	0.20
	PSO	48.00%	20.00%	71.73%	4.00%	46.50%	-0.31
	ABCO	48.00%	20.00%	70.00%	4.00%	46.44%	-0.34
	ASO	48.00%	20.00%	83.47%	4.00%	32.22%	0.06
	GSO	48.00%	20.00%	68.80%	4.27%	47.30%	-0.35
	RAND	48.00%	20.00%	66.20%	4.40%	50.97%	-0.42
1.0	GEFeS	48.00%	20.00%	96.40%	4.13%	21.81%	0.22
	PSO	48.00%	20.00%	74.40%	4.13%	45.56%	-0.24
	ABCO	48.00%	20.00%	69.47%	4.00%	46.16%	-0.35
	ASO	48.00%	20.00%	83.60%	4.00%	30.50%	0.06
	GSO	48.00%	20.00%	68.40%	4.27%	47.23%	-0.36
	RAND	48.00%	20.00%	64.67%	4.27%	49.78%	-0.44

Table D.12. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Hybrid Feature Set -566 Features, 175+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	92.00%	12.00%	100.00%	11.20%	52.59%	0.02
	PSO	92.00%	12.00%	95.87%	10.53%	51.38%	-0.28
	ABCO	92.00%	12.00%	94.93%	12.27%	51.50%	0.05
	ASO	92.00%	12.00%	96.00%	7.73%	56.50%	-0.31
	GSO	92.00%	12.00%	94.93%	12.53%	51.36%	0.08
	RAND	92.00%	12.00%	93.07%	11.87%	50.08%	0.00
0.1	GEFeS	92.00%	12.00%	100.00%	8.53%	13.53%	-0.20
	PSO	92.00%	12.00%	96.27%	8.80%	46.76%	-0.22
	ABCO	92.00%	12.00%	95.60%	12.00%	49.50%	0.04
	ASO	92.00%	12.00%	95.87%	6.67%	50.49%	-0.41
	GSO	92.00%	12.00%	94.67%	10.40%	50.69%	-0.10
	RAND	92.00%	12.00%	92.40%	11.47%	50.03%	-0.04
0.3	GEFeS	92.00%	12.00%	100.00%	8.67%	13.14%	-0.19
	PSO	92.00%	12.00%	97.07%	11.47%	46.36%	0.01
	ABCO	92.00%	12.00%	95.60%	11.07%	48.29%	-0.04
	ASO	92.00%	12.00%	96.27%	6.00%	39.12%	-0.45
	GSO	92.00%	12.00%	93.60%	11.73%	48.89%	0.00
	RAND	92.00%	12.00%	91.73%	11.33%	49.97%	-0.06
0.5	GEFeS	92.00%	12.00%	100.00%	8.40%	12.66%	-0.21
	PSO	92.00%	12.00%	95.47%	10.93%	45.78%	-0.05
	ABCO	92.00%	12.00%	96.27%	11.33%	47.46%	-0.01
	ASO	92.00%	12.00%	95.60%	7.07%	30.01%	-0.37
	GSO	92.00%	12.00%	93.33%	11.83%	47.93%	0.00
	RAND	92.00%	12.00%	91.33%	11.20%	50.44%	-0.07
0.7	GEFeS	92.00%	12.00%	100.00%	9.47%	12.39%	-0.12
	PSO	92.00%	12.00%	95.73%	10.53%	44.43%	-0.08
	ABCO	92.00%	12.00%	94.93%	12.67%	46.40%	0.09
	ASO	92.00%	12.00%	95.07%	7.73%	22.69%	-0.32
	GSO	92.00%	12.00%	93.33%	10.13%	46.81%	-0.14
	RAND	92.00%	12.00%	90.80%	9.73%	50.04%	-0.20
0.9	GEFeS	92.00%	12.00%	99.87%	8.67%	11.60%	-0.19
	PSO	92.00%	12.00%	94.67%	10.40%	44.00%	-0.10
	ABCO	92.00%	12.00%	94.00%	9.87%	45.15%	-0.16
	ASO	92.00%	12.00%	93.60%	6.93%	18.17%	-0.40
	GSO	92.00%	12.00%	94.13%	11.73%	45.32%	0.00
	RAND	92.00%	12.00%	89.07%	9.73%	50.05%	-0.22
1.0	GEFeS	92.00%	12.00%	100.00%	8.53%	11.78%	-0.20
	PSO	92.00%	12.00%	93.33%	10.13%	43.00%	-0.14
	ABCO	92.00%	12.00%	93.60%	10.00%	44.41%	-0.15
	ASO	92.00%	12.00%	93.33%	7.33%	16.20%	-0.37
	GSO	92.00%	12.00%	92.80%	8.00%	45.42%	-0.32
	RAND	92.00%	12.00%	90.93%	10.40%	49.59%	-0.14

Table D.13. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, LIWC Feature Set - 93 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	96.00%	93.00%	97.47%	83.50%	57.67%	-0.09
	PSO	96.00%	93.00%	94.63%	81.27%	55.66%	-0.14
	ABCO	96.00%	93.00%	93.80%	81.40%	57.74%	-0.15
	ASO	96.00%	93.00%	93.93%	91.03%	75.91%	-0.04
	GSO	96.00%	93.00%	93.33%	81.33%	57.85%	-0.15
	RAND	96.00%	93.00%	92.97%	80.87%	48.57%	-0.16
0.1	GEFeS	96.00%	93.00%	97.07%	78.30%	43.19%	-0.15
	PSO	96.00%	93.00%	94.47%	81.20%	52.76%	-0.14
	ABCO	96.00%	93.00%	94.50%	80.70%	52.62%	-0.15
	ASO	96.00%	93.00%	94.00%	90.87%	74.87%	-0.04
	GSO	96.00%	93.00%	93.97%	81.63%	56.88%	-0.14
	RAND	96.00%	93.00%	93.57%	80.60%	49.68%	-0.16
0.3	GEFeS	96.00%	93.00%	97.23%	79.00%	45.05%	-0.14
	PSO	96.00%	93.00%	95.03%	81.40%	51.72%	-0.13
	ABCO	96.00%	93.00%	94.40%	80.87%	54.44%	-0.15
	ASO	96.00%	93.00%	94.03%	90.40%	72.80%	-0.05
	GSO	96.00%	93.00%	93.93%	80.77%	54.91%	-0.15
	RAND	96.00%	93.00%	93.17%	80.73%	50.11%	-0.16
0.5	GEFeS	96.00%	93.00%	97.00%	79.93%	42.97%	-0.13
	PSO	96.00%	93.00%	94.10%	80.27%	53.80%	-0.16
	ABCO	96.00%	93.00%	94.43%	79.76%	53.84%	-0.16
	ASO	96.00%	93.00%	94.10%	89.97%	71.43%	-0.05
	GSO	96.00%	93.00%	93.83%	80.80%	53.66%	-0.15
	RAND	96.00%	93.00%	92.67%	79.77%	49.43%	-0.18
0.7	GEFeS	96.00%	93.00%	97.37%	79.00%	42.44%	-0.14
	PSO	96.00%	93.00%	94.47%	80.63%	51.47%	-0.15
	ABCO	96.00%	93.00%	94.37%	79.80%	53.76%	-0.16
	ASO	96.00%	93.00%	94.17%	89.17%	69.39%	-0.06
	GSO	96.00%	93.00%	93.37%	80.53%	53.91%	-0.16
	RAND	96.00%	93.00%	92.83%	80.43%	50.32%	-0.17
0.9	GEFeS	96.00%	93.00%	96.97%	79.17%	42.01%	-0.14
	PSO	96.00%	93.00%	94.27%	80.03%	51.22%	-0.16
	ABCO	96.00%	93.00%	94.30%	80.03%	51.40%	-0.16
	ASO	96.00%	93.00%	93.83%	88.40%	67.60%	-0.07
	GSO	96.00%	93.00%	93.57%	81.47%	53.12%	-0.15
	RAND	96.00%	93.00%	93.37%	80.80%	50.47%	-0.16
1.0	GEFeS	96.00%	93.00%	96.97%	78.73%	41.86%	-0.14
	PSO	96.00%	93.00%	94.30%	80.40%	51.72%	-0.15
	ABCO	96.00%	93.00%	94.60%	80.63%	51.72%	-0.15
	ASO	96.00%	93.00%	93.90%	87.70%	66.52%	-0.08
	GSO	96.00%	93.00%	93.80%	80.80%	54.12%	-0.15
	RAND	96.00%	93.00%	92.87%	79.77%	50.90%	-0.17

Table D.14. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Topic Modeling Feature Set - 45 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	85.00%	80.00%	96.90%	76.73%	76.96%	0.10
	PSO	85.00%	80.00%	94.87%	75.67%	73.26%	0.06
	ABCO	85.00%	80.00%	95.53%	76.13%	74.89%	0.08
	ASO	85.00%	80.00%	96.43%	77.93%	84.96%	0.11
	GSO	85.00%	80.00%	94.37%	75.50%	72.59%	0.05
	RAND	85.00%	80.00%	92.03%	74.50%	74.89%	0.01
0.1	GEFeS	85.00%	80.00%	96.10%	75.90%	70.15%	0.08
	PSO	85.00%	80.00%	93.67%	74.70%	70.52%	0.04
	ABCO	85.00%	80.00%	94.93%	75.53%	73.26%	0.06
	ASO	85.00%	80.00%	96.47%	78.00%	85.33%	0.11
	GSO	85.00%	80.00%	94.03%	75.30%	73.78%	0.05
	RAND	85.00%	80.00%	91.37%	73.97%	73.26%	0.00
0.3	GEFeS	85.00%	80.00%	96.53%	76.23%	71.04%	0.09
	PSO	85.00%	80.00%	93.77%	74.87%	70.74%	0.04
	ABCO	85.00%	80.00%	95.00%	75.60%	72.52%	0.06
	ASO	85.00%	80.00%	96.43%	78.00%	84.44%	0.11
	GSO	85.00%	80.00%	94.53%	75.%37	75.70%	0.05
	RAND	85.00%	80.00%	91.13%	73.60%	72.52%	-0.01
0.5	GEFeS	85.00%	80.00%	96.47%	76.30%	70.00%	0.09
	PSO	85.00%	80.00%	93.30%	74.33%	70.44%	0.03
	ABCO	85.00%	80.00%	94.73%	75.53%	72.44%	0.06
	ASO	85.00%	80.00%	96.60%	77.97%	83.33%	0.11
	GSO	85.00%	80.00%	93.50%	74.93%	73.04%	0.04
	RAND	85.00%	80.00%	91.40%	73.70%	72.44%	0.00
0.7	GEFeS	85.00%	80.00%	96.37%	76.47%	70.30%	0.09
	PSO	85.00%	80.00%	93.93%	74.47%	70.89%	0.04
	ABCO	85.00%	80.00%	95.33%	75.80%	71.70%	0.07
	ASO	85.00%	80.00%	96.60%	78.00%	83.19%	0.11
	GSO	85.00%	80.00%	94.03%	74.93%	71.19%	0.04
	RAND	85.00%	80.00%	90.63%	73.23%	71.70%	-0.02
0.9	GEFeS	85.00%	80.00%	96.30%	75.90%	70.30%	0.08
	PSO	85.00%	80.00%	93.23%	74.40%	71.85%	0.03
	ABCO	85.00%	80.00%	94.63%	75.20%	72.30%	0.05
	ASO	85.00%	80.00%	96.77%	78.00%	82.81%	0.11
	GSO	85.00%	80.00%	94.63%	75.50%	71.48%	0.06
	RAND	85.00%	80.00%	91.47%	74.00%	72.30%	0.00
1.0	GEFeS	85.00%	80.00%	96.57%	76.20%	69.85%	0.09
	PSO	85.00%	80.00%	94.97%	75.43%	69.19%	0.06
	ABCO	85.00%	80.00%	95.30%	75.67%	72.15%	0.07
	ASO	85.00%	80.00%	96.83%	78.00%	83.11%	0.11
	GSO	85.00%	80.00%	93.33%	74.90%	72.81%	0.03
	RAND	85.00%	80.00%	91.17%	73.77%	72.15%	-0.01

Table D.15. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Stylometry Feature Set - 428 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	93.50%	90.00%	98.70%	76.03%	49.55%	-0.10
	PSO	93.50%	90.00%	93.23%	76.10%	50.79%	-0.16
	ABCO	93.50%	90.00%	92.47%	76.07%	51.21%	-0.17
	ASO	93.50%	90.00%	94.83%	78.57%	57.25%	-0.11
	GSO	93.50%	90.00%	93.23%	76.03%	51.51%	-0.16
	RAND	93.50%	90.00%	90.93%	76.00%	50.07%	-0.18
0.1	GEFeS	93.50%	90.00%	99.23%	76.00%	34.35%	-0.09
	PSO	93.50%	90.00%	93.23%	76.03%	48.47%	-0.16
	ABCO	93.50%	90.00%	92.93%	76.10%	49.09%	-0.16
	ASO	93.50%	90.00%	94.83%	77.73%	54.05%	-0.17
	GSO	93.50%	90.00%	92.40%	76.10%	51.27%	-0.17
	RAND	93.50%	90.00%	91.73%	76.10%	50.40%	-0.17
0.3	GEFeS	93.50%	90.00%	99.07%	76.00%	32.54%	-0.10
	PSO	93.50%	90.00%	93.87%	76.00%	48.12%	-0.15
	ABCO	93.50%	90.00%	92.47%	76.07%	49.31%	-0.13
	ASO	93.50%	90.00%	94.77%	77.27%	47.10%	-0.12
	GSO	93.50%	90.00%	92.77%	76.07%	49.20%	-0.16
	RAND	93.50%	90.00%	92.20%	76.00%	49.70%	-0.17
0.5	GEFeS	93.50%	90.00%	99.07%	76.00%	29.88%	-0.10
	PSO	93.50%	90.00%	93.70%	76.23%	47.74%	-0.15
	ABCO	93.50%	90.00%	92.73%	76.07%	48.15%	-0.13
	ASO	93.50%	90.00%	95.03%	76.47%	41.29%	-0.17
	GSO	93.50%	90.00%	92.17%	76.07%	48.36%	-0.17
	RAND	93.50%	90.00%	91.27%	76.20%	49.82%	-0.18
0.7	GEFeS	93.50%	90.00%	99.00%	76.07%	26.00%	-0.10
	PSO	93.50%	90.00%	93.87%	76.10%	46.97%	-0.15
	ABCO	93.50%	90.00%	93.13%	76.07%	47.98%	-0.16
	ASO	93.50%	90.00%	95.30%	76.23%	36.21%	-0.13
	GSO	93.50%	90.00%	92.77%	76.03%	47.62%	-0.16
	RAND	93.50%	90.00%	91.63%	76.10%	50.19%	-0.17
0.9	GEFeS	93.50%	90.00%	98.93%	76.03%	23.16%	-0.10
	PSO	93.50%	90.00%	92.93%	76.00%	46.50%	-0.16
	ABCO	93.50%	90.00%	92.50%	76.00%	46.44%	-0.17
	ASO	93.50%	90.00%	95.87%	76.00%	32.22%	-0.13
	GSO	93.50%	90.00%	92.20%	76.07%	47.30%	-0.17
	RAND	93.50%	90.00%	91.30%	76.10%	50.97%	-0.18
1.0	GEFeS	93.50%	90.00%	99.10%	76.03%	21.81%	-0.10
	PSO	93.50%	90.00%	93.60%	76.03%	45.56%	-0.15
	ABCO	93.50%	90.00%	92.37%	76.00%	46.16%	-0.17
	ASO	93.50%	90.00%	95.90%	76.00%	30.50%	-0.13
	GSO	93.50%	90.00%	92.10%	76.07%	47.23%	-0.17
	RAND	93.50%	90.00%	91.17%	76.07%	49.78%	-0.18

Table D.16. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-50, Hybrid Feature Set - 566 Features, 175+(org = 200, adv = 200).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	99.00%	89.00%	100.00%	77.80%	52.59%	-0.12
	PSO	99.00%	89.00%	99.97%	77.63%	51.38%	-0.13
	ABCO	99.00%	89.00%	98.73%	78.07%	51.50%	-0.13
	ASO	99.00%	89.00%	99.00%	76.93%	56.50%	-0.14
	GSO	99.00%	89.00%	98.73%	78.13%	51.36%	-0.12
	RAND	99.00%	89.00%	98.27%	77.97%	50.08%	-0.13
0.1	GEFeS	99.00%	89.00%	100.00%	77.13%	13.53%	-0.12
	PSO	99.00%	89.00%	99.07%	77.20%	46.76%	-0.13
	ABCO	99.00%	89.00%	98.90%	78.00%	49.50%	-0.12
	ASO	99.00%	89.00%	98.97%	76.67%	50.49%	-0.14
	GSO	99.00%	89.00%	98.67%	77.60%	50.69%	-0.13
	RAND	99.00%	89.00%	98.10%	77.87%	50.03%	-0.13
0.3	GEFeS	99.00%	89.00%	100.00%	77.17%	13.14%	-0.12
	PSO	99.00%	89.00%	99.27%	77.87%	46.36%	-0.12
	ABCO	99.00%	89.00%	98.90%	77.77%	48.29%	-0.13
	ASO	99.00%	89.00%	99.07%	76.50%	39.12%	-0.14
	GSO	99.00%	89.00%	98.40%	77.93%	48.89%	-0.13
	RAND	99.00%	89.00%	97.93%	77.83%	49.97%	-0.14
0.5	GEFeS	99.00%	89.00%	100.00%	77.10%	12.66%	-0.12
	PSO	99.00%	89.00%	98.87%	77.73%	45.78%	-0.13
	ABCO	99.00%	89.00%	99.07%	77.83%	47.46%	-0.12
	ASO	99.00%	89.00%	98.90%	76.77%	30.01%	-0.14
	GSO	99.00%	89.00%	98.33%	77.97%	47.93%	-0.13
	RAND	99.00%	89.00%	97.83%	77.80%	50.44%	-0.14
0.7	GEFeS	99.00%	89.00%	100.00%	77.37%	12.39%	-0.12
	PSO	99.00%	89.00%	98.93%	77.63%	44.43%	-0.13
	ABCO	99.00%	89.00%	98.73%	78.17%	46.40%	-0.12
	ASO	99.00%	89.00%	98.77%	76.93%	22.69%	-0.14
	GSO	99.00%	89.00%	98.33%	77.53%	46.81%	-0.14
	RAND	99.00%	89.00%	97.70%	77.43%	50.04%	-0.14
0.9	GEFeS	99.00%	89.00%	99.97%	77.17%	11.60%	-0.12
	PSO	99.00%	89.00%	98.67%	77.60%	44.00%	-0.13
	ABCO	99.00%	89.00%	98.50%	77.47%	45.15%	-0.13
	ASO	99.00%	89.00%	98.40%	76.73%	18.17%	-0.14
	GSO	99.00%	89.00%	98.53%	77.93%	45.32%	-0.13
	RAND	99.00%	89.00%	97.27%	77.43%	50.05%	-0.15
1.0	GEFeS	99.00%	89.00%	100.00%	77.13%	11.78%	-0.12
	PSO	99.00%	89.00%	98.33%	77.53%	43.00%	-0.14
	ABCO	99.00%	89.00%	98.40%	77.50%	44.41%	-0.14
	ASO	99.00%	89.00%	98.33%	76.83%	16.20%	-0.14
	GSO	99.00%	89.00%	98.20%	77.00%	45.42%	-0.14
	RAND	99.00%	89.00%	97.73%	77.60%	49.59%	-0.14

Table D.17. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, LIWC Feature Set - 93 Features, 375+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	52.00%	44.00%	54.40%	6.27%	65.09%	-0.81
	PSO	52.00%	44.00%	50.40%	5.47%	60.72%	-0.91
	ABCO	52.00%	44.00%	53.20%	6.13%	60.00%	-0.84
	ASO	52.00%	44.00%	57.87%	13.87%	81.08%	-0.57
	GSO	52.00%	44.00%	51.73%	3.60%	61.00%	-0.92
	RAND	52.00%	44.00%	49.33%	4.53%	49.14%	-0.95
0.1	GEFeS	52.00%	44.00%	54.27%	3.47%	61.65%	-0.88
	PSO	52.00%	44.00%	51.33%	4.00%	59.50%	-0.92
	ABCO	52.00%	44.00%	50.27%	4.27%	59.46%	-0.94
	ASO	52.00%	44.00%	56.67%	14.00%	80.22%	-0.59
	GSO	52.00%	44.00%	50.80%	3.87%	60.22%	-0.94
	RAND	52.00%	44.00%	49.07%	4.67%	51.79%	-0.95
0.3	GEFeS	52.00%	44.00%	51.47%	1.47%	59.75%	-0.98
	PSO	52.00%	44.00%	51.20%	3.33%	57.63%	-0.94
	ABCO	52.00%	44.00%	50.67%	4.93%	58.67%	-0.91
	ASO	52.00%	44.00%	56.27%	13.73%	78.67%	-0.61
	GSO	52.00%	44.00%	50.67%	2.67%	57.63%	-0.97
	RAND	52.00%	44.00%	49.60%	2.27%	49.82%	-0.99
0.5	GEFeS	52.00%	44.00%	53.73%	1.60%	56.77%	-0.93
	PSO	52.00%	44.00%	51.87%	3.33%	56.20%	-0.93
	ABCO	52.00%	44.00%	50.53%	5.07%	57.35%	-0.91
	ASO	52.00%	44.00%	55.33%	12.67%	76.63%	-0.65
	GSO	52.00%	44.00%	48.67%	4.13%	58.14%	-0.97
	RAND	52.00%	44.00%	48.13%	2.40%	50.79%	-1.02
0.7	GEFeS	52.00%	44.00%	53.87%	2.27%	55.70%	-0.91
	PSO	52.00%	44.00%	47.87%	2.00%	55.66%	-1.03
	ABCO	52.00%	44.00%	50.93%	4.13%	55.63%	-0.93
	ASO	52.00%	44.00%	56.53%	12.40%	74.23%	-0.63
	GSO	52.00%	44.00%	48.67%	1.73%	58.96%	-1.02
	RAND	52.00%	44.00%	49.73%	2.53%	50.36%	-0.99
0.9	GEFeS	52.00%	44.00%	55.87%	0.67%	52.19%	-0.94
	PSO	52.00%	44.00%	50.27%	2.00%	53.41%	-0.99
	ABCO	52.00%	44.00%	49.20%	0.53%	53.37%	-1.04
	ASO	52.00%	44.00%	56.00%	12.27%	71.61%	-0.64
	GSO	52.00%	44.00%	50.80%	1.60%	55.66%	-0.99
	RAND	52.00%	44.00%	47.47%	3.07%	50.11%	-1.02
1.0	GEFeS	52.00%	44.00%	54.80%	0.27%	52.47%	-0.37
	PSO	52.00%	44.00%	52.40%	2.93%	55.13%	-0.93
	ABCO	52.00%	44.00%	51.47%	1.60%	53.51%	-0.97
	ASO	52.00%	44.00%	55.07%	12.93%	70.47%	-0.65
	GSO	52.00%	44.00%	48.27%	3.20%	55.23%	-1.00
	RAND	52.00%	44.00%	48.13%	2.40%	49.25%	-1.02

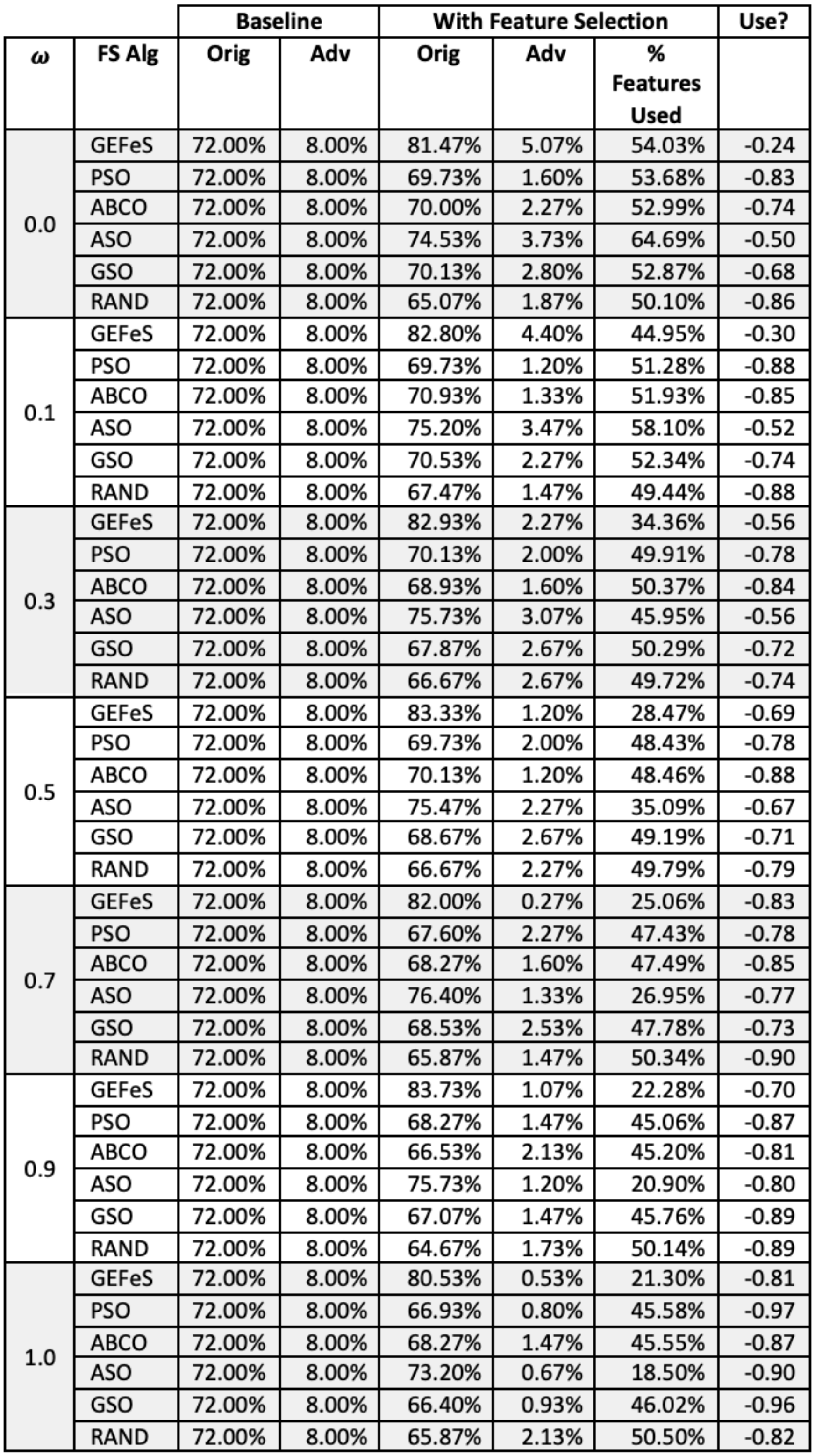
Table D.18. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Topic Modeling Feature Set - 45 Features, 375+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	32.00%	0.00%	42.13%	0.00%	87.78%	0.10
	PSO	32.00%	0.00%	35.33%	0.00%	80.59%	0.03
	ABCO	32.00%	0.00%	37.47%	0.00%	81.93%	0.05
	ASO	32.00%	0.00%	32.00%	0.00%	100.00%	0.00
	GSO	32.00%	0.00%	36.27%	0.00%	80.59%	0.04
	RAND	32.00%	0.00%	33.20%	0.00%	49.78%	0.01
0.1	GEFeS	32.00%	0.00%	41.73%	0.00%	85.70%	0.10
	PSO	32.00%	0.00%	35.20%	0.00%	79.41%	0.03
	ABCO	32.00%	0.00%	36.27%	0.00%	82.22%	0.04
	ASO	32.00%	0.00%	32.00%	0.00%	100.00 %	0.00
	GSO	32.00%	0.00%	34.67%	0.00%	80.52%	0.03
	RAND	32.00%	0.00%	29.73%	0.13%	48.52%	-0.02
0.3	GEFeS	32.00%	0.00%	42.27%	0.00%	86.15%	0.10
	PSO	32.00%	0.00%	37.22%	0.00%	79.63%	0.05
	ABCO	32.00%	0.00%	38.53%	0.00%	82.89%	0.07
	ASO	32.00%	0.00%	32.00%	0.00%	100.00%	0.00
	GSO	32.00%	0.00%	33.47%	0.13%	78.00%	0.01
	RAND	32.00%	0.00%	33.47%	0.00%	48.96%	0.01
0.5	GEFeS	32.00%	0.00%	42.40%	0.00%	86.30%	0.10
	PSO	32.00%	0.00%	36.53%	0.00%	78.30%	0.05
	ABCO	32.00%	0.00%	36.53%	0.00%	79.93%	0.05
	ASO	32.00%	0.00%	32.00%	0.00%	100.00%	0.00
	GSO	32.00%	0.00%	34.67%	0.00%	79.48%	0.03
	RAND	32.00%	0.00%	32.93%	0.13%	49.56%	0.01
0.7	GEFeS	32.00%	0.00%	42.80%	0.00%	86.74%	0.11
	PSO	32.00%	0.00%	37.07%	0.00%	77.93%	0.05
	ABCO	32.00%	0.00%	34.67%	0.00%	81.41%	0.03
	ASO	32.00%	0.00%	32.00%	0.00%	99.93%	0.00
	GSO	32.00%	0.00%	35.87%	0.00%	79.11%	0.04
	RAND	32.00%	0.00%	32.40%	0.00%	49.56%	0.00
0.9	GEFeS	32.00%	0.00%	43.87%	0.00%	84.67%	0.12
	PSO	32.00%	0.00%	36.13%	0.00%	80.15%	0.04
	ABCO	32.00%	0.00%	37.73%	0.00%	79.26%	0.06
	ASO	32.00%	0.00%	32.00%	0.00%	99.85%	0.00
	GSO	32.00%	0.00%	34.67%	0.00%	78.15%	0.03
	RAND	32.00%	0.00%	33.20%	0.13%	48.59%	0.01
1.0	GEFeS	32.00%	0.00%	41.60%	0.00%	84.22%	0.10
	PSO	32.00%	0.00%	37.73%	0.00%	78.30%	0.06
	ABCO	32.00%	0.00%	37.07%	0.00%	77.48%	0.05
	ASO	32.00%	0.00%	32.00%	0.00%	99.85%	0.00
	GSO	32.00%	0.00%	35.07%	0.13%	78.44%	0.03
	RAND	32.00%	0.00%	34.27%	0.27%	51.19%	0.02

Table D.19. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Stylometry Feature Set - 428 Features, 375+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	28.00%	12.00%	52.40%	0.67%	52.83%	-0.07
	PSO	28.00%	12.00%	31.73%	1.60%	52.55%	-0.73
	ABCO	28.00%	12.00%	31.73%	0.53%	53.79%	-0.82
	ASO	28.00%	12.00%	43.07%	19.07%	59.79%	1.13
	GSO	28.00%	12.00%	31.73%	0.40%	52.62%	-0.83
	RAND	28.00%	12.00%	32.27%	1.07%	50.23%	-0.76
0.1	GEFeS	28.00%	12.00%	52.40%	0.40%	49.31%	-0.10
	PSO	28.00%	12.00%	34.27%	0.67%	51.92%	-0.72
	ABCO	28.00%	12.00%	31.33%	1.07%	51.14%	-0.79
	ASO	28.00%	12.00%	46.00%	19.20%	54.95%	1.24
	GSO	28.00%	12.00%	31.47%	0.93%	52.40%	-0.80
	RAND	28.00%	12.00%	31.60%	0.93%	49.40%	-0.79
0.3	GEFeS	28.00%	12.00%	53.07%	0.13%	42.21%	-0.09
	PSO	28.00%	12.00%	31.87%	1.33%	49.87%	-0.75
	ABCO	28.00%	12.00%	34.00%	0.67%	50.03%	-0.73
	ASO	28.00%	12.00%	48.27%	19.20%	47.73%	1.32
	GSO	28.00%	12.00%	34.53%	1.20%	50.79%	-0.67
	RAND	28.00%	12.00%	30.13%	0.80%	50.21%	-0.86
0.5	GEFeS	28.00%	12.00%	55.73%	0.00%	36.11%	-0.01
	PSO	28.00%	12.00%	32.67%	0.93%	48.30%	-0.76
	ABCO	28.00%	12.00%	33.07%	1.33%	48.79%	-0.71
	ASO	28.00%	12.00%	47.07%	14.67%	41.86%	0.90
	GSO	28.00%	12.00%	32.53%	0.08%	49.03%	-0.77
	RAND	28.00%	12.00%	29.33%	0.80%	50.37%	-0.89
0.7	GEFeS	28.00%	12.00%	52.93%	0.00%	32.41%	-0.11
	PSO	28.00%	12.00%	34.13%	0.67%	46.50%	-0.73
	ABCO	28.00%	12.00%	32.40%	1.33%	47.02%	-0.73
	ASO	28.00%	12.00%	46.93%	17.60%	37.46%	0.31
	GSO	28.00%	12.00%	30.80%	1.47%	46.95%	-0.78
	RAND	28.00%	12.00%	31.47%	0.27%	50.02%	-0.85
0.9	GEFeS	28.00%	12.00%	54.40%	0.40%	27.90%	-0.02
	PSO	28.00%	12.00%	31.20%	0.80%	45.62%	-0.82
	ABCO	28.00%	12.00%	32.80%	1.20%	45.54%	-0.73
	ASO	28.00%	12.00%	47.73%	1.47%	32.78%	-0.17
	GSO	28.00%	12.00%	32.00%	0.67%	45.86%	-0.80
	RAND	28.00%	12.00%	30.13%	1.47%	49.93%	-0.80
1.0	GEFeS	28.00%	12.00%	54.00%	0.67%	26.20%	-0.02
	PSO	28.00%	12.00%	32.93%	1.20%	44.14%	-0.72
	ABCO	28.00%	12.00%	30.00%	1.07%	45.58%	-0.84
	ASO	28.00%	12.00%	48.53%	0.00%	30.49%	-0.27
	GSO	28.00%	12.00%	32.27%	1.07%	44.98%	-0.76
	RAND	28.00%	12.00%	30.80%	0.80%	49.91%	-0.83

Table D.20. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Hybrid Feature Set - 566 Features, 375+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	72.00%	8.00%	81.47%	5.07%	54.03%	-0.24
	PSO	72.00%	8.00%	69.73%	1.60%	53.68%	-0.83
	ABCO	72.00%	8.00%	70.00%	2.27%	52.99%	-0.74
	ASO	72.00%	8.00%	74.53%	3.73%	64.69%	-0.50
	GSO	72.00%	8.00%	70.13%	2.80%	52.87%	-0.68
	RAND	72.00%	8.00%	65.07%	1.87%	50.10%	-0.86
0.1	GEFeS	72.00%	8.00%	82.80%	4.40%	44.95%	-0.30
	PSO	72.00%	8.00%	69.73%	1.20%	51.28%	-0.88
	ABCO	72.00%	8.00%	70.93%	1.33%	51.93%	-0.85
	ASO	72.00%	8.00%	75.20%	3.47%	58.10%	-0.52
	GSO	72.00%	8.00%	70.53%	2.27%	52.34%	-0.74
	RAND	72.00%	8.00%	67.47%	1.47%	49.44%	-0.88
0.3	GEFeS	72.00%	8.00%	82.93%	2.27%	34.36%	-0.56
	PSO	72.00%	8.00%	70.13%	2.00%	49.91%	-0.78
	ABCO	72.00%	8.00%	68.93%	1.60%	50.37%	-0.84
	ASO	72.00%	8.00%	75.73%	3.07%	45.95%	-0.56
	GSO	72.00%	8.00%	67.87%	2.67%	50.29%	-0.72
	RAND	72.00%	8.00%	66.67%	2.67%	49.72%	-0.74
0.5	GEFeS	72.00%	8.00%	83.33%	1.20%	28.47%	-0.69
	PSO	72.00%	8.00%	69.73%	2.00%	48.43%	-0.78
	ABCO	72.00%	8.00%	70.13%	1.20%	48.46%	-0.88
	ASO	72.00%	8.00%	75.47%	2.27%	35.09%	-0.67
	GSO	72.00%	8.00%	68.67%	2.67%	49.19%	-0.71
	RAND	72.00%	8.00%	66.67%	2.27%	49.79%	-0.79
0.7	GEFeS	72.00%	8.00%	82.00%	0.27%	25.06%	-0.83
	PSO	72.00%	8.00%	67.60%	2.27%	47.43%	-0.78
	ABCO	72.00%	8.00%	68.27%	1.60%	47.49%	-0.85
	ASO	72.00%	8.00%	76.40%	1.33%	26.95%	-0.77
	GSO	72.00%	8.00%	68.53%	2.53%	47.78%	-0.73
	RAND	72.00%	8.00%	65.87%	1.47%	50.34%	-0.90
0.9	GEFeS	72.00%	8.00%	83.73%	1.07%	22.28%	-0.70
	PSO	72.00%	8.00%	68.27%	1.47%	45.06%	-0.87
	ABCO	72.00%	8.00%	66.53%	2.13%	45.20%	-0.81
	ASO	72.00%	8.00%	75.73%	1.20%	20.90%	-0.80
	GSO	72.00%	8.00%	67.07%	1.47%	45.76%	-0.89
	RAND	72.00%	8.00%	64.67%	1.73%	50.14%	-0.89
1.0	GEFeS	72.00%	8.00%	80.53%	0.53%	21.30%	-0.81
	PSO	72.00%	8.00%	66.93%	0.80%	45.58%	-0.97
	ABCO	72.00%	8.00%	68.27%	1.47%	45.55%	-0.87
	ASO	72.00%	8.00%	73.20%	0.67%	18.50%	-0.90
	GSO	72.00%	8.00%	66.40%	0.93%	46.02%	-0.96
	RAND	72.00%	8.00%	65.87%	2.13%	50.50%	-0.82

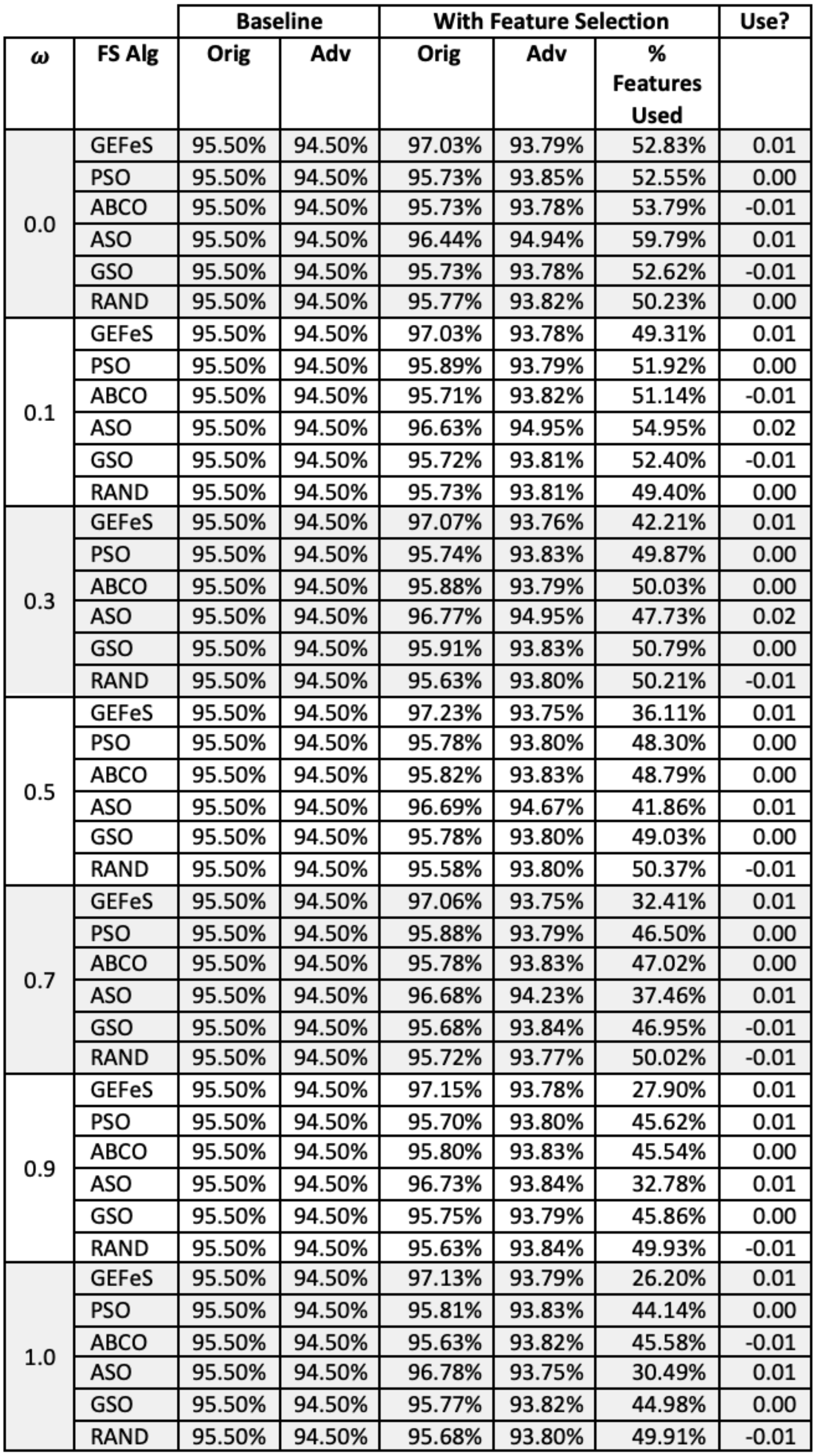
Table D.21. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, LIWC Feature Set - 93 Features, 375+(org = 400, adv = 400).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	96.75%	96.25%	96.85%	93.84%	65.09%	-0.02
	PSO	96.75%	96.25%	96.50%	93.69%	60.72%	-0.03
	ABCO	96.75%	96.25%	96.59%	93.65%	60.00%	-0.03
	ASO	96.75%	96.25%	97.17%	94.42%	81.08%	-0.01
	GSO	96.75%	96.25%	96.59%	93.58%	61.00%	-0.03
	RAND	96.75%	96.25%	96.25%	93.45%	49.14%	-0.03
0.1	GEFeS	96.75%	96.25%	96.88%	93.70%	61.65%	-0.03
	PSO	96.75%	96.25%	96.42%	93.46%	59.50%	-0.03
	ABCO	96.75%	96.25%	96.48%	93.60%	59.46%	-0.03
	ASO	96.75%	96.25%	97.04%	94.38%	80.22%	-0.02
	GSO	96.75%	96.25%	96.43%	93.50%	60.22%	-0.03
	RAND	96.75%	96.25%	96.24%	93.47%	51.79%	-0.03
0.3	GEFeS	96.75%	96.25%	96.67%	93.54%	59.75%	-0.03
	PSO	96.75%	96.25%	96.46%	93.47%	57.63%	-0.03
	ABCO	96.75%	96.25%	96.28%	93.42%	58.67%	-0.03
	ASO	96.75%	96.25%	96.98%	94.33%	78.67%	-0.02
	GSO	96.75%	96.25%	96.46%	93.46%	57.63%	-0.03
	RAND	96.75%	96.25%	96.45%	93.49%	49.82%	-0.03
0.5	GEFeS	96.75%	96.25%	96.88%	93.42%	56.77%	-0.03
	PSO	96.75%	96.25%	96.48%	93.44%	56.20%	-0.03
	ABCO	96.75%	96.25%	96.38%	93.53%	57.35%	-0.03
	ASO	96.75%	96.25%	96.86%	94.19%	76.63%	-0.02
	GSO	96.75%	96.25%	96.26%	93.48%	58.14%	-0.03
	RAND	96.75%	96.25%	96.26%	93.40%	50.79%	-0.03
0.7	GEFeS	96.75%	96.25%	96.72%	93.49%	55.70%	-0.03
	PSO	96.75%	96.25%	96.34%	93.48%	55.66%	-0.03
	ABCO	96.75%	96.25%	96.47%	93.54%	55.63%	-0.03
	ASO	96.75%	96.25%	96.92%	94.16%	74.23%	-0.02
	GSO	96.75%	96.25%	96.40%	93.47%	58.96%	-0.03
	RAND	96.75%	96.25%	96.27%	93.32%	50.36%	-0.04
0.9	GEFeS	96.75%	96.25%	96.79%	93.34%	52.19%	-0.03
	PSO	96.75%	96.25%	96.35%	93.33%	53.41%	-0.03
	ABCO	96.75%	96.25%	96.35%	93.31%	53.37%	-0.03
	ASO	96.75%	96.25%	96.83%	94.10%	71.61%	-0.02
	GSO	96.75%	96.25%	96.38%	93.30%	55.66%	-0.03
	RAND	96.75%	96.25%	96.16%	93.38%	50.11%	-0.04
1.0	GEFeS	96.75%	96.25%	96.77%	93.36%	52.47%	-0.03
	PSO	96.75%	96.25%	96.39%	93.30%	55.13%	-0.03
	ABCO	96.75%	96.25%	96.35%	93.23%	53.51%	-0.04
	ASO	96.75%	96.25%	96.73%	94.09%	70.47%	-0.02
	GSO	96.75%	96.25%	96.22%	93.40%	55.23%	-0.04
	RAND	96.75%	96.25%	96.10%	93.24%	49.25%	-0.04

Table D.22. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Topic Modeling Feature Set - 45 Features, 375+(org = 400, adv = 400).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	88.75%	86.75%	88.96%	86.33%	87.78%	0.00
	PSO	88.75%	86.75%	87.31%	85.10%	80.59%	-0.04
	ABCO	88.75%	86.75%	87.85%	85.51%	81.93%	-0.02
	ASO	88.75%	86.75%	88.75%	86.75%	100.00%	0.00
	GSO	88.75%	86.75%	87.94%	85.68%	80.59%	-0.02
	RAND	88.75%	86.75%	87.15%	85.08%	49.78%	-0.04
0.1	GEFeS	88.75%	86.75%	88.88%	86.27%	85.70%	0.00
	PSO	88.75%	86.75%	87.47%	85.27%	79.41%	-0.03
	ABCO	88.75%	86.75%	87.86%	85.59%	82.22%	-0.02
	ASO	88.75%	86.75%	88.75%	86.75%	100.00 %	0.00
	GSO	88.75%	86.75%	87.73%	85.56%	80.52%	-0.03
	RAND	88.75%	86.75%	86.56%	84.71%	48.52%	-0.05
0.3	GEFeS	88.75%	86.75%	88.95%	86.31%	86.15%	0.00
	PSO	88.75%	86.75%	87.73%	85.39%	79.63%	-0.03
	ABCO	88.75%	86.75%	87.88%	85.47%	82.89%	-0.02
	ASO	88.75%	86.75%	88.75%	86.75%	100.00%	0.00
	GSO	88.75%	86.75%	87.47%	85.38%	78.00%	-0.03
	RAND	88.75%	86.75%	87.07%	84.98%	48.96%	-0.04
0.5	GEFeS	88.75%	86.75%	88.94%	86.19%	86.30%	-0.01
	PSO	88.75%	86.75%	87.59%	85.31%	78.30%	-0.03
	ABCO	88.75%	86.75%	87.51%	85.23%	79.93%	-0.03
	ASO	88.75%	86.75%	88.75%	86.75%	100.00%	0.00
	GSO	88.75%	86.75%	87.42%	85.25%	79.48%	-0.03
	RAND	88.75%	86.75%	86.87%	84.82%	49.56%	-0.04
0.7	GEFeS	88.75%	86.75%	88.90%	86.23%	86.74%	0.00
	PSO	88.75%	86.75%	87.49%	85.18%	77.93%	-0.03
	ABCO	88.75%	86.75%	87.58%	85.41%	81.41%	-0.03
	ASO	88.75%	86.75%	88.75%	86.75%	99.93%	0.00
	GSO	88.75%	86.75%	87.47%	85.23%	79.11%	-0.03
	RAND	88.75%	86.75%	86.75%	84.73%	49.56%	-0.05
0.9	GEFeS	88.75%	86.75%	88.92%	86.18%	84.67%	0.00
	PSO	88.75%	86.75%	87.44%	85.19%	80.15%	-0.03
	ABCO	88.75%	86.75%	87.76%	85.41%	79.26%	-0.03
	ASO	88.75%	86.75%	88.75%	86.75%	99.85%	0.00
	GSO	88.75%	86.75%	87.18%	85.02%	78.15%	-0.04
	RAND	88.75%	86.75%	87.09%	85.03%	48.59%	-0.04
1.0	GEFeS	88.75%	86.75%	88.81%	86.21%	84.22%	-0.01
	PSO	88.75%	86.75%	87.28%	84.92%	78.30%	-0.04
	ABCO	88.75%	86.75%	87.57%	86.25%	77.48%	-0.03
	ASO	88.75%	86.75%	88.75%	86.75%	99.85%	0.00
	GSO	88.75%	86.75%	87.45%	85.27%	78.44%	-0.03
	RAND	88.75%	86.75%	87.03%	84.90%	51.19%	-0.04

Table D.23. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Stylometry Feature Set - 428 Features, 375+(org = 400, adv = 400).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	95.50%	94.50%	97.03%	93.79%	52.83%	0.01
	PSO	95.50%	94.50%	95.73%	93.85%	52.55%	0.00
	ABCO	95.50%	94.50%	95.73%	93.78%	53.79%	-0.01
	ASO	95.50%	94.50%	96.44%	94.94%	59.79%	0.01
	GSO	95.50%	94.50%	95.73%	93.78%	52.62%	-0.01
	RAND	95.50%	94.50%	95.77%	93.82%	50.23%	0.00
0.1	GEFeS	95.50%	94.50%	97.03%	93.78%	49.31%	0.01
	PSO	95.50%	94.50%	95.89%	93.79%	51.92%	0.00
	ABCO	95.50%	94.50%	95.71%	93.82%	51.14%	-0.01
	ASO	95.50%	94.50%	96.63%	94.95%	54.95%	0.02
	GSO	95.50%	94.50%	95.72%	93.81%	52.40%	-0.01
	RAND	95.50%	94.50%	95.73%	93.81%	49.40%	0.00
0.3	GEFeS	95.50%	94.50%	97.07%	93.76%	42.21%	0.01
	PSO	95.50%	94.50%	95.74%	93.83%	49.87%	0.00
	ABCO	95.50%	94.50%	95.88%	93.79%	50.03%	0.00
	ASO	95.50%	94.50%	96.77%	94.95%	47.73%	0.02
	GSO	95.50%	94.50%	95.91%	93.83%	50.79%	0.00
	RAND	95.50%	94.50%	95.63%	93.80%	50.21%	-0.01
0.5	GEFeS	95.50%	94.50%	97.23%	93.75%	36.11%	0.01
	PSO	95.50%	94.50%	95.78%	93.80%	48.30%	0.00
	ABCO	95.50%	94.50%	95.82%	93.83%	48.79%	0.00
	ASO	95.50%	94.50%	96.69%	94.67%	41.86%	0.01
	GSO	95.50%	94.50%	95.78%	93.80%	49.03%	0.00
	RAND	95.50%	94.50%	95.58%	93.80%	50.37%	-0.01
0.7	GEFeS	95.50%	94.50%	97.06%	93.75%	32.41%	0.01
	PSO	95.50%	94.50%	95.88%	93.79%	46.50%	0.00
	ABCO	95.50%	94.50%	95.78%	93.83%	47.02%	0.00
	ASO	95.50%	94.50%	96.68%	94.23%	37.46%	0.01
	GSO	95.50%	94.50%	95.68%	93.84%	46.95%	-0.01
	RAND	95.50%	94.50%	95.72%	93.77%	50.02%	-0.01
0.9	GEFeS	95.50%	94.50%	97.15%	93.78%	27.90%	0.01
	PSO	95.50%	94.50%	95.70%	93.80%	45.62%	0.01
	ABCO	95.50%	94.50%	95.80%	93.83%	45.54%	0.00
	ASO	95.50%	94.50%	96.73%	93.84%	32.78%	0.01
	GSO	95.50%	94.50%	95.75%	93.79%	45.86%	0.00
	RAND	95.50%	94.50%	95.63%	93.84%	49.93%	-0.01
1.0	GEFeS	95.50%	94.50%	97.13%	93.79%	26.20%	0.01
	PSO	95.50%	94.50%	95.81%	93.83%	44.14%	0.00
	ABCO	95.50%	94.50%	95.63%	93.82%	45.58%	-0.01
	ASO	95.50%	94.50%	96.78%	93.75%	30.49%	0.01
	GSO	95.50%	94.50%	95.77%	93.82%	44.98%	0.00
	RAND	95.50%	94.50%	95.68%	93.80%	49.91%	-0.01

Table D.24. A Comparison of Adversarial Author Identification with and without Feature Selection Using the CASIS-100, Hybrid Feature Set - 566 Features, 375+(org = 400, adv = 400).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	98.25%	94.25%	98.84%	94.07%	54.03%	0.00
	PSO	98.25%	94.25%	98.06%	93.80%	53.68%	-0.01
	ABCO	98.25%	94.25%	98.08%	93.84%	52.99%	-0.01
	ASO	98.25%	94.25%	98.41%	93.98%	64.69%	0.00
	GSO	98.25%	94.25%	98.12%	93.91%	52.87%	0.00
	RAND	98.25%	94.25%	97.80%	93.85%	50.10%	-0.01
0.1	GEFeS	98.25%	94.25%	98.92%	94.02%	44.95%	0.00
	PSO	98.25%	94.25%	98.09%	93.81%	51.28%	-0.01
	ABCO	98.25%	94.25%	98.16%	93.81%	51.93%	-0.01
	ASO	98.25%	94.25%	98.45%	93.97%	58.10%	0.00
	GSO	98.25%	94.25%	98.11%	93.84%	52.34%	-0.01
	RAND	98.25%	94.25%	97.95%	93.83%	49.44%	-0.01
0.3	GEFeS	98.25%	94.25%	98.92%	93.88%	34.36%	0.00
	PSO	98.25%	94.25%	98.11%	93.85%	49.91%	-0.01
	ABCO	98.25%	94.25%	98.01%	93.80%	50.37%	-0.01
	ASO	98.25%	94.25%	98.48%	93.94%	45.95%	0.00
	GSO	98.25%	94.25%	97.96%	93.88%	50.29%	-0.01
	RAND	98.25%	94.25%	97.86%	93.86%	49.72%	-0.01
0.5	GEFeS	98.25%	94.25%	98.94%	93.81%	28.47%	0.00
	PSO	98.25%	94.25%	98.08%	93.84%	48.43%	-0.01
	ABCO	98.25%	94.25%	98.11%	93.80%	48.46%	-0.01
	ASO	98.25%	94.25%	98.46%	93.88%	35.09%	0.00
	GSO	98.25%	94.25%	98.02%	93.89%	49.19%	-0.01
	RAND	98.25%	94.25%	97.88%	93.85%	49.79%	-0.01
0.7	GEFeS	98.25%	94.25%	98.83%	93.73%	25.06%	0.00
	PSO	98.25%	94.25%	97.92%	93.83%	47.43%	-0.01
	ABCO	98.25%	94.25%	98.00%	98.83%	47.49%	-0.01
	ASO	98.25%	94.25%	98.53%	93.83%	26.95%	0.00
	GSO	98.25%	94.25%	98.00%	93.88%	47.78%	-0.01
	RAND	98.25%	94.25%	97.84%	93.82%	50.34%	-0.01
0.9	GEFeS	98.25%	94.25%	98.93%	93.77%	22.28%	0.00
	PSO	98.25%	94.25%	97.98%	93.81%	45.06%	-0.01
	ABCO	98.25%	94.25%	97.86%	93.83%	45.20%	-0.01
	ASO	98.25%	94.25%	98.38%	93.72%	20.90%	0.00
	GSO	98.25%	94.25%	97.88%	93.78%	45.76%	-0.01
	RAND	98.25%	94.25%	97.77%	93.83%	50.14%	-0.01
1.0	GEFeS	98.25%	94.25%	98.73%	93.73%	21.30%	0.00
	PSO	98.25%	94.25%	97.89%	93.76%	45.58%	-0.01
	ABCO	98.25%	94.25%	97.99%	93.82%	45.55%	-0.01
	ASO	98.25%	94.25%	98.25%	93.72%	18.50%	-0.01
	GSO	98.25%	94.25%	97.86%	93.77%	46.02%	-0.01
	RAND	98.25%	94.25%	97.83%	93.85%	50.50%	-0.01

Appendix E

Genetic and Swarm Based Feature Selection with Respect to Adversarial Author Profiling Detailed Results

This following is a detailed explanation of the results from the experiments described in Chapter 8.

We generated the results using a cross-product of feature selection algorithms (i.e., GEFeS, PSO, ABCO, ASO, GSO, and random sampling) and feature sets (i.e., LIWC, Topic Modeling, Stylometry and Hybrid) with and without feature selection as well as with and without adversarial texts. The random sampling feature selection algorithm (RAND) serves as a baseline for feature selection. We also employed a Linear Support Vector Machine (LSVM) as the AIdS used in fitness evaluation.

We used five-fold crossover using four of the samples for training and the fifth for validation. We calculate accuracy using two methods: one method trains on four of the five folds and calculates accuracy using only the fifth sample with and without adversarial samples; the second method also trains on four of the folds, but calculates accuracy across all five folds, again with and without adversarial samples. We designate the former results as *100+(org = 25, adv = 25)*, and the latter results as *100+(org = 125, adv = 125)*, where *org* indicates the original (non-adversarial) samples and *adv* indicates the adversarial samples.

Tables E.1-E.4 show results from the experiments using *100+(org = 25, adv = 25)*, and tables E.5-E.8 show results from experiments using *100+(org = 125, adv = 125)*. In each of the table titles we indicate the feature set name (LIWC, Topic Modeling, Stylometry or Hybrid), the number of features and the accuracy configuration. Since we have two accuracy configurations and four feature sets, we represent the results in 8 (2 X 4) tables.

Each table lists the accuracies based on original and adversarial samples with and without feature selection, as well as two other metrics (mean percent features used and the *Use?* indicator value based on a risk weighting factor (ρ) of 0.5 to balance the risk). Each table follows the same format. The first column of the table, labeled ω, is the value of the weight of the penalty applied within the fitness function (see Equation 4.1). The second column, labeled *FS Alg*, indicates the feature selection algorithm used for the results. The next group of columns, labeled *Baseline*, represent those results achieved without feature selection. There are two columns within the *Baseline* group. The first, labeled *Orig*, is the accuracies achieved using no feature selection and no adversarial texts. The second, labeled *Adv* are the accuracies achieved with no feature selection used, but adversarial texts are introduced. The second group of columns, labeled *With Feature Selection* shows the results of using feature selection. This group has three columns. The first two columns are the same as the definition of the two *Baseline* columns (except with feature selection). The third column, labeled *% Features Used* indicates the mean percent of features used after feature selection was applied. Because of the stochastic nature of the feature selection algorithms, it is necessary to run the same configuration 30 times to get statistically useful results. Therefore, the feature selection columns represent the mean across the 30 runs. The final column, labeled *Use?* is the indicator calculated according to Table 8.1, as described previously.

The rows of each table are grouped by the seven ω values (0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0). Within each ω group there is a row for each feature selection algorithm (GEFeS, PSO, ABCO, ASO, GSO, RAND). Note that the *Baseline* values are the same for each feature set across all the rows because in the *Baseline* case there is no feature selection. So, neither ω, nor the feature selection algorithm affect the accuracy.

Given the extracted features, we followed the same procedure outlined in [19] to process the features, namely applying TF-IDF, standardization and normalization. Also, [19] found better results by applying this processing *before* combining the features into the hybrid feature set, so we followed this process. Note that the LIWC feature set has 93 features, Topic Modeling has 45 and Stylometry has 428 features, which yields a Hybrid feature set consisting of 566 features.

As previously mentioned, we ran each feature selection configuration 30 times and averaged the results. For each of these runs, we allowed a total of 15,000 fitness evaluations. Since the *Baseline* configurations are not stochastic in nature, these configurations only required one run. Also, while we trained the feature selection algorithms using the fitness value as explained in Equation 4.1, the results in the tables reflect unpenalized accuracy.

Table E.1 gives the measurements generated using the LIWC feature set and the PAN19-25 dataset training on four of the samples and testing on the fifth (i.e., *100+(org = 25, adv = 25)*). The baseline accuracies, without feature selection, are 40.00% with original samples and 32.00% when introducing adversarial samples. With feature selection as shown in the second group of columns has non-adversarial accuracies ranging from 50.27% (ω = 0.1 using RAND) to 70.00% (ω = 1.0 using GEFeS). Adversarial samples cause the accuracies to drop to a range of 4.00% for several values of ω with ASO to 7.60% (ω = 0.9 using RAND). The *Use?* values are all negative, meaning that feature selection is not favored due to the susceptibility it affords. These *Use?* values range from -0.59 (ω = 0.3 using RAND) to -0.18 (ω = 1.0 using GEFeS).

Table E.2 results are based on using the Stylometry feature set and the PAN19-25 dataset training on four of the samples and testing on the fifth (i.e., *100+(org = 25, adv = 25)*). The baseline accuracies are 44.00% without adversarial samples, and 32.00% with adversarial samples. With feature selection, the original samples yield accuracies ranging from 48.27% (ω = 0.9 using RAND) to 78.00% (ω = 1.0 using GEFeS). Introducing adversarial samples reduces the accuracies to a range of 4.13% (ω = 0.5 & 0.9 using ASO) to 7.47% (ω = 0.0 using ASO and ω = 0.1 using ABCO). Again, the *Use?* values are all negative indicating that feature selection with Stylometry exposes a vulnerability to adversarial attacks. The *Use?* values range from -0.76 (ω = 0.9 using RAND) to -0.08 (ω = 0.3 using GEFeS).

Table E.3 results reflect the use of the Topic Modeling feature set with the *100+(org = 25, adv = 25)*) evaluation configuration. The baseline accuracies are 32.00% without adversarial samples, and 0.00% with adversarial samples. The accuracies are also slightly lower when using feature selection ranging from 39.60% (ω = 0.0 using ASO) to 54.53% (ω = 0.5 using GEFeS) without adversarial samples, and 0.80% (ω = 0.9 using ASO) to 4.40% (ω = 0.0 using RAND and ω = 0.1 using GEFeS) with adversarial samples. In this case, the *Use?* values are all positive indicating that feature selection is useful for Topic Modeling. This is mainly due to the fact that the baseline adversarial accuracy is 0.00%, and with feature selection the accuracy is greater than 0.00%, which means that any improvement due to feature selection can only be helpful. The *Use?* values range from 0.08 (ω = 0.0 using ASO) to 0.23 (ω = 0.5 using GEFeS).

Table E.4 indicates results using the Hybrid feature set. The baseline accuracies are 40.00% for original samples and 16.00% for adversarial samples. Using Feature selection, the accuracies range from 0.16% (ω = 0.1 using RAND) to 84.31% (ω = 0.7 using GEFeS) for original samples and drop to a range of 3.07% (ω = 1.0 using GEFeS) to 6.27% (ω = 0.1 using ABCO) with adversarial samples. In this case, the *Use?* values are mixed. All the GEFeS values are positive, and all but one other *Use?* value (ω = 0.5 using ASO) is negative. These values range from -0.68 (ω = 1.0 using ASO) to 0.37 (ω = 0.5 using GEFeS).

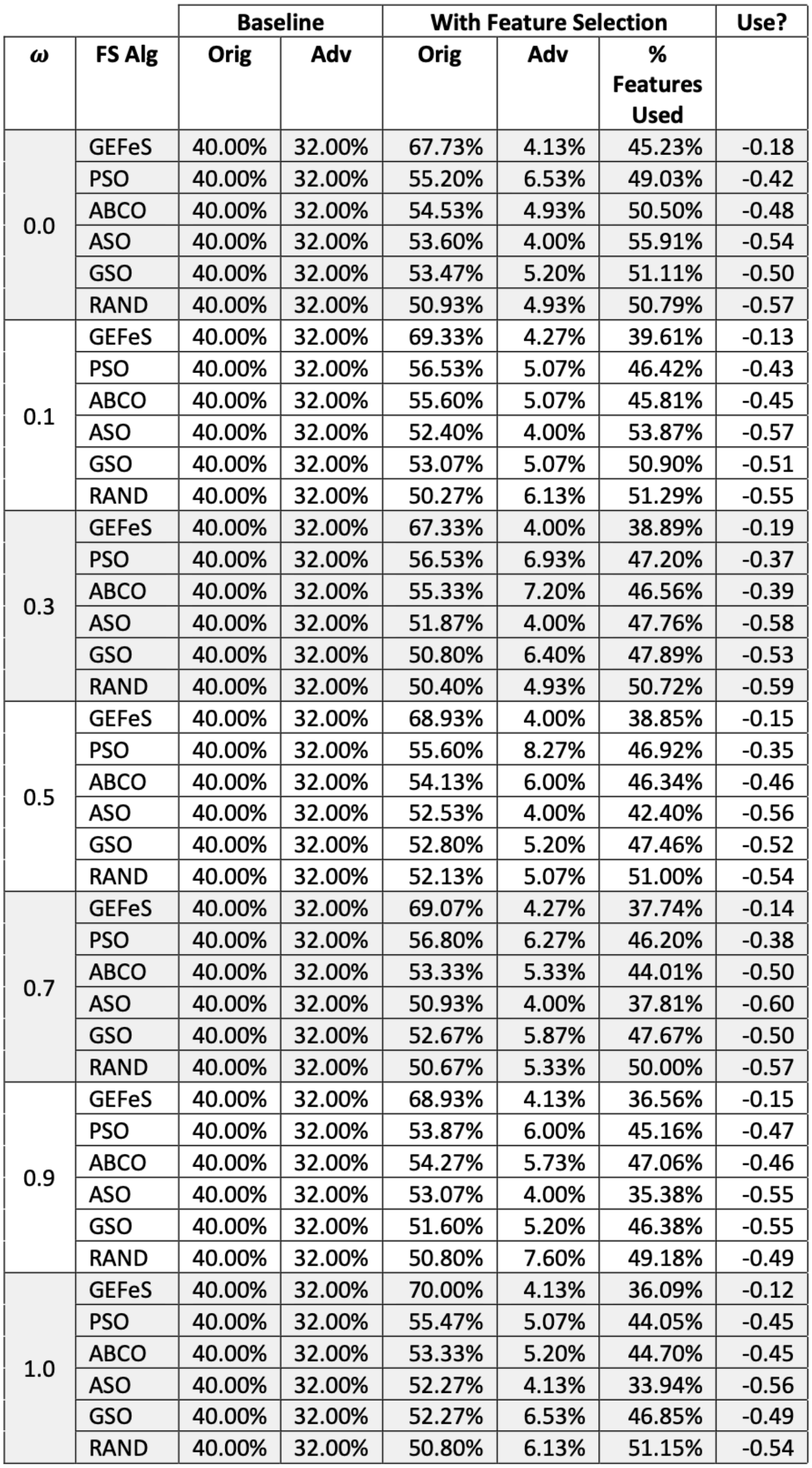
Recall that tables E.5-E.8 reflect the *100+(org = 125, adv = 125)* evaluation configuration. Table E.5 revisits the LIWC feature set using this evaluation configuration. The baseline values for Table E.5 are 87.20% and 85.60% corresponding to the original and adversarial samples. With feature selection, original sample accuracies range from 87.15% (ω = 0.7 and 0.9 using RAND) to 91.55% (ω = 0.1 using GEFeS). Adversarial samples drop the accuracies to 77.52% (ω = 1.0 using GEFeS) and 79.25% (ω = 0.3 using ASO). For the LIWC dataset in Table 8.5, the *Use*? values are slightly negative ranging from several instances of -0.09 to -0.03 (ω = 0.0 & 0.1 using GEFeS).

Table E.6, like Table E.2, uses the Stylometry feature set, but with the *100+(org = 125, adv = 125)* configuration. The baseline accuracy for original samples is 88.80%, and introducing adversarial samples lowers the accuracy to 86.40%. Feature selection accuracies for original samples range from 89.36% (ω = 0.9 using RAND) to 94.77% (ω = 1.0 using GEFeS), and for adversarial samples the accuracies drop to 79.28% (ω = 0.9 using ASO) to 81.17% (ω = 0.5 using GSO). The *Use?* values are all slightly non-positive with lows of -0.06 (ω = 0.7 & 1.0 using RAND and ω = 0.7 using ASO) to several instances of 0.0 using GEFeS (ω = 0.1, 0.3 & 0.5).

Table E.7 shows results for the Topic Modeling feature set, again using the *100+(org = 125, adv = 125)* configuration. The baseline accuracies are 79.20% and 72.80% for original and adversarial samples respectively. With feature selection using original samples, the accuracies range from 66.56% (ω = 1.0 using ASO) to 77.99% (ω = 0.0 using RAND), and with adversarial samples from 58.08% (ω = 1.0 using ASO) to 65.12% (ω = 0.0 using GEFeS). The *Use?* values are decidedly negative ranging from -0.36 (ω = 1.0 using ASO) to -0.16 (ω = 0.0 using GEFeS).

Finally, Table E.8 shows results for the Hybrid feature set using *100+(org = 125, adv = 125)*. The baseline accuracy for original samples is 88.00%, and for adversarial samples the accuracy is 83.20%. The accuracies for original samples using feature selection range from 90.69% (ω = 0.7 using RAND) to 96.67 (ω = 0.7 using GEFeS), and adversarial samples lowers the accuracies to 80.43% (ω = 0.7 using GEFeS) and 81.17% (ω = 0.1 using ABCO and ω = 0.5 using PSO). The *Use?* values are mostly slightly positive, with the exception of one negative value of -0.02 (ω = 1.0 using ASO) to 0.07 (ω = 0.1, 0.5, 0.7 & 0.9 using GEFeS).

Table E.1. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, LIWC Feature Set - 93 Features, 100+(org = 25, adv = 25).



		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	40.00%	32.00%	67.73%	4.13%	45.23%	-0.18
	PSO	40.00%	32.00%	55.20%	6.53%	49.03%	-0.42
	ABCO	40.00%	32.00%	54.53%	4.93%	50.50%	-0.48
	ASO	40.00%	32.00%	53.60%	4.00%	55.91%	-0.54
	GSO	40.00%	32.00%	53.47%	5.20%	51.11%	-0.50
	RAND	40.00%	32.00%	50.93%	4.93%	50.79%	-0.57
0.1	GEFeS	40.00%	32.00%	69.33%	4.27%	39.61%	-0.13
	PSO	40.00%	32.00%	56.53%	5.07%	46.42%	-0.43
	ABCO	40.00%	32.00%	55.60%	5.07%	45.81%	-0.45
	ASO	40.00%	32.00%	52.40%	4.00%	53.87%	-0.57
	GSO	40.00%	32.00%	53.07%	5.07%	50.90%	-0.51
	RAND	40.00%	32.00%	50.27%	6.13%	51.29%	-0.55
0.3	GEFeS	40.00%	32.00%	67.33%	4.00%	38.89%	-0.19
	PSO	40.00%	32.00%	56.53%	6.93%	47.20%	-0.37
	ABCO	40.00%	32.00%	55.33%	7.20%	46.56%	-0.39
	ASO	40.00%	32.00%	51.87%	4.00%	47.76%	-0.58
	GSO	40.00%	32.00%	50.80%	6.40%	47.89%	-0.53
	RAND	40.00%	32.00%	50.40%	4.93%	50.72%	-0.59
0.5	GEFeS	40.00%	32.00%	68.93%	4.00%	38.85%	-0.15
	PSO	40.00%	32.00%	55.60%	8.27%	46.92%	-0.35
	ABCO	40.00%	32.00%	54.13%	6.00%	46.34%	-0.46
	ASO	40.00%	32.00%	52.53%	4.00%	42.40%	-0.56
	GSO	40.00%	32.00%	52.80%	5.20%	47.46%	-0.52
	RAND	40.00%	32.00%	52.13%	5.07%	51.00%	-0.54
0.7	GEFeS	40.00%	32.00%	69.07%	4.27%	37.74%	-0.14
	PSO	40.00%	32.00%	56.80%	6.27%	46.20%	-0.38
	ABCO	40.00%	32.00%	53.33%	5.33%	44.01%	-0.50
	ASO	40.00%	32.00%	50.93%	4.00%	37.81%	-0.60
	GSO	40.00%	32.00%	52.67%	5.87%	47.67%	-0.50
	RAND	40.00%	32.00%	50.67%	5.33%	50.00%	-0.57
0.9	GEFeS	40.00%	32.00%	68.93%	4.13%	36.56%	-0.15
	PSO	40.00%	32.00%	53.87%	6.00%	45.16%	-0.47
	ABCO	40.00%	32.00%	54.27%	5.73%	47.06%	-0.46
	ASO	40.00%	32.00%	53.07%	4.00%	35.38%	-0.55
	GSO	40.00%	32.00%	51.60%	5.20%	46.38%	-0.55
	RAND	40.00%	32.00%	50.80%	7.60%	49.18%	-0.49
1.0	GEFeS	40.00%	32.00%	70.00%	4.13%	36.09%	-0.12
	PSO	40.00%	32.00%	55.47%	5.07%	44.05%	-0.45
	ABCO	40.00%	32.00%	53.33%	5.20%	44.70%	-0.45
	ASO	40.00%	32.00%	52.27%	4.13%	33.94%	-0.56
	GSO	40.00%	32.00%	52.27%	6.53%	46.85%	-0.49
	RAND	40.00%	32.00%	50.80%	6.13%	51.15%	-0.54

Table E.2. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Stylometry Feature Set - 428 Features, 100+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	44.00%	32.00%	74.67%	5.20%	47.82%	-0.14
	PSO	44.00%	32.00%	54.80%	6.40%	51.00%	-0.55
	ABCO	44.00%	32.00%	52.27%	5.73%	49.73%	-0.63
	ASO	44.00%	32.00%	59.73%	7.47%	54.74%	-0.41
	GSO	44.00%	32.00%	53.20%	6.93%	49.34%	-0.57
	RAND	44.00%	32.00%	50.13%	6.27%	49.63%	-0.66
0.1	GEFeS	44.00%	32.00%	76.93%	4.80%	36.15%	-0.10
	PSO	44.00%	32.00%	55.20%	5.47%	47.69%	-0.57
	ABCO	44.00%	32.00%	53.73%	7.47%	48.22%	-0.55
	ASO	44.00%	32.00%	61.73%	4.67%	48.14%	-0.45
	GSO	44.00%	32.00%	52.93%	5.07%	49.68%	-0.64
	RAND	44.00%	32.00%	50.00%	5.60%	49.19%	-0.69
0.3	GEFeS	44.00%	32.00%	77.33%	5.20%	33.70%	-0.08
	PSO	44.00%	32.00%	55.07%	5.20%	46.69%	-0.59
	ABCO	44.00%	32.00%	53.87%	6.27%	47.94%	-0.58
	ASO	44.00%	32.00%	62.27%	5.07%	37.21%	-0.43
	GSO	44.00%	32.00%	52.40%	5.87%	48.37%	-0.63
	RAND	44.00%	32.00%	52.80%	5.60%	49.23%	-0.63
0.5	GEFeS	44.00%	32.00%	76.93%	4.93%	29.20%	-0.10
	PSO	44.00%	32.00%	54.93%	5.20%	46.05%	-0.59
	ABCO	44.00%	32.00%	55.07%	4.80%	46.87%	-0.60
	ASO	44.00%	32.00%	61.20%	4.13%	29.11%	-0.48
	GSO	44.00%	32.00%	52.40%	6.93%	46.29%	-0.59
	RAND	44.00%	32.00%	49.33%	4.80%	50.32%	-0.73
0.7	GEFeS	44.00%	32.00%	76.93%	4.40%	24.98%	-0.11
	PSO	44.00%	32.00%	55.07%	6.67%	44.44%	-0.54
	ABCO	44.00%	32.00%	51.07%	6.13%	45.51%	-0.65
	ASO	44.00%	32.00%	61.33%	4.27%	23.30%	-0.47
	GSO	44.00%	32.00%	52.27%	4.93%	45.40%	-0.66
	RAND	44.00%	32.00%	49.87%	5.60%	50.29%	-0.69
0.9	GEFeS	44.00%	32.00%	75.07%	4.53%	22.35%	-0.15
	PSO	44.00%	32.00%	54.93%	6.00%	44.42%	-0.56
	ABCO	44.00%	32.00%	52.53%	6.13%	45.05%	-0.61
	ASO	44.00%	32.00%	62.27%	4.13%	19.50%	-0.46
	GSO	44.00%	32.00%	53.33%	4.93%	44.73%	-0.63
	RAND	44.00%	32.00%	48.27%	4.67%	49.94%	-0.76
1.0	GEFeS	44.00%	32.00%	78.00%	4.40%	20.42%	-0.09
	PSO	44.00%	32.00%	55.20%	5.33%	43.22%	-0.58
	ABCO	44.00%	32.00%	51.73%	4.93%	43.63%	-0.67
	ASO	44.00%	32.00%	63.47%	4.27%	22.66%	-0.42
	GSO	44.00%	32.00%	51.60%	4.93%	43.17%	-0.67
	RAND	44.00%	32.00%	49.33%	6.13%	49.59%	-0.69

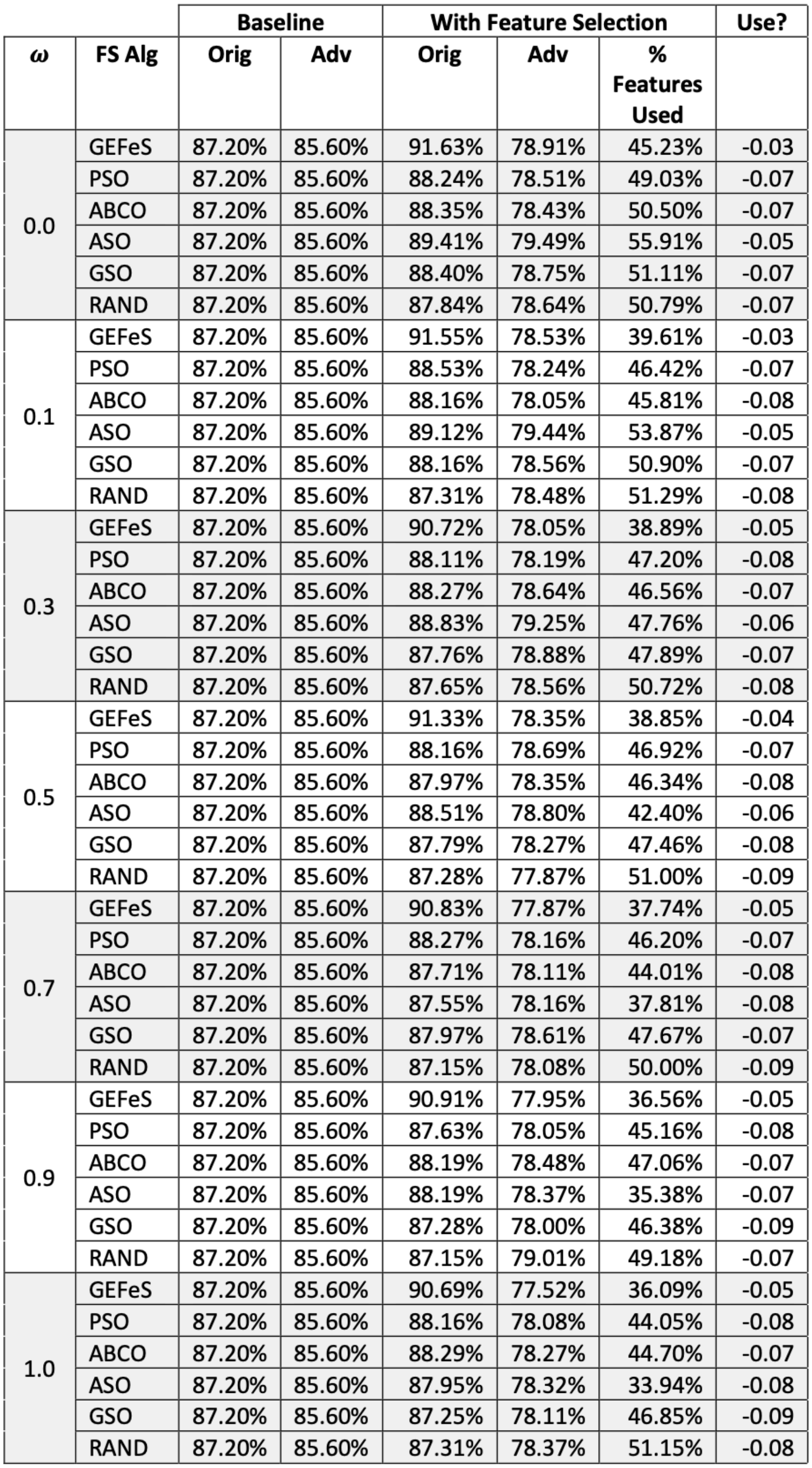
Table E.3. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Topic Modeling Feature Set - 45 Features, 100+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	32.00%	0.00%	52.80%	4.27%	53.85%	0.21
	PSO	32.00%	0.00%	45.20%	3.73%	53.85%	0.13
	ABCO	32.00%	0.00%	47.60%	3.47%	54.44%	0.16
	ASO	32.00%	0.00%	39.60%	3.87%	60.67%	0.08
	GSO	32.00%	0.00%	44.00%	3.73%	55.33%	0.12
	RAND	32.00%	0.00%	44.13%	4.40%	51.33%	0.12
0.1	GEFeS	32.00%	0.00%	53.47%	4.40%	49.19%	0.21
	PSO	32.00%	0.00%	45.47%	3.73%	52.44%	0.13
	ABCO	32.00%	0.00%	46.93%	3.33%	51.78%	0.15
	ASO	32.00%	0.00%	40.13%	4.40%	58.89%	0.08
	GSO	32.00%	0.00%	45.20%	3.33%	54.96%	0.13
	RAND	32.00%	0.00%	45.33%	3.47%	48.15%	0.13
0.3	GEFeS	32.00%	0.00%	52.67%	3.60%	47.78%	0.21
	PSO	32.00%	0.00%	46.00%	3.73%	50.07%	0.14
	ABCO	32.00%	0.00%	47.47%	3.87%	52.22%	0.15
	ASO	32.00%	0.00%	40.00%	3.60%	58.00%	0.08
	GSO	32.00%	0.00%	42.93%	3.07%	52.81%	0.11
	RAND	32.00%	0.00%	44.53%	3.33%	50.00%	0.13
0.5	GEFeS	32.00%	0.00%	54.53%	3.87%	48.07%	0.23
	PSO	32.00%	0.00%	45.87%	4.13%	53.04%	0.14
	ABCO	32.00%	0.00%	46.27%	3.33%	52.30%	0.14
	ASO	32.00%	0.00%	41.07%	2.80%	56.00%	0.09
	GSO	32.00%	0.00%	43.73%	4.00%	52.67%	0.12
	RAND	32.00%	0.00%	44.00%	3.07%	50.89%	0.12
0.7	GEFeS	32.00%	0.00%	53.07%	4.00%	48.59%	0.21
	PSO	32.00%	0.00%	46.93%	3.20%	50.81%	0.15
	ABCO	32.00%	0.00%	48.53%	3.07%	51.56%	0.17
	ASO	32.00%	0.00%	40.93%	2.13%	55.19%	0.09
	GSO	32.00%	0.00%	45.73%	3.07%	53.33%	0.14
	RAND	32.00%	0.00%	42.53%	3.07%	49.70%	0.11
0.9	GEFeS	32.00%	0.00%	52.80%	3.47%	48.07%	0.21
	PSO	32.00%	0.00%	46.67%	3.47%	52.67%	0.15
	ABCO	32.00%	0.00%	47.47%	3.47%	52.89%	0.15
	ASO	32.00%	0.00%	44.13%	0.80%	51.70%	0.12
	GSO	32.00%	0.00%	44.80%	4.27%	52.67%	0.13
	RAND	32.00%	0.00%	43.33%	3.07%	50.81%	0.11
1.0	GEFeS	32.00%	0.00%	53.33%	3.07%	47.26%	0.21
	PSO	32.00%	0.00%	48.80%	4.27%	50.89%	0.17
	ABCO	32.00%	0.00%	46.53%	2.67%	49.26%	0.15
	ASO	32.00%	0.00%	43.60%	1.20%	52.07%	0.12
	GSO	32.00%	0.00%	43.47%	3.73%	50.74%	0.11
	RAND	32.00%	0.00%	43.33%	3.73%	51.19%	0.11

Table E.4. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Hybrid Feature Set – 566 Features, 100+(org = 25, adv = 25).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	40.00%	16.00%	80.00%	4.93%	47.37%	0.31
	PSO	40.00%	16.00%	59.73%	4.00%	49.39%	-0.26
	ABCO	40.00%	16.00%	58.13%	4.80%	49.98%	-0.25
	ASO	40.00%	16.00%	70.80%	4.27%	49.65%	0.04
	GSO	40.00%	16.00%	57.07%	4.40%	49.98%	-0.30
	RAND	40.00%	16.00%	55.87%	3.87%	49.82%	-0.36
0.1	GEFeS	40.00%	16.00%	82.27%	3.47%	37.26%	0.27
	PSO	40.00%	16.00%	60.80%	4.13%	47.62%	-0.22
	ABCO	40.00%	16.00%	59.47%	6.27%	49.10%	-0.12
	ASO	40.00%	16.00%	70.40%	3.20%	41.52%	-0.04
	GSO	40.00%	16.00%	55.73%	4.67%	49.05%	-0.31
	RAND	40.00%	16.00%	16.00%	4.53%	49.73%	-0.35
0.3	GEFeS	40.00%	16.00%	82.00%	3.73%	33.52%	0.28
	PSO	40.00%	16.00%	59.60%	4.00%	47.43%	-0.26
	ABCO	40.00%	16.00%	55.73%	3.73%	47.95%	-0.37
	ASO	40.00%	16.00%	70.13%	3.60%	27.71%	-0.02
	GSO	40.00%	16.00%	56.93%	4.80%	47.82%	-0.28
	RAND	40.00%	16.00%	54.53%	4.93%	49.93%	-0.33
0.5	GEFeS	40.00%	16.00%	82.67%	4.80%	28.83%	0.37
	PSO	40.00%	16.00%	60.80%	5.87%	45.37%	-0.11
	ABCO	40.00%	16.00%	57.60%	5.60%	47.27%	-0.21
	ASO	40.00%	16.00%	68.13%	5.33%	19.19%	0.04
	GSO	40.00%	16.00%	56.40%	5.20%	47.61%	-0.30
	RAND	40.00%	16.00%	55.47%	3.87%	50.00%	-0.28
0.7	GEFeS	40.00%	16.00%	84.31%	2.93%	24.48%	0.29
	PSO	40.00%	16.00%	61.87%	5.07%	45.26%	-0.14
	ABCO	40.00%	16.00%	58.27%	4.27%	45.65%	-0.28
	ASO	40.00%	16.00%	65.47%	4.53%	14.53%	-0.08
	GSO	40.00%	16.00%	56.27%	5.20%	46.20%	-0.27
	RAND	40.00%	16.00%	53.60%	3.87%	50.37%	-0.42
0.9	GEFeS	40.00%	16.00%	84.00%	3.87%	22.40%	0.34
	PSO	40.00%	16.00%	61.20%	4.67%	43.86%	-0.18
	ABCO	40.00%	16.00%	58.13%	4.00%	44.81%	-0.30
	ASO	40.00%	16.00%	55.20%	4.27%	40.67%	-0.35
	GSO	40.00%	16.00%	56.13%	4.00%	44.73%	-0.35
	RAND	40.00%	16.00%	54.13%	4.27%	49.83%	-0.38
1.0	GEFeS	40.00%	16.00%	81.60%	3.07%	20.44%	0.23
	PSO	40.00%	16.00%	58.27%	4.00%	43.12%	-0.29
	ABCO	40.00%	16.00%	57.33%	4.13%	44.04%	-0.31
	ASO	40.00%	16.00%	41.87%	4.40%	49.80%	-0.68
	GSO	40.00%	16.00%	54.40%	5.33%	43.82%	-0.31
	RAND	40.00%	16.00%	54.53%	4.40%	49.76%	-0.36

Table E.5. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, LIWC Feature Set - 93 Features, 100+(org = 125, adv = 125).



		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	87.20%	85.60%	91.63%	78.91%	45.23%	-0.03
	PSO	87.20%	85.60%	88.24%	78.51%	49.03%	-0.07
	ABCO	87.20%	85.60%	88.35%	78.43%	50.50%	-0.07
	ASO	87.20%	85.60%	89.41%	79.49%	55.91%	-0.05
	GSO	87.20%	85.60%	88.40%	78.75%	51.11%	-0.07
	RAND	87.20%	85.60%	87.84%	78.64%	50.79%	-0.07
0.1	GEFeS	87.20%	85.60%	91.55%	78.53%	39.61%	-0.03
	PSO	87.20%	85.60%	88.53%	78.24%	46.42%	-0.07
	ABCO	87.20%	85.60%	88.16%	78.05%	45.81%	-0.08
	ASO	87.20%	85.60%	89.12%	79.44%	53.87%	-0.05
	GSO	87.20%	85.60%	88.16%	78.56%	50.90%	-0.07
	RAND	87.20%	85.60%	87.31%	78.48%	51.29%	-0.08
0.3	GEFeS	87.20%	85.60%	90.72%	78.05%	38.89%	-0.05
	PSO	87.20%	85.60%	88.11%	78.19%	47.20%	-0.08
	ABCO	87.20%	85.60%	88.27%	78.64%	46.56%	-0.07
	ASO	87.20%	85.60%	88.83%	79.25%	47.76%	-0.06
	GSO	87.20%	85.60%	87.76%	78.88%	47.89%	-0.07
	RAND	87.20%	85.60%	87.65%	78.56%	50.72%	-0.08
0.5	GEFeS	87.20%	85.60%	91.33%	78.35%	38.85%	-0.04
	PSO	87.20%	85.60%	88.16%	78.69%	46.92%	-0.07
	ABCO	87.20%	85.60%	87.97%	78.35%	46.34%	-0.08
	ASO	87.20%	85.60%	88.51%	78.80%	42.40%	-0.06
	GSO	87.20%	85.60%	87.79%	78.27%	47.46%	-0.08
	RAND	87.20%	85.60%	87.28%	77.87%	51.00%	-0.09
0.7	GEFeS	87.20%	85.60%	90.83%	77.87%	37.74%	-0.05
	PSO	87.20%	85.60%	88.27%	78.16%	46.20%	-0.07
	ABCO	87.20%	85.60%	87.71%	78.11%	44.01%	-0.08
	ASO	87.20%	85.60%	87.55%	78.16%	37.81%	-0.08
	GSO	87.20%	85.60%	87.97%	78.61%	47.67%	-0.07
	RAND	87.20%	85.60%	87.15%	78.08%	50.00%	-0.09
0.9	GEFeS	87.20%	85.60%	90.91%	77.95%	36.56%	-0.05
	PSO	87.20%	85.60%	87.63%	78.05%	45.16%	-0.08
	ABCO	87.20%	85.60%	88.19%	78.48%	47.06%	-0.07
	ASO	87.20%	85.60%	88.19%	78.37%	35.38%	-0.07
	GSO	87.20%	85.60%	87.28%	78.00%	46.38%	-0.09
	RAND	87.20%	85.60%	87.15%	79.01%	49.18%	-0.07
1.0	GEFeS	87.20%	85.60%	90.69%	77.52%	36.09%	-0.05
	PSO	87.20%	85.60%	88.16%	78.08%	44.05%	-0.08
	ABCO	87.20%	85.60%	88.29%	78.27%	44.70%	-0.07
	ASO	87.20%	85.60%	87.95%	78.32%	33.94%	-0.08
	GSO	87.20%	85.60%	87.25%	78.11%	46.85%	-0.09
	RAND	87.20%	85.60%	87.31%	78.37%	51.15%	-0.08

Table E.6. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Stylometry Feature Set - 428 Features, 100+(org = 125, adv = 125).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	88.80%	86.40%	94.40%	80.51%	47.82%	-0.01
	PSO	88.80%	86.40%	90.64%	80.96%	51.00%	-0.04
	ABCO	88.80%	86.40%	90.05%	80.75%	49.73%	-0.05
	ASO	88.80%	86.40%	91.15%	80.69%	54.74%	-0.04
	GSO	88.80%	86.40%	90.19%	80.93%	49.34%	-0.05
	RAND	88.80%	86.40%	89.76%	81.01%	49.63%	-0.05
0.1	GEFeS	88.80%	86.40%	94.72%	80.29%	36.15%	0.00
	PSO	88.80%	86.40%	90.75%	80.80%	47.69%	-0.04
	ABCO	88.80%	86.40%	90.35%	81.09%	48.22%	-0.04
	ASO	88.80%	86.40%	91.49%	80.08%	48.14%	-0.04
	GSO	88.80%	86.40%	90.08%	80.51%	49.68%	-0.05
	RAND	88.80%	86.40%	89.71%	80.83%	49.19%	-0.05
0.3	GEFeS	88.80%	86.40%	94.69%	80.27%	33.70%	0.00
	PSO	88.80%	86.40%	90.77%	80.80%	46.69%	-0.04
	ABCO	88.80%	86.40%	90.43%	80.91%	47.94%	-0.05
	ASO	88.80%	86.40%	91.52%	80.08%	37.21%	-0.04
	GSO	88.80%	86.40%	90.19%	80.88%	48.37%	-0.05
	RAND	88.80%	86.40%	90.35%	80.91%	49.23%	-0.05
0.5	GEFeS	88.80%	86.40%	94.67%	80.27%	29.20%	0.00
	PSO	88.80%	86.40%	90.64%	80.69%	46.05%	-0.05
	ABCO	88.80%	86.40%	90.64%	80.59%	46.87%	-0.05
	ASO	88.80%	86.40%	91.28%	79.87%	29.11%	-0.05
	GSO	88.80%	86.40%	90.27%	81.17%	46.29%	-0.04
	RAND	88.80%	86.40%	89.60%	80.69%	50.32%	-0.06
0.7	GEFeS	88.80%	86.40%	94.45%	79.95%	24.98%	-0.01
	PSO	88.80%	86.40%	90.56%	80.88%	44.44%	-0.04
	ABCO	88.80%	86.40%	90.00%	81.01%	45.51%	-0.05
	ASO	88.80%	86.40%	91.28%	79.87%	23.30%	-0.05
	GSO	88.80%	86.40%	90.08%	80.61%	45.40%	-0.05
	RAND	88.80%	86.40%	89.73%	80.88%	50.29%	-0.05
0.9	GEFeS	88.80%	86.40%	94.16%	80.05%	22.35%	-0.01
	PSO	88.80%	86.40%	90.53%	80.75%	44.42%	-0.05
	ABCO	88.80%	86.40%	90.19%	80.91%	45.05%	-0.05
	ASO	88.80%	86.40%	90.91%	79.28%	19.50%	-0.06
	GSO	88.80%	86.40%	90.24%	80.56%	44.73%	-0.05
	RAND	88.80%	86.40%	89.36%	80.64%	49.94%	-0.06
1.0	GEFeS	88.80%	86.40%	94.77%	80.05%	20.42%	-0.01
	PSO	88.80%	86.40%	90.67%	80.69%	43.22%	-0.05
	ABCO	88.80%	86.40%	89.97%	80.61%	43.63%	-0.05
	ASO	88.80%	86.40%	91.41%	79.57%	22.66%	-0.05
	GSO	88.80%	86.40%	90.00%	80.67%	43.17%	-0.05
	RAND	88.80%	86.40%	89.49%	80.85%	49.59%	-0.06

Table E.7. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Topic Modeling Feature Set - 45 Features, 100+(org = 125, adv = 125).

		Baseline		With Feature Selection			Use?
ω	**FS Alg**	**Orig**	**Adv**	**Orig**	**Adv**	**% Features Used**	
0.0	GEFeS	79.20%	72.80%	74.83%	65.12%	53.85%	-0.16
	PSO	79.20%	72.80%	70.93%	62.64%	53.85%	-0.24
	ABCO	79.20%	72.80%	73.73%	64.91%	54.44%	-0.18
	ASO	79.20%	72.80%	70.13%	62.99%	60.67%	-0.25
	GSO	79.20%	72.80%	72.32%	64.27%	55.33%	-0.20
	RAND	79.20%	72.80%	77.99%	65.04%	51.33%	-0.19
0.1	GEFeS	79.20%	72.80%	74.21%	64.04%	49.19%	-0.18
	PSO	79.20%	72.80%	71.76%	63.41%	52.44%	-0.22
	ABCO	79.20%	72.80%	72.77%	64.05%	51.78%	-0.20
	ASO	79.20%	72.80%	69.63%	62.48%	58.89%	-0.26
	GSO	79.20%	72.80%	72.35%	63.97%	54.96%	-0.21
	RAND	79.20%	72.80%	72.80%	64.43%	48.15%	-0.20
0.3	GEFeS	79.20%	72.80%	72.24%	62.43%	47.78%	-0.23
	PSO	79.20%	72.80%	71.07%	62.61%	50.07%	-0.24
	ABCO	79.20%	72.80%	73.20%	64.48%	52.22%	-0.19
	ASO	79.20%	72.80%	68.99%	61.71%	58.00%	-0.28
	GSO	79.20%	72.80%	71.95%	63.97%	52.81%	-0.21
	RAND	79.20%	72.80%	72.77%	64.53%	50.00%	-0.19
0.5	GEFeS	79.20%	72.80%	73.63%	63.49%	48.07%	-0.20
	PSO	79.20%	72.80%	72.24%	63.89%	53.04%	-0.21
	ABCO	79.20%	72.80%	72.03%	63.44%	52.30%	-0.22
	ASO	79.20%	72.80%	68.13%	60.48%	56.00%	-0.31
	GSO	79.20%	72.80%	71.55%	63.60%	52.67%	-0.22
	RAND	79.20%	72.80%	72.96%	64.77%	50.89%	-0.19
0.7	GEFeS	79.20%	72.80%	73.95%	64.13%	48.59%	-0.19
	PSO	79.20%	72.80%	72.16%	63.41%	50.81%	-0.22
	ABCO	79.20%	72.80%	72.43%	63.33%	51.56%	-0.22
	ASO	79.20%	72.80%	67.81%	60.05%	55.19%	-0.32
	GSO	79.20%	72.80%	72.11%	63.57%	53.33%	-0.22
	RAND	79.20%	72.80%	71.84%	63.95%	49.70%	-0.21
0.9	GEFeS	79.20%	72.80%	72.93%	63.07%	48.07%	-0.21
	PSO	79.20%	72.80%	72.56%	63.92%	52.67%	-0.21
	ABCO	79.20%	72.80%	72.69%	63.89%	52.89%	-0.20
	ASO	79.20%	72.80%	67.09%	58.43%	51.70%	-0.35
	GSO	79.20%	72.80%	72.00%	63.89%	52.67%	-0.21
	RAND	79.20%	72.80%	71.23%	63.17%	50.81%	-0.23
1.0	GEFeS	79.20%	72.80%	72.77%	62.72%	47.26%	-0.22
	PSO	79.20%	72.80%	73.01%	64.11%	50.89%	-0.20
	ABCO	79.20%	72.80%	71.63%	62.59%	49.26%	-0.24
	ASO	79.20%	72.80%	66.56%	58.08%	52.07%	-0.36
	GSO	79.20%	72.80%	71.09%	63.15%	50.74%	-0.23
	RAND	79.20%	72.80%	71.09%	63.17%	51.19%	-0.23

Table E.8. A Comparison of Adversarial Author Identification with and without Feature Selection Using the PAN19-25, Hybrid Feature Set – 566 Features, 100+(org = 125, adv = 125).

		Baseline		With Feature Selection			Use?
ω	FS Alg	Orig	Adv	Orig	Adv	% Features Used	
0.0	GEFeS	88.00%	83.20%	95.95%	80.93%	47.37%	0.06
	PSO	88.00%	83.20%	91.92%	80.77%	49.39%	0.02
	ABCO	88.00%	83.20%	91.57%	80.91%	49.98%	0.01
	ASO	88.00%	83.20%	94.16%	80.85%	49.65%	0.04
	GSO	88.00%	83.20%	91.39%	80.85%	49.98%	0.01
	RAND	88.00%	83.20%	91.15%	80.75%	49.82%	0.01
0.1	GEFeS	88.00%	83.20%	96.45%	80.69%	37.26%	0.07
	PSO	88.00%	83.20%	92.16%	80.83%	47.62%	0.02
	ABCO	88.00%	83.20%	91.81%	81.17%	49.10%	0.02
	ASO	88.00%	83.20%	94.08%	80.64%	41.52%	0.04
	GSO	88.00%	83.20%	91.39%	80.93%	49.05%	0.01
	RAND	88.00%	83.20%	90.91%	80.88%	49.73%	0.01
0.3	GEFeS	88.00%	83.20%	96.27%	80.61%	33.52%	0.06
	PSO	88.00%	83.20%	91.87%	80.75%	47.43%	0.01
	ABCO	88.00%	83.20%	91.09%	80.69%	47.95%	0.00
	ASO	88.00%	83.20%	94.03%	80.72%	27.71%	0.04
	GSO	88.00%	83.20%	91.39%	80.96%	47.82%	0.01
	RAND	88.00%	83.20%	90.91%	80.99%	49.93%	0.01
0.5	GEFeS	88.00%	83.20%	96.45%	80.88%	28.83%	0.07
	PSO	88.00%	83.20%	92.16%	81.17%	45.37%	0.02
	ABCO	88.00%	83.20%	91.49%	81.09%	47.27%	0.01
	ASO	88.00%	83.20%	93.63%	81.07%	19.19%	0.04
	GSO	88.00%	83.20%	91.28%	80.93%	47.61%	0.01
	RAND	88.00%	83.20%	91.09%	81.07%	50.00%	0.01
0.7	GEFeS	88.00%	83.20%	96.67%	80.43%	24.48%	0.07
	PSO	88.00%	83.20%	92.32%	80.96%	45.26%	0.02
	ABCO	88.00%	83.20%	91.55%	80.75%	45.65%	0.01
	ASO	88.00%	83.20%	93.09%	80.91%	14.53%	0.03
	GSO	88.00%	83.20%	91.23%	81.01%	46.20%	0.01
	RAND	88.00%	83.20%	90.69%	80.75%	50.37%	0.00
0.9	GEFeS	88.00%	83.20%	96.61%	80.59%	22.40%	0.07
	PSO	88.00%	83.20%	92.13%	80.83%	43.86%	0.02
	ABCO	88.00%	83.20%	91.63%	80.80%	44.81%	0.01
	ASO	88.00%	83.20%	90.96%	80.77%	40.67%	0.00
	GSO	88.00%	83.20%	91.15%	80.72%	44.73%	0.01
	RAND	88.00%	83.20%	90.75%	80.77%	49.83%	0.00
1.0	GEFeS	88.00%	83.20%	96.16%	80.45%	20.44%	0.06
	PSO	88.00%	83.20%	91.57%	80.72%	43.12%	0.01
	ABCO	88.00%	83.20%	91.44%	80.80%	44.04%	0.01
	ASO	88.00%	83.20%	88.37%	80.88%	49.80%	-0.02
	GSO	88.00%	83.20%	90.85%	81.04%	43.82%	0.01
	RAND	88.00%	83.20%	90.91%	80.88%	49.76%	0.01