Enhancing Aspect-Based Sentiment Analysis: Investigating Aspect Term Extraction and Annotation Schemes

by

Gabrielle Taylor

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

Auburn, Alabama
July 25, 2023

Keywords: natural language processing, topic modeling, aspect-based, aspect term extraction

Approved by

Xiao Qin, Chair, Alumni Professor of Computer Science and Software Engineering
Ashish Gupta, Co-Chair, Globe Life Professor of Analytics
Cheryl Seals, Charles W. Barkley Professor of Computer Science and Software Engineering
Santu Karmaker, Assistant Professor of Computer Science and Software Engineering

Abstract

We live in an age where there is constantly an increasing need for detailed information. From the company that needs to know what their users are feeling and thinking about their product to another crisis like a global pandemic, there is a growing need for information. We have seen various crises including mental health, natural disasters, and economic change. On the other hand, we desire the advancement of technology so that we can know more, experience more, and enjoy more. To achieve an improved experience, organizations need a way to understand their users. Computers aid in bridging the gap between organizations and their users' needs. For computers to gather detailed sentiment information from text, there is a necessity for an approach that analyzes the content of the text. Typically, Sentiment Analysis (SA) is implemented at a document level, sentence level, aspect level, fine-grained level, and emotion level. There is a range of depth to the levels of SA. However, one of the most fine-grained levels of SA is Aspect Based Sentiment Analysis (ABSA). It is one of the standard most in-depth approaches to SA. While the ABSA approaches lead to detailed information, they continue to produce limits in information by solely focusing on the polarity of an aspect term. Thus, the outputs of ABSA models are deficient. We thus need to design and develop approaches for to improve ABSA which is why we focus on Aspect Term Extraction

This dissertation argues that to provide detailed information from text, we must design techniques for ATE. In this regard, we design an ATE framework to address it's integral part of the development of the sentiment analysis branch of Natural Language Processing (NLP). We believe that a successful ATE approach using annotation schemes has not yet been elucidated.

In the first part of this dissertation, we provide the background of ABSA, why there is a need for ATE. We then provide one case study that exposes the limits of topic modeling as an approach to ATE in ABSA which has been previously used. Through our experiments, we develop a framework of Sentiment and Emotion Analysis (SEA) that presents the weakness of topic modeling that is commonly used for Aspect-Based Sentiment and Emotion Analysis. In the second part of this dissertation, we present a unique neural network-based approach to ATE using annotation schemes. Additionally, we present our own annotation scheme, which we all reverse IOB (or _IOB), as a successful approach to Aspect Term Extraction. The method models aspect extraction from the data and the emotion analysis of the aspect text. The output of our models produces results that make our approach a viable solution for producing detailed information. We conclude our dissertation with an opportunity for ABSA to be a tool for detection in online tasks. This dissertation exhibits how to develop approaches that contain these properties and which implementations will return accurate information.

Acknowledgments

The world of graduate school is a world of mystery, adventure, and danger. Similar to natural selection, only the strong survive. Graduate school is ambiguous, challenge, and forming. The ones that enter, never leave the same, they are formed, molded and transformed into more of their true selves. You learn who you are and what you are really made of. I had no idea what this journey would be like, but I am forever grateful. I must begin my acknowledgments with the one that brought me to graduate school: God. Years ago he put this desire in my heart. He gave me a love for computers at the age of eight. He gave me a love for languages at the age of twelve. All along the way to graduate school, he has shaped and molded me to be able to survive this journey. For this, I am forever grateful. This is his dissertation.

Before even choosing Auburn University, I met Dr. Seals through a friend from church that taught at Auburn, Dr. Atkins. Dr. Seals inspired me and encouraged me to pursue this Doctorate degree at Auburn University. I will never forget that meeting and her words. I applied to Auburn because these two women believed in me. I began my journey unsure of what to expect. My first year was a challenge but also very exciting. Soon after I began school, I learned that there were not any courses on Natural Langauge Processing, which was what I desired to pursue. I prayed for a way to learn Natural Language Processing and for a professor that would teach me. After my first year of graduate school, I was still set on pursuing Natural Langauge Processing wholeheartedly. My issue was that I did not know any professors that taught in that field that I could work under. However, I happened upon the greatest set of advisors: Dr. Xiao Qin and Dr. Ashish Gupta. I have them to thank for this incredible journey. They have given me the opportunity to become a researcher and have supported me in so many ways. Dr. Qin was the first person to take me under his

wing and give me the opportunity to explore my passion for Natural Language Processing. I am forever grateful for him taking me on as a student and believing in me. When he met with me, he would clearly explain my strengths and my opportunities for growth. He has always motivated and encouraged me in my research and given me the best direction to complete my degree. I met Dr. Gupta my second year of graduate school when he became my advisor. With his knowledge of Natural Language Processing, he gave me ideas and options and allowed me to choose my own research. He walked with me every step of the way and provided me with resources so that I could do my best. When my second year of grad school began, I discovered that a professor for Natural Language Processing was hired, Dr. Santu! My prayers were answered! I was overjoyed and chose to take a course of his every semester. I am incredibly grateful for the knowledge, insight, and challenge that Dr. Santu provided. He has helped me on projects, given me direction, and encouraged me in the field of NLP. I am deeply grateful that he teaches at Auburn University.

There were a number of people that have helped me over the past 5 years pursuing this Doctorate degree. My parents have provided wise words about balance and hard work. They have encouraged me, corrected me, inspired me, and prayed for me faithfully. The have walked me through my hardest moments and rejoiced with me at the smallest successes. Thank you for your love and patience. My sisters and brother, Karyn, Timira, and Christian have been my cheerleaders and listening ears. They have given me places to stay, food, and wisdom when I did not have faith. Timira has walked with me and lived with me through all of my ups and downs! She is such a hard worker and inspires me to do so. We have struggled through graduate school together, but looking at the way she has persevered has helped me. To my grandparents, Ellawese, Tommy, Gloria, John, and Brenda, thank you for your love and for paving the way for me in this way. You all are my heroes. My GC1 family has supported me, connected me, and provided food, love, and even dissertation examples for me. The way they pursue God and their dreams gives me the strength to do the same. To my best friends Kara, Brielle, Maurissa, Taylor, Gabbie, Lizzy, Kianee, Xergio, Gabe, and

Ida, I love you guys. I am so grateful for your patience and understanding and compassion as I have been on this journey. Your love, excitement, knowledge, and prayers were everything I needed to get here. I love you all forever.

To the Office of the Graduate School, every single one of you has blessed me immensely. Jay for hiring me and believing in me. I was so grateful to have a fellow Computer Scientist to chat with at work and keep me motivated. Dr. Witte, your love is unimaginable. Thank you for supporting me and giving to me and encouraging me every step of the way. This job was sent from God. Dr. Flowers thank you for always being there when I needed something and for helping me to see that I can do this. Ms. Sherry, you have helped me so much. I loved seeing you every day and the way you give wholeheartedly to every student. Thank you for taking care of me. Mrs. Libby thank you for being there for me during my hardest time. I appreciate your compassion and true support. I wouldn't be able to do this without you or anyone at the Office. It was truly a blessing and a dream to work there.

Thank you to my friends, coworkers and mentors at Adobe! That was truly the best working experience I have ever had! Thank you Franck Dernoncourt and Trung Bui for hiring me and teaching me about NLP along the way. Thank you to my coworkers Sebastian, Anthony, Chandler, Darian, Sean, and Steven. Thank you Helina for your mentorship and friendship, making Adobe fun, and helping me to believe in myself. I am grateful to the GEM Fellowship for allowing me to work at Adobe. I learned so much about NLP from my coworkers and mentors. We laughed and built innovative projects every day. Thank you all for teaching me what POS means and all the other simple new-to-NLP questions I had.

To my church family at Auburn Tuskegee, thank you for sticking in there with me, having grace with me, and taking things off of my plate. I would name each one of you but that would be another dissertation length of information. I love you all. To the staff and friends at NorthRiver, I have no words. LaToya, you love and support and the way that you celebrate me, I am forever grateful that God put us together. Jordan, you inspire me and have helped me believe in myself. Kendall, thank you for taking things off of my plate and

also making life so fun and adventurous! MaKenzie, the way you handle chaos has helped me choose balance and has motivated me that I can finish my dissertation. Nandi, Lee, Garrett, Amanda, Kristina W, and Chandler, I have loved working with you all. You have kept me sane and given me hope. Your support and love have meant everything to me. To Byrce Outten, thank you for being a friend to me. Grad school has been hard on both of us. Thank you for sticking in there with me. Your ability to make it through has helped me make it through. Thank you for your words and for never giving up on me.

To my roommates over the years, thank you. Erin, you gave me housing when I had none. Alexa, you helped me with Computer Science when I felt so hopeless and confused. Victoria, you are such a hard worker and that has helped me to grow in that area as well. Thank you for your love. Thank you to Mosi for introducing me to the Forest app! It truly helps me to work hard and stay focused. Thank you to Demi Ola for encouraging me and giving me vision during my hardest Dissertation week. It gave me all the motivation I needed. Thank you Carlos for all of the scriptures that have encouraged me to make it and helped me keep my focus on God. Thank you Jose for intentionally and unintentionally helping me to finish by the way you live your life. Thank you for the scriptures, songs, prayers, and inspiring talks. Thank you for being on this journey with me. To everyone that I have met along the way, that have helped me and inspired me, thank you from the bottom of my heart. Thank you, God for putting all of these amazing people that you created in my life. Forever grateful.

Table of Contents

Abstract ii

Acknowledgments iv

List of Figures xii

List of Tables xv

1 Introduction 1

1.1 Scope of the Dissertation Research 1

1.1.1 Sentiment and Emotion Analysis 2

1.1.2 Aspect Term Extraction 5

1.1.3 Misinformation on Social Media 6

1.2 Motivations 7

1.2.1 Motivations for Sentiment and Emotion Analysis 8

1.2.2 Motivations for Aspect Term Extraction 8

1.2.3 Motivations for Misinformation Detection 9

1.3 Contributions 9

1.4 Dissertation Organization 10

2 The History of Aspect-Based Sentiment Analysis 12

2.1 Introduction 12

2.2 Natural Language Processing 12

2.2.1 Sentiment Analysis 15

2.2.2 Emotion Analysis 18

2.3 Aspect Based Sentiment Analysis 19

2.3.1 Aspect and Opinion Term Extraction 21

2.3.2 Aspect Category Detection 22

2.3.3 Sentiment Classification 23
2.4 Aspect-Based Emotion Analysis 25
2.5 Summary 27
3 Sentiment and Emotion Analysis on Social Media Data 28
3.1 Problem Statement of Topic-Based Sentiment and Emotion Analysis 28
3.1.1 Why Sentiment and Emotion Analysis is Critical? 29
3.1.2 Research Goals and Problem Statement 30
3.1.3 Research Milestones 30
3.2 Research Methodology 31
3.2.1 Dataset 31
3.2.2 A Modeling Method 32
3.3 English Analysis 34
3.3.1 English Sentiment 34
3.3.2 English Emotions 39
3.3.3 English Themes 39
3.4 Spanish Analysis 44
3.4.1 Spanish Sentiment 45
3.4.2 Spanish Emotions 46
3.4.3 Spanish Themes 50
3.5 Language Analysis 52
3.6 Sentiment Analysis Between the English and Spanish Speakers 52
3.6.1 Emotion Analysis Between the English and Spanish Speakers 54
3.6.2 Themes Analysis Between the English and Spanish Speakers 59
3.7 Summary 60
4 Aspect Term Extraction using Annotation Schemes 62
4.1 Formulation of Research Tasks 62
4.2 Problem Statement and Research Questions 63

4.2.1 Problem Statement . . . 63
4.2.2 Research Questions . . . 64
4.3 Methodology . . . 65
4.3.1 Annotation Schemes . . . 67
4.4 Aspect Term Extraction Experiments . . . 69
4.4.1 Dataset Preparation for Aspect Term Extraction . . . 70
4.4.2 Compared Methods for Aspect Term Extraction . . . 72
4.4.3 Results of Aspect Term Extraction . . . 74
4.5 Summary . . . 84
5 AI-Enabled Misinformation Detection and Research Opportunities . . . 87
5.1 Misinformation Types . . . 89
5.1.1 Fake News . . . 90
5.1.2 Fake Reviews . . . 92
5.1.3 Misinformation in Microblogs . . . 93
5.1.4 Rumors . . . 95
5.1.5 Satire . . . 97
5.1.6 Clickbait . . . 99
5.2 Misinformation Detection Methods . . . 101
5.2.1 Natural Language Processing . . . 102
5.2.2 Machine Learning . . . 108
5.2.3 Data Mining . . . 117
5.3 Research Opportunities . . . 120
5.3.1 Research Opportunity 1: Corpora . . . 120
5.3.2 Research Opportunity 2: Non-English Languages . . . 121
5.3.3 Research Opportunity 3: Misinformation Infancy . . . 123
5.4 Summary . . . 124
6 Conclusions and Future Work . . . 128

6.1 Sentiment and Emotion Analysis 128
6.2 Aspect Term Extraction 129
6.3 Future Research Directions 131
6.3.1 Advanced Sentiment and Emotion Analysis 131
6.3.2 Optimizing Aspect-Based Sentiment Analysis 132
6.4 Summary 132
Bibliography 133

List of Figures

2.1 NLP to ABEA 13

2.2 Sentiment Analysis Levels 16

2.3 Sentiment Analysis Approaches 17

2.4 ABSA Subtasks 20

2.5 ATE Example 21

2.6 ACD Example 23

2.7 Example Ensemble Classifier 24

2.8 ABSA-ABEA Overlap 27

3.1 Language Analysis Process 34

3.2 English Jan-Apr Sentiment Chart 36

3.3 English May-Aug Sentiment Chart 37

3.4 English Sept-Dec Sentiment Chart 38

3.5 English Jan-Apr Emotion Chart 40

3.6 English May-Aug Emotion Chart 41

3.7 English Sept-Dec Emotion Chart 42

3.8 Spanish Jan-Apr Sentiment Chart 46

3.9 Spanish May-Aug Sentiment Chart 47

3.10 Spanish Sept-Dec Sentiment Chart 48

3.11 Spanish Jan-Apr Emotions Graph 52

3.12 Spanish May-Aug Emotions Chart 53

3.13 Spanish Sept-Dec Emotions Chart 54

3.14 Language Positive Comparison Chart 55

3.15 Language Negative Comparison Chart 56

3.16 Top 2 English-Speaker Emotions 57

3.17 Top 2 Spanish-Speaker Emotions 58

4.1 Our Aspect Term Extraction Methodology 66

4.2 Input and Output of ATE Model 69

4.3 AWARE Dataset Aspect Term Count 71

4.4 Aspect Term Extraction Dataset Format 72

4.5 Sample Sentence with Annotation Schemes 73

4.6 Topic Modeling Process 73

4.7 Precision Formula 74

4.8 Recall Formula 74

4.9 F1-Score Formula 74

4.10 Metrics Calculation Confusion Table 75

4.11 LLM Time Comparison 76

4.12 GPT F1 Scores Comparison 80

4.13 Electra F1 Scores Comparison 80

4.14 Roberta F1 Scores Comparison 80

4.15 Distilled BERT F1 Scores Comparison 81

4.16 Annotation Schemes F1 Scores Comparison 81

4.17 IO & reverse IOB F1 Scores Comparison 82

4.18 All Labels F1 Scores Comparison 83

4.19 Aspect Labels F1 Scores Comparison 83

4.20 IO & reverse IOB Labels F1 Scores Comparison 84

5.1 A variety of misinformation types along with the corresponding motivations to build detection systems 89

5.2 The taxonomy tree of the cutting-edge misinformation techniques. 101

List of Tables

2.1 COVID-19 NLP Paper Comparison Table 18

3.1 English Tweets 2020 Events 35

3.2 English Themes: January to April 44

3.3 English Themes: May to August 44

3.4 English Themes: September to December 45

3.5 English Top Theme Emotions 45

3.6 Top 5 Spanish-speaking Countries in Dataset 47

3.7 Spanish Tweets 2020 Events 49

3.8 Spanish Top Theme Emotions 51

3.9 Spanish Themes: January to April 51

3.10 Spanish Themes: May to August 51

3.11 Spanish Themes: September to December 55

3.12 English and Spanish Similar Themes 59

3.13 Language Comparison Top Theme Emotions 59

4.1 Aspect Results from Comparison Models 75

4.2 IOB Aspect Results 77

4.3 IO Aspect Results 77

4.4 BILOU Aspect Results 77

4.5 BIES Aspect Results 78

4.6 IE Aspect Results 78

4.7 IOE Aspect Results 79
4.8 BI Aspect Results 79
4.9 _IOB Aspect Results 79
5.1 Language Models/Text Processing Comparison Table 103
5.2 Machine Learning Comparison Table 126
5.3 Data Mining Comparison Table 127
5.4 Open Issues Summarization 127

Chapter 1

Introduction

In this dissertation research, we will present novel approaches to enhancing Aspect Based Sentiment Analysis. We begin this dissertation with the history of Aspect Based Sentiment Analysis (ABSA). We surmise that this level of background information will provide clarity on the research gap and direction that this dissertation addresses. This Introduction chapter highlights the motivations for the ABSA research thrusts - Sentiment and Emotion Analysis (SEA) and Aspect Term Extraction (ATE). SEA and ATE are subtasks of ABSA. Our aim is to enhance Aspect-Based Sentiment Analysis by investigating sentiment and emotion analysis approaches and aspect term extraction approaches using annotation schemes.

More specifically, this chapter is organized as follows. We begin by elaborating on the background and motivation of topic-based sentiment and emotion analysis in Section 1.1.1. Next, we share the background and motivation of aspect-term extraction in Section 1.1.2. We continue by introducing the motivation of adding misinformation as a research opportunity for future work in Section 1.1.3. After that, Section 1.3 concludes the contributions of this dissertation research. In the final section, we show the organization of this dissertation in Section 1.4.

1.1 Scope of the Dissertation Research

Within this dissertation, we will be discussing Aspect Based Sentiment Analysis (ABSA), methods, and improvements the aspect term extraction subtask. Our research on Sentiment and Emotion Analysis on social media is articulated in Section 1.1.1. Section 1.1.2 focuses on Aspect Term Extraction, the subtask of ABSA, and our method of implementation that

can improve detection. Finally, in Section 1.1.3 we introduce Misinformation on social media as an area of research opportunity for ABSA.

1.1.1 Sentiment and Emotion Analysis

In an effort to add an approach to ABSA, we first focus on Sentiment and Emotion Analysis and their corresponding topics. We select a hot topic for the dataset: COVID-19. The surge of COVID-19 beginning in 2020, brought a range of sentiments and emotions towards various aspects of the virus. We will discuss our model that analyzes the topics, sentiments, and emotions of our dataset. In this dissertation, we build the uniqueness of this study on three things: Twitter, English and Spanish, and COVID-19. As our first step, we analyzed the sentiments and emotions of English and Spanish tweets pertaining to COVID-19 from January 2020 to December 2020.

COVID-19, also known as the *Coronavirus*, is a virus that spreads by close contact between people and has spread throughout the world [273]. The symptoms range from mild to very extreme. Since 2020, it has developed into a number of variants of this virus [99, 2]. The spread of the Coronavirus has drastically changed the way of life for most individuals. Considering that close contact risks the spread of the virus, the world has had to limit contact to prevent the spread. This limitation led to many businesses closing, a drastic reduction in travel, and more preventative measures. World Health Organization reports that COVID-19 has led to more than 6.87 million deaths in the world. Many have suffered the devastating loss of family members, finances, and businesses. There is so much to analyze under this topic using Sentiment Analysis and Emotion Analysis.

Sentiment Analysis (SA) utilizes text analysis and computational techniques to classify the sentiment of the text input as positive, negative, or neutral. The task of gathering sentiment and opinions from data consistently improves many fields such as marketing, consumer reviews, and social media. *Emotion Analysis* (EA) is the process of identifying and analyzing the underlying emotions expressed in textual data [131]. EA extracts features

from text to gather subjective information and determine emotions. Sentiment analysis is not always enough to understand the specifics of what people truly feel; therefore, emotion analysis can retrieve further information from the users. The three traits of this study are Twitter, English and Spanish, and COVID-19. Performing sentiment and emotion analysis is desirable because it enables us to know the typical human responses so that we can adequately prepare for the future [45, 319]. This dissertation work can be helpful for government officials and medical professionals by: (1) learning how to use SA and EA during crises and (2) learning the emotions and needs coming from COVID-19. These individuals are the ones who determine the most appropriate way to communicate with individuals, keep them at ease, and help them follow the rules set in place. Sentiment analysis has helped individuals understand the perspective and opinions of individuals around the world about a specific topic [110]. Emotion analysis gives even more detailed information. Using these techniques during a global crisis is only logical. People felt a range of emotions during 2020 due to the Coronavirus affecting many lives.

The expansive growth of social media (e.g., Twitter and Facebook) brings about an incredible increase in information and opinions online. The constant stream of information creates a demand for processing that data. Utilizing Twitter, we systematically demonstrate that an in-depth analysis of the topics, sentiments, and emotions concerning the COVID-19 pandemic gives insight into future crises. Twitter has been an excellent resource for researchers, marketers, political campaigners, and even medical practitioners. It provides data to predict the flu, the stock market, spammers, and events [5, 55, 46, 283]. Twitter allows for real-time updates; an analysis of tweets can also show political influence or opinions [257]. This microblogging platform is one of the best resources for active, unfiltered data from real users worldwide. Twitter is the best platform to look into COVID-19 sentiments and emotions. Since 2020, microblogging platforms have been a source for communities to receive information about the virus and an outlet for people to share their thoughts and feelings. Microblogging platforms furnish an environment where individuals can freely share

their thoughts, opinions, and emotions about any topic. Twitter is undoubtedly one of the most popular microblogging platforms. We opt for the sentiment and emotion analysis of tweets because tweets typically portray the feelings or opinions of individuals [173]. Twitter provides an easy-to-use platform with short messages for people's thoughts. A growing number of users post a diversity of their thoughts on Twitter, which registers 500 million tweets a day [135]. Tweets are an incredible data source for researchers to understand and investigate the mind and thoughts of the world.

Twitter is an ideal resource for sentiment and emotion analysis. Analyzing the sentiment in tweets is more challenging but more beneficial than analyzing blogs or articles [316]. One challenge with this task is that a significant portion of tweets embraces real-time information [110]. Secondly, tweets are concise, which hinders detecting overall sentiment and emotions with little context. Thanks to potential benefits, there is a pressing demand to construct modern sentiment analysis systems using tweets. For instance, tweets contain a variety of topics and opinions, allowing researchers to grasp the majority topic sentiment by scanning a few thousand tweets [158].

We choose to supplement our COVID-19 tweets by exploring English and Spanish COVID-19 tweets to gather a more well-rounded analysis. We desire to see if the two language cultures have differing sentiments, emotions, and conversations. It enhances our analysis by comparing the two. We believe that this comparison is beneficial for government officials and medical professionals. The research of sentiment and emotion analysis on a multilingual dataset is challenging due to the lower resources in non-English languages. In order to advance research within sentiment and emotion analysis, we must continue to broaden beyond solely English data [79, 36]. Analyzing English sentiments and emotions is not enough when investigating a global pandemic. Kim *et al.* state that native non-English speakers have more influence than English speakers. They also learned that there are topics that native non-English users prefer to address in their native language rather than English

users on social media [160]. Therefore, our research will give insight into the conversations and sentiments that a wide range of people speaks on.

1.1.2 Aspect Term Extraction

Social media platforms like Instagram, Facebook, and Twitter host millions of users every single day. All of these businesses provide services with a strong focus on communication and a relationship. It is a fundamental resource for companies to use these platforms to better understand their customers. Natural language processing methods (NLP) have created an incredible opportunity for companies to analyze these communications platforms. Social media is a great help because people regularly post their opinions on all kinds of businesses and products. The analysis of social media data can provide valuable insights by detecting customer feedback [29]. As mentioned in the previous section, Sentiment Analysis (SA) identifies and extracts user opinions in one class, such as negative, positive, or neutral. However, it would be helpful to narrow and precisely describe the insights described in texts. Sentiment Analysis is mainly researched at three different levels: document, sentence, and aspect [139]. The analysis of sentiment at the document or sentence level assumes that there is only one topic to grasp sentiment from which is not typically the care. Therefore, analyzing sentiment at the aspect level will return more accurate results. Due to this, there now exists a sub-branch of sentiment analysis called Aspect-based Sentiment Analysis.

Aspect-Based Sentiment Analysis (ABSA) is a branch of SA that is more focused on which aspect of the sentence is the sentiment related to. As we use the term *aspects*, we mean the attributes or components of a product or service. It is the category, feature, or topic of a section of text. ABSA is a technique that locates the terms related to the aspects and recognizes the sentiment connected to each aspect. There are two main ingredients for an ABSA model: the aspect terms and the sentiment classification. These ingredients are essential to extract sentiment for each aspect from the given text. We focus on improving Aspect Term Extraction.

The standard ABSA needs labeled data containing aspect terms and aspect categories for each piece of text along with its sentiment score. However, the task of ABSA can be solved using an unsupervised approach which does not require labeled data and aspect terms. The unsupervised approach requires a set of aspect categories and the model identifies the terms and sentiments for those categories. There are many ways to tackle the task of ABSA.

Currently, ABSA is vital to companies and organizations because it can help automatically analyze customer data to gain powerful insights. ABSA does the difficult work of sifting through data. It's impossible for organizations to manually inspect thousands of tweets, posts, conversations, or customer reviews, especially if they want to analyze information on a granular level. The automatization aspect of ABSA saves money and time for companies that they can use on other tasks. Presently, people are very vocal about products, services, and their experiences on social media where they often share their authentic thoughts and leave feedback. The feedback provided leaves a wealth of insights for businesses to use to improve. Therefore, ATE is a crucial task to attend to. The user needs to know which aspects to focus on to know what sentiment or emotion is coming from.

In this dissertation, we build a novel approach to ATE by implementing annotation schemes. We believe that this approach enhances and enriches the classification of aspects and emotions. We will discuss our technique further in Chapter 4.

1.1.3 Misinformation on Social Media

In this section, we will discuss the background of *misinformation* as a motivation for the research in aspect-based sentiment analysis.

Misinformation, which is simply inaccurate information, has become more pervasive on the Internet by the day. The term *post-truth* describes the world we live in today. Post-truth is an environment where the facts are irrelevant compared to emotional appeals and personal beliefs; these more than facts are now shaping public opinions. The dilemma that post-truth culture has created has become a critical issue since it has affected politics,

voting, and the market. Publishing and sharing misinformation has constantly misled the viewpoints, beliefs, and opinions of online users. People today are more inclined to believe any information online, thereby making users vulnerable. As our society advances with technology and Internet interactions, misinformation detection systems will be vital so that individuals can distinguish legitimate information from misleading ones. James Gleick once said: "when information is cheap, attention becomes expensive". Therefore, we must be assiduous in our pursuit of detecting misinformation so that users can safely continue to use the Internet as a safe and effective tool.

In this dissertation, we shed a bright light on sentiment and emotion detection techniques that have been relatively unexplored as tools for misinformation. We start this study by categorizing the versatile types of misinformation, including fake news, fake reviews, misinformation in microblogs, rumors, satire, and clickbait. We also discuss an array of challenges, previous approaches, and future research opportunities. Through our background research on misinformation, we discovered the lacking approaches using sentiment analysis and emotion analysis. In this dissertation, we explore the advancement of sentiment and emotion analysis on social media so that it can be used in issues like misinformation.

1.2 Motivations

The performance of sentiment analysis has been significantly growing in the past decade. With the evolution of NLP research, the models for, popular subtasks, and sentiment analysis have benefited from those advancements which developed Aspect Based Sentiment Analysis. In this dissertation study, I propose to seamlessly integrate deep learning models using annotation schemes as an Aspect-Term Extraction approach to effectively detect aspect terms in a sentence.

Our Aspect Term Extraction model presented in Chapter 4 integrates the motivated issues discussed in Chapter 2. The techniques presented in Chapter 3 into one seamless

technique. Our technique aids in detecting helpful terms in a text. Three emerging trends below strongly motivate us to do this research.

- There is a need for more sentiment and emotion analysis on English and Spanish social media data (see Section 1.2.1).
- Annotation Schemes have not been used for Aspect Term Extraction within ABSA (see Section 1.2.2).
- There is a pressing demand for techniques to advance misinformation detection (see Section 1.2.3).

The details of our motivations for this research are laid out in sections 1.2.1-1.2.3

1.2.1 Motivations for Sentiment and Emotion Analysis

Sentiment and Emotion Analysis (SEA) generates feedback from the population, a much-needed tool as social media continues to rise. In our preliminary research on Sentiment and Emotion Analysis, discussed in Chapter 3, we use COVID-19 tweets as our dataset. We do this because COVID-19 is a crisis that has caused a range of opinions and opinions. The multilingual aspect of the dataset gives us a wider view of the world's opinions. It motivates us to find the aspects of the crisis. The sentiment analysis process and emotion analysis process provide us with useful results. These results can be supplemented with an effective aspect classifier.

1.2.2 Motivations for Aspect Term Extraction

Aspect Term Extraction generates specific results from selected users which provides doctors, researchers, and government officials with valuable data. For our research on ATE, discussed in detail in Chapter 4, we choose a unique dataset and a unique technique. For our dataset, we choose to explore a review dataset. We use the AWARE dataset that was built for Aspect Based Sentiment Analysis. Our unique technique is the implementation of

GPT, Electra, RoBERTa, and Distilled BERT using annotation schemes. This approach has not yet been explored for ABSA.

1.2.3 Motivations for Misinformation Detection

Misinformation is currently pervasive on the internet and has affected many lives. There are a number of techniques for misinformation but sentiment and emotion analysis are one of the least explored techniques in this area. Due to this, we will explore improving sentiment and emotion analysis and aspect extraction techniques on social media so that they can be tools for misinformation in the future.

1.3 Contributions

We embark on this dissertation research with a thorough background and history of ABSA (see Chapter 2. We also discuss open research issues identified for our research directions. As previously mentioned, we may the foundation of emotions in Chapter **??** by detailing the history and definitions of emotions.

In this dissertation, we continue in Chapter 3 by building a system to detect sentiments, emotions, and reactions toward a crisis. This kind of information provides insight into public concerns which can be helpful for officials when designing and implementing guidelines for communities in a crisis situation. In Chapter 3, we articulate the design of the sentiment and emotion analysis model for English and Spanish. We propose the analysis tool of using spikes in trends to correspond to an event that has taken place. Finally, we offer results that provide an understanding of the reactions of users toward COVID-19.

For our advanced ATE model and contributions, detailed in Chapter 4. We use annotation schemes to accurately identify aspects. This kind of information provides insight into public concerns which can be helpful for companies and organizations that desire feedback from their user reviews.

More specifically, let us tabulate the contributions of each research task below. First, we list key contributions made by our sentiment and emotion analysis studies, the details of which can be found in Chapter 3.

1. We articulate the design of the analysis model
2. We propose the verification tool of using spikes in trends to correspond to an event that has taken place.
3. We offer results that provide an understanding of the reactions of users towards COVID-19.

Second, the contributions of ATE in this dissertation are summarized as follows. Additionally, the details of which can be found in Chapter 4.

1. The deep learning classifiers with annotation schemes as effective solutions
2. The reverse IOB annotation scheme provides a more accurate classification for Aspect Term Extraction

1.4 Dissertation Organization

This dissertation is organized as follows. In the next Chapter, Chapter 2, related work reported in the literature is comprehensively and extensively reviewed. We begin by discussing the history of sentiment analysis. We then present the development of SEA and ABSA and research issues centered around them. Lastly, propose an idea to incorporate how to incorporate the three ideas.

In Chapter 3, we describe our approach to sentiment and emotion analysis with topics for English and Spanish. Moreover, we discuss the results of COVID-19 tweets and the applicability of the approach. We then conclude with the solutions and contributions.

In Chapter 4, we dive into building off SEA with a new model to successfully extract specific aspects. We propose the use of annotation schemes as our novel Aspect Term Extraction. Our approach achieves high-accuracy results.

In the last chapter (Chapter 6), we conclude this dissertation by summarizing an array of major research contributions. Moreover, we discuss future research directions from various perspectives that have not yet been fully addressed in Chapter 3 and Chapter 4.

Chapter 2

The History of Aspect-Based Sentiment Analysis

2.1 Introduction

In the past decade, a growing number of large companies have chosen to utilize sentiment analysis to process enormous amounts of data. The challenges these companies face with these models are the accuracy and the precise detailed results. In this chapter, we present a variety of previous research studies that are closely related to this dissertation. We begin this chapter by focusing on the existing methods for sentiment analysis and emotion analysis. We then expound on research related to aspect-based sentiment and emotion analysis. We discuss the subtasks, methods, and state-of-the-art (SOTA) solutions. Lastly, we highlight the ongoing challenge of misinformation that organizations face online.

More specifically, this chapter is organized as follows. We discuss in detail the background and methods of Sentiment Analysis and Emotion Analysis in Section 2.2. Lastly, we move on to detail the cutting-edge aspect-based sentiment and emotion analysis techniques in Section 2.4. Through this history, we will begin from the broad view of natural language processing to the most specific view of aspect-based emotion analysis which we can see in Figure 2.1

2.2 Natural Language Processing

This section focuses on how researchers use Natural Language Processing (NLP) techniques including sentiment and emotion analysis to analyze data, especially data for crises. Many researchers have applied Sentiment Analysis on Twitter data to understand individuals of a specific demographic. The research reveals behavioral patterns toward situations

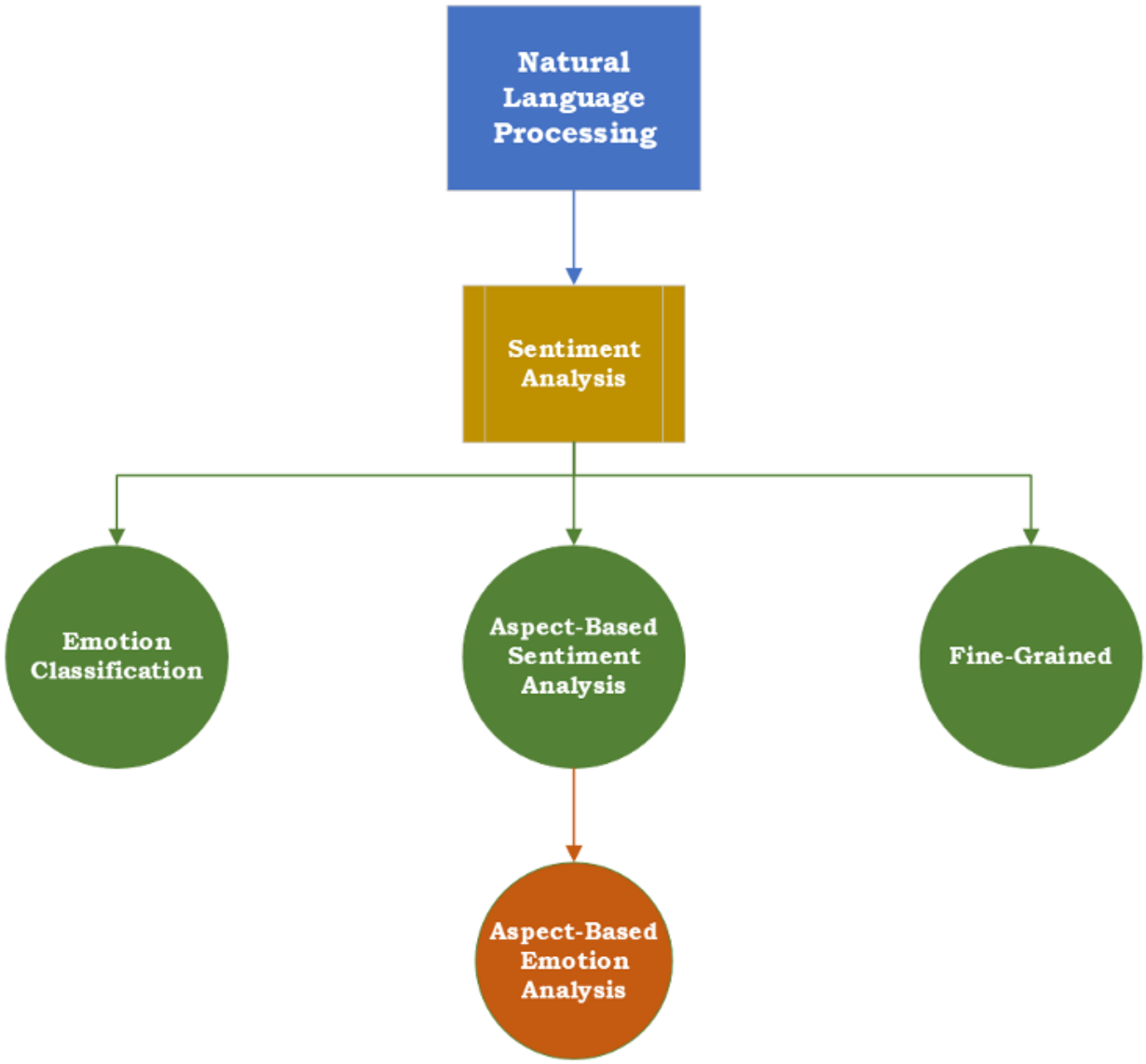

Figure 2.1: NLP to ABEA

around the world. In this section, we will discuss work related to our research. We will start by exploring NLP,sentiment analysis, and emotion analysis.

Natural Language Processing plays a crucial role in detecting misinformation. NLP, in combination with Machine Learning (ML), creates state-of-the-art detection systems. Natural language processing is an interdisciplinary field of computer science and linguistics using many machine learning algorithms. NLP furnishes an algorithmic method for computers to process and understand natural language. Generally speaking, four distinctive modules in NLP systems are data gathering, preprocessing data, training, and testing. There are several ways to gather data, including, but not limited to, web scraping, crowdsourcing, and open-source datasets. Preprocessing techniques, like tokenization, stemming, lemmatization, and parts-of-speech (POS)tagging, manipulate raw data to be further analyzed in optimized ways. Language models are one of the technical underpinnings in NLP used to understand the input enough to predict the following word given an input word. There are two types of language models: neural and statistical models. Neural language models use neural networks to predict the following letter, word, or phrase. Statistical language models like n-grams (see Section 4.1.1) are used to predict what word, letter, or phrase comes next in the sequence given the previous content. Depending on the desired goal, data then are divided into testing datasets and training datasets. The training data is input for classifiers, clusters, or neural networks. Model development happens in the training phase, and then the model tests by using the testing data. We can then analyze and make adjustments to the model based on testing results. All of these tools are very useful for linguistic understanding that leads to the detection of misinformation.

In what follows, we elaborate on the most popular NLP techniques for misinformation because NLP techniques exhibit strengths in detecting fake news. The data preprocessing phase uses NLP methods to understand the data before applying machine learning (see Section 4.2) or data mining (see Section 4.3) procedures. Table 4 summarizes all the NLP techniques applied to detect misinformation in the six domains. The five empty cells in

Table 4 suggest open issues and research opportunities. For instance, there is a potential opportunity to utilize paragraph vector schemes to detect fake news and reviews.

2.2.1 Sentiment Analysis

Sentiment Analysis (SA) is the most popular technique used to understand the sentiment or opinion of a person from their text [156, 180]. There are three different levels of sentiment analysis including document-level, sentence-level, and aspect-level [37, 152]. In Figure 2.2, we can see the different sentiment levels. There are a number of approaches to SA including lexicon-based, machine learning-based, and hybrid. The approaches are visible in Figure 2.3 Lexicon-based SA uses a corpus or a dictionary to determine the polarity of the word [1, 156]. It calculates the polarity based on the input string of words and determines the sentiment of the text. This method has been very useful in SA research, but it is not the most accurate method. It is a very static approach to understanding the sentiment of a text. Since human language is constantly changing and very complex, the advancement of accuracy for SA is needed. The gap that Lexicon-based SA produces is where the Machine Learning (ML) approach comes in. Sentiment Analysis using ML has covered more ground with the supervised learning method. Supervised learning utilizes math to develop classifiers that can take in data and labels and predict new data labels. Dhaoui *et al.* have reported that both approaches are similar in accuracy; they both achieve higher accuracy with positive sentiment classification than negative sentiment classification [92]. There are several popular tools for Sentiment Analysis; a few are: TextBlob, VADER, LIWC, and SentiWordNet [19]. These four tools are useful for getting the classification or polarity scores of words and sentences.

Sentiment Analysis of COVID-19. Interest in research related to the COVID-19 pandemic has grown increasingly. In Chapter 3, we discuss our approach to sentiment analysis related to COVID-19. Several researchers have implemented various methods to analyze information regarding COVID-19. A number of researchers solely use one language like ([193, 57, 196]). Manguri *et al.* gathered data by pulling it from Twitter through python

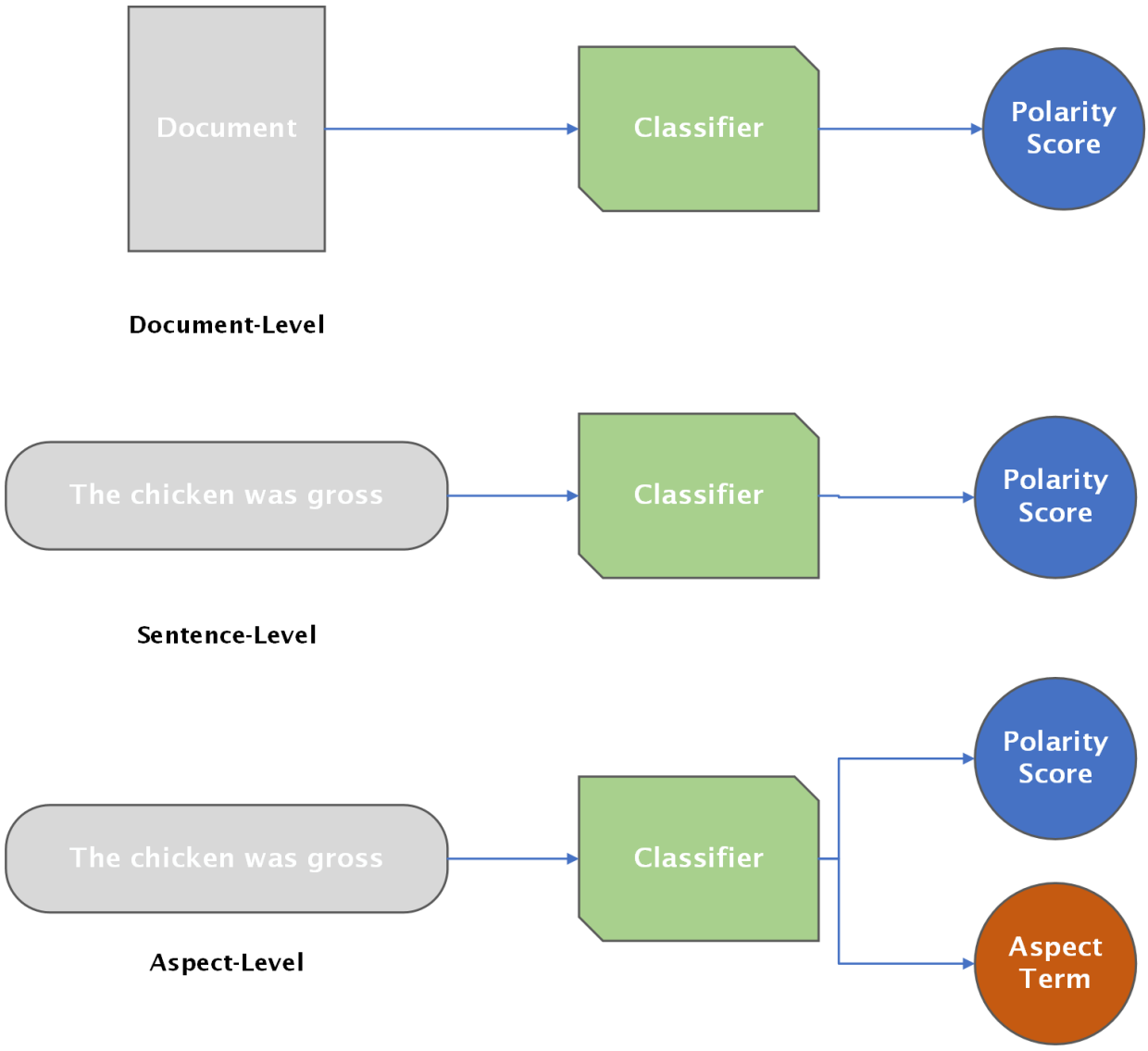

Figure 2.2: Sentiment Analysis Levels

programming, using the Tweepy library. They then used the TextBlob library in python to perform sentiment analysis. They gathered data for seven days, 09-04-2020 to 15-04-2020 [193]. Boon-Itt and Skunkan report that people have a negative outlook toward COVID-19. Their analysis revealed three themes relating to COVID-19: the COVID-19 pandemic emergency, how to control COVID-19, and reports on COVID-19 [57]. They used a data mining approach on Twitter data. They collected 107,990 COVID-related tweets between December 13, 2019, and March 9, 2020, on which they performed sentiment analysis and

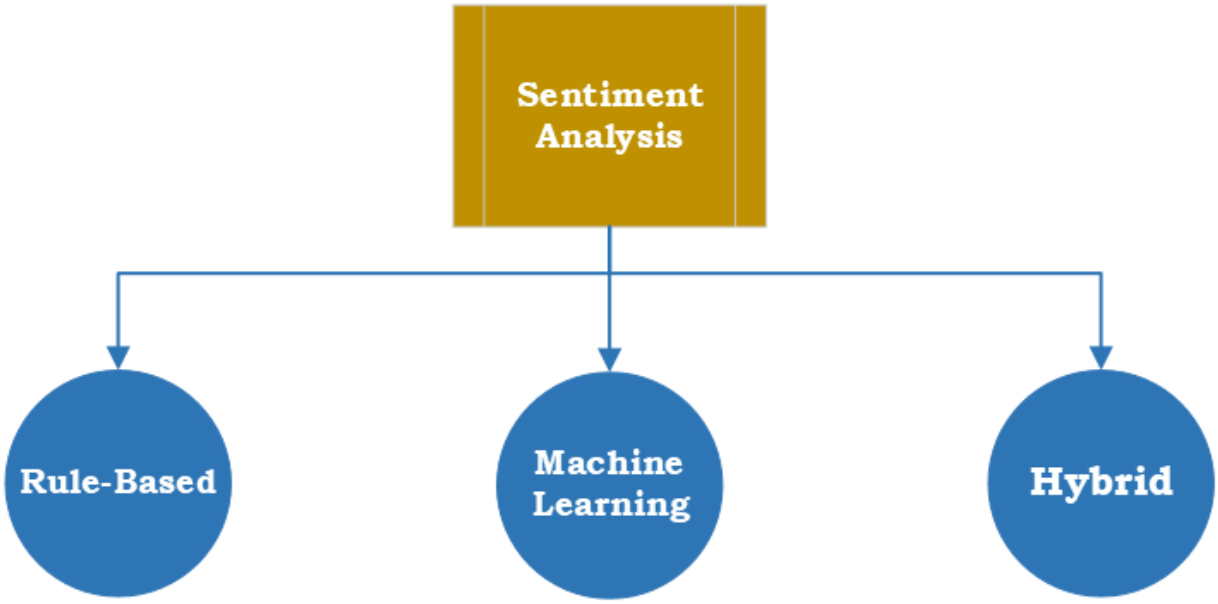

Figure 2.3: Sentiment Analysis Approaches

topic modeling. Boon-Itt and Skunkan used the Latent Dirichlet Allocation (LDA) algorithm for topic modeling. Medford *et al.* gathered their data using hashtags related to COVID for January. They gathered themes, topics, sentiments, and predominant emotions. They found that most of the tweets expressed fear, and 30% expressed surprise [196]. Gupta *et al.* developed a large dataset of 132 million tweets from January 2020 to January 2021 extracted by using key words [128]. They utilized NLP techniques and machine learning emotion algorithms to label each tweet. The tweets have 17 latent semantic attribute labels, generally including the binary, degree of emotion intensity, and sentiment and emotion categories. Gencoglu *et al.* used casual modeling to investigate COVID-19. Their approach led them to discover and quantify casual relationships between pandemic characteristics, Twitter activity, and public sentiment. Their results show that the variables affect public attention and sentiment [116]. Yin *et al.* found that there was more positive sentiment than negative sentiment during the two-week collection of tweets [301].

Other researchers analyzed multilingual data. Amara *et al.* select Facebook data and a multilingual dataset of 7 languages [24]. In another study, Kruspe *et al.* choose

to analyze tweets collected during the first months of the COVID-19 pandemic in Europe. They study the sentiment of the numerous European languages using a neural network and multilingual sentence embeddings. They noticed the correlation of lockdown announcements to the decline of sentiment among most countries analyzed [166]. Finally, Yang *et al.* use English and Arabic Twitter posts and Weibo messages to measure the sentiment during the pandemic. They labeled the tweets in 10 emotion categories. They then used a transformer framework to conduct multi-label sentiment classification. Their results show that positive sentiments increased over time, which foretold hope for an improved COVID-19 world [298].

Papers	**English Data**	**Multilingual Data**	**Topic Modeling**	**Sentiment Analysis**	**Emotion Analysis**	**1 Month**	**2-6 Months**	**12 Months**
[193]	X			X		X		
[57, 196]	X		X	X			X	
[166]		X		X			X	
[294]	X		X	X			X	
[129]	X		X	X			X	
[116]	X			X			X	
[301]	X		X	X		X		
[298]		X	X	X			X	
[24]		X	X				X	

Table 2.1: COVID-19 NLP Paper Comparison Table

2.2.2 Emotion Analysis

Emotion Analysis (EA) is a branch of Sentiment Analysis that analyzes and identifies the underlying emotions expressed in textual data, such as Ekman's six basic emotions. Paul Ekman identified the six basic emotions: anger, surprise, disgust, enjoyment, fear, and sadness [100]. Analyzing text to gather emotions is challenging since researchers typically rely on facial expressions to identify human emotions. The internet has provided a wide range of opportunities for emotion analysis. It has also provided more motivation to understand people's expression of emotions [269]. Similar to SA, there are lexical-based approaches and machine learning-based approaches. A few standard lexicons are EmoSenticNet and NRC-EmoLex, each containing thousands of words along with their related classifications [227,

200]. In addition, there are a variety of Machine Learning and Deep Learning methods, such as SVM, KNN, maximum entropy, CNN, LSTM, and Transformers [30, 4, 82]. EA is an exceptional field that provides a deeper insight into the thoughts and opinions of people.

2.3 Aspect Based Sentiment Analysis

Aspect Based Sentiment Analysis (ABSA) is a type of fine-grained Sentiment Analysis that extracts important aspects and classifies the related sentiment. We will continue to reference ABSA as we explore Aspect Based Emotion Analysis (ABEA) due to the close relation of the two tasks and the outweighing number of resources ABSA has compared to ABEA. We will begin by introducing the basic ideas behind aspect-based emotion analysis in Section 2.4, we briefly present the leading-edge techniques in the three major tasks in ABSA, namely, aspect and opinion term extraction (see Section 2.3.1), aspect category detection (see Section 2.3.2), and finally, aspect and sentiment and emotion classification (see Section 2.3.3-2.4).

ABSA is an important research field because it provides users with more descriptive information about a subject or product. ABSA is widely used for businesses and companies to understand the trends behind their sales. *Repustate* is a software company that utilizes ABSA in its API to help a wide range of businesses analyze sentiments across different platforms like product reviews, podcasts, and videos. The results from their API have given businesses information to make wise business decisions for their customers. ABSA can also be helpful in the medical field, specifically psychology. Researchers George *et al.* confirm that the application of ABSA has been useful in suicide prediction [117]. In their study, they developed a psychiatry-relevant lexicon and identified specific categories of words in order to classify suicidal and non-suicidal cases. They used ABSapp.

In order to build an effective model, researchers mainly pursue machine learning techniques. However, the challenge is that each model requires a large, domain-specific dataset.It is important that datasets are domain specific because the model is searching for patterns

to retrieve aspects. For example, it would be much easier to learn about user feedback about the features of the new iPhone if the dataset only contained reviews about the newest iPhone. However, if the dataset had reviews from a restaurant, the new Marvel movie, and the iPhone, it would be more of a challenge to gather aspects about so many differing features. Each aspect retrieved is associated with an aspect category. The identified categories verify the aspects and allow for an even more specific analysis of the dataset.

There are four main facets of ABSA: Aspect Term Extraction, Aspect Category Detection, Opinion Term Extraction, and Sentiment Classification which we display in Figure 2.4. We will elaborate on these facets and methods used in the next sections with the exception of Sentiment Classification. SC is the same as Sentiment Analysis and therefore, uses the same techniques that we have previously mentioned. There are a number of subtasks under ABSA that combine the facets. For example, Aspect Category Sentiment Classification (ASC) is a subtask of classifying the sentiment expressed in the aspect category. There are many ways to implement ABSA. While some researchers build end-to-end models for compound ABSA subtasks, others implement pipeline methods, and others implement multiple models for the single ABSA subtasks [310].

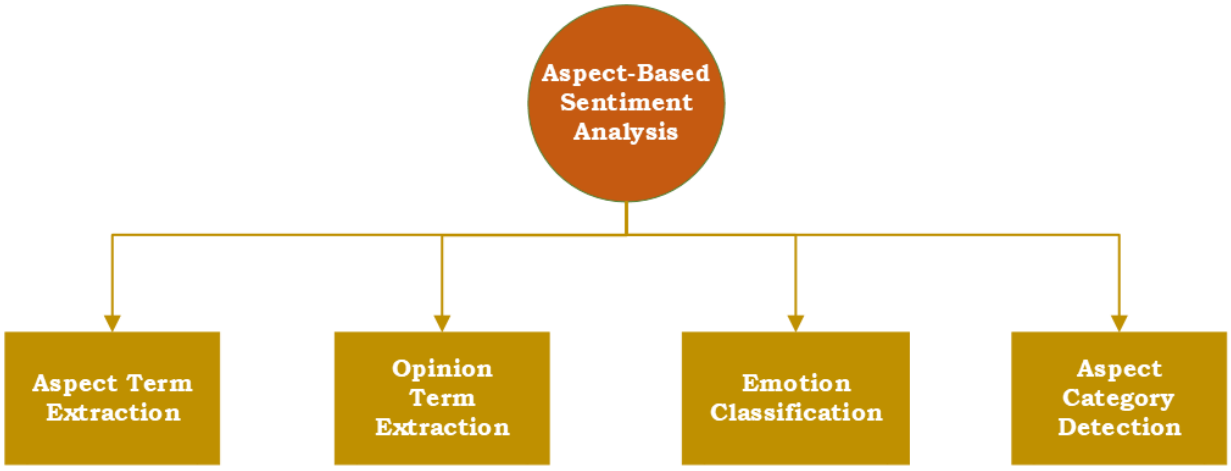

Figure 2.4: ABSA Subtasks

2.3.1 Aspect and Opinion Term Extraction

Aspect Term Extraction (ATE) aims to extract aspect terms from a sentence that people have expressed opinions on [70]. We can see an example of ATE in Figure 2.5. In the sentence *"The chicken was gross"*, ATE should extract the word "chicken". Analogous to all fields of machine learning, the task of ATE has developed over time. Beginning with rule-based methods and moving toward deep-learning methods. A number of researchers continue to use a range of approaches.

Opinion Term Extraction (OTE) aims to extract the opinion terms from a sentence that contains an object of focus. For example, in the sentence *"The chicken was gross"*, OTE should extract "gross". The process of OTE is very similar to ATE except the target word category varies. A variety of researchers pursue the extraction of aspects and opinions together using the same extraction technique [80, 287, 286]. We will examine approaches to ATE and OTE with the understanding that the methods are useful for both subprocesses of ABSA.

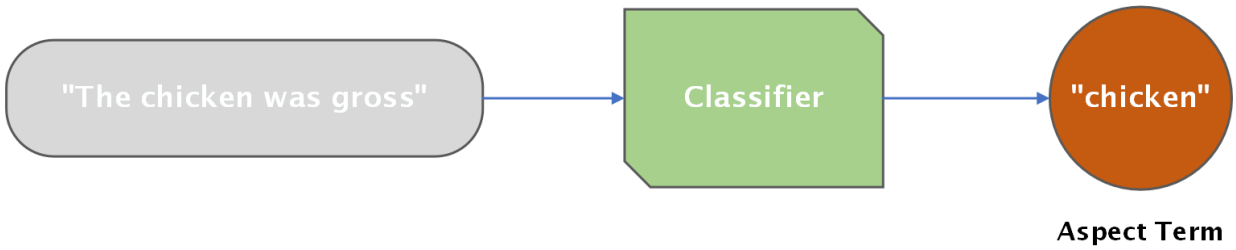

Figure 2.5: ATE Example

Venugopalan *et al.* determine that Aspect Term Extraction is the most challenging step in aspect-based sentiment analysis because its task is to discover the main features of the product in the reviews [274]. The task is essential to the progression of the ABSA process. Without the aspect term, the analysis will be inaccurate. The rule-based approach typically utilizes dependency parsing and a set of rules to make sense of the tree in order to locate the aspects [226]. Ku *et al.* proposed a rule-based approach that makes use of common-sense knowledge and sentence dependency trees to expose implicit and explicit aspects. They use

the Stanford Dependency Parser to build a dependency tree where rules about the subject and noun are utilized. They also used the Implicit Aspect Corpus, developed by Cruz-Garcia *et al.*, in order to extract implicit aspects and their categories [78]. The IACs are manually labeled. Lastly, they used SenticNet 3, a concept-level opinion lexicon, to extract sentiment words and determine aspect or aspect relationships. Augustyniak *et al.* tested ATE methods on a number of word embeddings, including word2vec, glove embeddings, numberbatch, fastText embeddings, and Amazon Reviews. The ATE method presented was a comprehensive analysis of various customizations of LSTM-based architecture. Their conclusion was that it is not always the biggest, most sophisticated, and most complex model that is the best for every situation. They surmise that it is better to start experiments with simple models and pre-trained word embeddings.

2.3.2 Aspect Category Detection

Aspect Category Detection (ACD) is one of the challenging sub-tasks of ABSA and has become popular due to the rapid growth in customer review data online. Given a predefined set of aspect categories and a set of review sentences, the goal of ACD is to detect implicit or explicit aspect categories from the sentences [232]. In the example presented in Figure 2.6, *The chicken is gross*, the aspect category is FOOD. Aspect Category Detection can be utilized for many tasks like a content recommendation, customer assistance in e-commerce, and efficient retrieval of scholarly documents [66]. These particular fields have a number of different aspects on which users have opinions. The more aspect categories, the more accurate the feedback or recommendation will be. The ACD systems are used to identify the key features that users care about in a product and summarize their feedback which provides essential insights that ease the decision-making process.

There are a vast number of approaches to ACD. Alvarez-López *et al.* used a linear SVM classifier, 12 unique predefined categories, and a word list of 2000 sentences, grouped in 350 reviews [23]. Ghadery *et al.* proposed an unsupervised method that utilizes clusters

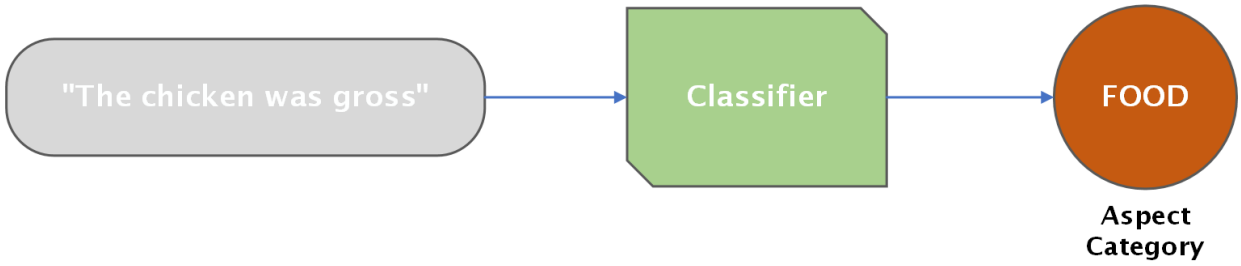

Figure 2.6: ACD Example

of unlabeled reviews and soft cosine similarity. The experimental results from this approach show that the proposed unsupervised approach outperforms several baseline methods. Given the set categories, they selected 5 seed words per category. They retrieved the similarity by calculating the average of soft cosine similarity values between the sentence and each of the seed words belonging to that category [118]. Yenkikar *et al.* selected category seed words, used Stanford CoreNLP to build a co-occurrence matrix and used a set of association rules to aspect categories [300]. Moreira *et al.* used four machine learning algorithms: Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). In their analysis, they concluded that the classifiers with the best general performance were SVM and RF [204]. Mamatha *et al.* built a Conditional Random Field (CRF) based model for category detection [192]. The model proved impressive results for each aspect category. A number of researchers use a set of rules and a simple Machine Learning algorithm or pursue more complex machine-learning methods. We have seen these approaches prove successful results.

2.3.3 Sentiment Classification

Ensemble classifiers are a new method within ABSA for sentiment classification. Xenos *et al.* implemented an ensemble classifier for sentiment polarity [291]. They used an ensemble of two Logistic Regression (LR) classifiers, each yielded one sentiment score for each sentiment label. The ensemble then computes a linear combination of the two scores and the label

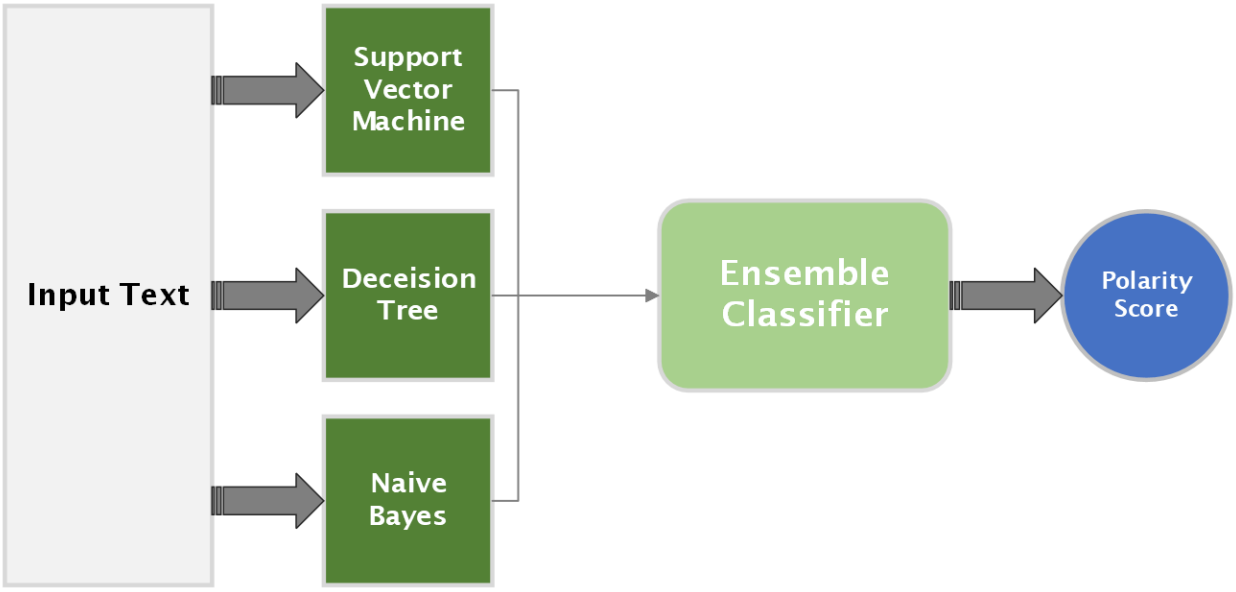

Figure 2.7: Example Ensemble Classifier

with the highest score is correct. Perikos *et al.* developed an ensemble classifier implementing Naive Bayes, MaxEnt, and SVM for sentiment classification. Their scores significantly improved by combining the modes than classifying the sentiment individually. Dashtipour *et al.* discovered that there was no research using the ensemble classifier method on Persian data. They built an ensemble classifier on a Persian dataset that achieved a 79.68% accuracy. Their classifier, shown in Figure 2.7, included a Convolution Neural Network (CNN), a Support Vector Machine, and a Multilayer Perceptron (MLP).

In this section, we will discuss models and efforts toward Aspect Based Sentiment Analysis and Aspect Based Emotion Analysis.

According to Do *et al.*, there is an extensive number of approaches to ABSA [94]. The pre-trained models (PTM) have proven effective in ABSA [81]. Dai *et al.* compare the PTM trees and the dependency parsing trees on multiple ABSA models. Their experiment shows two things. First, the tree produced from fine-tuned RoBERTa outperforms the dependency-parsed tree. Second, the RoBERTa-based model outperforms the previous SOTA results on six different datasets across four languages. They concur that this is the case because PTMs

implicitly incorporate task-oriented syntactic information [81]. Similarly, Do *et al.* analysis affirms that including pre-trained and fine-tuned word embeddings, and incorporating linguistic factors like part-of-speech and grammar rules, drastically improves a model's performance [94].

2.4 Aspect-Based Emotion Analysis

There are a handful of researchers that have pursued Aspect-Based Emotion Analysis. Padme *et al.* gather a review dataset and predict that the text not only contains sentiment information but also emotional information about the product [211]. Therefore, they extend the Latent Dirichlet Allocation (LDA) method and perform an emotion analysis on the dataset. They perform two tasks detection: Aspect detection from the reviews and emotion analysis. They used LDA to produce general topics and finalized the aspects by implementing the keyword spotting technique. They then analyze the emotions for an aspect from a review [211]. Another set of researchers, Ismail *et al.*, propose a framework for Emotion and Aspect Based Sentiment Analysis on data related to Covid-19. They use a variety of Machine Learning classifiers including Logistic Regression (LR), Linear SVC, Multinomial Naive Bayes, and Random Forests to classify the aspects as well as the emotion. They gathered their data from Twitter and performed preprocessing. They perform emotion analysis by using the names of machine learning classifiers and training them on an annotated training dataset labeled with the six basic emotions [143]. We see that a common successful technique is utilizing ML classifiers. García-Méndez *et al.* implement a stacking method of ML classifiers including NB, DT, Random Forest, and Stochastic Gradient Descent (SGD) which gave their system successful results.

Another set of researchers use Machine Learning for their Aspect Term Extraction using multiple Machine Learning models, and they use a package library *Text2Emotion* to for the emotion analysis [197]. Feizollah *et al.* implemented LDA for ATE and the National

Research Council of Canada Emotion Lexicon (NRC) for sentiment and emotion analysis [109]. Dehbozorgi *et al.* use the Part of Speech (POS) tags to identify the aspects and use *Text2Emotion* python package to identify the emotions [87]. Suciati *et al.* compared ML classifiers and Deep Learning classifiers that are given an aspect for sentiment and emotion analysis [260]. They were able to find that for certain aspects the models performed relatively the same. Devi and Deepa build a neural network with an Exponential Linear Unit activation function to classify tweets as "happy", "sad", or "fear". They used linguistic-based approaches by using keywords to label the tweets with emotion labels. All of the tweets were related to Covid deaths [90]. Polignano *et al.* used SVM for ATE and used word embedding centroids and the NRC lexicon for the emotion analysis [224]. Sirisha and Chandana produced an advanced model using aspect terms and sentences as input into a RoBerta and LSTM classifier to produce sentiment and emotion classification scores [255]. Jorvekar and Gangwar approach ABEA by using TF-IDF, NLP Features, and Co-relation features to extract features to classify the emotions with a Recurrent Neural Network (RNN) [151].

To summarize, there are a variety of approaches, most consisting of ML models or Lexicon-based approaches for emotion classification and aspect term extraction. A couple of researchers chose to focus on the emotion analysis portion and use an aspect word as input. There is so much to explore within ABEA. These papers are the first of many. We can see in Figure 2.8, the differences and similarities between ABSA and ABEA. Evidently, aspect term extraction is essential to improve ABEA as well.

Aspect-Based Emotion Analysis: an NLP technique that entails detecting emotions related to the aspect term in a text. The baseline subtasks of ABEA include (1) aspect-term extraction and (2) emotion analysis (or classification). The other subtasks of ABEA are (3) opinion-term extraction and (4) aspect category detection.

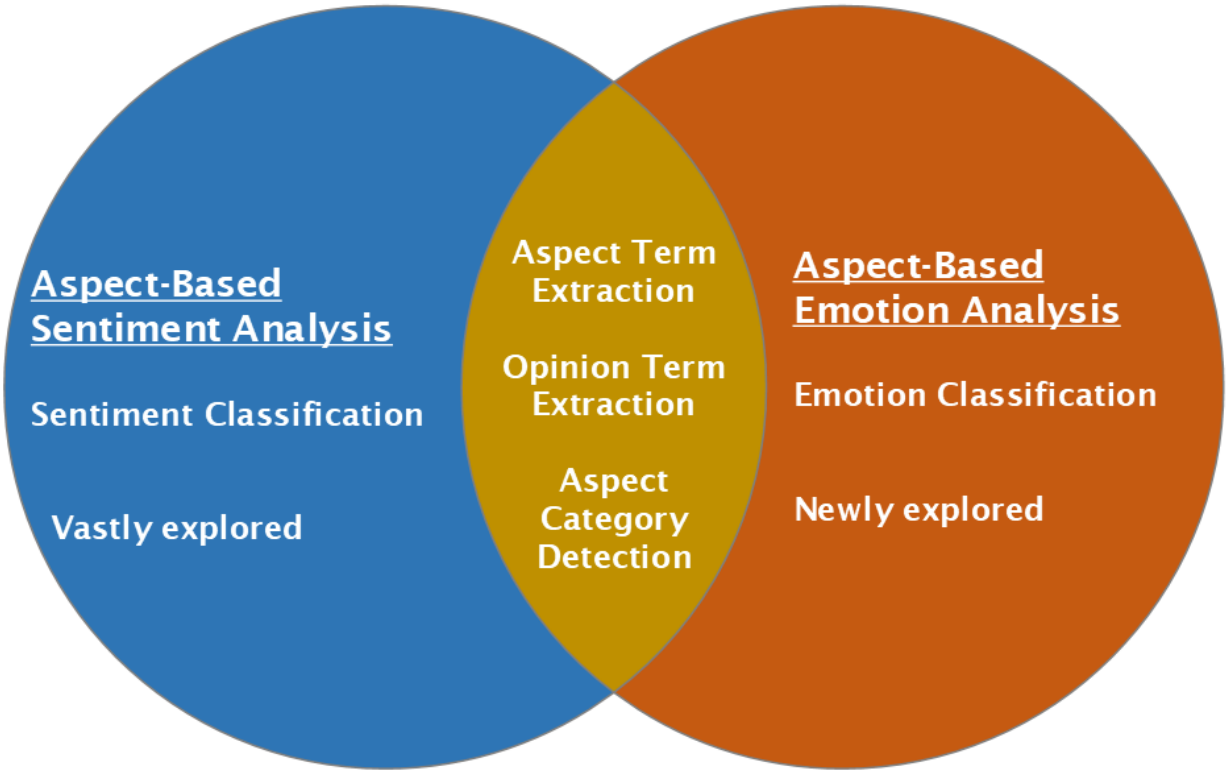

Figure 2.8: ABSA-ABEA Overlap

2.5 Summary

In this chapter, we were able to discuss how natural language processing led to aspect-based sentiment and emotion analysis. As online consumers grow, the need for detailed information increases along with it. Sentiment Analysis branched off into aspect-based sentiment analysis to provide more fine-grained details. ABSA provides information about the aspect term, sentiment, aspect category, and opinion term. Information is vital to companies improving their products. Companies use ABSA technology, like Repustate, to gather feedback from their users and customers. While ABSA provides great detail of the text, aspect-based emotion analysis provides even more. We believe that the aspect term extraction task is what gives a level of detail needed in the industry. In future chapters, we are excited to expound on this idea.

Chapter 3

Sentiment and Emotion Analysis on Social Media Data

COVID-19 has had a devastating impact not only on various health outcomes and the quality of life of individuals but also on their social well-being. Understanding the sentiment and emotions of people is an essential component for effective crisis management during health crises such as COVID-19. In this chapter, we perform a comparative study of longitudinal sentiment and emotion analysis of English and Spanish tweets pertaining to COVID-19. The aim of this study is to discover the varying sentiment and emotions of different languages by using transfer learning-based text analytic approaches. We verify the sentiments and emotions by uncovering major themes from tweets using topic modeling. Our modeling results are expected to offer deeper insights into various issues related to psychology, public health, and economics through social media interaction during COVID-19.

In this chapter, we offer modeling strategies and results that provide an understanding of the reactions of users toward COVID-19 in two different languages. The understanding of reactions in a crisis can be utilized by various stakeholders such as care providers, policymakers, and researchers. We then will explain our data and methodology in Section 3.2. Once we have an understanding of the process we will look at the results in Sections 3.3- 3.5 and then we will discuss our summary in Section 3.7. The related work of this chapter can be found in Section 2.

3.1 Problem Statement of Topic-Based Sentiment and Emotion Analysis

In this chapter, we will discuss our research exploring Sentiment and Emotion Analysis or *SEA*. We start this section by explaining why topic-based sentiment and emotion analysis

is critical in subsection 3.1.1. Then, we reveal our problem statements for this research in subsection 3.1.2

3.1.1 Why Sentiment and Emotion Analysis is Critical?

Sentiment and Emotion Analysis generates feedback from the population, a much-needed tool as social media continues to rise. For this research, we selected a multilingual dataset of COVID-19 tweets and used Machine Learning techniques to analyze that dataset. We desired to look at a dataset with a crisis because we knew there would be a wide range of emotions that we could detect. We are motivated to do this for three reasons: (1) COVID-19 is a crisis that has caused a range of opinions and opinions; however, we need to know which aspects cause those sentiments and emotions, (2) exploring multilingual data provides a larger view of the reactions in the world, (3) social media is a platform for people to express their feelings and opinions, making it a perfect resource for SEA. Performing topic-based sentiment and emotion analysis is desirable because it enables us to know the typical human responses so that we can adequately prepare for the future [45, 319]. This chapter can be helpful for government officials and medical professionals. Emotion analysis gives even more detailed information. Using these techniques during a global crisis is only logical. We desired to not only analyze English tweets but Spanish as well. We hoped to obtain a well-rounded analysis. We desire to see if the two language cultures have differing sentiments, emotions, and conversations. Research on topic-based sentiment and emotion analysis for non-English languages is challenging due to the lower resources. In order to advance research within SEA, we must continue to broaden beyond solely English data [79, 36]. Twitter is undoubtedly one of the most popular microblogging platforms. We opt for Tweets for our SEA because tweets typically portray the feelings or opinions of individuals [173].

3.1.2 Research Goals and Problem Statement

Research Goal. The overarching goal of this part of the dissertation study is two-fold. First, we discover the varying sentiment and emotions of different languages by using transfer learning-based text analytic approaches including BERT and BETO. Second, we offer deeper insights into various issues through the topics discovered related to psychology, public health, and economics through social media interaction during COVID-19.

More specifically, we devise a system to gather sentiments, emotions, and reactions toward a crisis for doctors, government officials, and researchers. This kind of information provides insight into public concerns which can be helpful for officials when designing and implementing guidelines for communities in a crisis situation.

Problem Statement. The main research challenges to be addressed in this part of the dissertation research are expressed in the format of the problem statement as follows:

1. *Research Problem 1.* How to design a model to detect topics, sentiments, and emotions from social media data?

2. *Research Problem 2.* Is it viable to devise a verification tool to link sentiment trends with news events?

3. *Research Problem 3.* What understandings behind the reactions of users towards COVID-19 can be derived from modeling results?

3.1.3 Research Milestones

We verify the sentiments and emotions by uncovering major themes from the English and Spanish tweets using topic modeling. The text analytic process includes various phases. There is a data processing phase, an LDA-based topic modeling phase, followed by a sentiment and emotion analysis phase. We found that the English speakers and Spanish speakers had drastically differing emotions towards COVID-19. We also found that the spikes in the sentiment and emotions data correlate to major events happening in the world. We show the

chronological development of conversation in Spanish and English around the pandemic and compare the types of conversations. Our analysis provides deeper insights into various issues related to psychology, public health, and economics through social media interaction during COVID-19. The knowledge discovery from our analysis can benefit health and government organizations by giving them insight into people's feelings, reactions, and opinions towards the Coronavirus and crises.

3.2 Research Methodology

In this section, we will explain the background details of this research. Our dataset in Subsection 3.2.1 and our modeling method in subsection 3.2.2.

3.2.1 Dataset

Our dataset collection originated from Banda from Georgia State University *et al.* [38]. They collected tweets using various keywords, which retrieved tweets since January 2020. The dataset currently includes the tweet identifiers, their respective language, and the location of all tweets. When we gathered the data, we gathered data for the entire year of 2020. We use this data for our research. We chose exclusively the year 2020 because it is a crisis year for everyone worldwide. Examining the entire year will give us a thorough understanding of what people truly feel and think. In order to construct our dataset, we took three steps:

1. We reduce the dataset to solely English and Spanish tweets
2. Verify that English tweets are from the U.S. and that Spanish tweets are from Spanish-speaking countries
3. Reduce to the same number of tweets from each week in the year
4. Normalize the number of tweets so that they are even between datasets

Due to our distribution verification for the tweets, our new dataset contains 150,000 English tweets and 150,000 Spanish tweets.

3.2.2 A Modeling Method

The process of our methodology contains four phases. First, we process the English and Spanish data separately. Secondly, we analyze the results and compare the separate datasets for our analysis of the two languages.

In our model, the effective tools we use are Latent Dirichlet Allocation and Pysentimiento. *Latent Dirichlet Allocation* (LDA) is a probabilistic model that analyzes words and then generates clusters of the words that are similarly associated [53]. The input for this model is clean text; it then outputs several clusters of topics found in the text. *Pysentimiento* is a multilingual Python toolkit for Sentiment and Emotion Analysis. It uses pre-trained Transformer models, like BERT, BERTweet, and BETO, to perform various SocialNLP tasks [229]. These models returned impressive results that assured us of their usefulness in our research. For example, BERTweet-based was the highest-performing model for English Sentiment and Emotion datasets, and BETO was the highest-performing model for Spanish text.

Furthermore, BETO is a BERT model trained on a Spanish dataset in order to perform streamlined NLP tasks. The Spanish dataset used contains 300 million lines, 3 billion tokens, and 18 billion characters. It performs better than the Best Multilingual BERT in most NLP tasks [63]. Through BETO, Pysentimiento is a valuable toolkit for sentiment and emotion classification tasks in English and Spanish texts. The emotion classification portion categorizes data using Ekman's six basic emotions: anger, disgust, fear, joy, sadness, and surprise [100]. We choose this impressive toolkit because it contains what other methods currently lack, the ability to tackle more than one language for sentiment and emotion analysis. Previously, we explored VADER and TextBlob, which are two of the leading libraries for sentiment analysis and NLP tasks [140, 56]. *VADER* is an English lexicon and

rule-based sentiment analysis tool. *TextBlob* is a library used for many primarily English-based NLP tasks. However, the tools were deficient in classifying sentiment and emotion for Spanish data. We believe the Pysentimiento tool to be a more advanced and beneficial tool for our research. BETO provides us with a more advanced method for gathering accurate opinions and feelings about the beginning of the COVID-19 pandemic.

In Section 3.2.1 we cover the data preparation. We then move the tweets to the Topic Modeling process. We utilize *Latent Dirichlet Allocation* (LDA) for topic modeling to infer which topics were most popular among the COVID-19 tweets. For our research, we selected bigrams and trigrams in order to grasp a full understanding of the prominent topics. By selecting bigrams and trigrams, the model analyzes the words in our input dataset by bigrams and trigrams. The model then returns clusters of similar bigrams and trigrams. We choose not to use the Bag of Words method in order to have a complete view of users' conversations. From those topics, we extract themes that are the overall analysis of conversation from a topic.

Next, we move into the SEA, where we gathered an overall analysis and then a subliminal analysis of topics. First, we gather an overall SEA of the English and Spanish datasets. We can see the range of emotions and their fluctuations throughout 2020. Secondly, we divide each dataset into topic clusters and perform emotion analysis on each topic cluster. The results show which emotions users have towards a topic. Our script utilizes the Pysentimiento library to process each English and Spanish dataset's tweets. The script then clusters the tweets into positive, negative, and emotion clusters. The emotion clusters correspond with Ekman's six basic emotions: anger, disgust, fear, joy, sadness, and surprise.

The final stage of our model is the analysis of the two languages. We analyze and compare the themes, sentiments, and emotions that we gather in this stage. The comparison allows us to make inferences about the similarities and differences in the results.

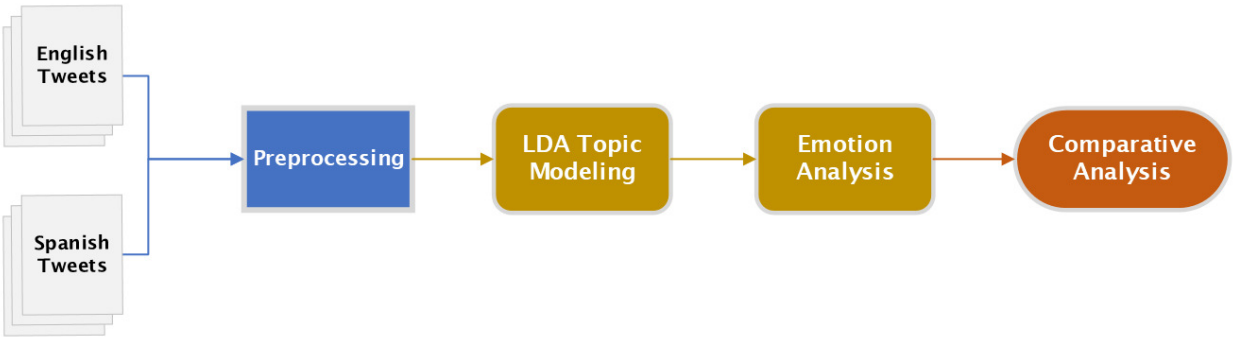

Figure 3.1: Language Analysis Process

3.3 English Analysis

After the data preparation and topic modeling procedures followed by a sentiment and emotion analysis, we thoroughly evaluate the results obtained from tweets written in English. The sentiment line charts presented in Section 3.3.1 offer us a visual of how the sentiment fluctuated throughout the year. The emotion bar charts depicted in Section 3.3.2 show which emotion most people were feeling. We extract an array of themes (see Section 3.3.3) and made various charts to present and understand the trends from our empirical study. The themes give us an explanation to validate the emotion and sentiment.

3.3.1 English Sentiment

We use line charts to display the sentiment over the year. In the three sentiment charts, we plot the number of positive tweets per day and the number of negative tweets per day throughout the year 2020. From the subfigures in Figures 3.2- 3.4, our first overall observation is that the negative sentiment far exceeds the positive sentiment throughout 2020. We infer that many people felt more negatively than positively about the global pandemic. As the year continues, we see the negative sentiment increasing. There is a significant spike during April and a continuation of spikes throughout the summer months. We validate the correctness of our results by relating the spikes with any significant events or changes during the year. We can see in Table 3.1, several events align with tweets in our dataset, validating the spike we

see. For example, on April 3, it was announced that the global number of coronavirus cases had surpassed 1 million. According to Johns Hopkins, nearly 53,000 people had died from the virus.

We deduce that this event could be a reason for the first spike in April. We notice that most of the spikes are around the beginning of the month. We grasp from the tweets that conversations about a vaccine in the later months are possibly the reason for the decrease in negative sentiment, narrowing the gap between the negative and positive sentiment. We conclude that the smaller the gap between the positive and negative, the less extreme the circumstances where more people are calming down.

Throughout the Spring and Summer, the discrepancy between negative and positive sentiment increased due to the escalating events. In contrast, such a negative-positive sentiment gap reduced in the Fall and Winter as talk of the vaccine surfaced. In a nutshell, we propose validating the sentiment analysis's correctness by aligning spikes with significant events. This proposed method is feasible because our findings demonstrate that we are able to match each spike with a significant event as a rationale behind the sentiment trending.

Month	**Event—Milestone**	**Tweet**
February	Spread of Covid in US First coronavirus death in Washington	*Coronavirus has been Declared a Global Emergency by WHO*
April	Global coronavirus cases surpasses 1 million Nearly 53,000 people had died from the virus	*We are staying home and its still increasing*
June	The WHO announced that all people should wear masks in public	*WHO waited until June 5 to inform the world that masks are recommended*
August	Several countries exceeded 1 or 2 million Covid cases	*The number of people infected with the Coronavirus across the world surpassed 17 million*
September	Book published revealing President Trump's knowledge of the deadly virus early on	*realDonaldTrump 187000 dead innocent Americans lives lost from Coronavirus and your gross incompetence and inaction*
November	Election Month	*Don't forget though He ruined the greatest opportunity to guarantee his reelection too. He could have rallied the country and fought Coronavirus with all the resources and expertise needed to prevent the spread of the virus And he chose not to. Not a leader VoteHimOut*

Table 3.1: English Tweets 2020 Events

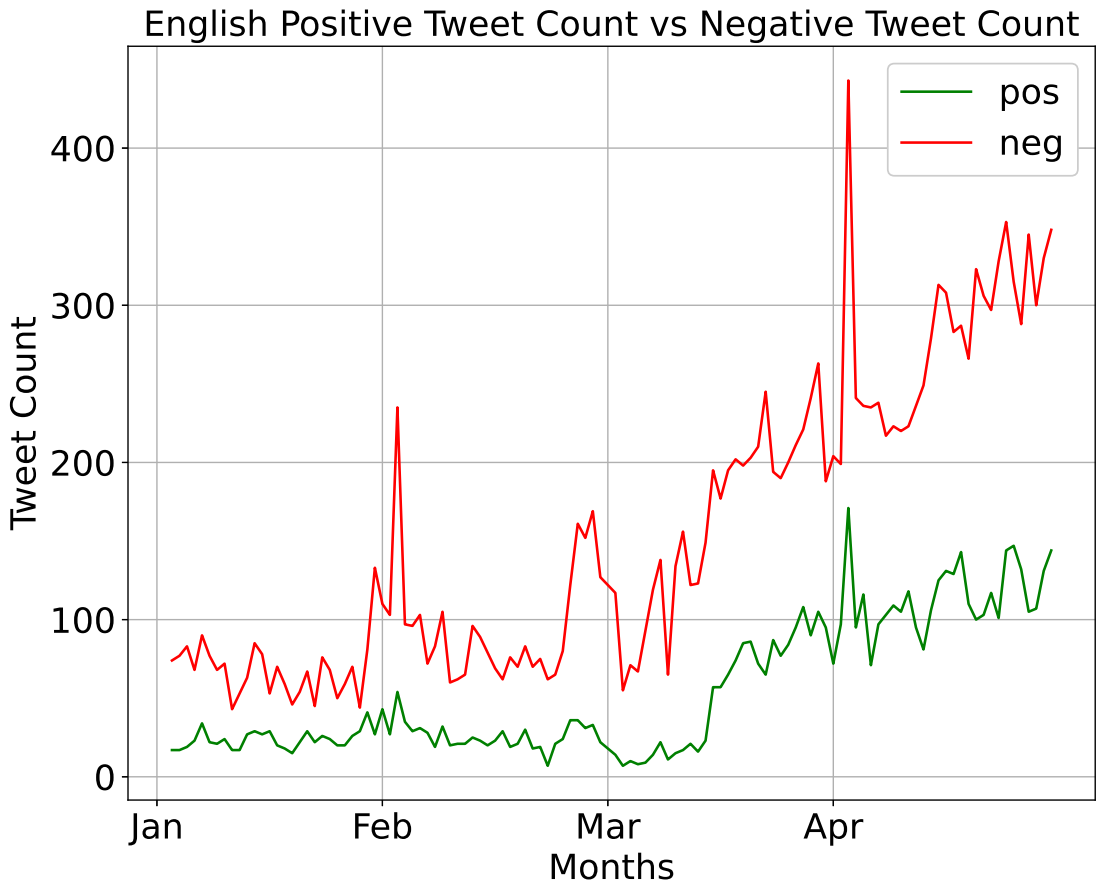

Figure 3.2: English Jan-Apr Sentiment Chart

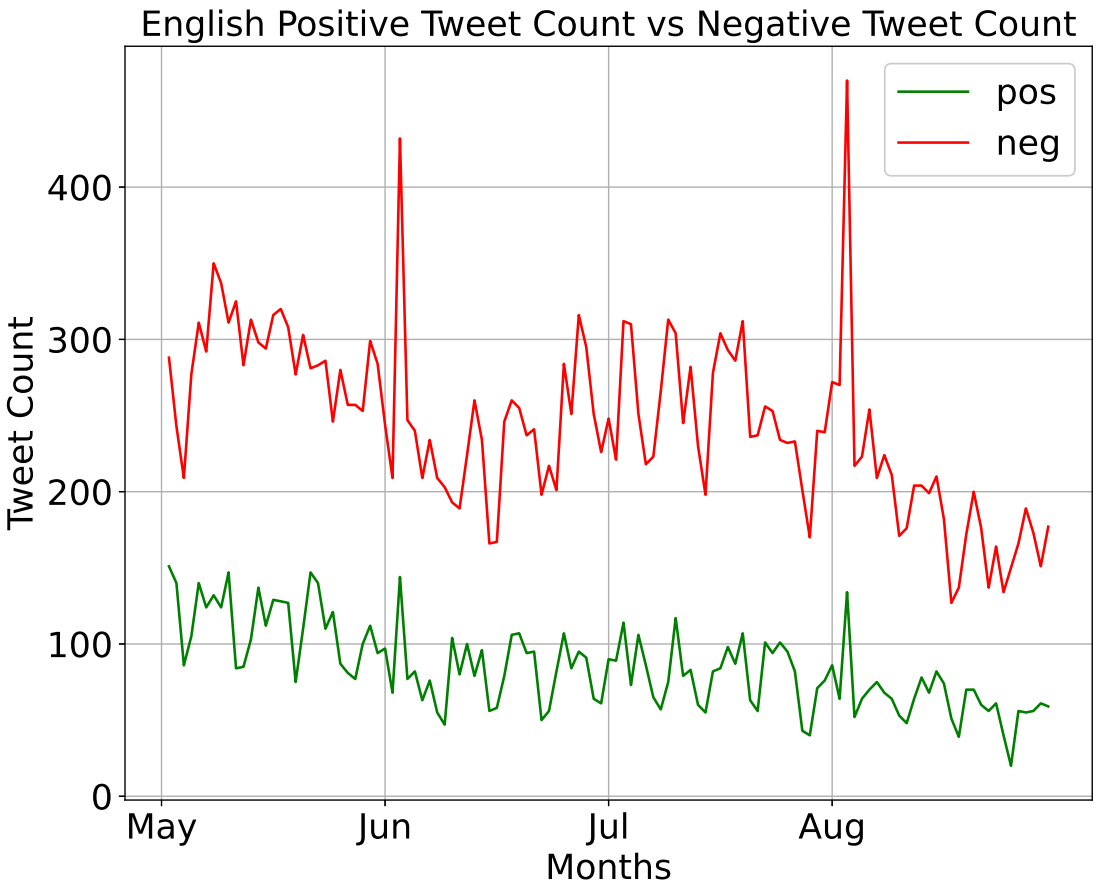

Figure 3.3: English May-Aug Sentiment Chart

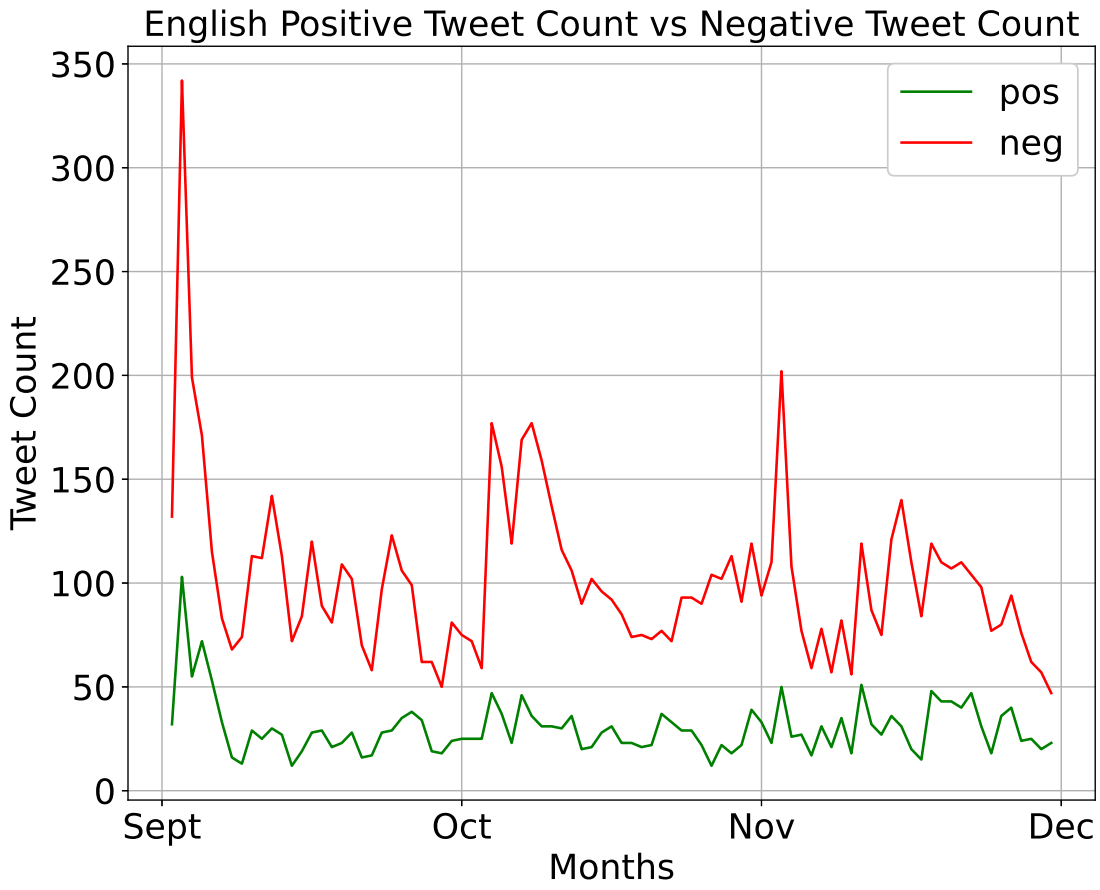

Figure 3.4: English Sept-Dec Sentiment Chart

3.3.2 English Emotions

The line charts plotted in Figures 3.5-3.7 show the range of emotions that people felt in the different time frames of the year. In the results of the English emotions, we discover that throughout the entire year, fear and disgust are consistently the highest emotions among peers. This pattern correlates to the trend observed from the sentiment results depicted in Figures 3.2-3.4. Recall that (see Section 3.3.1) the negative sentiment was drastically higher than the positive sentiment throughout the year. We now see that the negative sentiment correlate to the fear and disgust emotions captured in Figures 3.5-3.7). The emotions heighten the most during the summer months and mellow down as the year ends. During the summer of 2020, the number of COVID-19 cases significantly increased, which justifies the surge of emotion. Similar to Section 3.3.1, the tweets reveal that the discussion surrounding the vaccine from September to December is what led to the decrease in intense emotions. Surprisingly, we find that joy exceeds sadness and anger the majority of the time. We conclude that these results show that people can still find joy and hope during a challenging time.

3.3.3 English Themes

Utilizing LDA, we categorize the tweets into five topics using bigrams and trigrams to develop those topic clusters. Next, we infer an array of themes from those topics. Table 3.2 summarizes a list of the most similar word tokens and deduced themes from the tokens. The themes gleaned from January to April mainly discuss:

1. The long-term effects of COVID-19.

2. The ways people are fighting COVID-19.

3. The news contributes to the reactions to the pandemic.

The themes of May to August largely consist of:

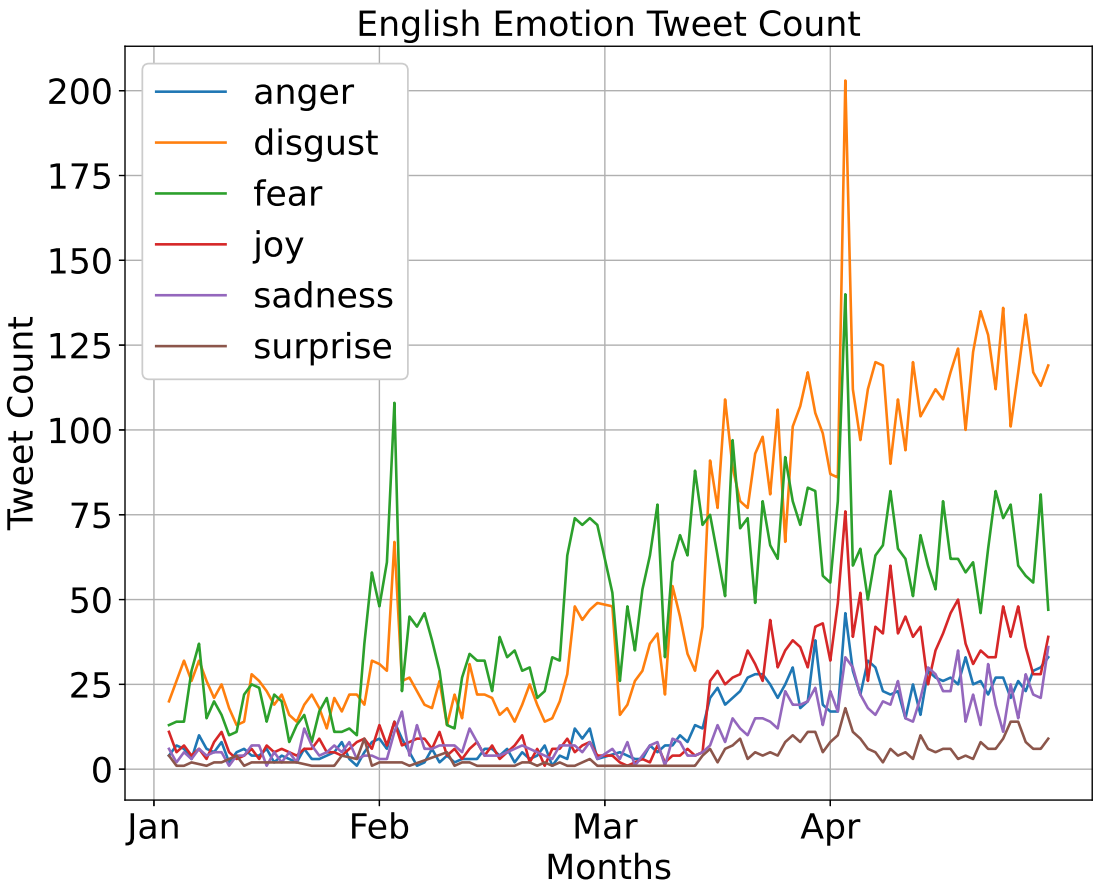

Figure 3.5: English Jan-Apr Emotion Chart

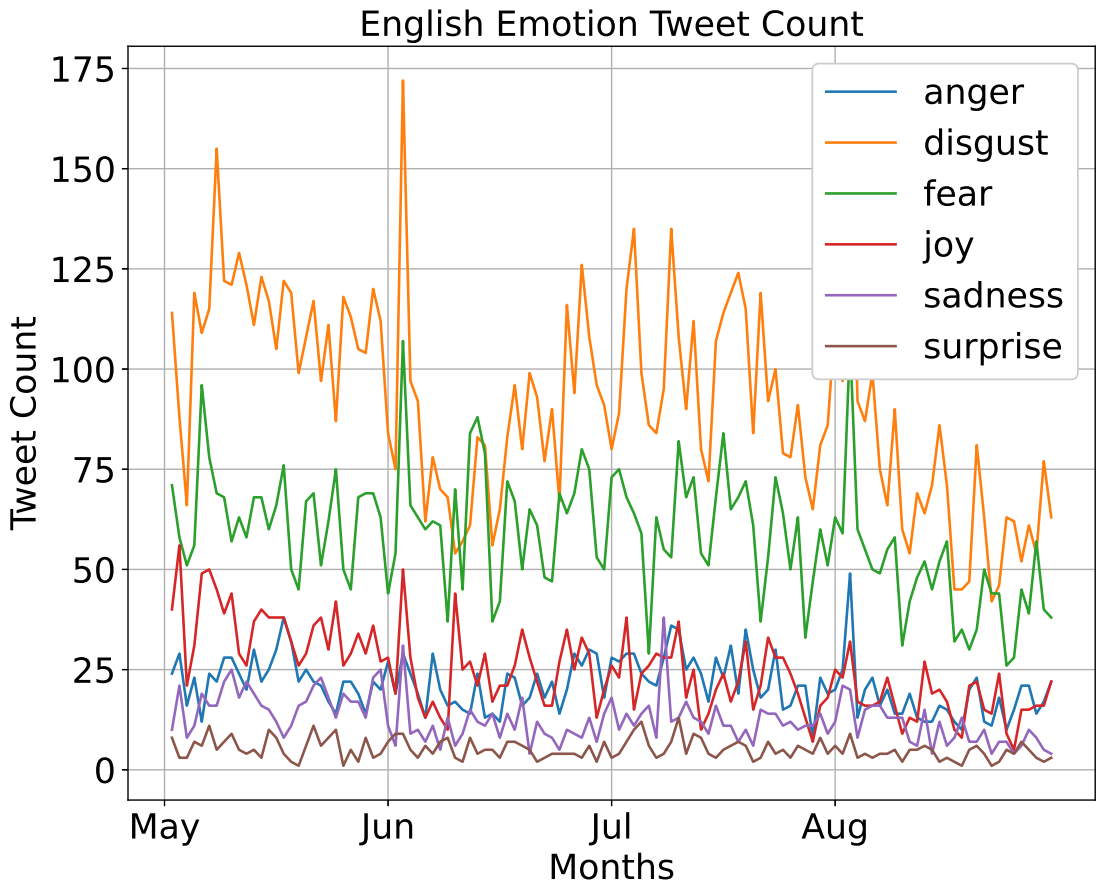

Figure 3.6: English May-Aug Emotion Chart

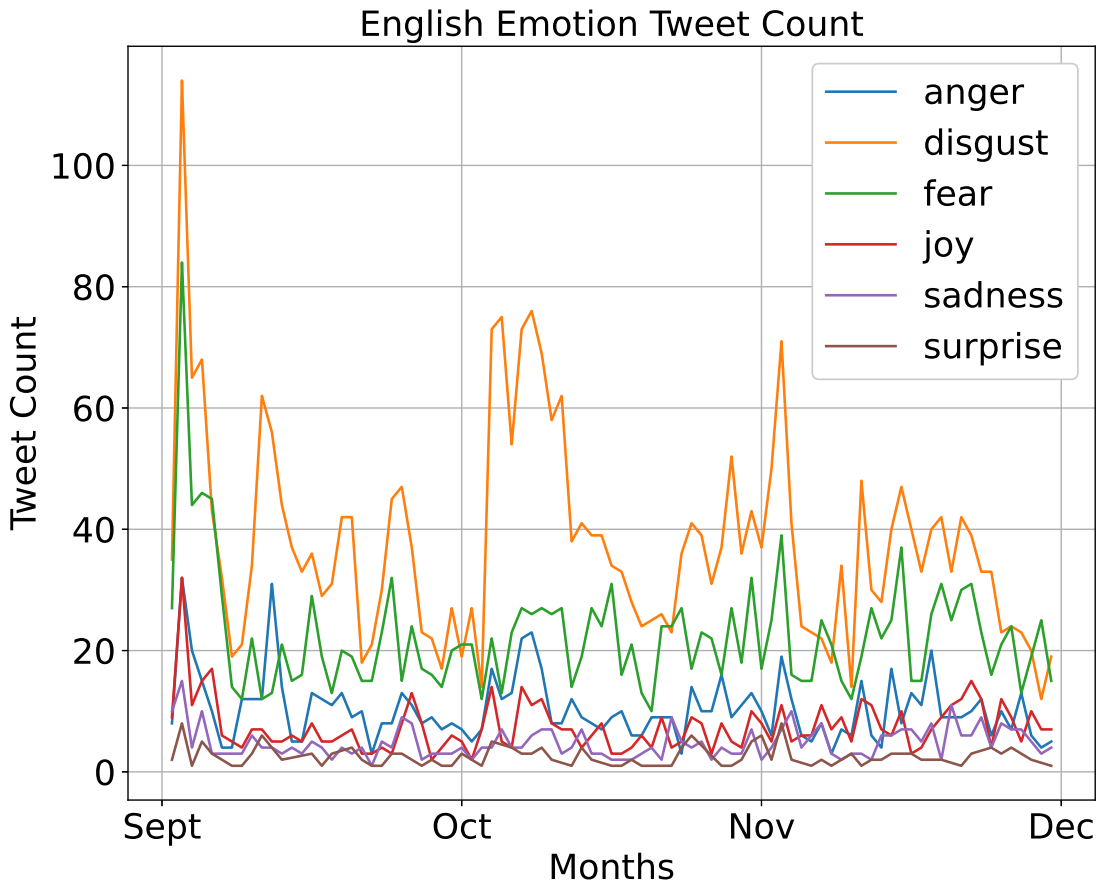

Figure 3.7: English Sept-Dec Emotion Chart

1. The worldwide reality of this outbreak.

2. The tasks of health officials.

3. Constant news updates.

The months of September through December consider themes concerning:

1. The government's involvement.

2. The ultimate impacts of COVID-19.

3. The number of recovered patients.

There were three main themes that we see consistent themes throughout 2020. The first theme we observe is the impacts and long-term effects of COVID-19 on the world. The topics show that COVID-19 impacted the population, the economy, and numerous countries and created a new normal. COVID-19 has changed the world. The results unveil that a second theme that appears throughout the year is people's concern for accurate news. We notice the concerns about the trustworthiness of the news sources. False information during a global pandemic can be damaging; therefore, the news source concern was a vital topic. The third theme present throughout the year is the preventative measures to fight COVID-19. The pandemic brought about wearing masks, washing hands, social distancing, and the lockdown. Health officials and the government created these methods to prevent the spread. These methods, along with the death toll, are what created a new normal for our world. One of the impacts of the lockdown is mental health challenges. It is evident from the analysis that people were generally concerned with COVID-19's impact at the beginning of the year. However, as the year went on, people were more aware that this was a global issue and the need for health officials worldwide to be united in fighting the Coronavirus. Near the end of the year, people began discussing recovered patients.

In Table 3.5, we see that throughout the various topics discussed during the year, the primary emotion that English speakers felt is fear. In the table, the *# Tweets* represents

the number of tweets labeled as fear. The *% Tweets* represents the number of tweets out of the entire tweet counts for that part of the year. We observe that many of the tweets are not labeled with emotion. Fear is the highest labeled emotion, containing no more than 13% of the entire English dataset. Overall, we conclude that the English themes have some consistent patterns throughout the year, but as time continues, people gain more insight into how the virus has changed the world.

ID	**Key Terms**	**Theme Deduction**	**Tokens**
1	- social distancing, stay safe, covid vaccines - news coronavirus, bbc news, fake news, covid update	Safety Measures News Importance	46.6%
2	- coronavirus outbreak, china coronavirus, covid crisis - covid vaccine, wash hand, fight covid, face masks - tested positive, spread covid, mental health	Outbreak Crisis Fighting Covid Affecting Health	53.3%
3	- health care, task force, stop spread - small businesses, around world, many people	Preventative Strategy World Effects	56.6%
4	- cruise ship, small business, confirm cases - public health, immune system, covid relief	Ripple Effects Health Concerns	50%
5	- white house, united states, president trump - health organization, world health, health workers	Government Involved Global Health	36.6%

Table 3.2: English Themes: January to April

ID	**Key Terms**	**Theme Deduction**	**Tokens**
1	- social distancing, contact tracing, response covid - south korea, hong kong	Spread Control Outbreak Worldwide	23.3%
2	- covid crisis, impact covid, around world - public health, wear masks, save lives, loved ones	Crisis Impact Safety Precautions	50%
3	- health officials, covid patients, confirmed covid	Managing Patients	33.3%
4	- news coronavirus, white house, bbc news - nursing homes, spread coronavirus, covid lockdown - back school, new normal	News Updates Dangerous Effects Life Adjusting	43.3%
5	- covid response, long term, covid vaccine - good news, fake news, new york times	Future Effects News Importance	30%

Table 3.3: English Themes: May to August

3.4 Spanish Analysis

Now we are positioned to analyze the Spanish data, aiming to compare the sentiment differences among tweets written in two languages (see also Section 3.5). Section 3.4.1 dives into the sentiment analysis of the Spanish tweets posted throughout the year. Next, the

ID	**Key Terms**	**Theme Deduction**	**Tokens**
1	- white house, joe biden, supreme court - health officials, public health, around world	Government Concern Health Updates	23.3%
2	- second wave, covid infection, young people - covid update, via yahoo, bbc news	New Effects Online Updates	26.6%
3	- coronavirus task force, fight covid, contact tracing - american people, loved ones, people die	Fighting Covid American Deaths	60%
4	- donald trump, trump administration, covid response - herd immunity, social distancing, covid restrictions - new york times, washington post, good news	Presidential Response Virus Protection News Sources	43.3%
5	- mental health, impact covid, new normal - americans died, death toll, death population - cases death recovered, death recovered, recovery rate	Covid Changes Increase Death Recovery Impact	56.6%

Table 3.4: English Themes: September to December

Months	**Top Emotion**	**# Tweets**	**% Tweets**
Jan-Apr	Fear	2512	4.6%
May-Aug	Fear	2983	3.8%
Sept-Dec	Fear	941	4.4%

Table 3.5: English Top Theme Emotions

Spanish emotion analysis is in Section 3.4.2. Finally, Section 3.4.3 discusses a set of themes inferred from the topics.

3.4.1 Spanish Sentiment

The line charts display the sentiments drawn from tweets written by the Spanish speakers in Figures 3.8- 3.10. In the January to April chart, plotted in Figure 3.8, we notice that the negative sentiment far exceeds the positive sentiment. Interestingly, the gap between the negative and positive sentiments significantly widens over the summer months (May to August); the gap then reduces during the end of the year (September to December). Overall we can see that consistently throughout 2020, positivity was very low within the Spanish-speaker community.

Beginning in February, we see that the negative sentiment trend drastically fluctuates. In the following months, the negative sentiment surges as the pandemic escalates. An intriguing observation is that the spikes occur at the beginning of each month. We speculate that these spikes correspond with events during that time. We investigate the users' countries

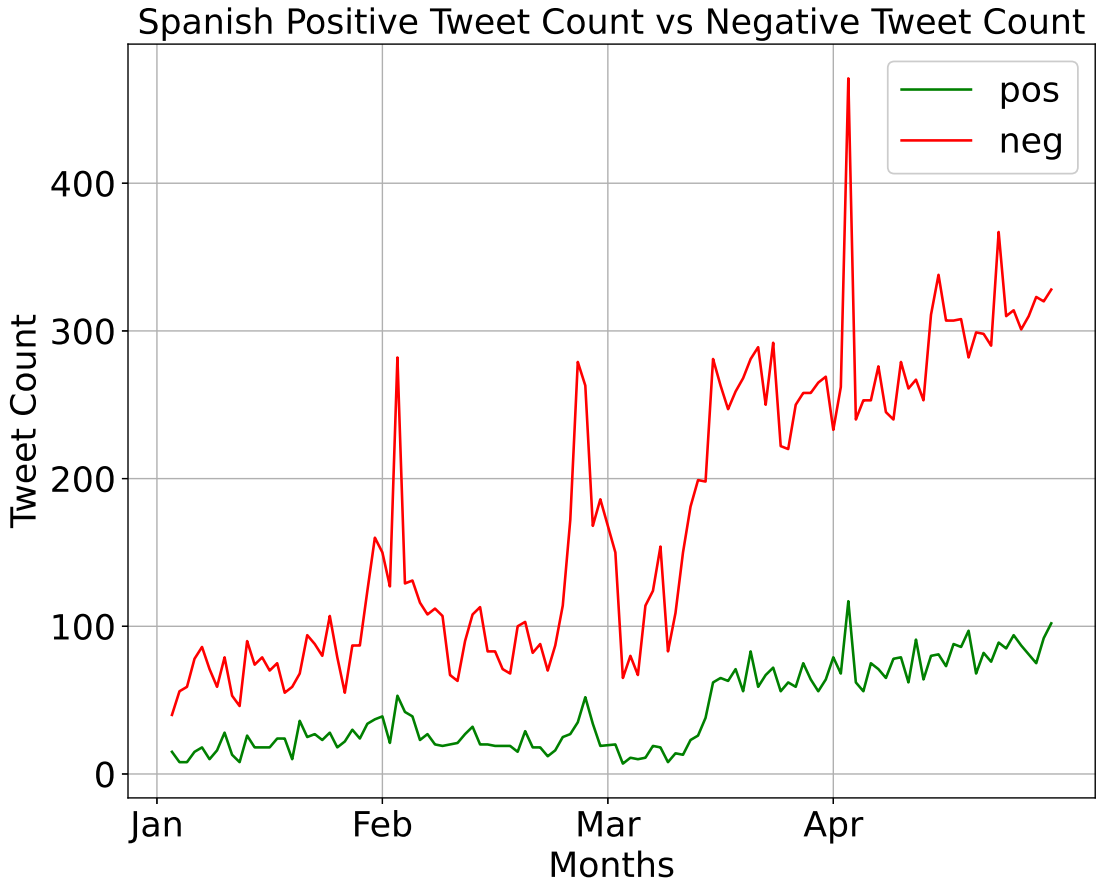

Figure 3.8: Spanish Jan-Apr Sentiment Chart

and consider the events during that time that could potentially correspond to the spikes. Table 3.6 shows the top five labeled countries within the dataset, which we rely on to gather event information during the spiked months. Table 3.7 displays the events or milestones of those months. We then analyze a set of tweets during the time of the spikes, which we compare to the events. We confirm that the spikes match the events of the Spanish-speaking countries. Furthermore, the spike dates correspond with the events.

3.4.2 Spanish Emotions

Figures 3.11-3.13 depict the fluctuation of the emotions within the Spanish-speaker community throughout the year. The emotions charts are very telling for Spanish speakers.

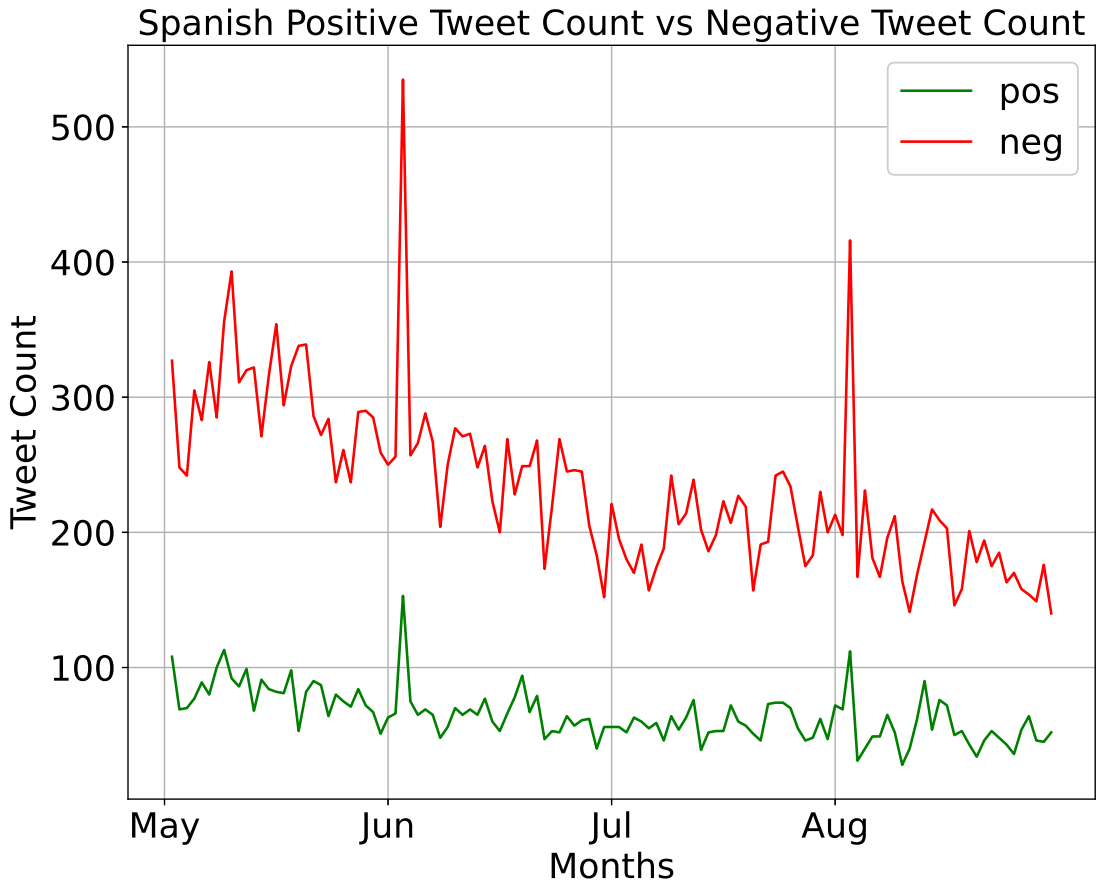

Figure 3.9: Spanish May-Aug Sentiment Chart

#	**Country**
1	Mexico
2	Venezuela
3	Spain
4	Argentina
5	Columbia

Table 3.6: Top 5 Spanish-speaking Countries in Dataset

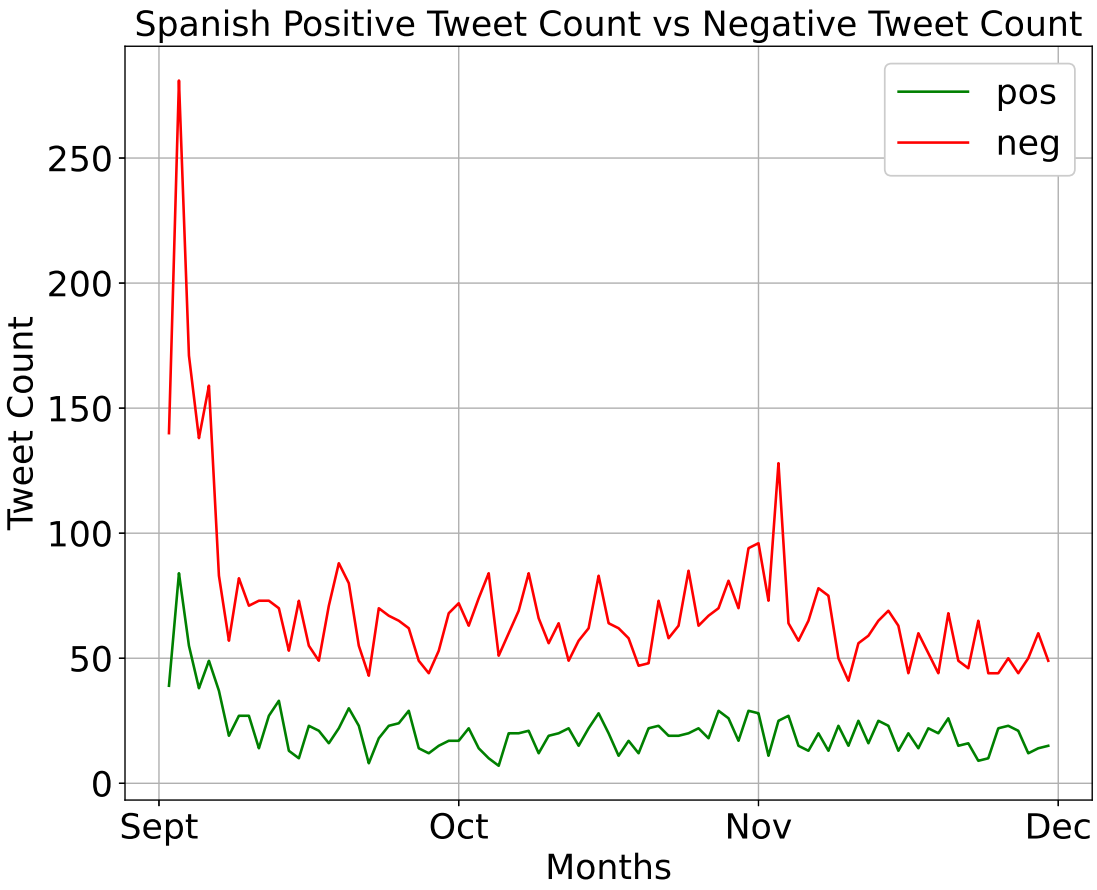

Figure 3.10: Spanish Sept-Dec Sentiment Chart

Month	**Event—Milestone**	**Tweet**
February	First case of Covid-19 in Mexico brought by a Mexican tourist that visited Italy weeks earlier	*Mood me siento mal y me pongo a rezar pensando que sea que el coronavirus nos alcanzó aquí*
March	First Covid death in Mexico city Enters Phase 2: Lockdown Latin American stock markets fall more than 20% throughout the month	*Los más triste de la cuarentena es no ver a mi amorcito por un chingo de tiempo te odio coronavirus*
April	Health emergency because of accelerated cases Enters Phase 3	*La gobernadora de Sonora Claudia Artemisa Pavlovich planteó a las autoridades federales su intención de restringir el acceso a la entidad por la frontera norte con EEUU ante la emergencia sanitaria causada por el virus*
June	1.6 million cases in Latin America	*40 millones han perdido sus trabajos 100000 están muertos por coronavirus Millones pasan hambre*
August	Mexico is the 3rd country with the most deaths by Covid	*coronavirus Las noticias en el mundo son desalentadoras nos estamos acercando a los 700000 muertos no se ha superado sino que prácticamente va y vuelve en todos los países*
September	Unemployment around 149 million jobs lost in the region	*COVID19 Proteger a los trabajadores en el lugar de trabajo La COVID19 provoca una inmensa pérdida de ingresos provenientes del trabajo en todo el mundo OIT*

Table 3.7: Spanish Tweets 2020 Events

All of the emotions are at their highest during the summer months. However, anger is consistently the highest felt emotion throughout the year. Similar to the sentiment results plotted in Figures 3.8-3.10, the emotions had significant peaks at the beginning of the months. We discover that there was a significant continuous peaking of emotion during the first third of the year between March and April, especially anger.

Similar to the English emotion analysis elaborated in Section 3.3.2, Table 3.8 reveals that throughout the various topics discussed during the year, the main emotion that Spanish speakers felt is *surprise*. In the table, the *# Tweets* represents the number of tweets labeled as *surprise* or *anger*. The *% Tweets* represents the number of tweets out of the entire tweet counts for that part of the year. We observe that many tweets are not labeled with an emotion. *Surprise* is the highest labeled theme emotion, which contains no more than 3% of the entire Spanish dataset.

3.4.3 Spanish Themes

Using LDA's Topic Modeling, as discussed in Chapter 2, we retrieve five topics. From those topics, we infer informative themes. The themes are translated from Spanish to English for this research. Within the first third of the year, most Spanish speakers' conversations revolved around topics about *combating Covid with sanitation, vaccination, and cures*. Another main conversation was *Covid's impact worldwide*. During the second third of the year, conversations surround *the second wave of the pandemic, Spanish-speaking countries affected, and the new normal brought by Covid*. With the final third of the year, the conversations revolve around *vaccine trials worldwide, the government's involvement, and patient recovery*.

We notice some contrasts in themes and some similarities throughout the year. The similar themes relate to *the rise in Covid cases and deaths, the government's involvement, and world impact*. We see more conversations on how to protect people during the beginning of the year, and at the end of the year, the conversations focused on recovery and vaccinations.

Months	**Top Emotion**	**# Tweets**	**% Tweets**
Jan-Apr	Anger	151	.24%
May-Aug	Surprise	722	.85%
Sept-Dec	Surprise	309	1.5%

Table 3.8: Spanish Top Theme Emotions

ID	**Key Terms**	**Theme Deduction**	**Tokens**
1	- ministerio salud, personal salud, salud publica	Health Concerns	36.6%
	- positivos covid, mas casos, mas muertos	Many Positive	
2	- coronavirus espana, coronavirus mexico	Affecting Countries	30%
	- emergencia sanitaria, crisis sanitaria, combatir covid	Health Fight	
3	- muertes covid, epidemia covid, simiedo coronavirus	Pandemic Fear	53.3%
	- stema salud, frente coronavirus, prevenir coronavirus	Prevent Spread	
	- redes sociales, firma peticion, peticion via	Social Petitions	
4	- mundo coronavirus, bbc news, news mundo	Worldwide News	53.3%
	- medidas preventias, cura coronavirus	Fighting Covid	
	- contagiados nuevos muertes, muertes muertes hoy	Death Rates	
5	- vacunas covid, lucha coronavirus, mata coronavirus	Fight Coronavirus	56.6%
	- mas mil, fallecidos coronavirus, adultos mayores	Older People	
	- america latina, coronavirus argentina, corea sur	World Impacted	

Table 3.9: Spanish Themes: January to April

ID	**Key Terms**	**Theme Deduction**	**Tokens**
1	- nuevos casos coronavirus, positivo covid, aumento casos	Cases Growth	56.6%
	- fallecidos covid, contagio covid, pacientes covid	New Deaths	
2	- casos confirmados, casos nuevos, mil casos	Increasing Cases	40%
	- muertes coronavirus, nuevas muertes, mil muertos	Increasing Deaths	
3	- coronavirus argentina, coronavirus mundo, coronavirus espana	Spanish Countries	46.6%
	- mil muertes, mas millones, brote coronavirus	Millions Affected	
4	- plena pandemia, crisis coronavirus, segunda ola	Second Wave	46.6%
	- prueba coronavirus, sistema salud, vacunas covid, coronavirus gobierno	Health Decisions	
5	- nueva normalidad, nuevo coronavirus, mundo coronavirus	World Normal	60%
	- news mundo, bbc news, redes sociales	various news	
	- crisis sanitaria, secretaria salud, evitar propagacion	Prevent Spread	

Table 3.10: Spanish Themes: May to August

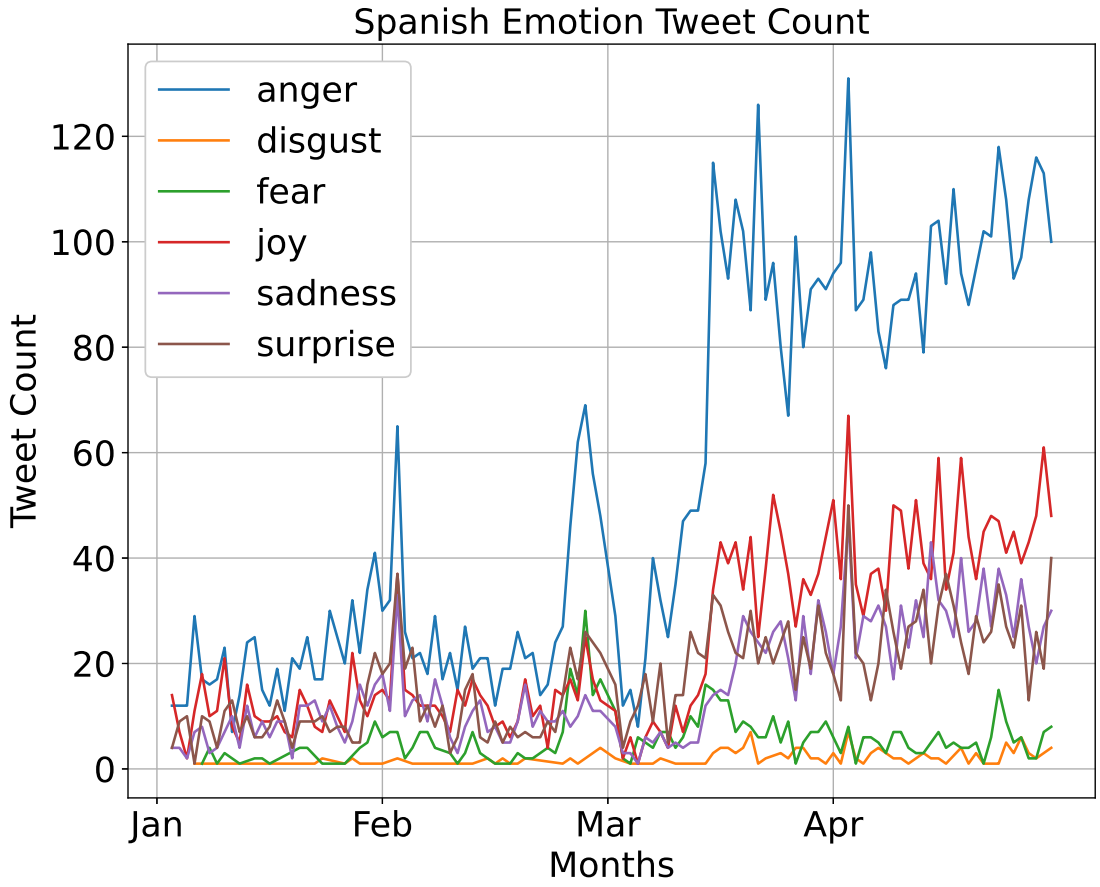

Figure 3.11: Spanish Jan-Apr Emotions Graph

3.5 Language Analysis

3.6 Sentiment Analysis Between the English and Spanish Speakers

Here we compare the sentiment between the English and Spanish speakers to evaluate if people from the different groups feel differently throughout the Covid year. We attempt to justify the rationale behind such differences. We see in Figure 3.14 that the positive sentiment was much higher among English speakers than it was among Spanish speakers. The speakers felt the most positive during the months of April to June. As the year continued, the positivity slowly declines. We examine the tweets posted from April to June to gauge what people feel positive towards. However, Figure 3.15 unveils that the negative sentiments

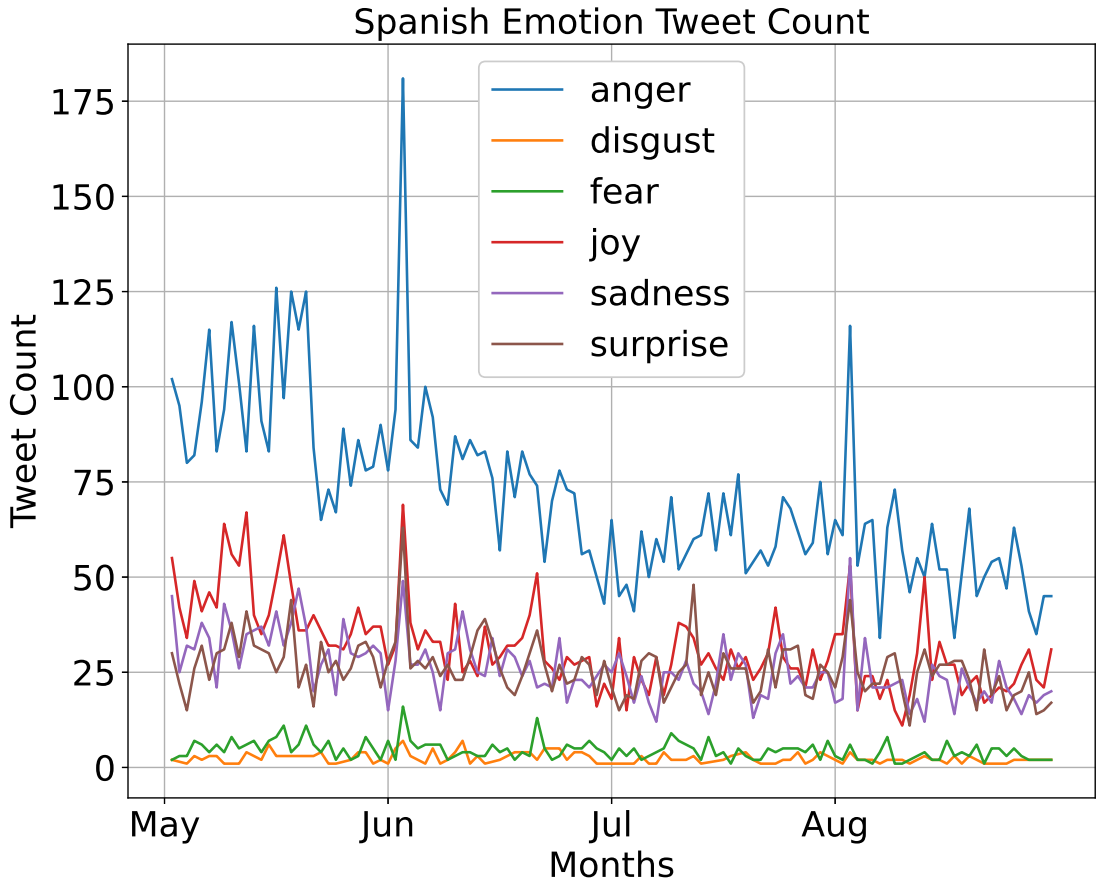

Figure 3.12: Spanish May-Aug Emotions Chart

between English and Spanish speakers almost share an identical trend. Similar to the positive comparison, the speakers felt more sentiment around the months of April to June. It is evident that there was a topic of conversation that people felt more intensely about; this topic of conversation became a driving force behind a growing number of tweets of sentiment around those times.

We infer that Spanish speakers responded with an overall more negative sentiment than the English speakers because the Spanish speakers have a lower positive sentiment and a high negative sentiment. In contrast, English speakers have a high positive sentiment and a high negative sentiment. Both speakers had the same negative feelings about the global

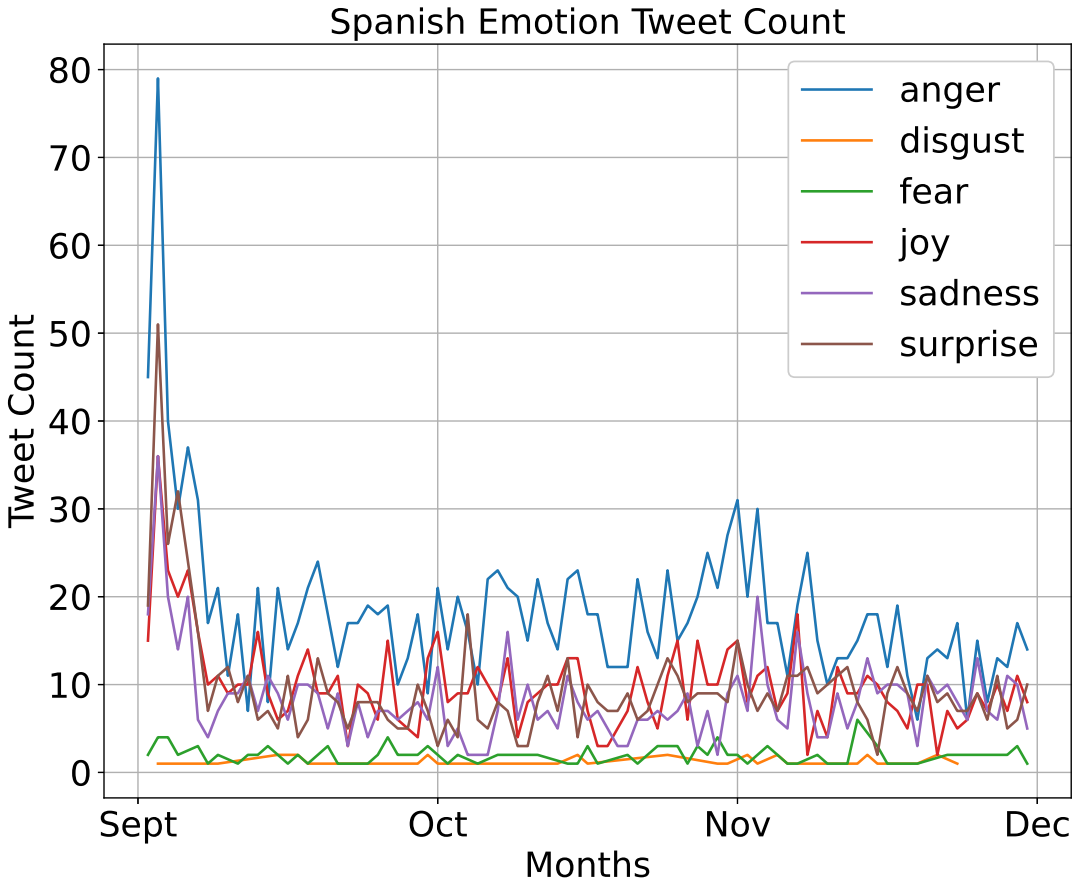

Figure 3.13: Spanish Sept-Dec Emotions Chart

pandemic throughout the year. We believe that the sentiment decreased during the end of the year because people were coming to terms with the new normal.

3.6.1 Emotion Analysis Between the English and Spanish Speakers

We observe from the emotion figures (see Figs. 3.16 and 3.17) that English speakers felt more fear and disgust; however, Spanish speakers felt more anger and joy. The results signify that the highest English emotion is significantly higher than the highest Spanish emotion throughout the year. Therefore, we can infer that English speakers experienced more intensity with their top emotions than Spanish speakers. Figure 3.16 unveils that fear and disgust for the English speakers are consistently close in intensity; in contrast, there is

ID	Key Terms	Theme Deduction	Tokens
1	- nuevos casos, nmas informacion, reporta nuevos - salud publica, ensayos vacuna, comunidad madrid, nivel mudial - ultimas horas, ultimos dias, muertos coronavirus	Information updates World Vaccine Final Moments	53.3%
2	- nuevos casos, casos activos, casos fallecidos - pacientes covid, medidads sanitarias, medidas prevencion	Many Cases Protect Patients	53.3%
3	- covid mexico, coronavirus argentina, coronavirus espana, - donald trump, secretaria salud, personas fallecidas	Spanish Impact Government Tracking	40%
4	- segunda ola, mas muertos, aumento casos - vacunas covid, ensayos clinicos, nivel nacional	Second Wave Effects Trials Developing	43.3%
5	- medidas preventivas, vacuna coronavirus, personas recuperadas - ola covid, coornavirus aire, mas contagios	New Recoveries Airborne Covid	60%

Table 3.11: Spanish Themes: September to December

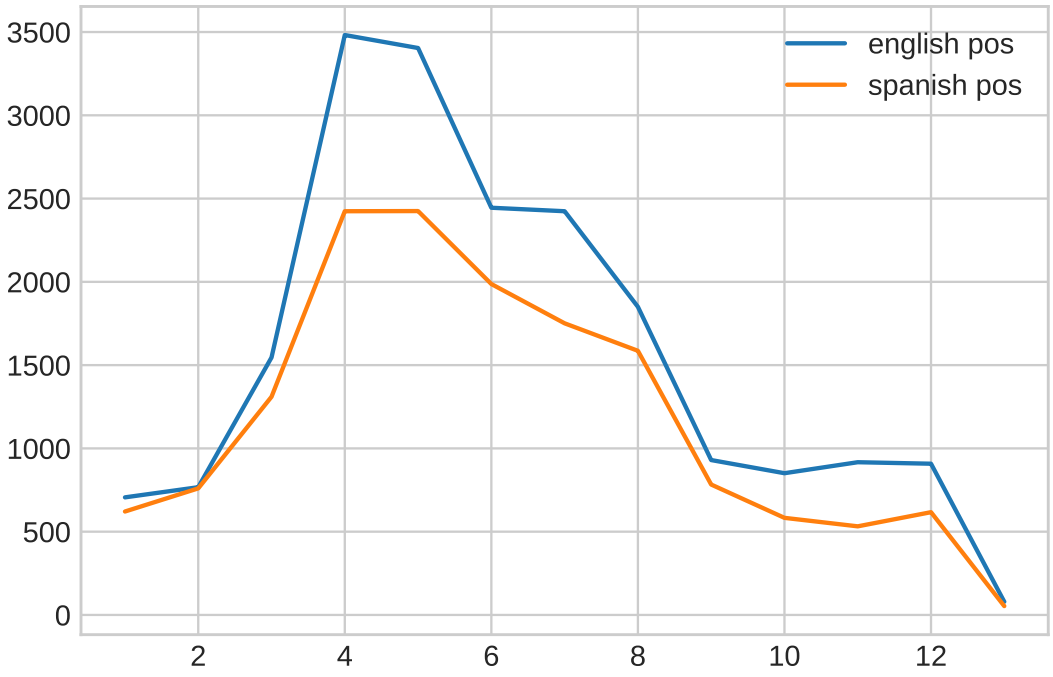

Figure 3.14: Language Positive Comparison Chart

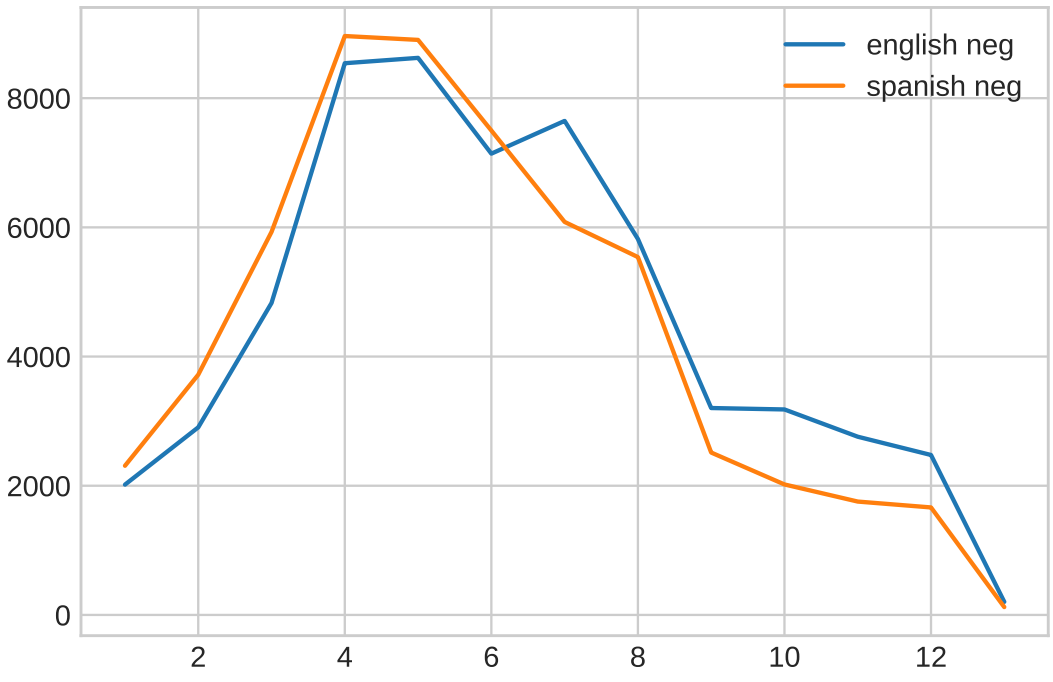

Figure 3.15: Language Negative Comparison Chart

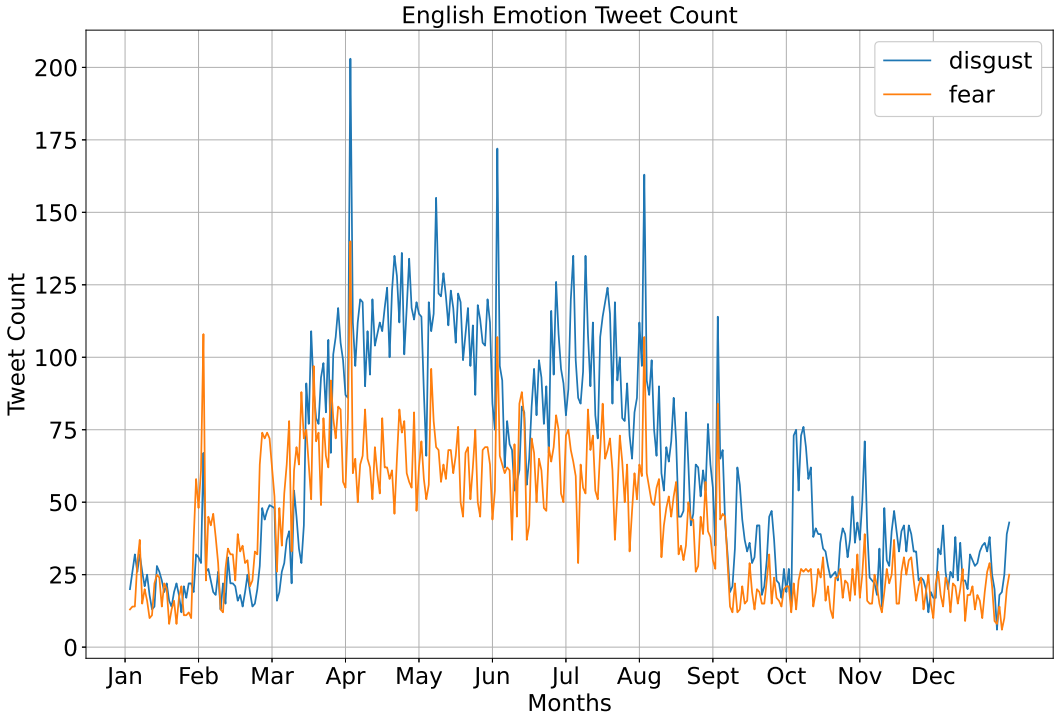

Figure 3.16: Top 2 English-Speaker Emotions

a significant difference between anger and joy for the majority of the year for the Spanish speakers, as shown in Figure 3.17. We gather that the negative emotions will have similar intensities. However, negative and positive emotions are less likely to have the same level of intensity in a crisis like the Coronavirus pandemic. Table 3.13 reveals that the # of tweets and % of tweets are higher for the top English-speaker emotion than the top Spanish-speaker emotion. We conclude that the English speakers felt more intensely about the events occurring than the Spanish speakers. Although the Spanish speakers felt a high amount of anger, their second-highest emotion is joy, which leads us to believe that Spanish speakers were calmer throughout the 2020 pandemic.

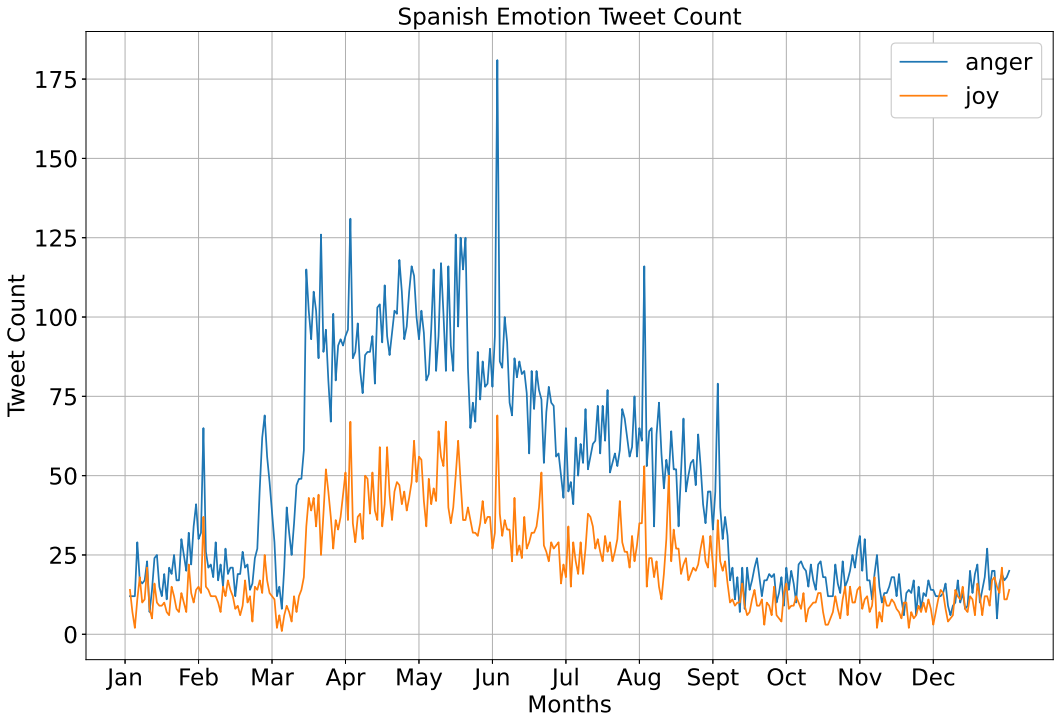

Figure 3.17: Top 2 Spanish-Speaker Emotions

3.6.2 Themes Analysis Between the English and Spanish Speakers

When it comes to the language comparison analysis, we take a close look at the topics produced by LDA. In Table 3.12, we see similar themes between the languages. The similar positive themes throughout the year on both sides contain hope for the vaccine and positive words about the safety measures. Overall, we analyze that the English and Spanish speakers felt similarly hopeful and protective during the 2020 global pandemic. On the other hand, we find two negative themes discussed amongst the English and Spanish speakers. The first one is concern about the increasing about of death, and the second is a concern for other countries.

Importantly, we observe variances between the two speakers. More specifically, the English speakers seemed to be more concerned about politics and the Trump administration. On the other hand, the Spanish speakers were more concerned with the effects of the pandemic as it affects the economy.

ID	**English Themes**	**Spanish Themes**
Jan-Apr	Preventative Strategy	Prevent Spread
	Fighting Covid	Fighting Covid
	Effects on World	World Impacted
May-Aug	Outbreak Worldwide	Increasing Cases
	Future Affects	World Normal
Sept-Dec	Virus Protection	Trials Developing
	Recovery Impact	New Recoveries

Table 3.12: English and Spanish Similar Themes

Months	**Language**	**Top Emotion**	**# Tweets**	**% Tweets**
Jan-Apr	English	Fear	2512	4.6%
May-Aug	English	Fear	2983	3.8%
Sept-Dec	English	Fear	941	4.4%
Jan-Apr	Spanish	Anger	151	.24%
May-Aug	Spanish	Surprise	722	.85%
Sept-Dec	Spanish	Surprise	309	1.5%

Table 3.13: Language Comparison Top Theme Emotions

3.7 Summary

In this chapter, we gathered insight into the reactions toward COVID-19 in 2020 by analyzing the topics, sentiments, and emotions extracted from tweets related to Covid-19. The analysis framework embraces four integrated phases. First, we separated and preprocessed the English and Spanish tweets. Second, we used LDA to glean topics and infer themes from those topics from the English and Spanish tweets. Third, we used *Pysentimiento* to gather the overall sentiment emotions from each language while subsequently employing Pysentimiento to retrieve the emotions from each topic cluster. Lastly, we undertook a comparative analysis of the English and Spanish tweets. Our dataset derives from a dataset created by researchers at Georgia State University. The dataset contains Covid-related tweets collected using corresponding keywords. Then, we further constructed our balanced dataset by processing the data, giving us 150,000 English tweets and 150,000 Spanish tweets.

Overall, we discover that the different groups and cultures had a variety of emotions towards the COVID-19 pandemic. The groups both mainly felt negative sentiments throughout the year, but the specifics of emotions and conversations were diverse. Even so, the themes reveal that both groups agree that the pandemic has completely changed the world and created a new normal. Furthermore, we present three specific key observations from our research:

1. The English and Spanish speakers had drastically different emotion reactions towards COVID-19

2. The English speakers had stronger negative-emotion reactions throughout COVID-19 than the Spanish speakers.

3. Analyzing the emotions of the topic clusters gives more insightful information about the individuals

The spikes corresponding with events can be helpful in future research. We can assume that the time of the spike is an important time period to analyze. Topic modeling is a great resource to retrieve useful information to determine the reason for the spike. The negative sentiment surged over the summer due to the second wave which brought out more fear and anger in people. The vaccine brought hope which lead to a reduction in high negative sentiment. We also observed the more intense reaction of the English speakers from the greater number of tweets that reflected negative emotions compared to the Spanish speaker number of tweets. We do theorize that this can also be due to the classification method. We surmise that these observations can provide the healthy groundwork for more research towards sentiment and emotion analysis across languages in crises.

We believe that this study has provided insights into the differing responses between English and Spanish speakers during the COVID-19 global pandemic. We believe that this research will continue to be useful for medical professionals who handle multiple groups in crises, psychology, public health, and economics research.

Our aim for this chapter within this dissertation was to take a step towards aspect-based emotion analysis. Within this study, we used LDA for information extraction and Pysentimiento for sentiment and emotion classification. Due to the generality of LDA, the information extracted is general topics within a dataset. In the next chapter, we proceed to a more detailed challenge which is the aspect level of information from a sentence. We implement a design for aspect-based emotion analysis, and we aim for this next chapter to include applying advanced algorithms to further improve the approach within this chapter. We will discuss the details of our next method in Chapter 4.

Chapter 4

Aspect Term Extraction using Annotation Schemes

In this chapter, we offer a novel approach to Aspect-Term Extraction (ATE) – a vastly unexplored research area. The aim of this part of the dissertation research is to spark more efforts toward the task of ABSA. The results we provide give more guidance and direction that can lead to more fine-grained details for companies and organizations that desire user feedback. The results of this study are slated to facilitate more data-driven decisions within an organization, especially where resources are a constraint.

This chapter has six major sections in ABSA. We begin by explaining the background and formal definitions of Aspect-Based Emotion Analysis to lay even more of a foundation for our research and efforts in Section 4.1. We then delve into our problem statement and research questions in Section 4.2 in order to clarify our intentions and the questions to be answered. In Section 4.4.1 we provide an in-depth analysis of our dataset and in Section 4.3 our methodology. Next, we provide a thorough analysis of the experiment and results in Sections 4.4-**??**. Lastly, we will discuss our summary in Section 4.5. For a more thorough background to this research, Chapter 2 contains detailed information on ABSA. As we continue to explain our research, we would like to remind you that we are motivated to do this research due to the gap in research efforts toward ABSA. We desire that our contribution motivates other researchers to continue within this field.

4.1 Formulation of Research Tasks

Over the past decade, the evolution of Aspect-Based Sentiment Analysis - ABSA - has skyrocketed; it is now an increasingly researched field [245]. ABSA is a sub-field of Natural Language Processing that divides data into aspects and extracts the sentiment. As

we discussed in Chapter 2, Sentiment Analysis (SA) is broken down into document-level, sentence-level, and aspect-level. ABSA on the other hand is a customer-centric approach to understanding consumer emotion, while topic-based sentiment analysis. ABSA is known to provide rich information from data than sentiment analysis [83]. In this chapter, we shed bright light on a subtask of aspect-based sentiment analysis: *aspect term extraction*. As we explained in Chapter 2, aspect term extraction or *ATE* detects the opinion terms. Throughout this dissertation research, we choose to pursue tackling this essential subtask.

Aspect Term Extraction: ATE is a subtask that involves identifying the specific aspects s in text that people are expressing emotions towards. For example, in a restaurant review, the aspects might be food, atmosphere, or customer service.

Aspect-Level Emotion Classification: ALEC is a subtask that involves determining the emotions expressed towards the aspect identified in the aspect term extraction phase. Emotions identified may be joy, sadness, anger, fear, disgust, surprise, and to name just a few.

4.2 Problem Statement and Research Questions

In this section, we articulate a list of research problems to be tackled in this chapter. After presenting the problem statement, we elaborate on three research questions to be addressed in this part of the dissertation study.

4.2.1 Problem Statement

The idea of gaining specific and automated insight from text is becoming essential to the success of organizations. Organizations are continually attempting to determine customers' predictive behaviors and needs. In spite of the fact that ABSA is a fairly new area of research, ABSA has a healthy number of experiments using rule-based, machine learning-based, and deep learning-based approaches. Now there is a need, however, to go further by gleaning more details from customer reviews and social media posts that can be useful for

companies and organizations to make more data-driven decisions. The demand for detailed information presents a new set of challenges and opportunities. We believe that aspect-based emotion analysis is slated to bridge this gap in research. Unfortunately, little is explored in regard to Aspect-Based Emotion Analysis. As discussed in Chapter 2 there are few papers that tackle this problem, even so, the current models have pursued simple rule-based methods [211, 143]. If the development of ABSA has grown in value for the industry sectors and research communities.

The problem statement for this part of the dissertation research is tabulated below.

Problem Statement: Aspect-Based Sentiment Analysis (ABSA) is an emerging field in Natural Language Processing (NLP). In this dissertation, we address the challenges in developing an effective Aspect Term Extraction system with Annotation Schemes that can provide a deeper understanding of user opinions expressed and contribute to enhanced decision-making.

4.2.2 Research Questions

We present a list of research questions as the first step toward solving the aforementioned problem. In what follows, we highlight an overarching goal of this part of the dissertation studies, followed by the three specific research questions to be addressed in this chapter.

A Research Goal: ATE as a part of ABSA is able to give companies and researchers useful information about the people they service. This information can help them to make accurate data-driven decisions that improve the experience of their customers. We believe that improving ATE can help improve customer experiences and help people that need it. The overarching goal of this study is to fill in the technological gap in ABSA tackling ATE using annotation schemes.

Throughout this chapter, we address the following research issues in the context of Aspect-Based Sentiment Analysis.

- *Research Question 1:* How does the granularity of an annotation scheme impact the performance of aspect term extraction models?
- *Research Question 2:* What are the strengths and limitations of different annotation schemes for aspect term extraction?
- *Research Question 3:* What are the challenges in designing an annotation scheme for aspect term extraction?

4.3 Methodology

Our methodology for this research embraces three main parts, including (1) data preparation, (2) aspect term extraction. Figure 4.1 depicts a visual presentation of our methodology. We discuss our dataset preparation in the next section, Section 4.4.1. The diagram shows that the dataset is first preprocessed and broken down into sentences, and annotation schemes. Lastly, the sentences and annotation schemes are passed into the aspect term extractor model.

Large Language Models

For this research, we choose four known large language models, namely, GPT, Electra, RoBERTa, and Distilled BERT. We explain each LLM along with a rationale behind electing the model.

Generative Pre-trained Transformer (or GPT) is a large language model based on the Transformer architecture. It was trained on large amounts of unlabeled text data to learn statistical patterns and structures of language. Although relatively new, the GPT models have performed well on a variety of tasks including text generation, language translation, and text classification.

ELECTRA stands for Efficiently Learning an Encoder that Classifies Token Replacements Accurately. ELECTRA is a language model that uses a novel pre-training objective

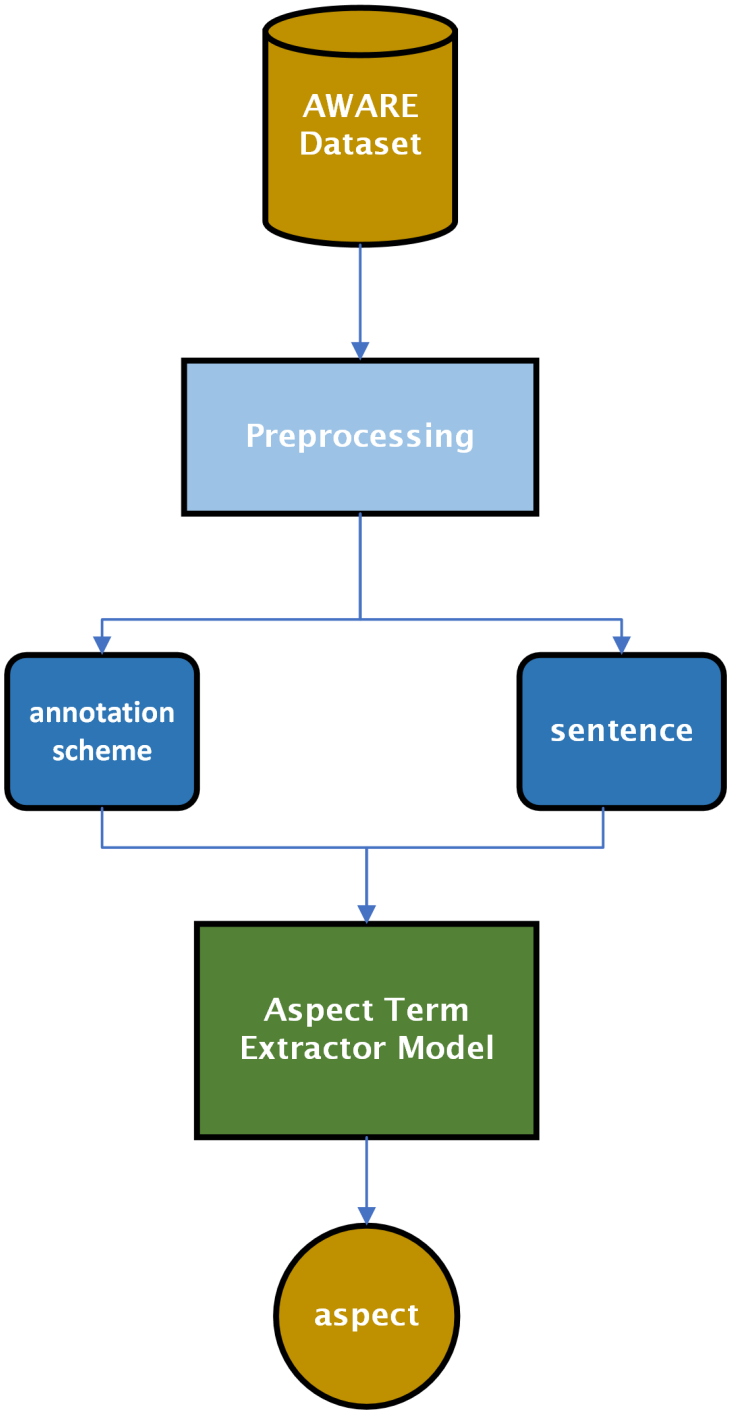

Figure 4.1: Our Aspect Term Extraction Methodology

called a discriminator which predicts whether a token has been replaced in a given sentence. This approach enables faster training and achieves competitive performance compared to

other models. We infer that this model would perform well on aspect term extraction because of our token classification approach.

RoBERTa stands for Robustly Optimized BERT Approach. RoBERTa is built with the foundation of a BERT (Bidirectional Encoder Representations from Transformers) model. It is a more advanced BERT due to the ability to use larger batch sizes and training on more data which enhances the performance across NLP tasks.

Distilled BERT, a compressed version of the BERT model, aims to curb the computational and memory requirements of BERT while preserving most of its performance. DistilBERT achieves this by applying a knowledge distillation technique, where it is trained to mimic the behavior of the larger BERT model. DistilBERT provides a more efficient solution for various natural language processing tasks. We select this model to compare its performance as a faster LLM against the other high-performing LLMs on our dataset. Ideally, we have a variety of LLMs and offer a variety of potentially successful solutions. Each of these language models has been influential in advancing the field of natural language.

4.3.1 Annotation Schemes

Annotation schemes are a set of tags allocated to a word in a sentence to represent the word and positional information. There are seven well-known annotation schemes implemented in this part of the study. The seven schemes are denoted as IO, IE, BI, IOB, IOE, BIES, and BILUO, respectively. We then develop our own annotation scheme: reverse IOB (or _IOB). In what follows, we define these schemes in the context of the Aspect Term Extraction [71].

1. IO: a simple scheme where each word from a sentence is assigned a tag. An inside tag is given the letter "I" and an outside tag is given the letter "O". The "I" tag is for aspect terms, and the "O" tag is for all non-aspect words. There is one non-aspect label and one aspect label.

2. IE: It labels the start of an aspect word with "I-aspect", the end of the aspect word with "E-aspect" and non-aspect words with the tag "E-O" and "I-O". There are two non-aspect labels and two aspect labels.

3. BI: This scheme labels the beginning of an aspect word with a "B-aspect" and the middle and end of it with "I-aspect. It also labels the beginning of non-aspect words with "B-O" and all other words with "I-O". There are two non-aspect labels and two aspect labels.

4. IOB: This approach is one of the most popular schemes; often referred to as the "BIO" scheme. It assigns a tag, a "B" for the begging of an aspect word, an "I" for the inside of an aspect term, and an "O" for non-aspect words. There is one non-aspect label and two aspect labels.

5. IOE: This technique assigns an "I" for the inside of an aspect term, an "O" for non-aspect words, and an "E" for the end of an aspect word. There is 1 non-aspect label and 2 aspect labels.

6. BIES: For aspect words, this scheme uses "B-aspect" for the beginning of an aspect word, "I-aspect" for the inside of an aspect word, "E-aspect" for the end of an aspect word, and "S-aspect" for a single unit aspect word. This scheme also uses the tag "B-O" to label the beginning of non-aspect words, the tag "I-O" to label the inside of non-entity words, and the tag "S-O" for single non-aspect tokens. This scheme is considered one of the most complex schemes. There are four non-aspect labels and 4 aspect labels.

7. BILUO: This annotation scheme is also known as "IOBES", which tags a "B" for the beginning of an aspect word, an "I" for the inside of an aspect word, an "L" for the last of an aspect word, a "U" as a single-token unit aspect word, and an "O" for all words outside of the aspect word. There is one non-aspect label and four aspect labels.

Figure 4.2: Input and Output of ATE Model

8. _IOB: Our annotation strategy is the reverse of IOB. It tags a "B" for the beginning of a non-aspect word, an "I" for the inside of a non-aspect word, and an "O" for the aspect word. There are two non-aspect labels and one aspect label.

Here we show a sample of pseudocode of how we implemented the annotation schemes for the sentences in our dataset.

Algorithm 1: IOB Annotation Code

```
Data: inputsentence
Result: IOB − labels
check ← 0;
tag ←' ' iob scheme ← [];
for word in sentence do
    if word ≠aspect term then
        tag ← O;
    else
        check ← check + 1;
        if check ≥ 1 then
            tag ← I − ASPECT;
        else
            tag ← B − ASPECT;
        end
    end
end
```

4.4 Aspect Term Extraction Experiments

In this section, we describe each of the experiments and reveal the results. For our ATE system, we choose to implement a four large language models in order to clearly analyze if the

Algorithm 2: _IOB Annotation Code

```
Data: inputsentence
Result: _IOB − labels
check ← 0;
tag ←' ';
_iobscheme ← [];
for word in sentence do
    if word aspectterm then
        tag ← O − ASPECT;
        check ← 0;
    else
        check ← check + 1;
        if check ≥ 1 then
            tag ← I;
        else
            tag ← B;
        end
    end
end
```

annotation scheme improves the results. We surmise that implementing various annotation schemes tends to enhance aspect term extraction results. We compare our method with two others including LDA and BERTopic because LDA has been a common approach to ATE [211, 109].

4.4.1 Dataset Preparation for Aspect Term Extraction

To evaluate Aspect-Based Sentiment Analysis, we adopt a reliable ABSA dataset. The dataset tested in this study is the *Aspect-Based Sentiment Analysis Dataset of Apps Reviews for Requirements Elicitation (AWARE)*, which is a dataset built for ABSA containing app reviews of three different domains such as Productivity, Social Networking, and Games. Within the dataset, there are 11323 sentences where each sentence is labeled with a boolean value indicating whether the sentence exhibits a positive or negative sentiment. Additionally, each sentence is labeled with an Aspect Term and an Aspect Category.

In Figure 4.3, we display the variation of aspect term counts in the dataset. There are significantly more occurrences of one aspect term than two, and these factors are important to note as features of our dataset.

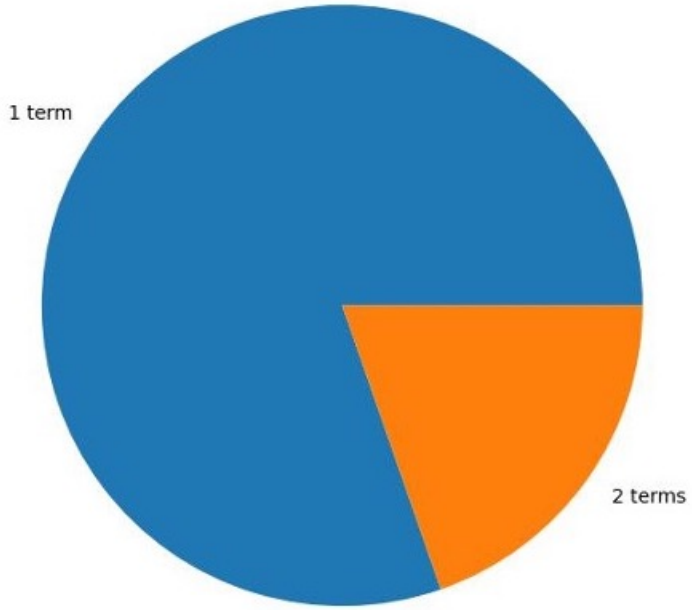

Figure 4.3: AWARE Dataset Aspect Term Count

For aspect term extraction, we decide to apply the annotation schemes offering simple labels or tags attached to textual elements of a sentence [141, 165]. Within the NLP arena, annotation schemes have mainly been used to tackle Named Entity Recognition (NER) tasks. NER identifies features of interest such as names of people, places, and organizations, in addition to dates and currency. It is a challenge to know which annotation scheme should be implemented and deployed [21]. We implement seven well-known annotation schemes as a new approach to facilitate aspect term extraction: the seven schemes are denoted as IO, IE, BI, IOB, IOE, BIES, and BILUO, respectively [21, 71]. We then add our own newly designed annotation scheme referred to as *reverse IOB* (or _IOB). Reverse IOB is the reverse of the original annotation scheme: IOB. For reverse IOB, there is one annotation label (O-aspect)

and two non-aspect labels (B and I). We test this scheme out along with the other seven existing annotation schemes.

In order to incorporate these annotation schemes, we apply SpaCy, an open-source library for NLP, and our own logic to label each sentence with an annotation scheme which we explained in Algorithm 2 [272]. Our dataset contains columns for each annotation scheme which hold a list of labels, one label per word in the corresponding sentence. In Figure 4.4, we show a sample version of our dataset format. Figure 4.5 sketches the differences in the annotation scheme labels. The annotation schemes provide a uniqueness to the dataset.

	Sentence #	Word	IO	IOB	IOE	BI	IE	BILUO	BIES	Word Idx	Tag Idx
0	Sentence 0	One	O	O	O	I-O	I-O	O	B-O	2495	0
1	Sentence 0	improvement	O	O	O	I-O	I-O	O	B-O	2185	0
2	Sentence 0	I	O	O	O	I-O	I-O	O	B-O	2312	0
3	Sentence 0	want	O	O	O	I-O	I-O	O	B-O	3582	0
4	Sentence 0	to	O	O	O	I-O	I-O	O	B-O	5067	0

Figure 4.4: Aspect Term Extraction Dataset Format

4.4.2 Compared Methods for Aspect Term Extraction

In this section, we discuss the baseline methods that are compared against our method. We choose to compare our model with LDA and BERTopic thanks to the common use of LDA in previous ABSA papers. Our belief is that these approaches are good and broad for the task of ATE. We observe from Figure 4.6 how LDA and BERTopic function.

Latent Dirichlet Allocation (LDA): is a statistical model that is commonly deployed for topic modeling. LDA, which finds hidden topics within a set of documents, assumes that there is a common topic amongst a group of documents – and it can be found distributed over the words in the document. LDA discovers the topic by looking for patterns in documents.

We utilize a Python library to implement LDA. For this model, we only have to implement the count for the topics in order to retrieve the number of expected aspect terms.

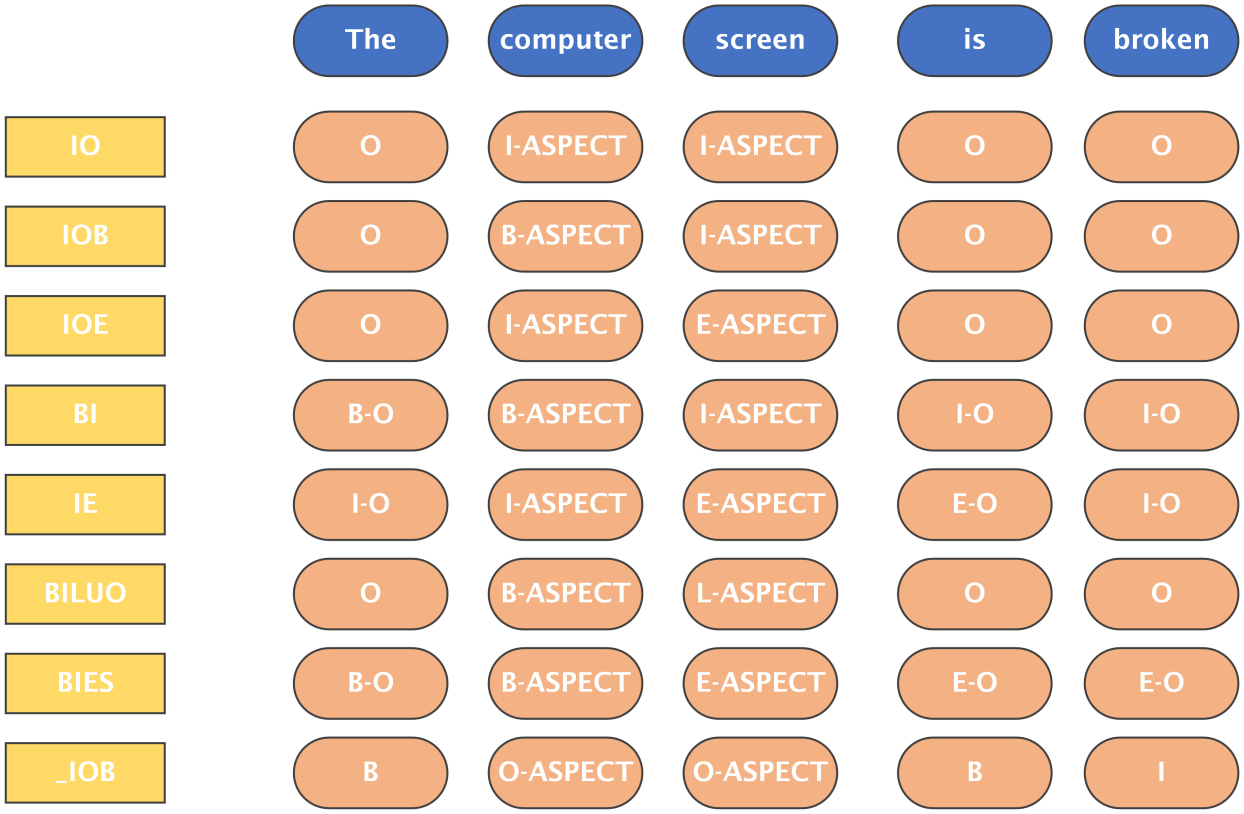

Figure 4.5: Sample Sentence with Annotation Schemes

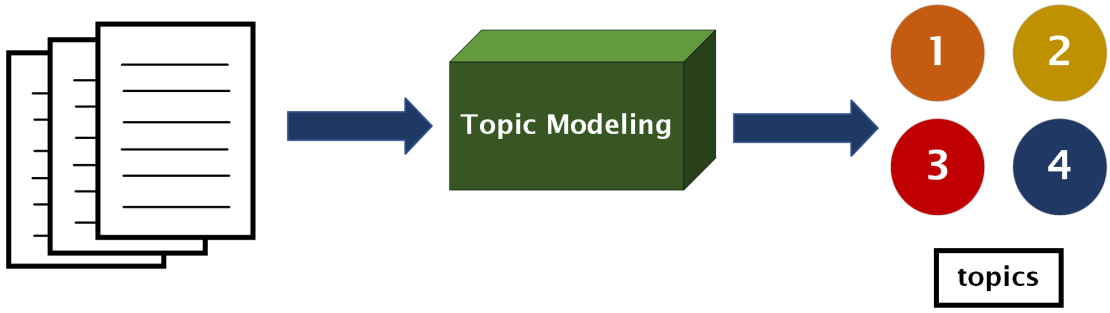

Figure 4.6: Topic Modeling Process

We then compare the output terms with the aspect terms from the dataset to verify the accuracy of LDA for ATE.

BERTopic: is a newer topic modeling technique that applies a pre-trained BERT model and dimensionality reduction to discover topics. Documents pass through BERTopic are embedded using BERT and; then, the dimensionality of the vectors is reduced. Next,

the vectors are put into clusters and those clusters are labeled by using the most frequent words. The main words are considered the topics of that cluster.

We implement BERTopic by the virtue of the Python library. For the parameters, we pass in 'English' for the language, 2191 for the *top n words*, and (1,2) for the *n gram range*. Our dataset contains terms of size 1 and size 2; therefore, we choose unigrams and bigrams. Once the model extracts all of the terms, we implement our own metric to check the accuracy of the extracted terms.

4.4.3 Results of Aspect Term Extraction

In this section, we evaluate the results collected from the aspect term extraction module. The goal is to compare our ATE model with the previous approaches driven by this dataset. We also compare the results of the annotation schemes against each other to rank the tested competitors. We adopt Precision, Recall, and F1-Score as our evaluation metrics. In Figures4.7- 4.10, we display the formulas for calculating the metrics quantifying the models' results.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Figure 4.7: Precision Formula

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Figure 4.8: Recall Formula

$$F1 = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

Figure 4.9: F1-Score Formula

Actual		Predicted	
		Negative	**Positive**
	Negative	*True Negative*	*False Positive*
	Positive	*False Negative*	*True Positive*

Figure 4.10: Metrics Calculation Confusion Table

Clustering Model Results

Table 4.1 shows the results of the clustering models. We observe that overall the clustering models, LDA and BERTopic perform low for aspect term extraction. The topic modeling approach to ATE is not very precise as desired due to the fact that LDA and BERTopic gather general topics over a set of documents. In order for the aspect term extraction module to have accurate results, the method must analyze each sentence individually followed by extracting aspect terms. We acknowledge that LDA and BERTopic do offer insightful and useful information for determining an aspect term. We propose to utilize these methods as baseline methods in comparison with advanced approaches for ATE.

Model	**Precision**	**Recall**	**F1-Score**
LDA	0.03	0.03	0.03
BERTopic	0.23	0.23	0.23

Table 4.1: Aspect Results from Comparison Models

LLM Model Results

We begin our analysis by comparing the large language models' runtimes. We gauge the time in minutes for each model to run to train for each annotation scheme. In Figure 4.11, the bars unveil the average number of minutes to train the model. The Electra model took the longest to train on our dataset. Comparatively, the Distilled Bert model consistently ran significantly faster than the other models. The GPT and RoBERTa models have similar runtime performance.

In our analysis, we gather the f1, accuracy, precision, and recall measures for each annotation label of each annotation scheme for each large language model. Given the number of labels, annotation schemes, and models, we collect a total of 500 data results tabulated in Tables 4.2 - Table 4.9.

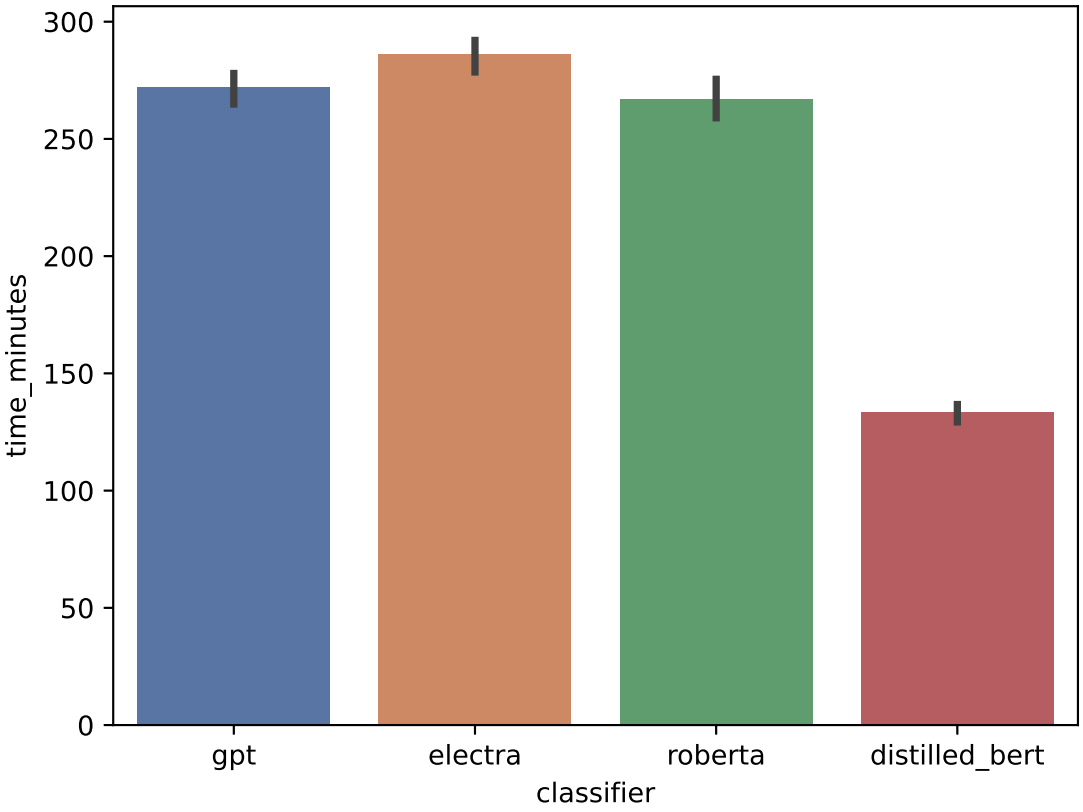

Figure 4.11: LLM Time Comparison

Figures 4.12- 4.15 unravel the F1 scores for the different llms. Across the different models, we observe a consistent pattern with the differences among the annotation schemes. It is clear that IOB, IO, and _IOB continuously outperform the other five annotation schemes. From another angle, Figure 4.16 shows f1 scores of all the annotation schemes across all the models. Therefore, IO and _IOB are the best-performing annotation schemes. Going further, Figure 4.17 displays a more focused comparison of IO and _IOB: the _IOB scheme has a

scheme	classifier	label	f1-score	recall	precision	accuracy
IOB	gpt	I-ASPECT	0.645	0.476	1	0.4697
IOB	gpt	B-ASPECT	0.69857	0.5367	1	0.38716
IOB	gpt	O	0.983	0.9884	0.9793	0.9573
IOB	electra	I-ASPECT	0.549450549	0.3787	1	0.4018
IOB	electra	B-ASPECT	0.6002	0.4287	1	0.299
IOB	electra	O	0.982	0.99	0.9755	0.9575
IOB	roberta	I-ASPECT	0.308	0.182	1	0.298
IOB	roberta	B-ASPECT	0.5468	0.376	1	0.238
IOB	roberta	O	0.988	0.993	0.983	0.97
IOB	distilled_bert	I-ASPECT	0.50537	0.338	1	0.329
IOB	distilled_bert	B-ASPECT	0.525	0.356	1	0.2478
IOB	distilled_bert	O	0.9847	0.9935	0.976	0.9639

Table 4.2: IOB Aspect Results

scheme	classifier	label	f1-score	recall	precision	accuracy
IO	gpt	I-ASPECT	0.7603	0.6133	1	0.4662
IO	gpt	O	0.982	0.988	0.976	0.9534
IO	electra	I-ASPECT	0.7427	0.5907	1	0.433
IO	electra	O	0.9817	0.9954	0.9686	0.9601
IO	roberta	I-ASPECT	0.703	0.542	1	0.3951
IO	roberta	O	0.9834	0.9924	0.9745	0.9603
IO	distilled_bert	I-ASPECT	0.5564	0.3856	1	0.24889
IO	distilled_bert	O	0.9853	0.99361	0.9772	0.9651

Table 4.3: IO Aspect Results

scheme	classifier	label	f1-score	recall	precision	accuracy
BILOU	gpt	B-ASPECT	0.2978	0.175	1	0.2904
BILOU	gpt	I-ASPECT	0	0	0	0.4285
BILOU	gpt	L-ASPECT	0.4028	0.2522	1	0.2706
BILOU	gpt	U-ASPECT	0.6208	0.4501	1	0.3182
BILOU	gpt	O	0.9864	0.988	0.9848	0.9619
BILOU	electra	B-ASPECT	0.4758	0.3122	1	0.319
BILOU	electra	I-ASPECT	0	0	0	0.4558
BILOU	electra	L-ASPECT	0.5592	0.3881	1	0.3881
BILOU	electra	U-ASPECT	0.5917	0.4202	1	0.2845
BILOU	electra	O	0.9876	0.9959	0.9795	0.9718
BILOU	roberta	B-ASPECT	0.4246	0.2695	1	0.2587
BILOU	roberta	I-ASPECT	0	0	0	0.8
BILOU	roberta	L-ASPECT	0.5442	0.3738	1	0.3936
BILOU	roberta	U-ASPECT	0.24294	0.273	1	0.18
BILOU	roberta	O	0.9905	0.994	0.987	0.9754
BILOU	distilled_bert	B-ASPECT	0.4545	0.2941	1	0.2536
BILOU	distilled_bert	I-ASPECT	0	0	0	0.2738
BILOU	distilled_bert	L-ASPECT	0.4102	0.258	1	0.291
BILOU	distilled_bert	U-ASPECT	0.3487	0.2112	1	0.1375
BILOU	distilled_bert	O	0.9871	0.9932	0.9811	0.9681

Table 4.4: BILOU Aspect Results

higher f1 score than IO; therefore, we can affirm that _IOB is the best performer among all the tested annotation techniques.

For a more detailed view, we compare all of the labels across the eight annotation schemes. Figure 4.18 reveals the f1 scores of each label: each of the labels for non-aspect

scheme	classifier	label	f1-score	recall	precision	accuracy
BIES	gpt	B-ASPECT	0.3148	0.1868	1	0.306
BIES	gpt	I-ASPECT	0	0	0	0.6607
BIES	gpt	E-ASPECT	0.6687	0.502	1	0.536
BIES	gpt	S-ASPECT	0.7163	0.558	1	0.4329
BIES	gpt	B-O	0.8788	0.8197	0.9475	0.6685
BIES	gpt	I-O	0.9741	0.9633	0.9851	0.9165
BIES	gpt	E-O	0.4616	0.3119	0.8871	0.1805
BIES	electra	B-ASPECT	0.4855	0.3205	1	0.3675
BIES	electra	I-ASPECT	0	0	0	0.3783
BIES	electra	E-ASPECT	0.5752	0.4037	1	0.4264
BIES	electra	S-ASPECT	0.714	0.5552	1	0.4123
BIES	electra	B-O	0.8908	0.8567	0.9335	0.7047
BIES	electra	I-O	0.9694	0.954	0.9852	0.8998
BIES	electra	E-O	0.7856	0.68	0.9301	0.496
BIES	roberta	B-ASPECT	0.565	0.3939	1	0.4748
BIES	roberta	I-ASPECT	0	0	0	0.8
BIES	roberta	E-ASPECT	0.5574	0.386	1	0.4341
BIES	roberta	S-ASPECT	0.7725	0.6293	1	0.4805
BIES	roberta	B-O	0.9205	0.8817	0.9629	0.765
BIES	roberta	I-O	0.9681	0.9519	0.9847	0.8957
BIES	roberta	E-O	0.7869	0.6856	0.923	0.5
BIES	distilled_bert	B-ASPECT	0.56	0.389	1	0.349
BIES	distilled_bert	I-ASPECT	0	0	0	0.3552
BIES	distilled_bert	E-ASPECT	0.5847	0.4131	1	0.4349
BIES	distilled_bert	S-ASPECT	0.6569	0.489	1	0.352
BIES	distilled_bert	B-O	0.858	0.8127	0.9094	0.6408
BIES	distilled_bert	I-O	0.966	0.951	0.982	0.892
BIES	distilled_bert	E-O	0.7426	0.628	0.906	0.4379

Table 4.5: BIES Aspect Results

scheme	classifier	label	f1-score	recall	precision	accuracy
IE	gpt	E-ASPECT	0.566	0.3948	1	0.492
IE	gpt	I-ASPECT	0.734	0.579	1	0.435
IE	gpt	I-O	0.981	0.986	0.977	0.951
IE	gpt	E-O	0.452	0.3007	0.915	0.174
IE	electra	E-ASPECT	0.519	0.35	1	0.456
IE	electra	I-ASPECT	0.705	0.544	1	0.382
IE	electra	I-O	0.976	0.975	0.977	0.931
IE	electra	E-O	0.723	0.6116	0.885	0.416
IE	roberta	E-ASPECT	0.603	0.431	1	0.585
IE	roberta	I-ASPECT	0.768	0.623	1	0.474
IE	roberta	I-O	0.977	0.976	0.977	0.933
IE	roberta	E-O	0.742	0.648	0.868	0.447
IE	distilled_bert	E-ASPECT	0.458	0.297	1	0.377
IE	distilled_bert	I-ASPECT	0.608	0.436	1	0.294
IE	distilled_bert	I-O	0.975	0.973	0.977	0.928
IE	distilled_bert	E-O	0.698	0.575	0.887	0.384

Table 4.6: IE Aspect Results

terms is high performing whereas the aspect labels are lower. We infer that the non-aspect labels are consistently high due to the ease of accurately determining of a word is not an aspect term. Within a sentence, the majority of the words are not aspect terms, making the possibility of a word having a non-aspect term label high. As we focus on the aspect terms, we discover from Figure 4.19, that there is a difference amongst the aspect terms.

scheme	classifier	label	f1-score	recall	precision	accuracy
IOE	gpt	E-ASPECT	0.708	0.548	1	0.4306
IOE	gpt	I-ASPECT	0.341	0.205	1	0.344
IOE	gpt	O	0.984	0.99	0.979	0.961
IOE	electra	E-ASPECT	0.612	0.441	1	0.341
IOE	electra	I-ASPECT	0.527	0.358	1	0.381
IOE	electra	O	0.986	0.996	0.976	0.969
IOE	roberta	E-ASPECT	0.5968	0.425	1	0.313
IOE	roberta	I-ASPECT	0.353	0.214	1	0.478
IOE	roberta	O	0.985	0.992	0.978	0.9636
IOE	distilled_bert	E-ASPECT	0.483	0.319	1	0.241
IOE	distilled_bert	I-ASPECT	0.454	0.294	1	0.267
IOE	distilled_bert	O	0.985	0.994	0.977	0.966

Table 4.7: IOE Aspect Results

scheme	classifier	label	f1-score	recall	precision	accuracy
BI	gpt	B-ASPECT	0.195	0.108	1	0.748
BI	gpt	I-ASPECT	0.372	0.228	1	0.186
BI	gpt	I-O	0.977	0.968	0.986	0.927
BI	gpt	B-O	0.857	0.777	0.957	0.617
BI	electra	B-ASPECT	0.72	0.562	1	0.402
BI	electra	I-ASPECT	0.559	0.388	1	0.424
BI	electra	I-O	0.975	0.978	0.973	0.933
BI	electra	B-O	0.859	0.8007	0.927	0.634
BI	roberta	B-ASPECT	0.785	0.646	1	0.518
BI	roberta	I-ASPECT	0.641	0.4719	1	0.4803
BI	roberta	I-O	0.969	0.966	0.973	0.911
BI	roberta	B-O	0.901	0.859	0.947	0.722
BI	distilled_bert	B-ASPECT	0.612	0.441	1	0.318
BI	distilled_bert	I-ASPECT	0.516	0.348	1	0.369
BI	distilled_bert	I-O	0.348	0.972	0.974	0.923
BI	distilled_bert	B-O	0.834	0.768	0.912	0.588

Table 4.8: BI Aspect Results

scheme	classifier	label	f1-score	recall	precision	accuracy
_IOB	gpt	O-ASPECT	0.783	0.643	1	0.505
_IOB	gpt	B	0.9006	0.86	0.944	0.723
_IOB	gpt	I	0.971	0.966	0.976	0.914
_IOB	electra	O-ASPECT	0.7602	0.613	1	0.452
_IOB	electra	B	0.976	0.9809	0.972	0.937
_IOB	electra	I	0.875	0.844	0.908	0.681
_IOB	roberta	O-ASPECT	0.806	0.675	1	0.54
_IOB	roberta	B	0.888	0.867	0.909	0.712
_IOB	roberta	I	0.974	0.976	0.972	0.929
_IOB	distilled_bert	O-ASPECT	0.732	0.578	1	0.415
_IOB	distilled_bert	B	0.847	0.807	0.891	0.625
_IOB	distilled_bert	I	0.974	0.981	0.968	0.933

Table 4.9: _IOB Aspect Results

S-aspect (of BIES annotation scheme) and O-aspect (of _IOB annotation scheme) are the best-performing labels. The difference we notice between the high-performing aspect labels and the lower-performing aspect labels is that the higher labels are labeling one aspect term. The S-aspect is for a single aspect term. The O-aspect labels a word as an aspect term and is the only label amongst _IOB that labels an aspect term. We deduce that the schemes

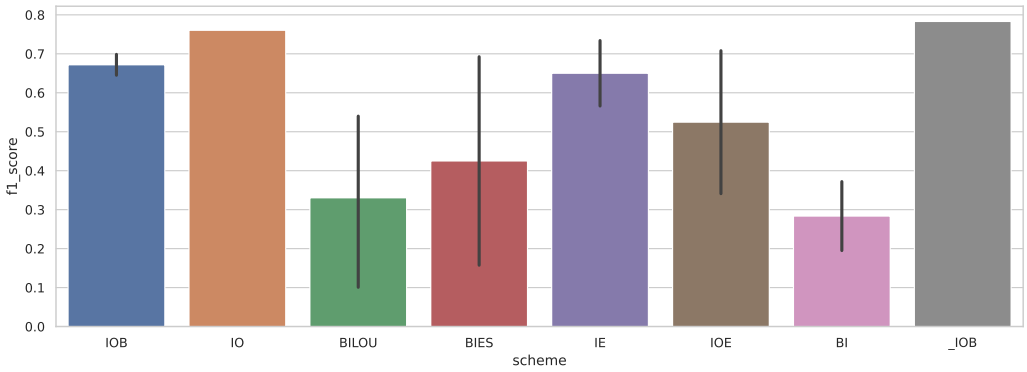

Figure 4.12: GPT F1 Scores Comparison

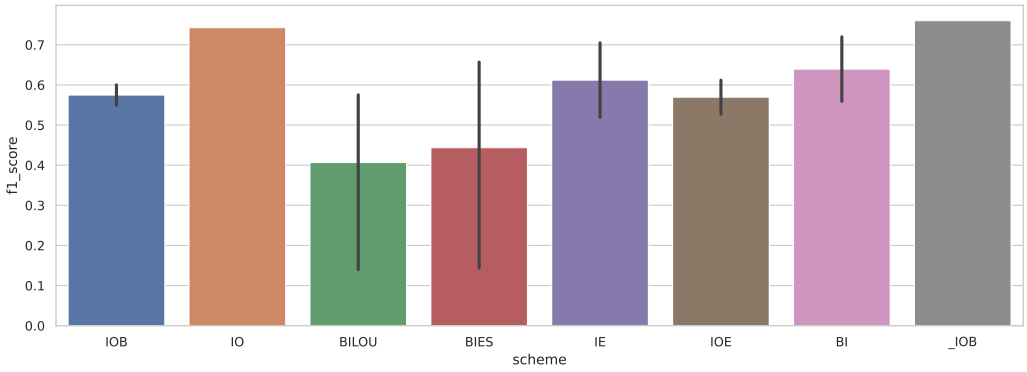

Figure 4.13: Electra F1 Scores Comparison

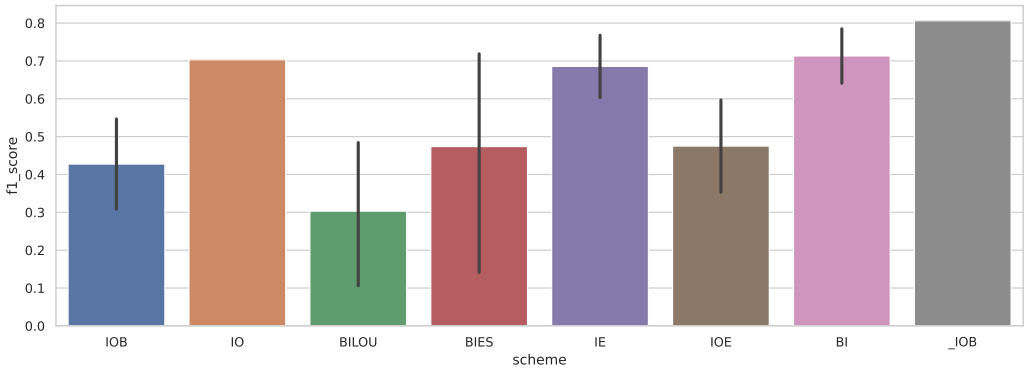

Figure 4.14: Roberta F1 Scores Comparison

with more labels for aspect terms have lower performance due to the complexity of finding the first and second aspect terms. As we saw in Figure 4.3, our dataset contains mainly

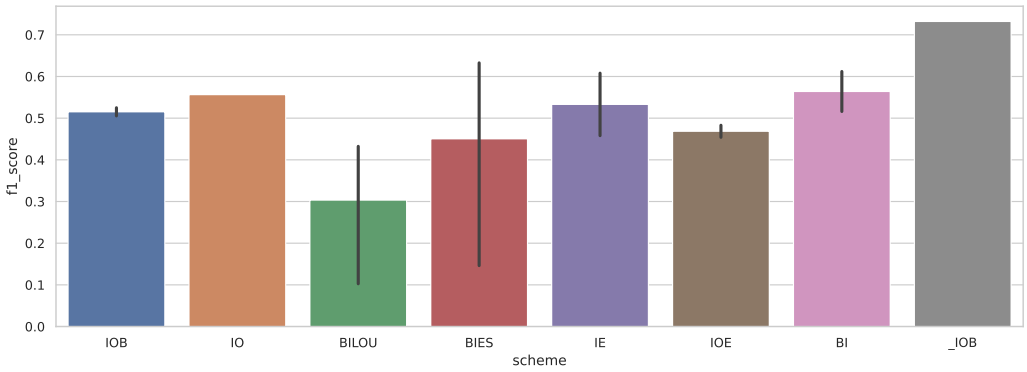

Figure 4.15: Distilled BERT F1 Scores Comparison

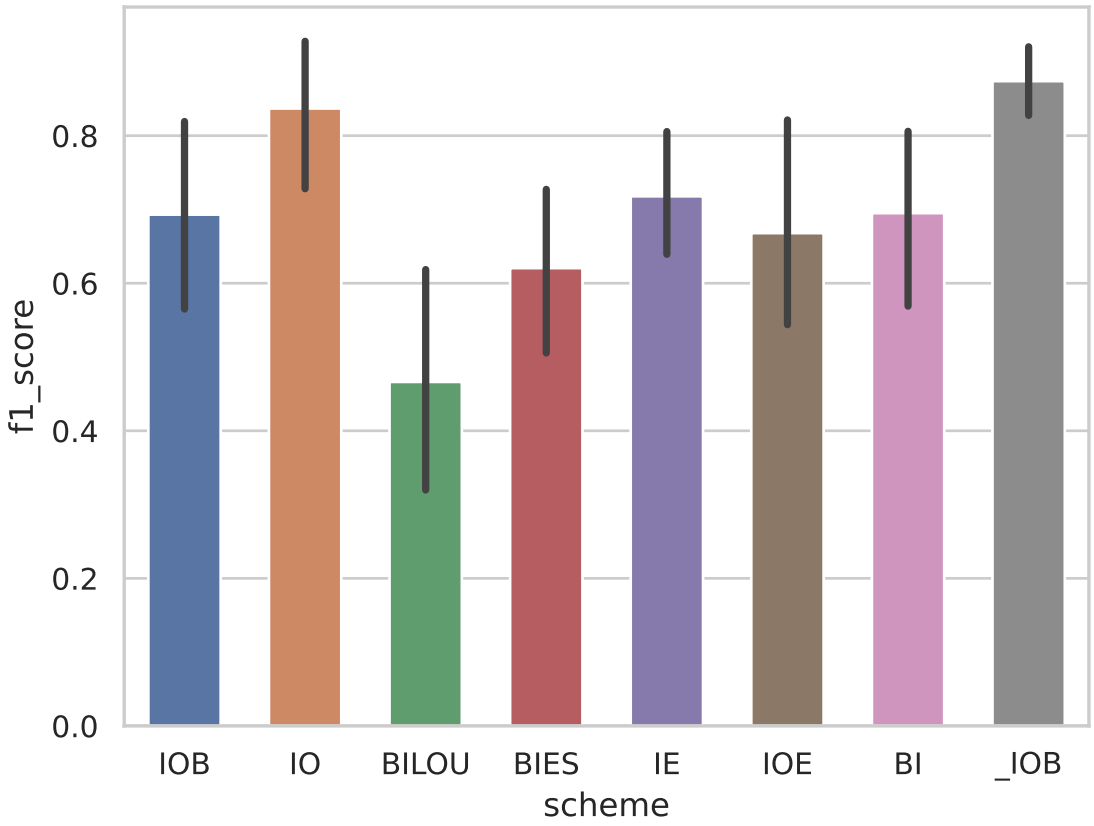

Figure 4.16: Annotation Schemes F1 Scores Comparison

singular aspect terms and a small percentage of two aspect terms. Due to this factor, we see a lower accuracy for I-aspect and L-aspect. This trend is expected because if the I-aspect

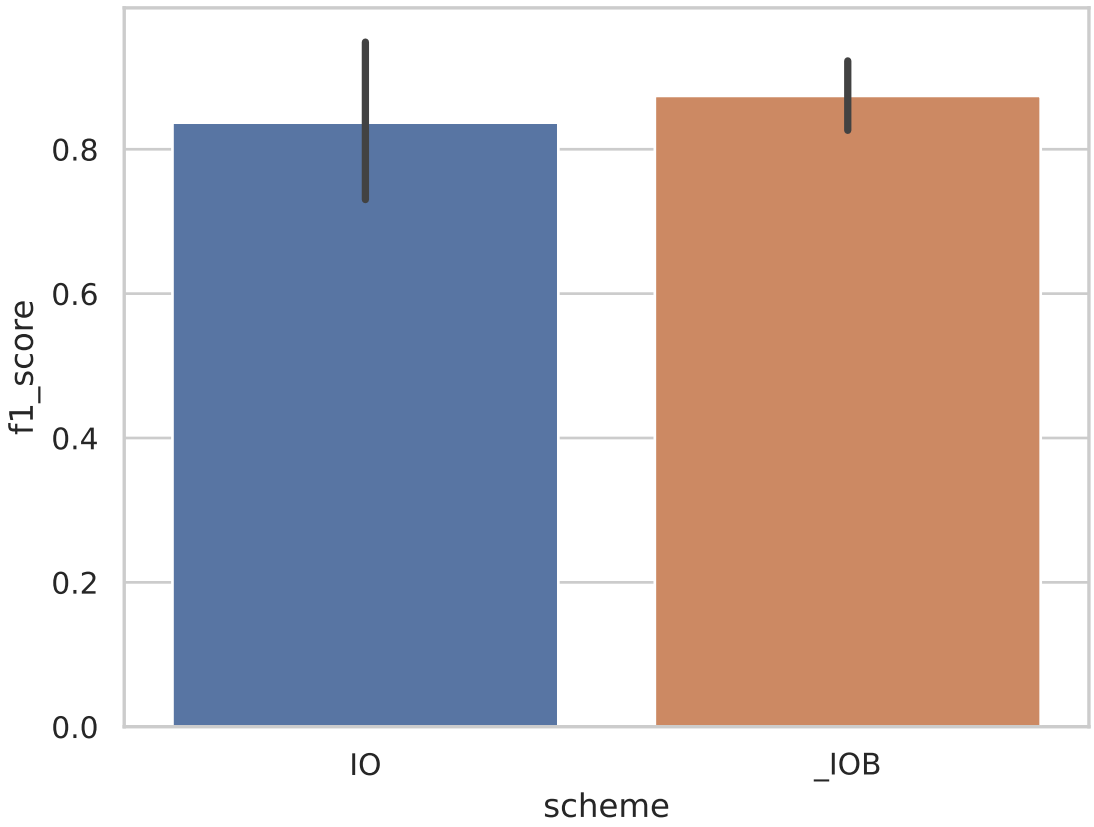

Figure 4.17: IO & reverse IOB F1 Scores Comparison

and L-aspect labels are to be used for BILOU, there must be three aspect terms. B-aspect would label the beginning term, I-aspect would label the middle term, and L-aspect would label the last term. For our dataset, the sentences with two aspect terms were labeled with B-aspect and L-aspect within BILOU. We reckon that with the small number of sentences with two aspect terms, the use of L-aspect is lower and the performance is lower. I-aspect is used in a number of annotation schemes including BIES, BI, IOB, IOE, and IE. I-aspect is used in each of these as the label of the second aspect term or the singular aspect term. We believe that because of the multiple uses of I-aspect, the performance is deteriorated.

Lastly, we analyze the differences in performance of the f1-scores of IO and _IOB labels since these schemes performed the best. In Figure 4.20, we see the similarity in scores of

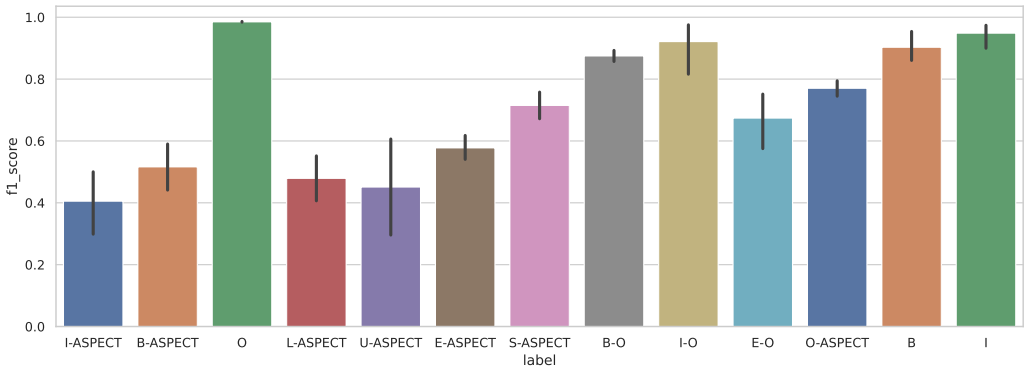

Figure 4.18: All Labels F1 Scores Comparison

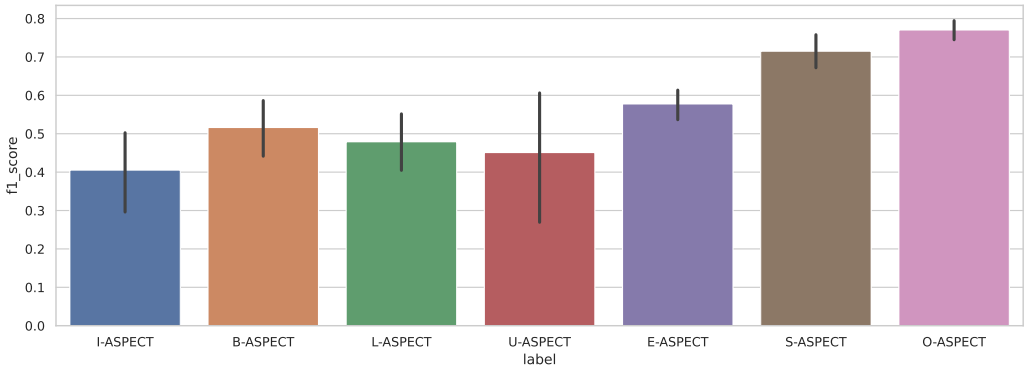

Figure 4.19: Aspect Labels F1 Scores Comparison

the non-aspect labels. We surmise that the _IOB non-aspect label scores slightly lower than the IO O label because there are two labels that must determine the difference between non-aspect terms. The O label does not have to differentiate between the non-aspect terms. For the aspect labels, clearly, the O-aspect label of _IOB enjoys better performance than that of the I-aspect label of IO.

From our analysis of the empirical results, we conclude that the singular label for aspect terms performs better for aspect labeling and extraction. Importantly, we learn from the analysis of the annotation schemes that the models with multiple labels for non-aspect terms deliver stunning performance in terms of classification. We notice amongst previous research the success of the IOB annotation scheme and the IO scheme. We devise our reverse IOB

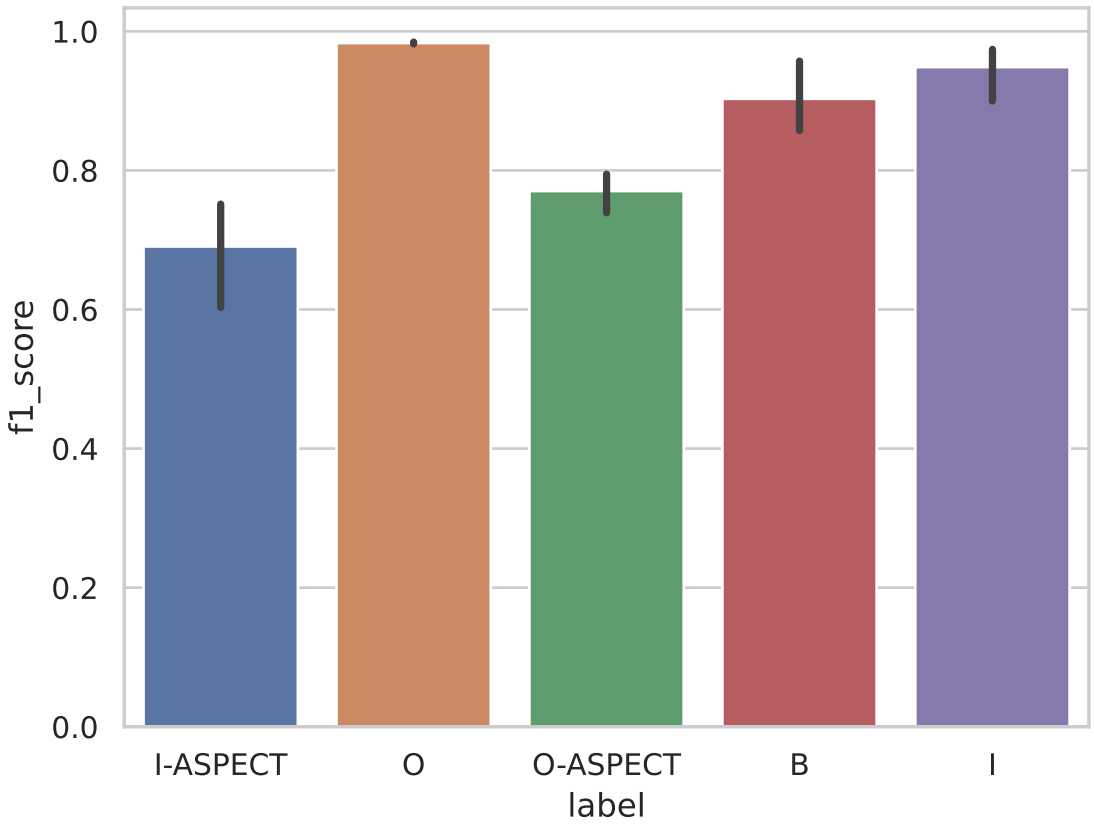

Figure 4.20: IO & reverse IOB Labels F1 Scores Comparison

(_IOB) annotation scheme because we are motivated by the success of IO and IOB having one label for aspect terms – and we are inspired by the success of the other annotation schemes with more complexity at a non-aspect term level.

Through our a thorough research, we confirm that our _IOB annotation scheme is a successful strategy to fulfill aspect term extraction tasks.

4.5 Summary

The goal of this part of the dissertation research is to generate more efforts toward Aspect-Based Sentiment Analysis. ABSA provides robust insight into customers using their own feedback. Throughout this chapter, we have discussed the background of Aspect-Based

and Sentiment Analysis, our motivations and problem statement, the dataset, our methodology, and the results. Aspect term extraction detects the aspect term in a sentence. For this dissertation, we adopted four models including GPT, Electra, RoBERTa, and Distilled BERT to tackle ATE. The eight annotation schemes provided the layer of complexity needed to detect the aspect terms with token labeling. The annotation scheme with the best performance is '_IOE' with 'IO' as a runner-up.

Throughout this part of the dissertation studies, we answered the following research questions in the context of the ABSA model. Below we revisit the questions and provide our findings.

Research Questions and Answers:

- *Research Question 1:* How does the granularity of an annotation scheme impact the performance of aspect term extraction models?

  Research Answer 1: We believe that the more granular and complex the annotation scheme, the worse the performance. We believe in a balance between simplicity and complexity. Depending on the type of extraction, the right annotation scheme can greatly improve the extraction. Clearly, _IOB performed better. It was complex with the non-aspect labels yet simple with the aspect label.

- *Research Question 2:* What are the strengths and limitations of different annotation schemes for aspect term extraction?

  Research Answer 2: The strengths of IO and _IOB are that they have singular label for aspect terms. The limitations of the other schemes are the level of complexity for this task limits the models' ability to extract. It is learning too much that skews the results.

- *Research Question 3:* What are the challenges in designing an annotation scheme for aspect term extraction?

Research Answer 3: Datasets are complex and this dataset was multi-domain. An annotation scheme has to work if the user uses a noun, adjective, and noun or verb. There is a variety of semantic combinations and locations therefore, the annotation scheme has to be simple enough to find simple enough to identify the aspect but complex enough to know which words are not aspects.

In summary, we argue that the aspect term extraction process can be bolstered by annotation schemes for ABSA. More specifically, the experimental results confirm that our _IOB delivers promising performance in terms of extracting aspects on the dataset. Our findings unveil that the RoBERTa classifier performs better than the competitors in terms of classification. We are hopeful that in the foreseeable future, researchers will build off of our research to shed bright light and insight on aspect term extraction.

Chapter 5

AI-Enabled Misinformation Detection and Research Opportunities

This chapter is to highlight the opportunity for Aspect-Based Emotion Analysis within the Misinformation Detection task by discussing the evolution of AI-enabled techniques within misinformation detection. We believe that ABSA can be a great resource and opportunity for misinformation detection. Misinformation has become increasingly prevalent on the Internet by the day and progressively more threatening. Individuals that are inaccurately informed tend to make misinformed decisions which have led to voting scandals, traffic accidents, and even health concerns. We are motivated to address a research gap by analyzing misinformation detection's overall progress and exposing the weaknesses that provide research opportunities. Our findings will further advance the work of misinformation detection with ABSA as an opportunity to fill the gap. Notably, we discuss the significance of misinformation detection systems and present the problems resulting from misinformation, the techniques for detection, and open issues within this research. Misinformation is becoming an issue that requires more attention and improved systems. We believe that our systematic review and synthesis of state-of-art research in this part of the dissertation will cultivate a path for these developments. Through our in-depth description of misinformation and techniques, we will expose the research opportunities.

Misinformation, which is fake or inaccurate information that is unintentionally spread, has become more pervasive on the Internet by the day [290]. The term *post-truth* is a word that describes the world we live in today. Post-truth is an environment where the facts are irrelevant compared to emotional appeals and personal beliefs; these more than facts are now shaping public opinions [195]. The dilemma that post-truth culture has created has become a critical issue since it has affected politics, voting, and the market. Publishing

and sharing misinformation has constantly misled the viewpoints, beliefs, and opinions of online users. People today are more inclined to believe any information online, thereby making users vulnerable. As our society advances with technology and Internet interactions, misinformation detection systems will be vital so that individuals can distinguish legitimate information from misleading ones. James Gleick once said, "when information is cheap, attention becomes expensive". Therefore, we must be assiduous in our pursuit of detecting misinformation so that users can safely continue to use the Internet as a safe and effective tool. In this dissertation, we shed a bright light on cutting-edge techniques that can detect various types of misinformation on a wide range of platforms.

We start this chapter by categorizing the versatile types of misinformation, including fake news, fake reviews, misinformation in microblogs, rumors, satire, and clickbait. We also discuss an array of challenges and future research opportunities. More specifically, gathering data for fake news, fake reviews, and misinformation in microblogs can be challenging because we have to acquire a large ground truth dataset to train models to detect misinformation. Another issue is the lack of misinformation detection research in non-English languages. Building datasets and creating detection methods pose a difficult task. We review specific methods that can improve misinformation detection like natural language processing, machine learning, and data mining techniques. In Section 5.1, we give an overview of the types of misinformation. In Section 5.2, we give an overview of the classification strategies along with their strengths and weaknesses. In Section 5.3, we lay out the research opportunities and propose potential solutions for those open issues. This comprehensive review is helpful for future research as a resource for successful solutions and possible ideas.

Within this section, we discussed the various types of misinformation and the adverse effects on internet users. Figure 5.1 outlines the definitions and motivations of the six misinformation types, which spur our efforts to survey cutting-edge methods adopted in a growing number of modern detection systems elaborated in the next section.

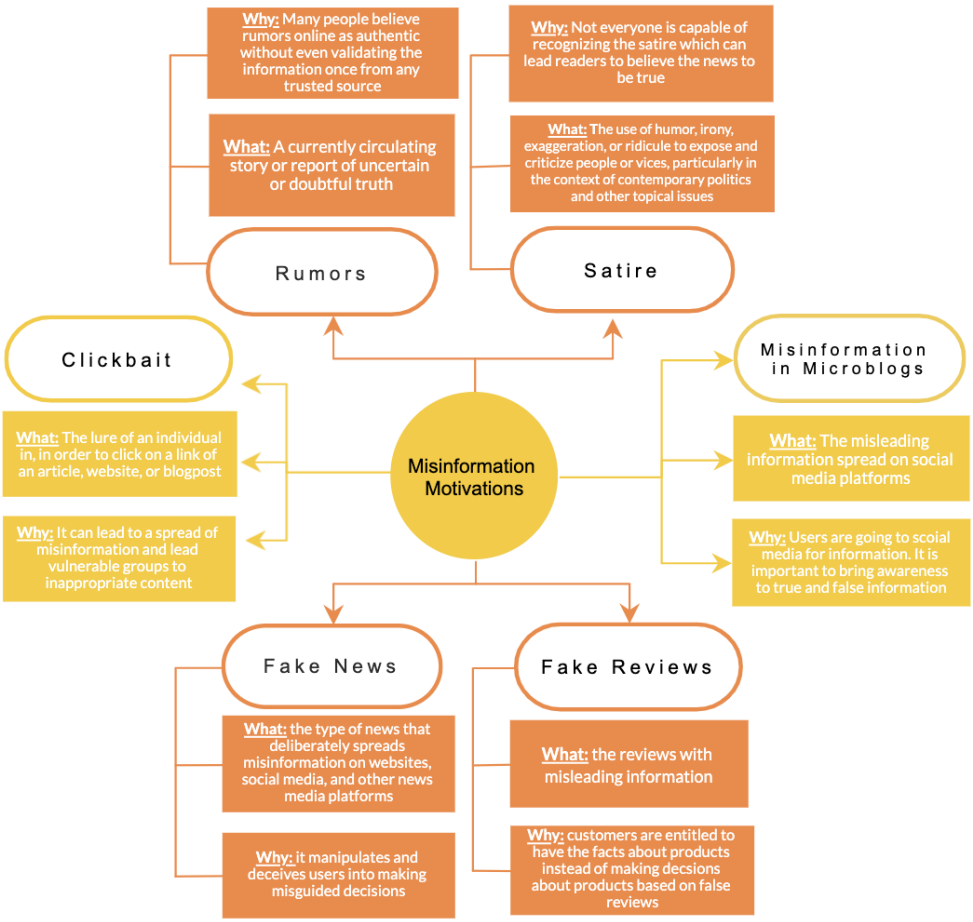

Figure 5.1: A variety of misinformation types along with the corresponding motivations to build detection systems

5.1 Misinformation Types

Oxford dictionary defines misinformation as false or inaccurate information, especially that which is deliberately intended to deceive. Misinformation exhibits a wide variety of formats which we shed light on in this section. The prominent six types of misinformation are fake news (see Section 3.1), fake reviews (see Section 3.2), misinformation in microblogs (see Section 3.3), rumors (see Section 3.4), satire (see Section 3.5), and clickbait (see Section 3.6). In each section, we define the type of misinformation, explain the negative effects, and explain why there is a need for better detection systems.

5.1.1 Fake News

Fake News is a type of news that deliberately spreads misinformation on websites, social media, and other news media platforms. It influences many people on a wide range of subjects, from politics to health to advertisement. Users find it challenging to distinguish authentic news articles from fake news articles. Moreover, more often than not, fewer people check for the validity of the information that they consume. The issue is growing worse due to the escalating number of users of social media and online forums [6]. Some people intentionally spread counterfeit articles by way of social media [7]. It is the ordinary users that generate the most root tweets about fake news. The primary purpose of spreading fake news is to influence popular belief on specific issues [8]. There is even a growing number of fake news posts that capture more views than real news. The spread of fake news on social media became a concern after the 2016 election. There is a large percentage of registered voters on Twitter that engage with fake news. The engagement is concerning because the democratic system only works effectively if the citizens are correctly informed [9]. Fake news can directly affect people and corporations due to false information spreading about their products or names [10, 11]. This dangerous problem manipulates and deceives users into making misguided decisions. Fake news detection systems help users become aware of potentially deceptive news. Predictions of fake news are based on the analysis of proven truthful and deceptive news [12]. There are linguistical differences between fake and legitimate news content that aids with detection [13]. According to a recent study (see also [12]), there are at least three distinct sub-tasks in fake news detection: a) fabrication, b) hoaxing, and c) satire detection. The ultimate goal of fake news detection is to free individuals to enjoy and utilize the Internet and social media effectively without fear of being misled. The detection systems protect people from hoaxes, exaggerations, and misleading opinions mascaraed as facts. People can then consume accurate information about health, events in the world, politics, and everything in between. As such, there is an increasing demand for verification systems that automatically process information.

Fake News Techniques. Among N-gram models, unigrams and bigrams are two of the most frequently used NLP methods in detecting fake news. Fake news detection is quite similar. Bhatt et al. used a unigram model to help convert the text to vectors for statistical information [54]. These features compare similarities between an article's body and its title and classify the article's stance. The stance then leads to the detection of fake news. Unigram and bigram models play a critical role in identifying misinformation in microblogs. With regards to detecting fake news, Qian et al. applied the skip-gram algorithm to create sentence representations and thereby an entire article representation [86]. That representation trained a neural network for detection. When it comes to fake news, Miller et al. strive to classify articles by analyzing headline article relations. Miller and his collaborators used each word in the article body and headline and obtained 100-dimension to 300-dimension pre-trained GloVe word embeddings to build an embedding matrix [90]. Ordinal has multiple categories with orders, an example of which are test grades ranging from A to F. Bourgonje et al. created a system to detect the stance of headlines by looking at the text in the articles to detect fake news detection or clickbait. In this study, they used a Logistic Regression to classify headline article pairs into the category of agree, disagree, or discuss. The Multinomial LR has the best results compared to a binary classifier [6]. Bhatt et al. tackled a subtask of fake news identification, stance detection. They looked at headlines and articles' bodies to detect stances of agree, disagree, discuss, or unrelated. A deep recurrent model produced the neural embeddings, an n-gram model produced the statistical features, and they hand-crafted a few external features. Then, a deep neural layer is responsible for combining all the features. This model classified the headline-body news pairs as agree, disagree, discuss, or unrelated. The evidence indicates that this novel model outperforms all the state-of-the-art techniques [54]. Miller et al. had success in classifying fake news articles using multiple bidirectional LSTMs and an attention mechanism to predict [90]. Nigam et al. developed a method for automating fake news detection for various events using classification. In this approach, a classifier predicts whether a piece of news is fake based on data sources, using a purely NLP

perspective. They used Twitter data and extracted key phrases from the news events to use as input to the detection system. After preprocessing, the model classified the data using features like sentiment score, number of tweets, number of followers, number of hashtags, user verification, and retweets. The Naive Bayes classifier was employed to detect fake news [201]. Zhang et al. piloted the aforementioned idea and developed a system called FEND. The clustering technique was seamlessly integrated with the text analytical approaches in FEND to detect fake news [203].

5.1.2 Fake Reviews

Fake reviews are reviews that contain misleading information. Many online consumers habitually read reviews prior to making any purchasing or investment decisions to verify the purchase's value. It is prudent for users to do so. This tendency has given other individuals a vulnerability to exploit by writing fake reviews to influence consumers. There are a number of reviewers who frequently post these fake reviews on online shopping sites. Some reviews aim to make people believe that a product is terrific and thereby gain more customers. However, other reviews intend to damage the reputation of a product by discouraging customers from purchasing it. The fake reviews problem has escalated in recent years. Some companies hire fake reviewers (or opinion spammers) that post fake reviews to promote or discredit products or businesses. Deploying fake review detection systems is a pressing task; the reason is two-fold. First, users and customers deserve facts about products, restaurants, and businesses. Second, business owners deserve to have the truth about their businesses and products spread. Surprisingly, evidence shows that companies are hiring these fake reviewers to write convincing fake reviews; it is now a business decision that companies invest in [14]. A *paid poster* is an additional title for individuals who work to post fake comments, reviews, and articles. These paid posters intend to influence the opinion of other people towards certain social events or business markets [15]. This new business of fake reviews can negatively affect various online communities. These businesses give their fake reviewers structure for what to

write and when to post their reviews. Fake reviewers earn more money if their well-written reviews receive more attention than the other fake reviews. Detection experts discovered various features that facilitate detecting fake reviews. They discovered that there are some linguistic differences between fake and genuine reviewers [16]. Sample features indicating signs of a fake review include, but are not limited to, review creation time, review ratings, and duplicated reviews. Supervised learning has been one of the main approaches for solving the detection problem [16]. Although a handful of fake review detection algorithms appear in the literature [17, 18], detecting fake reviews faces a few grand challenges. First, it is a challenging task to gather or simulate fake reviews. Synthesized fake reviews are unlikely to resemble genuine fake reviews. Secondly, paid opinion spammers tend to produce very believable fake content [14]. Third, sometimes fake reviewers are even given a template for how to write convincing reviews [15]. The issues mentioned above are becoming more and more sophisticated and, therefore, more challenging to detect fake reviews. Nevertheless, the task of detecting fake reviews has become more and more necessary.

Fake Reviews Techniques. In the realm of fake review detection, n-grams can discover similarities between fake reviews and authentic reviews. Sometimes there are certain words that fake reviewers commonly use or the review is almost identical to another. Mukherjee et al. used unigrams and bigrams as feature sets [65]. These word embeddings governed the initialization of the weights in the model's embedded layer, which in turn helped classify the articles. Barry utilized word embeddings in the research on fake review detection to evaluate the importance of word order in sentiment classification [94].

5.1.3 Misinformation in Microblogs

Misinformation is widespread on microblogs [19], which are popular mediums to share thoughts or instant messages quickly. Microblogs, most of which are considered social media, have become compelling grounds for public relations, marketing, and politics. More people are on microblogs getting information than learning information from in-person interactions,

newspapers, or television [20]. In today's world, online social media plays a vital role during real-world events, especially during crisis events. Authorities can use social media can be used during events for disaster management or by destructive groups to spread misinformation [21]. Social spammers, one category of misinformation in microblogs, are constantly on social media and websites spamming comment sections and inboxes. There is a pressing need for the automatic detection of spammers in social media [22]. To make misinformation more legitimate, spammers are getting into link farming, where spammers gain an excessive number of followers on social media to increase credibility. Link farming offers spammers the power to misinform more users and seem more legitimate [23]. Another form of misinformation on microblogs is astroturfing- an act of disguising the sponsors of a message or organization to make the message appear as if it came from participants. Astroturfing for political campaigns on microblogging platforms is very common. For political motives, individuals or organizations use multiple accounts to create the appearance of widespread support for a candidate or an opinion [24]. Twitter is a breeding ground for misinformation. Twitter is frequently the place where many news stories first break. However, a few issues surface with Twitter as a news source. First, many users are competing to be the first to post a newsworthy story. Secondly, there is a lack of structure and filtering on Twitter. These issues lead people to post quickly without first verifying the news. So, there is a reduction in the accuracy of the information spread, which is a significant issue [25]. Retweets are what make misinformation spread so rapidly. For example, during Hurricane Sandy in 2012, many people spread fake images on Twitter about the disaster. A raft of fake images of the hurricane circulated on Twitter; most of those tweets were retweets [21]. The fast-expanding world of social media fuels the spreading of misinformation, thereby disrupting people's ordinary lives. People are creating new ways of spreading misinformation in microblogs every day. As such, it is urgent to aid misinformation detection in microblogs and to implement early detection in social media [26]. These detection systems can enable users to make informed decisions while on social networks [27]. In a nutshell, social media

development has revolutionized the way people communicate, share information, and make decisions, but it also provides an ideal platform for publishing and spreading misinformation [28]. Consequently, an increasing number of misinformation detection systems were devised and tailored for microblogs. For example, Qazvinian et al. tested the effectiveness of the three categories of features for detecting rumors: content-based, network-based, and microblog-specific memes for correctly identifying misinformation [29]. A detailed analysis of the cutting-edge techniques for pinpointing misinformation in microblogs is in Section 4. Misinformation in microblogs is a deep and complex topic, but tackling this issue will genuinely help social media consumers.

Misinformation in Microblogs Techniques. Unigram and bigram models play a critical role in identifying misinformation in microblogs. Qazvinian et al. gathered unigrams and bigrams of the text in the tweets and the URL from the tweets [29]. Four features train the unigram and bigram models to detect misinformation in tweets. Similarly, Alkhodair et al. used skip-grams for the detection of misinformation in microblogs [88]. A skip-gram model takes in a corpus of text, which converted the sequence of words into a sequence of vectors. Those vectors are then used as input into an LSTM neural network to predict which class of misinformation the text fell into [88].

5.1.4 Rumors

According to the Oxford dictionary, a rumor is a currently circulating story or report of uncertain or doubtful truth. Alternatively, a rumor is a statement whose true value is unverifiable [30]. Evidence shows that rumors are ubiquitous on the internet [31, 32, 236]. News articles, websites, and social media constantly spread rumors that are true, partially true, or completely false. Nevertheless, rumors - aiming to mislead and misinform readers -can dangerously affect our world. Most people have started believing information online is authentic without even validating the information once from any trusted source. Individuals have the right to know whether the data they are receiving is reliable or not.

Many times, information shared at the time of emergencies originated from an inaccurate information source. Spreading misinformation then has an enormous amount of negative impact [33]. False rumors are damaging as they cause public panic and social unrest. For example, in 2015, a rumor was spread of shootouts and kidnappings by drug gangs near schools on Twitter and Facebook. This rumor caused severe chaos in the city, leading to several car crashes by parents rushing to pick up their children from school. More recently, misinformation about COVID-19 has become a challenging issue on social media during the 2020 pandemic [34]. There was false information about cures, laws, and conspiracy theories. During a pandemic, people typically begin to panic and become more desperate for any information. The desperation leads to the indulgence of misinformation on social media. Rumors about COVID-19 have confused people, making it more challenging for healthcare professionals to spread accurate information. Thereby, automatically predicting the accuracy of information is becoming increasingly paramount. Debunking rumors at an early stage of diffusion is particularly crucial to minimizing their harmful effects [31]. Identifying rumors, especially for breaking and developing news events, is a challenging task due to the absence of sufficient information about the exact rumor stories [35]. Fortunately, people tend to stop spreading a rumor if they learn that the information is not true [36]. We already have sites like snopes.com and factcheck.org that check for rumors and misinformation. However, detecting rumors concerning newly released information is still an open and challenging issue [35].

Rumors Techniques. Bahuleyan and Vechtomova focused on detecting rumors related to breaking news. More specifically, a unigram model is trained by scanning Twitter data to determine the stance of a tweet [35]. In a rumor-detection study [28], Yuan et al. analyzed tweets, retweets, source tweets, and user connections to identify rumors. They mapped the word embeddings to see the proximity of user tweets. All of the word embeddings, trained by a skip-gram model, are initialized with the 300-dimensional word vectors. The trained

embeddings are classifiable. Yu et al. address misinformation and rumors detection by applying paragraph vectors [26]. The goal of this study was the early detection of misinformation about events on microblogs. The microblog posts cluster based on their similar events, and then the paragraph vector develops each event's representation. Sentiment analysis detects an underlying specific tone that rumors commonly have. Algorithms focused on sentiment analysis rely on a classifier trained from a collection of data, and then the classifier deploys to process real data [126]. Ma et al. explored a way of applying RNNs to detect rumors within microblogs. The RNNs are used to learn the hidden representations of differing information in posts on microblogs over a certain amount of time. In this work, they collected rumor posts over a length of time and comments as source data. With multiple hidden layers, the RNN trains to detect if a rumor was true or untrue [31]. Jin et al. explored a way of adopting TF-IDF for rumor detection [232]. In this intriguing study, true rumor articles are in the form of vector representation. Each element in the vector has a TF-IDF score. TF-IDF can be useful for text matching, where both tweets and verified rumor articles are in the form of a v-dimensional vector. The v values in v-dimensional vectors are the size of a corpus dictionary. The TF-IDF scores in each vector represent the tweet or verified rumor texts. After the TF-IDF-enabled system computes the similarity between a tweet and a verified rumor, it returns a matching score. The experimental results suggest that the TF-IDF scheme outperforms the conventional semantic-embedding and lexicon-based methods in detecting rumors [232].

5.1.5 Satire

Oxford dictionary states that satire is the use of humor, irony, exaggeration, or ridicule to expose and criticize people or vices, particularly in the context of contemporary politics and other topical issues. Unlike comedy which evokes laughter as an end in itself, satire mocks with laughter [37]. News satire is a genre of deceptive news that intends to dispense satire in the form of legitimate news articles. These articles differ from "fake news", which intends

to mislead people by providing untrue facts. However, satirical news ridicules and criticizes something through satirical comments or fictitious stories. News satire, in which writers want the user to discover the falseness, incorporates irony in an attempt to provide humorous insight [38]. Satire is a type of implicit communication that intentionally reveals cues that show that it is untrue. Other types of fabrications aim to instill a false sense of truth in the reader [39]. One specific type of satire is sarcasm, a speech act where speakers implicitly convey their message. It is a sophisticated form of speech act that online writers widely use. Due to its ambiguity, sarcasm is sometimes difficult for humans to detect. Sarcasm recognition can be very beneficial in Natural Language Processing applications like dialogue systems or customer review analysis. In textual communication, it is more challenging to detect the sentiment of sarcasm due to the absence of non-verbal cues that might hint at the sarcasm [40]. Satire detection is essential [41] because not everyone can recognize satire, and this problem could lead readers to believe the news to be trustworthy. The satire could be too subtle, or the readers may not have the contextual or cultural background to understand [42]. Satirical news, embracing unique features, has a similar format and style as journalistic reporting, which helps with detection [39]. The ability to appropriately detect and deal with satire is enormously beneficial to a wide variety of fields [43]. For example, Rubi et al. created an algorithm to identify satirical news in order to minimize the potential deceptive impact of satire, thereby scaffolding sentiment analysis applications [39]

Satire Techniques. When it comes to satire, detection is possible by analyzing headlines. For example, Burfoot and Baldwin leveraged unigrams to find tokens that help classify an article as satirical [38]. Satire articles commonly use profanity and slag which are not typical for authentic news articles. In the realm of satire detection, Barbieri et al. deployed skip-grams of sizes 1 and 2 as word-based features for detecting satire between languages including English, Spanish, and Italian [43]. At the heart of this study, attention is the key to pinpointing the most useful paragraphs to detect satire. Attention is a machine learning method used to focus on important details in a sentence or an image. The findings from using

the attention method suggest that psycho-linguistic, writing stylistic, and structural features are the most beneficial for detection at the paragraph level. Yang explored 100-dimension GloVe embeddings for satire detection [42]. Satire is a popular form of misinformation that can be detected using sentiment analysis [37]. one research opportunity is using K-nearest neighbor (KNN) to detect satire. Rubin et al. built an SVM-based algorithm to detect satire with five predictive features, namely, absurdity, humor, grammar, negative affect, and punctuation. Rubin's team tested various combinations on several news articles. The empirical results suggest that the best predicting feature combination is absurdity, grammar, and punctuation, which detects satirical news with 90 percent precision and 84 percent recall. This system can aid in minimizing the potential deceptive impact of satire [39]. Similarly, Nafis et al. advocated SVM classification for detecting satire. The classifier analyzed the data and detected patterns useful for classification, and then predicted which class the input should be classified into [40]. In regard to satire detection, Ahmad et al. attempted to incorporate TF-IDF as a feature extraction strategy into an SVM-based detection system. The TF-IDF features trained the classifier, which they built using the LIBSVM classification tool to predict the article's satire or validity. The accuracy of the system was dependent on the type of feature extraction method [233].

5.1.6 Clickbait

Clickbait is a method of luring an individual to click on a link to an article, website, or blog post. Many times, clickbait is a scam with a compelling title that typically does not match the content information. Clickbait employs the cognitive phenomenon known as the curiosity gap, where the headlines make the reader curious enough to click the link and fill their curiosity gap [44]. Loewenstein's information gap theory of curiosity explains that people feel a gap between what they know and what they desire to know. First, there is a painful gap in knowledge that the headline provides, and then there is the need to fill this gap by clicking on the material [45]. Clickbait eventually causes disappointment because the

content rarely lives up to the expectation made within the headline. With the enormous amount of clickbait online, it is becoming increasingly important to develop techniques that automatically detect and combat clickbait [46]. Social media has recently become a platform for news sources and has led to the rise of clickbait posts. Click-baiting is the new marketing technique [47] on social media to acquire popular views. Publishers desire increased traffic on their web pages in order to boost income from their advertisements. Generally, clickbait utilizes exciting words and exaggerated tones to attract people [48]. Clickbait generally falls into two categories, namely, ambiguous and misleading [49]. Here are four clickbait title examples: 1. You Won't Believe What These Dogs Are Doing! [50] 2. What OJs Daughter Looks Like Now is Incredible! [50] 3. If you've ever used Google Docs for anything important, you should know about this [51] 4. The Hot New Phone Everybody Is Talking About [45] There is a pressing need to ensure the safety of social media usage. In particular, various social groups view the spread of adult content in social networks as undesirable. This content may even pose a serious threat to other vulnerable groups like children [52]. Besides, clickbait tends to cause fake news to spread on the Internet since many users forward clickbait without reading the contents. Therefore, it is crucial and demanding to develop technologies to detect clickbait [48].

Clickbait Techniques. Now clickbait (see, for example, [45]) is also detected using unigrams and bigrams. Similarly, in a clickbait detection study [103], Pandey et al. employed 300-dimensional word vectors in a BiLSTM model in combination with glove embeddings to achieve a very high detection accuracy. Ordinal has multiple categories with orders, an example of which are test grades ranging from A to F. Bourgonje et al. created a system to detect the stance of headlines by looking at the text in the articles to detect fake news detection or clickbait. In this study, they used a Logistic Regression to classify headline article pairs into the category of agree, disagree, or discuss. The Multinomial LR has the best results compared to a binary classifier [6].

5.2 Misinformation Detection Methods

Now we are positioned to shed a bright light on methods of detecting misinformation. With the flood of information arising from the Internet, traditional fact-checking and deception detection is inadequate. We are in pressing need of strong detection techniques because humans cannot detect misinformation on their own. There are two types of approaches to detection, namely linguistic approaches and network approaches. The linguistic approach extracts and analyzes the content of a deceptive message to associate language patterns with deception. In a network approach, network information (e.g., message metadata or network queries) can provide information on deception. Both types of misinformation detection schemes incorporate machine learning techniques [76], and include: natural language processing (see Section 5.2.1), machine learning (see Section 5.2.2), and data mining (see Section 5.2.3). Figure 5.2 depicts a taxonomy tree organizing all the misinformation detection techniques articulated in this section.

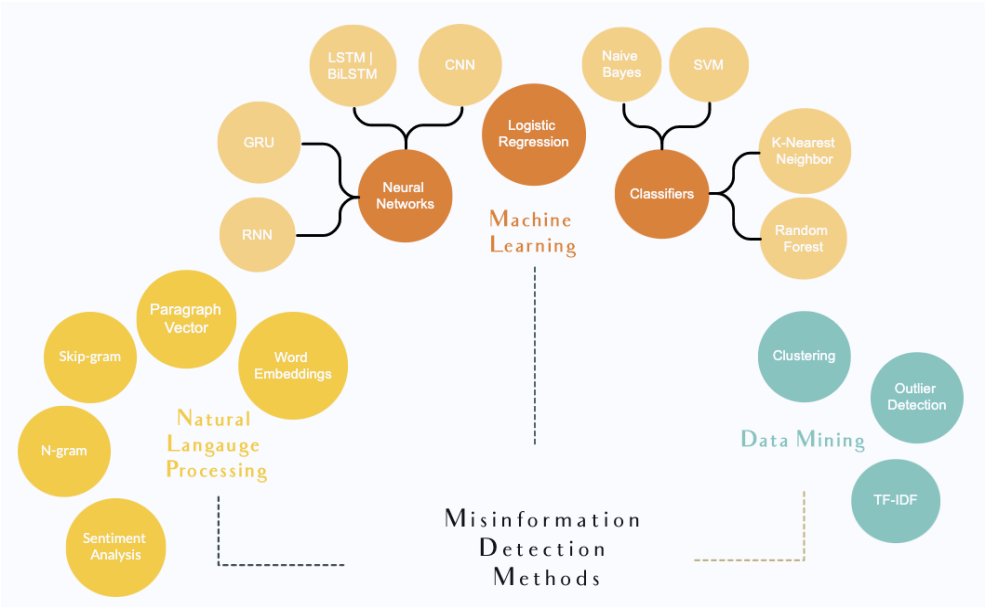

Figure 5.2: The taxonomy tree of the cutting-edge misinformation techniques.

5.2.1 Natural Language Processing

Natural Language Processing (NLP) plays a crucial role in detecting misinformation. NLP, in combination with Machine Learning (ML), creates state-of-the-art detection systems. *Natural language processing* is an interdisciplinary field of computer science and linguistics using many machine learning algorithms. NLP furnishes an algorithmic method for computers to process and understand natural language. Generally speaking, four distinctive modules in NLP systems are data gathering, preprocessing data, training, and testing. There are several ways to gather data, including, but not limited to, web scraping, crowdsourcing, and open-source datasets. Preprocessing techniques, like tokenization, stemming, lemmatization, and parts-of-speech (POS) tagging, manipulate raw data to be further analyzed in optimized ways. Language models are one of the technical underpinnings in NLP used to understand the input enough to predict the following word given an input word. There are two types of language models: neural and statistical models. Neural language models use neural networks to predict the following letter, word, or phrase. Statistical language models like n-grams (see Section 5.2.1) are used to predict what word, letter, or phrase comes next in the sequence given the previous content. Depending on the desired goal, data then are divided into testing datasets and training datasets. The training data is input for classifiers, clusters, or neural networks. Model development happens in the training phase, and then the model tests by using the testing data. We can then analyze and make adjustments to the model based on testing results. All of these tools are very useful for linguistic understanding that leads to the detection of misinformation.

In what follows, we elaborate on the most popular NLP techniques for misinformation because NLP techniques exhibit strengths in detecting fake news. The data preprocessing phase uses NLP methods to understand the data before applying machine learning (see Section 5.2.2) or data mining (see Section 5.2.3) procedures. Table 5.1 summarizes all the NLP techniques applied to detect misinformation in the six domains. The five empty cells in

Table 5.1 suggest open issues and research opportunities. For instance, there is a potential opportunity to utilize paragraph vector schemes to detect fake news and reviews.

	Fake News	**Fake Reviews**	**Misinformation in Microblogs**	**Rumors**	**Satire**	**Clickbait**
Unigram	[50], [266], [14],[157], [130], [203], [159], [84], [134], [205], [315], [222]	[206], [155], [112], [175], [249], [281], [39], [32], [91], [27]	[230], [133],[189],[132], [167], [190], [285]	[230], [133], [189], [132], [34], [190]	[60], [207], [113], [295], [268], [120]	[314], [77], [282], [22], [184], [215], [238]
Bigram	[266], [157], [48], [217], [125]	[206], [155], [112], [249], [281]	[230], [133], [189], [132], [285]	[230], [133], [189], [132]	[113], [295], [268], [42]	[314], [22], [184], [215]
Skipgram	[231], [125], [297]		[304]	[20], [304], [126]	[42]	[184], [238]
ParagraphVector			[303]	[303]	[295]	
Word Embeddings	[199], [64], [72], [72], [225]	[44], [218], [182]		[3], [69], [17], [51], [223], [275]	[295]	[214], [77], [142], [317], [7], [178]
Sentiment Analysis	[242], [159], [315], [252],[137],[127]	[155],[250], [220], [302], [101]	[252], [285], [169], [144]	[202], [256]	[89]	[282], [215]

Table 5.1: Language Models/Text Processing Comparison Table

N-Gram

An N-gram model is a statistical language model, which aims to learn the probability function of sequences of words. The model can simultaneously learn the similarity between any pair of words and the probability of word sequences. The goal of n-grams is to make predictions by determining the probability of the next word. N-gram models can help a wide range of applications, including speech recognition, language translation, and information retrieval, to name just a few [47]. For instance, a sequence-to-sequence framework makes use of N-gram models to predict the following sentence given the previous sentence or sentences [276]. Among N-gram models, *unigram* and *bigrams* are two of the most frequently used NLP methods in detecting fake news, fake reviews, misinformation on microblogs, and clickbait. A unigram (a.k.a., Bag-Of-Words or BOW) is a type of statistical language model that analyzes each word's number without considering word order. On the other hand, a bigram analyzes the distribution of a sequence of words by looking at each two-word sequence.

The model assigns a numeric value once it calculates the probability for each word or word sequence.

In the realm of fake review detection, n-grams can discover similarities between fake reviews and authentic reviews. Sometimes there are certain words that fake reviewers commonly use or the review is almost identical to another. Mukherjee *et al.* used unigrams and bigrams as feature sets [206]. These features are used for classification, to classify a review as fake or non-fake. Fake news detection is quite similar. Bhatt *et al.* used a unigram model to help convert the text to vectors for statistical information [50]. These features compare similarities between an article's body and its title and classify its stance. The stance then leads to the detection of fake news. Unigram and bigram models play a critical role in identifying misinformation in microblogs. Qazvinian *et al.* gathered unigrams and bigrams of the text in the tweets and the URL from the tweets [230]. Four features train the unigram and bigram models to detect misinformation in tweets. Similarly, Bahuleyan and Vechtomova focused on detecting rumors related to breaking news. More specifically, a unigram model is trained by scanning Twitter data to determine the stance of a tweet [34].

When it comes to satire, detection is possible by analyzing headlines. For example, Burfoot and Baldwin leveraged unigrams to find tokens that help classify an article as satirical [60]. Satire articles commonly use profanity and slag which are not typical for authentic news articles. Now clickbait (see, for example, [25]) is also detected using unigrams and bigrams. After employing unigrams and bigrams as features for the Naïve Bayes classifier [25], Anand *et al.* compared article headlines to classify articles. The findings indicate that the headlines are too short for bigrams.

Skip-Gram

Skip-Gram is another type of language model that predicts the nearby words given a target word. This model implements the concept of *distributional hypothesis*, which suggests that the more semantically similar two words are, the more distributionally similar they

will be; therefore, the words will occur in similar linguistic contexts. A skip-gram is the reverse of *Continuous BOW (CBOW)* where the context words in a sentence are the input, predicting the target word. The skip-gram model advocates for a simple neural network with one hidden layer to obtain the context words' probability. These weights used to correctly predict the context words are what make up the word vectors. When a word vector is on a graph, the words with similar meanings are closer together, and the words with different meanings will be further apart. The skip-gram model learns word vector representations that are good at predicting nearby words [198].

With regards to detecting fake news, Qian *et al.* applied the skip-gram algorithm to create sentence representations and thereby an entire article representation [231]. That representation trained a neural network for detection. Similarly, Alkhodair *et al.* used skip-grams for the detection of misinformation in microblogs [20]. A skip-gram model takes in a corpus of text, which converted the sequence of words into a sequence of vectors. Those vectors are then used as input into an LSTM neural network to predict which class of misinformation the text fell into [20].

In a rumor-detection study [304], Yuan *et al.* analyzed tweets, retweets, source tweets, and user connections to identify rumors. They mapped the word embeddings to see the proximity of user tweets. All of the word embeddings, trained by a skip-gram model, are initialized with the 300-dimensional word vectors. The trained embeddings are classifiable. In the realm of satire detection, Barbieri *et al.* deployed skip-grams of sizes 1 and 2 as word-based features for detecting satire between languages including English, Spanish, and Italian [42]. Bojanowski *et al.* created a skip-gram model-based approach to generate word representations to evaluate word similarity and analogy tasks in nine different languages. The vectors trained by the skip-gram model achieve state-of-the-art performance on word-similarity analysis [54]. The aforementioned skip-gram models can be adopted to devise rumor and satire detection systems. There are several ways to use skip-gram for detection.

Paragraph Vector

A paragraph vector is an unsupervised algorithm that learns fixed-length pieces of text such as sentences, paragraphs, and documents [172]. A vector representation of those pieces of text is used for training to predict words in paragraphs. Although the paragraph vector is not as popular as the previously mentioned NLP techniques (see Sections 5.2.1 and 5.2.1), the paragraph vector is known as an effective way of solving misinformation detection problems.

Yu *et al.* addresses misinformation and rumors detection by applying paragraph vectors [303]. The goal of this study was the early detection of misinformation about events on microblogs. The microblog posts cluster based on their similar events, and then the paragraph vector develops each event's representation. In another study, Yang *et al.* used paragraph vectors to investigate the difference between paragraph-level features and document-level features on satirical news detection [295]. At the heart of this study, *attention* is the key to pinpointing the most useful paragraphs to detect satire. *Attention* is a machine learning method used to focus on important details in a sentence or an image. The findings from using the attention method suggest that psycho-linguistic, writing stylistic, and structural features are the most beneficial for detection at the paragraph level.

Word Embeddings

Global Vector or GloVe is an unsupervised learning algorithm that generates word embeddings. The phrase *word embeddings* and *word vectors* are used interchangeably. GloVe word embeddings capture the global and local statistics of a corpus. In order to build the word embeddings, there are calculations. It calculates the frequency of words that co-occur with one another in a given corpus [221]. We witness that many use GloVe word embeddings in misinformation detection techniques. For example, Yang explored 100-dimension GloVe embeddings for satire detection [295].

When it comes to fake news, Miller *et al.* strive to classify articles by analyzing headline article relations [199]. Miller and his collaborators used each word in the article body and

headline and obtained 100-dimension to 300-dimension pre-trained GloVe word embeddings to build an embedding matrix. These word embeddings governed the initialization of the weights in the model's embedded layer, which in turn helped classify the articles. Barry utilized word embeddings in the research on fake review detection to evaluate the importance of word order in sentiment classification [44]. Researchers trained the GloVe embeddings on 42 billion tokens, with a vocabulary of 1.9 million words, so these vectors had a dimension of 300. The vectors then served as input to train the LSTM network. Finally, the LSTM model was able to determine the fidelity of a given review [44].

Abulaish *et al.* began their approach to rumor detection with a small set of seed rumor words, accompanied by applying word embedding to represent each category of seed words as numeric vectors [3]. They trained three different classification models for rumor detection using the 200-dimensional GloVe word vectors trained over the Twitter dataset. The evidence shows that this approach outperforms a state-of-the-art rumor detection method. Similarly, in a clickbait detection study [214], Pandey *et al.* employed 300-dimensional word vectors in a BiLSTM model in combination with glove embeddings to achieve a very high detection accuracy.

Sentiment Analysis

Sentiment Analysis (a.k.a, *Opinion Mining*) is one of the most widely studied topics in NLP because of its goal to understand the sentiment in written text [121][201]. Sentiment analysis aims to classify the sentiment of the text as positive, negative, or neutral. Therefore, this technique can gather an understanding of the opinion and detect the emotion behind the data.

The following three steps summarize a generic preprocessing procedure. First, we tokenize the text into single words. Second, we use n-grams, POS, stemming, spelling, semantic analysis, abbreviations as word processing methods. We gather emoticon dictionaries and acronym dictionaries are useful when data from media platforms. Researchers typically

build sentiment lexicons which essentially are dictionaries of words for sentiment analysis. Third, we analyze the words to identify a given text's polarity level: positive, neutral, or negative, which is the sentiment. The third step can be accomplished by clustering [191], classification [201], language models [88], or neural networks [212].

Growing evidence demonstrates that sentiment analysis can be a valuable tool for detecting misinformation. For instance, satire is a popular form of misinformation that can be detected using sentiment analysis [268]. Moreover, sentiment analysis detects an underlying specific tone that rumors commonly have. Algorithms focused on sentiment analysis rely on a classifier trained from a collection of data, and then the classifier deploys to process real data [26]. One of the biggest challenges in sentiment analysis lies in the dependence on languages. Tools like word embeddings, sentiment lexicons, and annotated data are language-specific. It is a time-consuming task to constantly have to optimize models for a variety of different languages [61]. Sentiment detection requires rich supervised training and evaluation resources and powerful models [258].

5.2.2 Machine Learning

Machine learning or *ML* is a subset of artificial intelligence that uses algorithms to analyze data and make predictions. The ML systems learn from training data by identifying patterns, which facilitate in making decisions and predictions. According to Choudrie *et al.*, Machine Learning is an effective technique in classifying COVID-19 information and COVID-19 misinformation [73]. There are three main branches of machine learning: supervised learning, unsupervised learning, and reinforcement learning. *Supervised learning* is an ML technique that trains a model for identification by giving the model data and labels to learn. Once the model learns which label matches which data, the model tests the unlabeled data. *Unsupervised learning* is another ML technique that analyzes data and finds hidden patterns in raw, unlabeled data. *Reinforcement learning* or *RL* is a technique of training a model to

make a sequence of decisions. RL is typically employed to teach a system to beat a game, learn a task, or make a decision.

Supervised learning is the most popular practice research for misinformation detection. For a supervised learning approach to misinformation detection to perform best, one ought to collect an extensive collection of accurate information and misinformation. Labeling the gathered data also takes an extensive amount of time and work. For these reasons, there is much room for growth in the unsupervised learning approaches to NLP problems [210]. Similarly, reinforcement learning applied to misinformation detection is still in its infancy.

In what follows, we discuss popular methods that are beneficial for detecting misinformation. The most commonly used classification models in detecting fake news are Support Vector Machine (SVM) and Naive Bayes Classifier (NBC). We review supervised learning approaches like regression and classification and review the usefulness of neural networks for detection. Table 5.2 holds the references to ten widely used machine-learning methods applied to six types of misinformation. With this table in place, we intuitively observe which methods are the most popular and effective. Table 5.2 also indicates which methods provide an opportunity for future research. For example, one research opportunity is using K-nearest neighbor (KNN) to detect satire.

Logistic Regression

Logistic regression or *LR* is a supervised learning algorithm widely adopted for classification. This algorithm implements the linear regression function. The logistic sigmoid function takes the output from the linear regression function and uses a logarithm to condense the output to a range between 0 and 1. There are multiple types of logistic regression: Binary, Multinomial, and Ordinal. *Binary* only has two options of classification. For example, accurate news (1) and fake news (0) are two options; another example is True (1) or False (0). *Multinomial* has three or more categories. For example, emotion categories include anger,

bad, disgust, fear, happiness, sadness, surprise. *Ordinal* has multiple categories with orders, an example of which are test grades ranging from A to F.

Bourgonje *et al.* created a system to detect the stance of headlines by looking at the text in the articles to detect fake news detection or clickbait. In this study, they used a Logistic Regression to classify headline article pairs into the category of 'agree,' 'disagree,' or 'discuss.' The Multinomial LR the best results compared to a binary classifier [58]. Shivagangadhar *et al.* took on the task of identifying fake and accurate online reviews. An LR classifier was trained on thousands of fake and accurate reviews using features such as unigram, bigram, and review length. The classifier was trained separately for each feature. The findings show that the Logistic Regression classifiers provide the highest accuracy in comparison to the other classifiers for detecting if a review was fake or true [249].

Neural Networks

Neural Networks are a collection of algorithmic solutions that vaguely imitate the architecture of the human brain. These networks are composed of pseudo neurons that allow them to detect patterns and solve artificial intelligence (AI) problems. Raw input enters the network, which detects patterns; the patterns ultimately allow for labeling or clustering of the data. There is a hoard of popular deep learning algorithms that use neural networks customized for misinformation detection. We discover that misinformation detection's most popular ones are recurrent neural networks, gated recurrent units, long short-term memory, bi-directional long-short term memory, and convolutional neural network.

(1) Recurrent Neural Network. The *Recurrent Neural Network* (RNN) is a machine learning algorithm that remembers past decisions and makes decisions based on what a model has learned. The previous output is the input in a hidden state. RNNs are very popular in the field of NLP. The algorithm reads the whole text from beginning to end, which is inefficient in processing long texts.

Bhatt *et al.* tackled a subtask of fake news identification, stance detection. They looked at headlines and articles' bodies to detect stances of agree, disagree, discuss, or unrelated. A deep recurrent model produced the neural embeddings, an n-gram model produced the statistical features, and they hand-crafted a few external features. Then, a deep neural layer is responsible for combining all the features. This model classified the headline-body news pairs as agree, disagree, discuss, or unrelated. The evidence indicates that this novel model outperforms all the state-of-the-art techniques [50]. Similarly, Ma *et al.* explored a way of applying RNNs to detect rumors within microblogs. The RNNs are used to learn the hidden representations of differing information in posts on microblogs over a certain amount of time. In this work, they collected rumor posts over a length of time and comments as source data. With multiple hidden layers, the RNN trains to detect if a rumor was true or untrue [188].

(2) Gated Recurrent Unit. *Gated Recurrent Unit* (GRU) is a type of recurrent neural network that utilizes gating, which controls what passes through and what gets blocked. There are two gates in a GRU, namely, update and reset. Similar to logistic regression, the sigmoid function is a centerpiece here. This algorithm decides when a hidden state should be updated and when the state should reset. GRUs can keep information for a long time while pruning information that is irrelevant to predictions. This algorithm is more advanced than the simple RNN model and can retain more relevant information and make better predictions.

Bajaj built a classifier that detects fake news based only on the content of source articles. In this pilot study, GloVe word embeddings are performing as an input into the GRU model. This approach was the most successful approach among the peers [35]. Furthermore, Liu *et al.* used a GRU and attention combination to create sentence vectors and word vectors. A word-level bidirectional GRU is an integral approach to determining the hidden representation of words. Similarly, the sentence-level bi-directional GRU originates the hidden representation of the sentences. Using these vectors in the model is proved successful in detecting spam in online reviews [184].

(3) Long Short-Term Memory and Bi-directional Long Short-Term Memory. *Long Short-Term Memory* (LSTM) is another type of RNN, which can remember information for long periods. LSTM is similar to GRU but even more complex because an LSTM has three gates: input, output, and forget. *Bi-directional Long Short-Term Memory* (BiLSTM) is an implementation of LSTM; however, the input sequence is read forward and then backward. BiLSTM learns to optimize prediction accuracy. This method is efficient and practical because sometimes, to understand a particular word, one must know the previous word and the upcoming word.

Miller *et al.* had success in classifying fake news articles using multiple bidirectional LSTMs and an attention mechanism to predict [199]. Conforti *et al.* tackled stance detection and compared headlines and articles. The proposed procedure constitutes four steps. First, the headline is encoded using a BiLSTM. Second, each sentence of an article converts into an embedded representation. Third, the article is then encoded using a BiLSTM. Finally, the article is read backward sentence by sentence. This method offers the model a better understanding of articles, thereby improving the predictions [75]. Similarly, Ajao *et al.* built a framework for detecting fake news on Twitter by a CNN-LSTM model for identifying features in a tweet. They added an LSTM model for sequence classification, followed by classifying the tweets as fake news or true news [16].

(4) Convolutional Neural Network. *Convolutional Neural Network* (CNN) is a neural network that contains several layers of convolutions applying activation functions to the results. CNN is popular in the computer vision or image processing field; recently, CNN is a powerful and prominent tool for NLP tasks. A matrix is used as input for the input layer to compute the output. Instead of image pixels, the input for NLP tasks is sentences or documents represented as a matrix. Each row of the matrix corresponds to one token, typically a word. Typically, these vectors are word embeddings like word2vec or GloVe.

Zhang *et al.* advocated for character-level convolutional networks for text classification. The findings unravel that convolutional networks can achieve what the SOTA models achieve.

They compared ConvNets to BOWs models, n-grams, and their TFIDF variants and deep learning models. ConvNet is an effective method, the performance of which largely depends upon data sizes [311]. In other research, Kim built a CNN model on top of a word2vec model and then tested seven NLP tasks' performance. The model performed better than the SOTA on four out of the seven tasks, including sentiment analysis and question classification. The unsupervised pre-training of word vectors is a vital piece [161]. Yu *et al.* proposed another detection approach using CNN. They present a convolutional approach referred to as CAMI for misinformation identification. The CAMI technique leverages CNNs to extract key features scattered among an input sequence to improve early detection. They used Large datasets to test for validation purposes; the results demonstrate that CAMI effectively detects misinformation and completes early detection tasks [303].

Classifiers

Classifiers, an essential branch of machine learning, classify data into usable categories to better understand the data. A few standard classifiers are Naive Bayes, support vector machine, k-nearest neighbor, and random forest. In what follows, we shed light on the application of state-of-the-art classifiers to detect misinformation.

(1) Naive Bayes Recall that *Naive Bayes* is a probabilistic machine learning algorithm that classifies based on the Bayes theorem where features are assumed to be independent of each other. Naive Bayes has become a popular solution when it comes to spam detection.

Nigam *et al.* developed a method for automating fake news detection for various events using classification. In this approach, a classifier predicts whether a piece of news is fake based on data sources, using a purely NLP perspective. They used Twitter data and extracted key phrases from the news events to use as input to the detection system. After preprocessing, the model classified the data using features like sentiment score, number of tweets, number of followers, number of hashtags, user verification, and retweets. The Naive Bayes classifier was employed to detect fake news [208]. Similarly, the Agarwalla *et al.* group

tried various classification techniques to classify fake articles. For example, the Naive Bayes classifier combined with Lidstone smoothing, SVM, and Logistic Regression. The results show that Naive Bayes with Lidstone smoothing exhibits the highest accuracy [10].

Dayani *et al.* looked into rumor detection on Twitter, where the Naive Bayes classifier gauged the relationship between the events [85]. Granik *et al.* use Naive Bayes classification to detect fake news by applying the probability of a previous event and comparing former events with a current one. Multinomial Naive Bayes quantifies the news accuracy [122]. There are many ways to use Naive Bayes for classification, and there has been much success with this algorithm.

(2) Support Vector Machine *Support Vector Machine* (SVM) is a supervised machine learning model that uses learning algorithms to classify data into two groups. An SVM model learns labeled training data for categories and then categorizes new data by finding the hyper-plane that differentiates the two classes very well. SVM is typically adopted to analyze data to tackle classification and regression problems.

Rubin *et al.* built an SVM-based algorithm to detect satire with five predictive features, namely, absurdity, humor, grammar, negative affect, and punctuation. Rubin's team tested various combinations on several news articles. The empirical results suggest that the best predicting feature combination is absurdity, grammar, and punctuation, which detects satirical news with 90 percent precision and 84 percent recall. This system can aid in minimizing the potential deceptive impact of satire [242]. Similarly, Nafis *et al.* advocated SVM classification for detecting satire. The classifier analyzed the data and detected patterns useful for classification, and then predicted which class the input should be classified into [207]. SVM is one of the most widely used algorithms for classification.

(3) K-Nearest Neighbor *K-Nearest Neighbor* (KNN) is a statistical method of determining which data points are closest. KNN is a simple yet effective classification algorithm in supervised learning. KNN is a prominent tool for recognizing patterns and data mining.

The data is preprocessed, converted into vectors, and then plotted on a graph. The proximity alludes to the similarity between points, and then those points are classified; thus, KNN is all about proximity-determining classification.

Recent studies demonstrated that KNN had achieved high accuracy in fake news detection [14]. For instance, Elmurngi *et al.* used KNN to classify movie reviews into positive or negative classes. Since the method bases classification on the closeness of objects, vectors are a centerpiece to determine the movie reviews' closeness. In this scheme, the reviews convert into vectors in an initial phase. Depending on the vectors' closeness, the evaluated reviews will fall into the categories of real or fake ones [101]. Overall, KNN is a valuable and straightforward method for detection purposes.

(4) Random Forest *Random Forest* (RF) is a classification algorithm that relies on decision trees to categorize data. Decision trees are a tool that uses an upside-down tree-like model to display potential decisions and the outcome of those decisions. In a very recent study, Bhutani *et al.* deployed the random forest technique to classify news articles [52]. They run RF in combination with a Naive Bayes classifier to build a training model in an initial step. According to the findings, the RF classifier outperforms the Naive Bayes classifier in detecting fake news articles. Similarly, Gupta and Kadam compared SVM and RFs performance in detecting fake news [130]. The RF was very accurate and even gave estimates of which variables were important in the classification. RF received an 86 percent accuracy on the test set, which was higher than the SVM. Random Forest provides impressive results with classification and therefore is resourceful in misinformation detection.

Large Language Models

They encourage more researchers to engage in the study of enhancing news recommendation performance through the utilization of large language models such as ChatGPT. There is a need for researchers to address the social issues arising from the dissemination of fake news articles when employing large language models like ChatGPT. It is crucial to

enhance the trustworthiness and reliability of language models to mitigate the impact of fake news [176]. This was an experimental observational study conducted in February 2023 evaluating the quality of the responses provided by ChatGPT (Version 3.5. OpenAI Inc, San Francisco, CA, USA) to 20 medical questions from diverse fields. When asked multiple medical questions, ChatGPT provided answers of limited quality for scientific publication. More importantly, ChatGPT provided deceptively real references. Users of ChatGPT should pay particular attention to the references provided before integration into medical manuscripts. This study highlights an important shortcoming of LLMs, namely their risk of being "confidently wrong". While such models have now reached astounding performance in simulating human conversations, the next stages in their development must improve information validity and responsible deployment [123]. In this paper, we investigate the impact of knowledge integration into PLMs for fake news detection. Our experiments show that knowledge-enhanced models can significantly improve fake news detection on LIAR where the KB is relevant and up-to-date. The mixed results on COVID-19 highlight the reliance on stylistic features and the importance of domain-specific and current KBs [284]. In this paper, we propose a novel transformer-based language model fine-tuning approach for these fake news detection. the predicted features extracted by the universal language model RoBERTa and domain-specific model CT-BERT are fused by one multiple-layer perception to integrate fine-grained and high-level specific representations. Quantitative experimental results evaluated on the existing COVID-19 fake news dataset show its superior performances compared to the state-of-the-art methods among various evaluation metrics. Furthermore, the best weighted average F1 score achieves 99.02% [67]. Dulhanty *et al.* take advantage of bidirectional cross-attention between claim-article pairs via pair encoding with self-attention, we construct a large-scale language model for stance detection by performing transfer learning on a RoBERTa deep bidirectional transformer language model, and were able to achieve state-of-the-art performance (weighted accuracy of 90.01%) on the Fake News Challenge Stage 1 (FNC-I) benchmark [97]

5.2.3 Data Mining

Data Mining is the process of examining big data and discovering patterns and anomalies to generate new useful information. The technique is an intersection of machine learning, statistics, and computer science. Data mining transforms big data into understandable and functional information, which can help with misinformation detection using clustering, outlier detection, and TF-IDF, for example. In this section, we will explain the usefulness of these approaches for misinformation detection.

Table 5.3 portrays three of the most common data mining techniques deployed for misinformation detection. Each reference reports a unique application of a data mining technique. Like Tables 5.1 and 5.2, Table 5.3 is beneficial for creative ways to observe data and uncover patterns. For example, one future research direction might be exploring the possibility of integrating outlier detection schemes into satire detection systems.

Clustering

Clustering techniques can be adopted as a technique to categorize news events, where each cluster represents a news event. First, the model clusters authentic news articles by similar topics. The basic idea of determining a new article's credibility is determining its topic and placing it in a cluster. Any article that does not fall into an existing cluster is considered fake news.

Zhang *et al.* piloted the aforementioned idea and developed a system called *FEND*. The clustering technique was seamlessly integrated with the text analytical approaches in FEND to detect fake news [308]. Scraped fake news and ground truth events were two core components in FEND to build a corpus for training and testing purposes. The FEND system procures legitimate news and refines news from a word-based dataset to a topic-based dataset, thereby detecting fake news through a two-layered filter - a fake topic detector and a fake event detector. The first phase extracts events from real news; the events serve as training data to create topic clusters. The second phase classifies unknown events into a

topic cluster. The FEND system makes use of text clustering and classification approaches in conjunction with a lexical database. Each news article is put into a topic cluster and compared against events of real news. If a news article falls outside any cluster, then the news is marked as a potential fake news piece. FEND achieves a high detection accuracy percentage.

In another study, Widyantoro *et al.* developed a model to explain the credibility of a tweet by examining the opinions of other users. The researchers analyzed whether users gave support or opposition and whether those users have a credible reputation. Therefore, this approach to misinformation detection assumes that the more users support a claim, the more valid the claim is. In order to display the validity of a tweet, the researchers ranked the tweets hierarchically from supportive to opposing tweets. This model leverages a hierarchical clustering scheme to group user opinions [285]. An opinion is placed in a cluster at every level if the content is similar to the opinions. In their evaluation, they found that their approach achieved the accuracy of ground truth. Clustering is very useful as an unsupervised approach to understanding data.

Outlier Detection

Outlier detection or *anomaly detection* is a process of finding data that deviate from expected results [248]. We refer to this type of data as outliers or anomalies, which may indicate an error in experiments or information sources.

In one fake-news-detection study, Peng researched a neural network-integrated with a bimodal distribution removal (BDR) algorithm to remove outliers [219]. In this approach, the system trains by examining the behaviors of fluctuating variances, followed by identifying the variances as outliers. The empirical results reveal that the BDR algorithm is a promising approach to detecting misinfomation [219]. In another study on spam detection, Fayazbakhsh *et al.* investigated the idea of local outlier factor (LOF), which is a density-based outlier detection algorithm. The LOF algorithm computes the local density difference of a given

data point concerning its neighbors. In light of LOF, outliers are the data points that have a significantly lower density than their neighbors. The LOF algorithm was applied to analyze product reviews to pinpoint outliers, which fall into the spam classification [105].

Although the BDR [219] and LOF [105] algorithms focus on two different application domains, they share similar design philosophies. More specifically, BDR and LOF's basic ideas are to map misinformation detection problems (e.g., fake news and spam review) into outlier detection problems, in which outliers are envisioned as misinformation. The evidence indicates that it is prudent and practical to construct novel misinformation detection solutions by employing cutting-edge outlier detection algorithms.

TF-IDF

Term frequency-inverse document frequency (TF-IDF) is a statistical measure that signifies the word's importance in the document and corpus by multiplying how many times a word appears in a document and the inverse of the frequency of that word across a set of documents. Locating the specific words can be utilized in detecting misinformation [147][13].

Jin *et al.* explored a way of adopting TF-IDF for rumor detection [147]. In this intriguing study, true rumor articles are in the form of vector representation. Each element in the vector has a TF-IDF score. TF-IDF can be useful for text matching, where both tweets and verified rumor articles are in the form of a v-dimensional vector. The v values in v-dimensional vectors are the size of a corpus dictionary. The TF-IDF scores in each vector represent the tweet or verified rumor texts. After the TF-IDF-enabled system computes the similarity between a tweet and a verified rumor, it returns a matching score. The experimental results suggest that the TF-IDF scheme outperforms the conventional semantic-embedding and lexicon-based methods in detecting rumors [147]. In regards to satire detection, Ahmad *et al.* attempted to incorporate TF-IDF as a feature extraction strategy into an SVM-based detection system. The TF-IDF features trained the classifier, which they built using the

LIBSVM classification tool to predict the article's satire or validity. The accuracy of the system was dependent on the type of feature extraction method [13].

The previously mentioned systems, aiming to detect rumor [147] and satire [13], have kindred attributes and differences. The primary difference between the two types of systems is that the rumor system is solely TF-IDF based, and the satire system uses machine learning in addition to TF-IDF. The systems similarities reside in TF-IDF which allowed both systems to outperform the semantic-embedding and lexicon-based methods. The results from these two studies suggest that misinformation detection systems are improving with the implementation of TF-IDF as a feature extraction method.

5.3 Research Opportunities

Misinformation detection is a vibrant field with multiple topics and numerous benefits. Although this field is quickly advancing, many open issues do call for attention. This section will discuss the issues that are unsolved within the field and propose potential solutions for the issues. Table 5.4 below summarizes the open issues and potential solutions that we propose in the following sections.

5.3.1 Research Opportunity 1: Corpora

Currently, there is a scarcity of deceptive news available as corpora for predictive modeling. This scarcity is a major stumbling block in this field of natural language processing and deception detection [241]. Rubin *et al.* suggest that there are nine items needed to build an excellent deceptive news detection corpus, namely, 1) availability of truthful and deceptive instances, 2) digital textual format accessibility, 3) verifiability of "ground truth," 4) homogeneity in lengths, 5) homogeneity in writing matter, 6) predefined timeframe, 7) the manner of news delivery, 8) pragmatic concerns, and 9) language and culture [241]. It is

relatively straightforward to address digital textual formats, homogeneity in lengths, and homogeneity in writing. However, one may face challenges to tackle the availability of truthful and deceptive instances and addressing language and culture.

Prior studies confirm that predictive methods can discover patterns in positive and negative data objects in news articles [241]. It becomes non-trivial to acquire counterparts to authentic news. Very recently, Wang has confirmed that there are limitations in statistical approaches to combating fake news due to the lack of labeled benchmark datasets [278].

An initial step towards tackling this issue is to construct a large-scale database customized for labeled news articles. Ideally, crowdsourcing initiatives can bring in data so that the database will be contributed by and shared among researchers and scientists focusing on developing fake news detection techniques. We propose an aggregation of efforts from the research community to perform data collection and labeling tasks. We recommend leveraging gamification strategies, which are likely to foster user contributions by game-like interactions. The data-sharing platform - serving as a crowdsourced system - is expected to be readily expanded to other types of misinformation.

In the next section, we will discuss the challenges that come with language and culture. Addressing this aspect will significantly improve misinformation detection techniques.

5.3.2 Research Opportunity 2: Non-English Languages

English is currently the leading language in misinformation detection and a few other languages like Spanish, Italian, and Mandarin. Typically in detection with non-English languages, little consideration is given to language and cultural differences [240]. The differences that language and culture bring could alter what features are examined and consequently change misinformation detection methods. To advance the field of misinformation detection, the research community ought to pay particular attention to developing multilingual deception techniques. In such a vastly growing field, researchers are facing challenges on multiple fronts.

One challenge lies in the scarcity of labeled misinformation in non-English languages. For instance, Monteiro *et al.* investigated the scarcity of labeled fake news datasets for the Portuguese language. The scarcity of these datasets prevents training classifiers from detecting fake news in Portuguese. As a solution, a reference corpus - made up of true and fake news - was introduced for Portuguese. After the corpus was analyzed to discover the linguistic characteristics, the machine learning techniques were applied to detect fake news in the Portuguese [203]. The empirical study demonstrated that the machine-learning-enabled system delivered outstanding detection results.

The second challenge is that different languages have various grammar structures. One candidate approach to tackle this problem is to incorporate grammar analysis into detection systems. For example, Seo *et al.* brought to light that there is less research on fake news detection that uses grammar on deep neural networks [247]. A novel fake news detection model, called FAGON, was constructed to advocate for grammatical transformation on a deep neural network. FAGON determines if an article is accurate or not by learning grammatical differences on a neural network. The results confirmed that the FAGON model effectively detects fake news in the Korean language [247]. It could be intriguing to explore the possibility of expanding the FAGON model to detect misinformation in languages other than Korean.

To address the above two challenges in non-English languages, we propose a possible way of investigating and improving non-English misinformation detection research. We aim to build a system that can analyze the grammatical structure while considering the culture of a language. Building a system that can do this will create ample opportunities for future research in misinformation detection. Once grammar and culture are understood, the analysis for misinformation should be similar to English techniques.

As an assessment plan, we will carry out the following three steps to evaluate our proposed system capable of analyzing grammatical structures of non-English languages: We intend to build a dataset with labeled articles translated into multiple languages. We will

benchmark a handful of fake news detection techniques driven by the dataset of news articles written in multiple languages. We quantitatively compare the detection techniques across various types of articles in multiple languages. In the next section, we will consider the areas of misinformation detection that are not as developed. We believe that improving these areas will allow other areas of misinformation detection to progress as well.

5.3.3 Research Opportunity 3: Misinformation Infancy

The known detection methods for rumors and clickbait are still in their infancy [320][314]. Therefore, there are plenty of opportunities and demands to develop new detection techniques in the fields of rumors and clickbait.

Growing evidence indicates the need for improving the existing rumor detection schemes. For example, Ma *et al.* stated that typically users depend on common sense and rumor-reporting websites like snopes.com and factcheck.org to distinguish rumors from factual events. However, since people run the rumor reporting websites, the websites do not have the best coverage on a topic, and rumor detection has a significant delay [188]. The delay and lack of coverage on rumor topics are probably due to the lack of advancement in automated rumor detection systems. According to Zubiaga *et al.*, there are three challenges when it comes to rumor detection. First, if the content is the only feature used, the accuracy will be low. Second, rumors typically are tracked using keywords; however, there is inconsistency among user writings, making it non-trivial to track rumors of the same event with different words. Third, there is limited availability of datasets to facilitate performance evaluation [320].

Now let us elaborate on the challenges of devising clickbait detection systems. The state-of-the-art clickbait detection approaches are generally categorized into two camps, namely, lexical-based and machine learning-based algorithms [314]. The lexical algorithms look for semantic similarities between the headline and the content of an article. Since headlines are customarily concise, there is a roadblock to computing the similarities between the headline and the content accurately. Additionally, some clickbait headlines are relevant to the

content, and some headlines are entirely irrelevant to the content. Therefore, the results of this detection technique are typically inaccurate due to the complex challenges. The machine learning approach improves detection accuracy by leveraging features. Unfortunately, many features used in these algorithms are domain-specific, demanding a significant amount of time. Furthermore, the machine-learning-based approach adds restrictions because the solution is domain-specific in nature [314].

We believe that there are many ideas and new techniques that can be gathered from the details we provide about misinformation detection. We would like to highlight our aim to expose an opportunity to use aspect-based emotion analysis for misinformation detection. ABSA retrieves vital information from a sentence on a detailed level which can be useful for detection. ABSA can gather features used for a model to detect if the input text is valid or false. We are excited about the idea of ABSA and misinformation detection. We hope that our research provides insight for this advancement in technology.

5.4 Summary

In this chapter, we have gone in-depth on misinformation and current detection methods. There are six common types of misinformation that we discussed, including fake news, fake reviews, misinformation in microblogs, rumors, satire, and clickbait. These types of misinformation are weaved throughout the internet from social media to blogs to review websites. There have been numerous experiments and built techniques to handle the issue of misinformation. In our comparison tables of previous research papers, we can clearly see consistent gaps. The main gaps are in the area of Satire, Clickbait, and Misinformation in Microblogs. We also notice an opportunity in the technique of sentiment analysis for these three misinformation areas. We surmise that ABSA can be useful in detecting these types of misinformation due to its focus on specific words and the detailed meaning behind the word. We would like to inspire other researchers to take on this intriguing task in advancing the

fields of ABSA and misinformation detection as both are very important fields in the world today. We believe that ABSA is an excellent future direction for misinformation detection.

	Fake News	**Fake Reviews**	**Misinformation in Microblogs**	**Rumors**	**Satire**	**Clickbait**
Logistic Regression	[58], [35], [124], [157], [159], [278], [175], [263], [251]	[249], [281], [246], [40]	[62], [6]	[145], [111], [170], [312], [119], [299], [306], [209], [181], [17], [177]	[235], [259], [267], [234], [174], [187]	[58], [214], [215], [142], [7], [321]
RNN	[35], [297], [219], [251]	[49], [233], [246]	[188], [62], [309]	[188], [293], [62], [223], [146], [98]		[25], [77], [321], [65]
GRU	[35], [84], [185], [243]	[184]	[189], [188], [190], [179]	[188], [190], [293], [179], [223]	[295], [86]	[317], [162], [65], [96]
LSTM	[35], [50], [102], [247], [222], [16], [289, **?**] [251], [243]	[27], [44], [153], [277], [49], [265], [271]	[188], [179], [309]	[188], [163], [293], [179], [17], [51], [168], [146], [98]	[86]	[214], [77], [321], [65]
Bi-LSTM	[35], [199], [84], [205], [75], [278], [64], [72], [225]	[265], [184], [246], [307], [182]	[309]	[126], [292], [186]	[194]	[321], [178], [162]
CNN	[35], [252], [278], [16], [185], [280]	[27], [49], [265], [279], [184], [261], [246]	[303], [252], [188], [304], [62], [309]	[188], [304], [68], [62], [69], [17], [15], [223], [146]	[86], [7]	[12], [314], [321], [178], [114], [305]
Naive Bayes	[159], [122], [52], [251], [244]	[249], [281], [101], [44], [246], [40], [74]	[28]	[202], [145], [119], [299], [306], [17], [85], [177]	[113], [267], [174], [187]	[321]
SVM	[157], [130], [125], [137], [278], [251], [243], [244]	[155], [249], [281], [39], [91], [27], [101], [44], [279], [261], [246], [40], [33]	[189], [132], [167], [190], [62], [28], [95]	[189], [132], [190], [202], [145], [3, 62], [170], [119], [17], [275], [177]	[235], [60], [207], [113], [295], [86], [268], [43], [259], [234], [174], [187]	[214], [282], [22], [282], [321]
KNN	[159], [244]	[101]	[6], [95]	[202], [145], [95], [85], [177]		
Random Forest	[48], [157], [130], [52], [115]	[40], [74], [33]	[189], [62], [179], [6]	[189], [62], [179], [170], [119], [299], [306]	[235], [268], [267], [234], [187]	[321]
Large Language Model	[18], [11], [262], [41], [97]		[136]		[296]	[318]

Table 5.2: Machine Learning Comparison Table

	Fake News	**Fake Reviews**	**Misinformation in Microblogs**	**Rumors**	**Satire**	**Clickbait**
Clustering	[308], [124], [217], [76], [138], [150]	[239], [213], [9], [108], [107], [31], [164]	[285], [149], [148], [150], [253], [31], [264], [183], [8]	[62], [264], [103], [270], [313], [288]	[254]	[228], [59]
Outlier Detection	[219], [154], [171]	[106], [239], [93], [104], [191]	[191], [216], [236], [248]	[62], [103], [236]		[59]
TF-IDF	[14], [52], [219], [237], [44]	[39], [91], [44], [246]	[133], [285], [179], [95], [147],[236]	[62], [179], [95], [292], [147], [236]	[207], [268], [13]	[214], [7]

Table 5.3: Data Mining Comparison Table

Areas	**Open Issues**	**Candidate Solutions**
Corpora	The lack of labeled benchmark datasets.	To construct a large-scale database customized for labeled news articles.
Non-English Language	The scarcity of labeled misinformation in non-English languages	To build a reference corpus for non-English languages.
Non-English Language	Various grammar structures	To incorporate grammar analysis into detection systems.
Misinformation Infancy	Rumor and clickbait detection are still in their infancy	To build a generic misinformation detection platform where domain-specific modules can be plugged in.

Table 5.4: Open Issues Summarization

Chapter 6

Conclusions and Future Work

In this chapter, we aim to conclude our dissertation by reviewing our findings. We were able to discuss the history of aspect-based sentiment analysis in Chapter 2 and how it lead to a research gap in the need for improvement of aspect term extraction. In Chapter 3, we built a model for sentiment and emotion analysis and analyzed tweets related to COVID-19. This model led us to build an aspect term extraction model which we discussed in Chapter 4. In the next sections, we will explain our conclusion for our research, and in Section 6.3 we will discuss our future research in this field.

6.1 Sentiment and Emotion Analysis

In this dissertation, specifically Chapter 3, we gathered insight into the reactions toward COVID-19 in 2020 by analyzing the topics, sentiments, and emotions extracted from tweets related to Covid-19. The analysis framework embraces four integrated phases. First, we separated and preprocessed the English and Spanish tweets. Second, we used LDA to glean topics and infer themes from those topics from the English and Spanish tweets. Third, we used *Pysentimiento* to gather the overall sentiment emotions from each language while subsequently employing Pysentimiento to retrieve the emotions from each topic cluster. Lastly, we undertook a complete language-comparison analysis of the English and Spanish tweets. Our dataset derives from a dataset created by researchers at Georgia State University. The dataset contains Covid-related tweets collected using corresponding keywords. Then, we further constructed our balanced dataset by processing the data, giving us 150K English and 150K Spanish tweets.

Overall, we discovered that the different groups and cultures had a variety of emotions towards the COVID-19 pandemic. The groups both mainly felt negative sentiments throughout the year, but the specifics of emotions and conversations were diverse. Even so, the themes reveal that both groups agree that the pandemic has completely changed the world and created a new normal. Furthermore, we present three specific key observations from our research:

1. The spikes in emotion and sentiment correlate to major events during that time.
2. The negative sentiment was highest during the second wave of COVID-19; however, it decreased as conversations about a vaccine appeared.
3. The English speakers had a more intense reaction throughout COVID-19 than the Spanish speakers.

The spikes corresponding with events can be helpful in future research. We can assume that the time of the spike is a critical time period to analyze. Topic modeling is a great resource to retrieve helpful information to determine the reason for the spike. The negative sentiment surged over the summer due to the second wave, which brought out more fear and anger. The vaccine brought hope which led to a reduction in high negative sentiment. We also detected a more intense reaction from the English speakers due to the more significant number of negative tweets in comparison to the Spanish speakers' number of tweets. We do theorize that this can also be due to the classification method. We surmise that these observations can provide the healthy groundwork for more research towards sentiment and emotion analysis in crises.

6.2 Aspect Term Extraction

Our research for the Aspect Term Extraction using annotation schemes proved excellent results. We tackled ATE as an ABSA subtask because of its high level of importance to ABSA.

For our ATE method, we implemented a variety of classifiers including GPT, Electra, RoBERTa, and Distilled BERT with annotation schemes and compared it to two topic modeling techniques, LDA and BERTopic. The RoBERTa model proved to produce valuable results. The annotation schemes provided the classifier with a better understanding of the formatting structure of the given sentence, thus adding them as features to our model and it resulted in a better aspect extraction score. We proved the success of using annotation schemes as a viable option for ATE. Each annotation scheme performed well in classifying the aspects. Our new annotation scheme, reverse IOB (or _IOB) outperformed every other annotation scheme the each classifier. The dataset only contained aspect terms of size 1 and 2; therefore, it was limited on which annotation scheme labels it needed. For example, it did not need the 'I' annotation label because there was no inside of an aspect term, only the beginning and last word.

An annotation scheme can be used to improve Aspect Term Extraction. We observed that each annotation scheme performed very well on the different models. We compare the results with the previous approaches on the AWARE dataset, we conclude that our approach outperforms the previous leading-edge methods. The annotation schemes performed very differently. The differences highlighted that the _IOB – the best annotation scheme – drastically improves ATE results for our ABSA approach.

From these experiments, we discovered our unique annotation along with a large language model proves to be an excellent model for aspect-based sentiment analysis. We believe that this will provide insight for companies, organizations, and government officials. Furthermore, we present three specific key conclusions from our ABSA research:

1. Aspect Term Extraction performs well with our _IOB annotation scheme

2. The RoBERTa classifier using with produces the best results for ATE with each annotation scheme.

3. The best annotation schemes for ATE contain simple architecture, not a complex one.

We also desire that readers keep in mind that each annotation scheme consisted of different sets of labels; this essentially created eight different training/testing partitions. In other words, each annotation scheme produced a different testing set. Therefore, the accuracy numbers for different annotation schemes are not directly comparable and should not be taken at face value. Rather, results for each annotation scheme should be used as baselines for future research using that particular scheme.

6.3 Future Research Directions

In this section, we will discuss the future directions of our research. Throughout this dissertation, our models and analysis have provided useful insight for future research. We are excited about the future of the sentiment analysis field.

6.3.1 Advanced Sentiment and Emotion Analysis

In our future research direction, we will improve upon our sentiment and emotion analysis on social media research by building an advanced End-to-End Multilingual Sentiment and Emotion Analysis model. We will examine why different cultures respond to the same event by gathering specific aspects that clarify the responses. In a second research direction, we aim for this model to boost the classification accuracy of Spanish data. We believe that the Spanish results in Chapter 3 may be at a disadvantage due to the significantly fewer classified tweets. Both the English and Spanish datasets began with 150,000 tweets; however, a large percentage of Spanish tweets were labeled neutral. This research has provided insights into the differing responses between English and Spanish speakers during the global pandemic. This research will continue to be helpful for medical professionals who handle multiple groups in crises, psychology, public health, and economics research. We maintain the belief that detailed information about sentiment and emotions is valuable details for a vast number of companies and organizations.

6.3.2 Optimizing Aspect-Based Sentiment Analysis

For our future work with ABSA, we aim to optimize our model in a number of ways. One method that we would like to improve aspect-based sentiment analysis is using our ATE method to find fine-grained emotions. We hope to build a technique based on our annotation scheme to improve the accuracy of emotion classification for ABEA. The second way that we aim to move forward is by building an End-to-End ABEA model for aspect term extraction and emotion classification. We believe that a united model will be more useful for future researchers and engineers. The last way we aim to advance ABSA is by implementing a model for misinformation classification. We aim to test if the ABSA subtasks can be a great method for detecting clickbait and satire. These new directions for ABSA will provide exciting new information to the NLP community.

6.4 Summary

This dissertation has provided a wide range of information about sentiment and emotion. The field of Sentiment Analysis is specific but has a number of intricate parts. We were able to highlight parts of sentiment analysis research gaps and tackle them. Our work with ABSA is exciting and we hope that future research done in this field brings about newer models and more innovation. We hope that our research provides more insight into this sub-field and sparks interest in building more models for ABSA. Throughout this dissertation, we were able to see that aspect-based sentiment analysis is useful for gaining detailed insights into sentiment expression towards specific aspects. We believe that our contributions provide a step for the field of sentiment analysis to continue growing and impacting many lives.

Bibliography

[1] A. Abirami and V. Gayathri. A survey on sentiment analysis methods and approach. In *2016 Eighth International Conference on Advanced Computing (ICoAC)*, pages 72–76. IEEE, 2017.

[2] L. J. Abu-Raddad, H. Chemaitelly, and A. A. Butt. Effectiveness of the bnt162b2 covid-19 vaccine against the b. 1.1. 7 and b. 1.351 variants. *New England Journal of Medicine*, 385(2):187–189, 2021.

[3] M. Abulaish, N. Kumari, M. Fazil, and B. Singh. A graph-theoretic embedding-based approach for rumor detection in twitter. In *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 466–470, 2019.

[4] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen. Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829, 2021.

[5] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu. Predicting flu trends using twitter data. In *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)*, pages 702–707. IEEE, 2011.

[6] D. D. Acula, L. A. C. Oblan, T. B. Pedroso, K. J. V. Riosa, and M. A. R. Tolibas. Implementing fact-checking in journalistic articles shared on social media in the philippines using knowledge graphs. In *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, pages 462–466. IEEE, 2018.

[7] P. Adelson, S. Arora, and J. Hara. Clickbait; didn't read: Clickbait detection using parallel neural networks, 2017.

[8] K. S. Adewole, T. Han, W. Wu, H. Song, and A. K. Sangaiah. Twitter spam account detection based on clustering and classification methods. *The Journal of Supercomputing*, pages 1–36, 2018.

[9] M. Adike and V. Reddy. Detection of fake review and brand spam using data mining. *Int J Recent Trends Eng Res*, 2(7):251–256, 2016.

[10] K. Agarwalla, S. Nandan, V. A. Nair, and D. D. Hema. Fake news detection using machine learning and natural language processing.

[11] A. Aggarwal, A. Chauhan, D. Kumar, S. Verma, and M. Mittal. Classification of fake news by fine-tuning deep bidirectional transformers based language model. *EAI Endorsed Transactions on Scalable Information Systems*, 7(27):e10–e10, 2020.

[12] A. Agrawal. Clickbait detection using deep learning. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pages 268–272. IEEE, 2016.

[13] T. Ahmad, H. Akhtar, A. Chopra, and M. W. Akhtar. Satire detection from web documents using machine learning methods. In *2014 International Conference on Soft Computing and Machine Intelligence*, pages 102–105. IEEE, 2014.

[14] H. Ahmed, I. Traore, and S. Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pages 127–138. Springer, 2017.

[15] M. Ahsan, M. Kumari, and T. Sharma. Detection of context-varying rumors on twitter through deep learning. *Int. J. Adv. Sci. Technol*, 128:45–58, 2019.

[16] O. Ajao, D. Bhowmik, and S. Zargari. Fake news identification on twitter with hybrid cnn and rnn models. In *Proceedings of the 9th International Conference on Social Media and Society*, pages 226–230, 2018.

[17] M. S. Akhtar, A. Ekbal, S. Narayan, V. Singh, and E. Cambria. No, that never happened!! investigating rumors on twitter. *IEEE Intelligent Systems*, 33(5):8–15, 2018.

[18] M. Al-Yahya, H. Al-Khalifa, H. Al-Baity, D. AlSaeed, and A. Essam. Arabic fake news detection: comparative study of neural networks and transformer-based approaches. *Complexity*, 2021:1–10, 2021.

[19] D. Alessia, F. Ferri, P. Grifoni, and T. Guzzo. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 2015.

[20] S. A. Alkhodair, S. H. Ding, B. C. Fung, and J. Liu. Detecting breaking news rumors of emerging topics in social media. *Information Processing & Management*, 57(2):102018, 2020.

[21] N. Alshammari and S. Alanazi. The impact of using different annotation schemes on named entity recognition. *Egyptian Informatics Journal*, 22(3):295–302, 2021.

[22] A. Alshehri, A. Nagoudi, A. Hassan, and M. Abdul-Mageed. Think before your click: Data and models for adult content in arabic twitter. *The 2nd Text Analytics for Cybersecurity and Online Safety (TA-COS-2018), LREC*, 2018.

[23] T. Alvarez-López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castano. Gti at semeval-2016 task 5: Svm and crf for aspect detection and unsupervised aspect-based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 306–311, 2016.

[24] A. Amara, M. A. H. Taieb, and M. B. Aouicha. Multilingual topic modeling for tracking covid-19 trends based on facebook data analysis. *Applied Intelligence*, pages 1–22, 2021.

[25] A. Anand, T. Chakraborty, and N. Park. We used neural networks to detect clickbaits: You won't believe what happened next! In *European Conference on Information Retrieval*, pages 541–547. Springer, 2017.

[26] A. F. Anta, L. N. Chiroque, P. Morere, and A. Santos. Sentiment analysis and topic detection of spanish tweets: A comparative study of of nlp techniques. *Procesamiento del lenguaje natural*, 50:45–52, 2013.

[27] T. V. M. Aono. Fake review detection focusing on emotional expressions and extreme rating. 2019.

[28] S. Aphiwongsophon and P. Chongstitvatana. Identify misinformation on twitter with machine learning.

[29] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko. Comprehensive analysis of aspect term extraction methods using various text embeddings. *Computer Speech & Language*, 69:101217, 2021.

[30] N. Azam, B. Tahir, and M. A. Mehmood. Sentiment and emotion analysis of text: a survey on approaches and resources. In *7th International Conference on Language and Technology (CLT)*, pages 87–94, 2020.

[31] A. G. Babu, S. S. Kumari, and K. Kamakshaiah. An experimental analysis of clustering sentiments for opinion mining. In *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*, pages 53–57, 2017.

[32] S. Baccianella, A. Esuli, and F. Sebastiani. Multi-facet rating of product reviews. In *European conference on information retrieval*, pages 461–472. Springer, 2009.

[33] P. R. Badri Satya, K. Lee, D. Lee, T. Tran, and J. Zhang. Uncovering fake likers in online social networks. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2365–2370, 2016.

[34] H. Bahuleyan and O. Vechtomova. Uwaterloo at semeval-2017 task 8: Detecting stance towards rumours with topic independent features. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 461–464, 2017.

[35] S. Bajaj. The pope has a new baby! fake news detection using deep learning, 2017.

[36] A. Balahur and J. M. Perea-Ortega. Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Information Processing & Management*, 51(4):547–556, 2015.

[37] P. Balaji, O. Nagaraju, and D. Haritha. Levels of sentiment analysis and its challenges: A literature review. In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, pages 436–439. IEEE, 2017.

[38] J. M. Banda, R. Tekumalla, G. Wang, J. Yu, T. Liu, Y. Ding, and G. Chowell. A Twitter Dataset of 150+ million tweets related to COVID-19 for open research, Apr. 2020. This dataset will be updated bi-weekly at least for these updates.

[39] R. Banerjee, S. Feng, J. S. Kang, and Y. Choi. Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1469–1473, 2014.

[40] S. Banerjee, A. Y. Chua, and J.-J. Kim. Using supervised learning to classify authentic and fake online reviews. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, pages 1–7, 2015.

[41] Y. Bang, E. Ishii, S. Cahyawijaya, Z. Ji, and P. Fung. Model generalization on covid-19 fake news detection. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*, pages 128–140. Springer, 2021.

[42] F. Barbieri, F. Ronzano, and H. Saggion. Do we criticise (and laugh) in the same way? automatic detection of multi-lingual satirical news in twitter. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[43] F. Barbieri, F. Ronzano, and H. Saggion. Do we criticise (and laugh) in the same way? automatic detection of multi-lingual satirical news in twitter. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, page 1215–1221. AAAI Press, 2015.

[44] J. Barry. Sentiment analysis of online reviews using bag-of-words and lstm approaches. In *AICS*, pages 272–274, 2017.

[45] R. Batra and S. M. Daudpota. Integrating stocktwits with sentiment analysis for better prediction of stock price movement. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–5. IEEE, 2018.

[46] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, volume 6, page 12, 2010.

[47] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin. A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155, 2003.

[48] P. Bharadwaj and Z. Shao. Fake news detection with semantic features and text mining. *International Journal on Natural Language Computing (IJNLC) Vol*, 8, 2019.

[49] R. Bhargava, A. Baoni, and Y. Sharma. Composite sequential modeling for identifying fake reviews. *Journal of Intelligent Systems*, 28(3):409–422, 2019.

[50] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal. Combining neural, statistical and external features for fake news stance identification. In *Companion Proceedings of the The Web Conference 2018*, pages 1353–1357, 2018.

[51] U. Bhattacharjee, P. Srijith, and M. S. Desarkar. Term specific tf-idf boosting for detection of rumours in social networks. In *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, pages 726–731. IEEE, 2019.

[52] B. Bhutani, N. Rastogi, P. Sehgal, and A. Purwar. Fake news detection using sentiment analysis. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5. IEEE, 2019.

[53] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[54] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, 2017.

[55] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

[56] V. Bonta and N. K. N. Janardhan. A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2):1–6, 2019.

[57] S. Boon-Itt and Y. Skunkan. Public perception of the covid-19 pandemic on twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4):e21978, 2020.

[58] P. Bourgonje, J. M. Schneider, and G. Rehm. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 84–89, 2017.

[59] P. Braun, A. Cuzzocrea, L. M. V. Doan, S. Kim, C. K.-S. Leung, J. F. A. Matundan, and R. R. Singh. Enhanced prediction of user-preferred youtube videos based on cleaned viewing pattern history. In *KES*, 2017.

[60] C. Burfoot and T. Baldwin. Automatic satire detection: Are you having a laugh? In *Proceedings of the ACL-IJCNLP 2009 conference short papers*, pages 161–164, 2009.

[61] E. F. Can, A. Ezen-Can, and F. Can. Multilingual sentiment analysis: An rnn-based framework for limited data. *arXiv preprint arXiv:1806.04511*, 2018.

[62] J. Cao, J. Guo, X. Li, Z. Jin, H. Guo, and J. Li. Automatic rumor detection on microblogs: A survey. *arXiv preprint arXiv:1807.03505*, 2018.

[63] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*, 2020.

[64] A. K. Chaudhry, D. Baker, and P. Thun-Hohenstein. Stance detection for the fake news challenge: identifying textual relationships with deep neural nets. *CS224n: Natural Language Processing with Deep Learning*, 2017.

[65] S. Chawda, A. Patil, A. Singh, and A. Save. A novel approach for clickbait detection. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1318–1321. IEEE, 2019.

[66] S. U. S. Chebolu, P. Rosso, S. Kar, and T. Solorio. Survey on aspect category detection. *ACM Computing Surveys (CSUR)*, 2022.

[67] B. Chen, B. Chen, D. Gao, Q. Chen, C. Huo, X. Meng, W. Ren, and Y. Zhou. Transformer-based language model fine-tuning methods for covid-19 fake news detection. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*, pages 83–92. Springer, 2021.

[68] Y. Chen, J. Sui, L. Hu, and W. Gong. Attention-residual network with cnn for rumor detection. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1121–1130, 2019.

[69] Y.-C. Chen, Z.-Y. Liu, and H.-Y. Kao. Ikm at semeval-2017 task 8: Convolutional neural networks for stance detection and rumor verification. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 465–469, 2017.

[70] Z. Chen and T. Qian. Enhancing aspect term extraction with soft prototypes. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2107–2117, 2020.

[71] H.-C. Cho, N. Okazaki, M. Miwa, and J. Tsujii. Named entity recognition with multiple segment representations. *Information Processing & Management*, 49(4):954–965, 2013.

[72] S. Chopra, S. Jain, and J. M. Sholar. Towards automatic identification of fake news: Headline-article stance detection with lstm attention models. In *Stanford CS224d Deep Learning for NLP final project*, 2017.

[73] J. Choudrie, S. Banerjee, K. Kotecha, R. Walambe, H. Karande, and J. Ameta. Machine learning techniques and older adults processing of online information and misinformation: A covid 19 study. *Computers in Human Behavior*, page 106716, 2021.

[74] N. S. Chowdhary and A. A. Pandit. Fake review detection using classification. *International Journal of Computer Applications*, 180(50):16–21, 2018.

[75] C. Conforti, M. T. Pilehvar, and N. Collier. Towards automatic fake news detection: cross-level stance detection in news articles. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 40–49, 2018.

[76] N. J. Conroy, V. L. Rubin, and Y. Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015.

[77] K. COrnn. Clickbait article detection using deep learning: These results will shock you!

[78] I. Cruz, A. F. Gelbukh, and G. Sidorov. Implicit aspect indicator extraction for aspect based opinion mining. *Int. J. Comput. Linguistics Appl.*, 5(2):135–152, 2014.

[79] A. Cui, M. Zhang, Y. Liu, and S. Ma. Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis. In *Asia information retrieval symposium*, pages 238–249. Springer, 2011.

[80] H. Dai and Y. Song. Neural aspect and opinion term extraction with mined rules as weak supervision. *arXiv preprint arXiv:1907.03750*, 2019.

[81] J. Dai, H. Yan, T. Sun, P. Liu, and X. Qiu. Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta. *arXiv preprint arXiv:2104.04986*, 2021.

[82] A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker. Emotion classification in a resource constrained language using transformer-based approach. *arXiv preprint arXiv:2104.08613*, 2021.

[83] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528, 2003.

[84] R. Davis and C. Proctor. Fake news, real consequences: Recruiting neural networks for the fight against fake news, 2017.

[85] R. Dayani, N. Chhabra, T. Kadian, and R. Kaushal. Rumor detection in twitter: An analysis in retrospect. In *2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS)*, pages 1–3. IEEE, 2015.

[86] S. De Sarkar, F. Yang, and A. Mukherjee. Attending sentences to detect satirical fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3371–3380, 2018.

[87] N. Dehbozorgi and D. P. Mohandoss. Aspect-based emotion analysis on speech for predicting performance in collaborative learning. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–7. IEEE, 2021.

[88] K. Denecke. Using sentiwordnet for multilingual sentiment analysis. In *2008 IEEE 24th International Conference on Data Engineering Workshop*, pages 507–512, 2008.

[89] M. Desai and M. A. Mehta. Techniques for sentiment analysis of twitter data: A comprehensive survey. In *2016 International Conference on Computing, Communication and Automation (ICCCA)*, pages 149–154, 2016.

[90] T. Devi and N. Deepa. A novel intervention method for aspect-based emotion using exponential linear unit (elu) activation function in a deep neural network. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1671–1675. IEEE, 2021.

[91] R. K. Dewang and A. K. Singh. Spam review detection through lexical chain based semantic similarity algorithm (lcbss) for negative reviews. *International Journal of Engineering and Technology (IJET)*, 8(6):2946–2955, 2016.

[92] C. Dhaoui, C. M. Webster, and L. P. Tan. Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, 2017.

[93] S. K. Dixit and A. J. Agrawal. Survey on review spam detection. 2013.

[94] H. H. Do, P. Prasad, A. Maag, and A. Alsadoon. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, 118:272–299, 2019.

[95] P. Donda and D. UshaRani. Detecting rumor in social networks using svm and comparing with supervised techniques. *International Journal of Research*, 6(7), 2019.

[96] M. Dong, L. Yao, X. Wang, B. Benatallah, and C. Huang. Similarity-aware deep attentive model for clickbait detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 56–69. Springer, 2019.

[97] C. Dulhanty, J. L. Deglint, I. B. Daya, and A. Wong. Taking a stance on fake news: Towards automatic disinformation assessment via deep bidirectional transformer language models for stance detection. *arXiv preprint arXiv:1911.11951*, 2019.

[98] C. T. Duong, Q. V. H. Nguyen, S. Wang, and B. Stantic. Provenance-based rumor detection. In *Australasian Database Conference*, pages 125–137. Springer, 2017.

[99] D. Duong. What's important to know about the new covid-19 variants?, 2021.

[100] P. Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[101] E. Elmurngi and A. Gherbi. Detecting fake reviews through sentiment analysis using machine learning techniques. *IARIA/data analytics*, pages 65–72, 2017.

[102] S. Esmaeilzadeh, G. X. Peh, and A. Xu. Neural abstractive text summarization and fake news detection. *arXiv preprint arXiv:1904.00788*, 2019.

[103] A. E. Fard, M. Mohammadi, Y. Chen, and B. Van de Walle. Computational rumor detection without non-rumor: A one-class classification approach. *IEEE Transactions on Computational Social Systems*, 6(5):830–846, 2019.

[104] S. Z. Farooq and H. A. Khanday. Spam detection : A review. 2016.

[105] S. K. Fayazbakhsh and J. Sinha. Review spam detection: a network-based approach. *Final project report: CSE*, 590, 2012.

[106] S. K. Fayazbakhsh and J. Sinha. Spam detection : A network-based approach. 2012.

[107] A. Fayazi, K. Lee, J. Caverlee, and A. Squicciarini. Uncovering crowdsourced manipulation of online reviews. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 233–242, 2015.

[108] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In *Seventh international AAAI conference on weblogs and social media*, 2013.

[109] A. Feizollah, N. B. Anuar, R. Mehdi, A. Firdaus, and A. Sulaiman. Understanding covid-19 halal vaccination discourse on facebook and twitter using aspect-based sentiment analysis and text emotion analysis. *International Journal of Environmental Research and Public Health*, 19(10):6269, 2022.

[110] R. Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[111] W. Ferreira and A. Vlachos. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 1163–1168, 2016.

[112] T. Fornaciari and M. Poesio. Identifying fake amazon reviews as learning from crowds. 2014.

[113] A. Frain and S. Wubben. Satiriclr: a language resource of satirical news articles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4137–4140, 2016.

[114] J. Fu, L. Liang, X. Zhou, and J. Zheng. A convolutional neural network for clickbait detection. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 6–10. IEEE, 2017.

[115] M. Gahirwal, S. Moghe, T. Kulkarni, D. Khakhar, and J. Bhatia. Fake news detection. *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(1):817–819, 2018.

[116] O. Gencoglu and M. Gruber. Causal modeling of twitter activity during covid-19. *Computation*, 8(4):85, 2020.

[117] A. George, D. Johnson, G. Carenini, A. Eslami, R. Ng, and E. Portales-Casamar. Applications of aspect-based sentiment analysis on psychiatric clinical notes to study suicide in youth. *AMIA Summits on Translational Science Proceedings*, 2021:229, 2021.

[118] E. Ghadery, S. Movahedi, H. Faili, and A. Shakery. An unsupervised approach for aspect category detection using soft cosine similarity measure. *arXiv preprint arXiv:1812.03361*, 2018.

[119] G. Giasemidis, C. Singleton, I. Agrafiotis, J. R. Nurse, A. Pilgrim, C. Willis, and D. V. Greetham. Determining the veracity of rumours on twitter. In *International Conference on Social Informatics*, pages 185–205. Springer, 2016.

[120] J. Golbeck, M. Mauriello, B. Auxier, K. H. Bhanushali, C. Bonk, M. A. Bouzaghrane, C. Buntain, R. Chanduka, P. Cheakalos, J. B. Everett, et al. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science*, pages 17–21, 2018.

[121] M. Graff, S. Miranda-Jiménez, E. S. Tellez, and D. Moctezuma. Evomsa: A multilingual evolutionary approach for sentiment analysis. *arXiv preprint arXiv:1812.02307*, 2018.

[122] M. Granik and V. Mesyura. Fake news detection using naive bayes classifier. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, pages 900–903. IEEE, 2017.

[123] J. Gravel, M. D'Amours-Gravel, and E. Osmanlliu. Learning to fake it: limited responses and fabricated references provided by chatgpt for medical questions. *Mayo Clinic Proceedings: Digital Health*, 1(3):226–234, 2023.

[124] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378, 2019.

[125] G. Guibon, L. Ermakova, H. Seffih, A. Firsov, and G. Le Noé-Bienvenu. Multilingual fake news detection with satire. 2019.

[126] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li. Rumor detection with hierarchical social attention network. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 943–951, 2018.

[127] A. Gupta, H. Li, A. Farnoush, and W. Jiang. Understanding patterns of covid infodemic: A systematic and pragmatic approach to curb fake news. *Journal of business research*, 2021.

[128] R. Gupta, A. Vishwanath, and Y. Yang. Global reactions to covid-19 on twitter: A labelled dataset with latent topic, sentiment and emotion attributes. 2021.

[129] R. K. Gupta, A. Vishwanath, and Y. Yang. Covid-19 twitter dataset with latent topics, sentiments and emotions attributes. *arXiv preprint arXiv:2007.06954*, 2020.

[130] Y. N. A. GUPTA and Y. KADAM. Stance detection for the fake news challenge. 2017.

[131] N. M. Hakak, M. Mohd, M. Kirmani, and M. Mohd. Emotion analysis: A survey. In *2017 international conference on computer, communications and electronics (COMPTELIX)*, pages 397–402. IEEE, 2017.

[132] S. Hamidian and M. Diab. Rumor identification and belief investigation on twitter. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 3–8, 2016.

[133] S. Hamidian and M. T. Diab. Rumor detection and classification for twitter data. *arXiv preprint arXiv:1912.08926*, 2019.

[134] A. Hanselowski, A. PVS, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych. A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*, 2018.

[135] S. Haustein, T. D. Bowman, K. Holmberg, A. Tsou, C. R. Sugimoto, and V. Larivière. Tweets as impact indicators: Examining the implications of automated "bot" accounts on t witter. *Journal of the Association for Information Science and Technology*, 67(1):232–238, 2016.

[136] E. Hoes, S. Altay, and J. Bermeo. Using chatgpt to fight misinformation: Chatgpt nails 72% of 12,000 verified claims. 2023.

[137] B. D. Horne and S. Adali. This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[138] S. Hosseinimotlagh and E. E. Papalexakis. Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In *Proceedings of the Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*, 2018.

[139] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

[140] C. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[141] N. Ide and H. Bunt. Anatomy of annotation schemes: mapping to graf. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 247–255, 2010.

[142] V. Indurthi and S. R. Oota. Clickbait detection using word embeddings. *arXiv preprint arXiv:1710.02861*, 2017.

[143] H. Ismail, A. Khalil, N. Hussein, and R. Elabyad. Triggers and tweets: Implicit aspect-based sentiment and emotion analysis of community chatter relevant to education post-covid-19. *Big Data and Cognitive Computing*, 6(3):99, 2022.

[144] S. Jain, V. Sharma, and R. Kaushal. Towards automated real-time detection of misinformation on twitter. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2015–2020. IEEE, 2016.

[145] S. Jayaprakash, P. Sumathi, and S. Suganya. Identification of rumor-spreading platform in microblogging systems. *Advances in Natural and Applied Sciences*, 11(6 SI):167–174, 2017.

[146] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816, 2017.

[147] Z. Jin, J. Cao, H. Guo, Y. Zhang, Y. Wang, and J. Luo. Detection and analysis of 2016 us presidential election related rumors on twitter. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*, pages 14–24. Springer, 2017.

[148] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang. News credibility evaluation on microblog with a hierarchical propagation model. In *2014 IEEE International Conference on Data Mining*, pages 230–239. IEEE, 2014.

[149] Z. Jin, J. Cao, Y. Zhang, and J. Luo. News verification by exploiting conflicting social viewpoints in microblogs. In *Thirtieth AAAI conference on artificial intelligence*, 2016.

[150] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia*, 19(3):598–608, 2016.

[151] M. G. N. Jorvekar and M. Gangwar. Aspect based emotion detection and topic modeling on social media reviews. 2021.

[152] M. Joshi, P. Prajapati, A. Shaikh, and V. Vala. A survey on sentiment analysis. *International Journal of Computer Applications*, 163(6):34–38, 2017.

[153] M. Juuti, B. Sun, T. Mori, and N. Asokan. Stay on-topic: Generating context-specific fake restaurant reviews. In *European Symposium on Research in Computer Security*, pages 132–151. Springer, 2018.

[154] A. K, D. P, and L. V. L. Emotion cognizance improves health fake news identification, 2019.

[155] J. S. Kang, P. Kuznetsova, M. Luca, and Y. Choi. Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1443–1448, 2013.

[156] H. Kaur, V. Mangat, et al. A survey of sentiment analysis techniques. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 921–925. IEEE, 2017.

[157] J. Y. Khan, M. Khondaker, T. Islam, A. Iqbal, and S. Afroz. A benchmark study on machine learning methods for fake news detection. *arXiv preprint arXiv:1905.04749*, 2019.

[158] A. M. Khattak, R. Batool, F. A. Satti, J. Hussain, W. A. Khan, A. M. Khan, and B. Hayat. Tweets classification and sentiment analysis for personalized tweets recommendation. *Complexity*, 2020, 2020.

[159] U. Khurana and B. O. K. Intelligentie. *The linguistic features of fake news headlines and statements*. PhD thesis, Master's thesis, University of Amsterdam, 2017.

[160] S. Kim, I. Weber, L. Wei, and A. Oh. Sociolinguistic analysis of twitter in multilingual societies. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 243–248, 2014.

[161] Y. Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[162] P. Klairith and S. Tanachutiwat. Thai clickbait detection algorithms using natural language processing with machine learning techniques. In *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, pages 1–4. IEEE, 2018.

[163] E. Kochkina, M. Liakata, and I. Augenstein. Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm. *arXiv preprint arXiv:1704.07221*, 2017.

[164] S. Kokate and B. Tidke. Fake review and brand spam detection using j48 classifier. *IJCSIT Int J Comput Sci Inf Technol*, 6(4):3523–3526, 2015.

[165] P. Kordjamshidi, M.-F. Moens, and M. van Otterlo. Spatial role labeling: Task definition and annotation scheme. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 413–420. European Language Resources Association (ELRA), 2010.

[166] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu. Cross-language sentiment analysis of european twitter messages during the covid-19 pandemic. *arXiv preprint arXiv:2008.12172*, 2020.

[167] K. K. Kumar and G. Geethakumari. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(1):1–22, 2014.

[168] S. Kumar and K. M. Carley. Tree lstms with convolution units to predict stance and rumor veracity in social media conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5047–5058, 2019.

[169] E. Kušen and M. Strembeck. Politics, sentiments, and misinformation: An analysis of the twitter discussion on the 2016 austrian presidential elections. *Online Social Networks and Media*, 5:37–50, 2018.

[170] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108. IEEE, 2013.

[171] L. V. S. Lakshmanan, M. Simpson, and S. Thirumuruganathan. Combating fake news: A data management and mining perspective. *Proc. VLDB Endow.*, 12(12):1990–1993, Aug. 2019.

[172] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196, 2014.

[173] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, F. Sanz, et al. Detecting signs of depression in tweets in spanish: behavioral and linguistic analysis. *Journal of medical Internet research*, 21(6):e14199, 2019.

[174] O. Levi, P. Hosseini, M. Diab, and D. A. Broniatowski. Identifying nuances in fake news vs. satire: Using semantic and linguistic cues. *arXiv preprint arXiv:1910.01160*, 2019.

[175] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective positive-unlabeled learning. In *2014 IEEE international conference on data mining*, pages 899–904. IEEE, 2014.

[176] X. Li, Y. Zhang, and E. C. Malthouse. A preliminary study of chatgpt on news recommendation: Personalization, provider fairness, fake news. *arXiv preprint arXiv:2306.10702*, 2023.

[177] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng. Rumor identification in microblogging systems based on users' behavior. *IEEE Transactions on Computational Social Systems*, 2(3):99–108, 2015.

[178] F. Liao, H. H. Zhuo, X. Huang, and Y. Zhang. Federated hierarchical hybrid networks for clickbait detection. *arXiv preprint arXiv:1906.00638*, 2019.

[179] D. Lin, B. Ma, D. Cao, and S. Li. Chinese microblog rumor detection based on deep sequence context. *Concurrency and Computation: Practice and Experience*, 31(23):e4508, 2019.

[180] B. Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.

[181] F. Liu, A. Burton-Jones, and D. Xu. Rumors on social media in disasters: Extending transmission to retransmission. In *PACIS*, page 49, 2014.

[182] W. Liu, W. Jing, and Y. Li. Incorporating feature representation into bilstm for deceptive review detection. *Computing*, 102(3):701–715, 2020.

[183] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1867–1870, 2015.

[184] Y. Liu and B. Pang. Opinion spam detection based on annotation extension and neural networks. *Computer and Information Science*, 12(2):87, 2019.

[185] Y. Liu and Y.-F. B. Wu. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[186] Z. Liu, S. Goel, M. Y. Raghuprasad, and S. Muresan. Columbia at semeval-2019 task 7: Multi-task learning for stance classification and rumour verification. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1110–1114, 2019.

[187] Z. Liu, S. Shabani, N. G. Balet, and M. Sokhn. Detection of satiric news on social media: analysis of the phenomenon with a french dataset. In *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6. IEEE, 2019.

[188] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha. Detecting rumors from microblogs with recurrent neural networks. 2016.

[189] J. Ma, W. Gao, and K.-F. Wong. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics, 2017.

[190] J. Ma, W. Gao, and K.-F. Wong. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics, 2018.

[191] S. Mahajan and D. P. K. Rana. Spam detection on social network through sentiment analysis. 2017.

[192] M. Mamatha, R. Thejeshwini, J. Thriveni, and K. Venugopal. Supervised aspect category detection of co-occurrence data using conditional random fields. In *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 1–4. IEEE, 2019.

[193] K. H. Manguri, R. N. Ramadhan, and P. R. M. Amin. Twitter sentiment analysis on worldwide covid-19 outbreaks. *Kurdistan Journal of Applied Research*, pages 54–65, 2020.

[194] R. McHardy, H. Adel, and R. Klinger. Adversarial training for satire detection: Controlling for confounding variables. *arXiv preprint arXiv:1902.11145*, 2019.

[195] L. McIntyre. *Post-truth*. MIt Press, 2018.

[196] R. J. Medford, S. N. Saleh, A. Sumarsono, T. M. Perl, and C. U. Lehmann. An "infodemic": leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. In *Open forum infectious diseases*, volume 7, page ofaa258. Oxford University Press US, 2020.

[197] P. Mehra. Unexpected surprise: Emotion analysis and aspect based sentiment analysis (absa) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, 45:101063, 2023.

[198] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[199] K. Miller and A. Oswalt. Fake news headline classification using neural networks with attention. Technical report, tech. rep., California State University, year, 2017.

[200] S. M. Mohammad and P. D. Turney. Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465, 2013.

[201] M. D. Molina-González, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Urena-López. Cross-domain sentiment analysis using spanish opinionated words. In *International Conference on Applications of Natural Language to Data Bases/Information Systems*, pages 214–219. Springer, 2014.

[202] T. Mondal, P. Pramanik, I. Bhattacharya, N. Boral, and S. Ghosh. Analysis and early detection of rumors in a post disaster scenario. *Information Systems Frontiers*, 20(5):961–979, 2018.

[203] R. A. Monteiro, R. L. Santos, T. A. Pardo, T. A. de Almeida, E. E. Ruiz, and O. A. Vale. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *International Conference on Computational Processing of the Portuguese Language*, pages 324–334. Springer, 2018.

[204] P. A. T. Moreira. Supervised machine learning approach for aspect-category detection. In *16th Doctoral Symposium in Informatics Engineering*, page 150, 2021.

[205] D. Mrowca, E. Wang, and A. Kosson. Stance detection for fake news identification. *Stanford University, California, US, rep.*, 2017.

[206] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance. What yelp fake review filter might be doing? In *Seventh international AAAI conference on weblogs and social media*, 2013.

[207] S. T. O. P. T. Nafis and S. Khanna. An improved method for detection of satire from user-generated content. 2015.

[208] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using em. *Machine learning*, 39(2-3):103–134, 2000.

[209] O. Oh, M. Agrawal, and H. R. Rao. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *Mis Quarterly*, pages 407–426, 2013.

[210] C. Olah. Understanding lstm networks, 2015.

[211] S. B. Padme and P. V. Kulkarni. Aspect based emotion analysis on online user-generated reviews. In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–5. IEEE, 2018.

[212] S. Paliwal, S. Kumar Khatri, and M. Sharma. Sentiment analysis and prediction using neural networks. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1035–1042, 2018.

[213] A. C. Pandey and D. S. Rajpoot. Spam review detection using spiral cuckoo search clustering method. *Evolutionary Intelligence*, 12(2):147–164, 2019.

[214] S. Pandey and G. Kaur. Curious to click it?-identifying clickbait using deep learning and evolutionary algorithm. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1481–1487. IEEE, 2018.

[215] O. Papadopoulou, M. Zampoglou, S. Papadopoulos, and I. Kompatsiaris. A two-level classification approach for detecting clickbait posts using text-based features. *arXiv preprint arXiv:1710.08528*, 2017.

[216] E. E. Papalexakis. Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. 2018.

[217] S. B. Parikh and P. K. Atrey. Media-rich fake news detection: A survey. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 436–441. IEEE, 2018.

[218] D. Park, S. Kim, J. Lee, J. Choo, N. Diakopoulos, and N. Elmqvist. Conceptvector: text visual analytics via interactive lexicon building using word embedding. *IEEE transactions on visualization and computer graphics*, 24(1):361–370, 2017.

[219] L. Peng. Fake news detection: Sequence models annual bio-inspired conference 2018.

[220] Q. Peng and M. Zhong. Detecting spam review through sentiment analysis. *JSW*, 9(8):2065–2072, 2014.

[221] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[222] O. T. Pfohl and F. Legros. Stance detection for the fake news challenge with attention and conditional encoding. *CS224n: Natural Language Processing with Deep Learning*, 2017.

[223] L. Poddar, W. Hsu, M. L. Lee, and S. Subramaniyam. Predicting stances in twitter conversations for detecting veracity of rumors: A neural approach. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 65–72. IEEE, 2018.

[224] M. Polignano, P. Basile, M. De Gemmis, and G. Semeraro. An emotion-driven approach for aspect-based opinion mining. In *IIR*, 2018.

[225] K. Popat, S. Mukherjee, A. Yates, and G. Weikum. Declare: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint arXiv:1809.06416*, 2018.

[226] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh. A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, pages 28–37, 2014.

[227] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay. Enhanced senticnet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2):31–38, 2013.

[228] A. Pujahari and D. S. Sisodia. Clickbait detection using multiple categorisation techniques. *Journal of Information Science*, 0(0):0165551519871822, 0.

[229] J. M. Pérez, J. C. Giudici, and F. Luque. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks, 2021.

[230] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[231] F. Qian, C. Gong, K. Sharma, and Y. Liu. Neural user response generator: Fake news detection with collective user intelligence. In *IJCAI*, pages 3834–3840, 2018.

[232] S. Ramezani, R. Rahimi, and J. Allan. Aspect category detection in product reviews using contextual representation. In *Proceedings of ACM SIGIR Workshop on eCommerce (SIGIR eCom'20)*, 2020.

[233] S. J. Rao, S. Bhat, S. Agrawal, and D. Anandakrishnan. Resilience of authorship detection methods and hoax detection techniques to machine generated text.

[234] A. Reganti, T. Maheshwari, A. Das, and E. Cambria. Open secrets and wrong rights: automatic satire detection in english text. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 291–294, 2017.

[235] A. N. Reganti, T. Maheshwari, U. Kumar, A. Das, and R. Bajpai. Modeling satire in english text for automatic detection. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 970–977. IEEE, 2016.

[236] P. Resnick, S. Carton, S. Park, Y. Shen, and N. Zeffer. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. 2014.

[237] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel. A simple but tough-to-beat baseline for the fake news challenge stance detection task, 2017.

[238] M. M. U. Rony, N. Hassan, and M. Yousuf. Baitbuster: A clickbait identification framework. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[239] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi. Deceptive review detection using labeled and unlabeled data. *Multimedia Tools and Applications*, 76(3):3187–3211, 2017.

[240] V. L. Rubin. Pragmatic and cultural considerations for deception detection in asian languages. 2014.

[241] V. L. Rubin, Y. Chen, and N. J. Conroy. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015.

[242] V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection*, pages 7–17, 2016.

[243] N. Ruchansky, S. Seo, and Y. Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806, 2017.

[244] M. J. C. Samonte. Polarity analysis of editorial articles towards fake news detection. In *Proceedings of the 2018 International Conference on Internet and e-Business*, pages 108–112, 2018.

[245] K. Schouten and F. Frasincar. Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830, 2015.

[246] Z. Sedighi, H. Ebrahimpoor-Komleh, A. Bagheri, and L. Kosseim. Opinion spam detection with attention-based neural networks. In *The Thirty-Second International Flairs Conference*, 2019.

[247] Y. Seo and C.-S. Jeong. Fagon: Fake news detection model using grammatic transformation on neural network. In *2018 Thirteenth International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, pages 1–5. IEEE, 2018.

[248] S. Shah and M. Goyal. Anomaly detection in social media using recurrent neural network. In *ICCS*, 2019.

[249] K. Shivagangadhar, H. Sagar, S. Sathyan, and C. Vanipriya. Fraud detection in online reviews using machine learning techniques. *International Journal of Computational Engineering Research (IJCER)*, 5(5):52–56, 2015.

[250] T. Shivaprasad and J. Shetty. Sentiment analysis of product reviews: a review. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pages 298–301. IEEE, 2017.

[251] K. Shu, D. Mahudeswaran, and H. Liu. Fakenewstracker: a tool for fake news collection, detection, and visualization. *Computational and Mathematical Organization Theory*, 25(1):60–71, 2019.

[252] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*, 2018.

[253] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36, 2017.

[254] A. Sinha, P. Patekar, and R. Mamidi. Unsupervised approach for monitoring satire on social media. In *Proceedings of the 11th Forum for Information Retrieval Evaluation*, FIRE '19, page 36–41, New York, NY, USA, 2019. Association for Computing Machinery.

[255] U. Sirisha and S. C. Bolem. Aspect based sentiment & emotion analysis with roberta, lstm. *International Journal of Advanced Computer Science and Applications*, 13(11), 2022.

[256] V. Sivasangari, A. K. Mohan, K. Suthendran, and M. Sethumadhavan. Isolating rumors using sentiment analysis. *Journal of Cyber Security and Mobility*, 7(1):181–200, 2018.

[257] T. A. Small. What the hashtag? a content analysis of canadian politics on twitter. *Information, communication & society*, 14(6):872–895, 2011.

[258] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.

[259] A. Stöckl. Detecting satire in the news with machine learning. *arXiv preprint arXiv:1810.00593*, 2018.

[260] A. Suciati and I. Budi. Aspect-based sentiment analysis and emotion detection for code-mixed review. *International Journal of Advanced Computer Science and Applications*, 11(9), 2020.

[261] C. Sun, Q. Du, and G. Tian. Exploiting product related review features for fake review detection. *Mathematical Problems in Engineering*, 2016, 2016.

[262] B. Taboubi, M. A. B. Nessir, and H. Haddad. icompass at checkthat! 2022: combining deep language models for fake news detection. *Working Notes of CLEF*, 2022.

[263] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*, 2017.

[264] Y. Tanaka, Y. Sakamoto, and T. Matsuka. Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. In *2013 46th Hawaii International conference on system sciences*, pages 649–658. IEEE, 2013.

[265] K. Taşağal and Ö. Uçar. Detection of fake user reviews with deep learning. *International Journal of Research in Engineering and Applied Sciences (IJREAS)*, 8(12), 2018.

[266] A. Thota, P. Tilak, S. Ahluwalia, and N. Lohia. Fake news detection: A deep learning approach. *SMU Data Science Review*, 1(3):10, 2018.

[267] P. P. Thu and T. N. Aung. Effective analysis of emotion-based satire detection model on various machine learning algorithms. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pages 1–5. IEEE, 2017.

[268] P. P. Thu and N. New. Implementation of emotional features on satire detection. In *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 149–154. IEEE, 2017.

[269] V. Tripathi, A. Joshi, and P. Bhattacharyya. Emotion analysis from text: A survey. *Center for Indian Language Technology Surveys*, 2016.

[270] D. Trpevski, W. K. Tang, and L. Kocarev. Model for rumor spreading over networks. *Physical Review E*, 81(5):056102, 2010.

[271] Ö. Uçar and K. Taşağal. Using lstm neural networks for fake online user review detection.

[272] Y. Vasiliev. *Natural language processing with Python and spaCy: A practical introduction*. No Starch Press, 2020.

[273] T. P. Velavan and C. G. Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3):278, 2020.

[274] M. Venugopalan and D. Gupta. An unsupervised hierarchical rule based model for aspect term extraction augmented with pruning strategies. *Procedia Computer Science*, 171:22–31, 2020.

[275] A. P. B. Veyseh, M. T. Thai, T. H. Nguyen, and D. Dou. Rumor detection in social networks via deep contextual modeling. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 113–120, 2019.

[276] O. Vinyals and Q. Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[277] C.-C. Wang, M.-Y. Day, C.-C. Chen, and J.-W. Liou. Detecting spamming reviews using long short-term memory recurrent neural network framework. In *Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government*, pages 16–20, 2018.

[278] W. Y. Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

[279] X. Wang, K. Liu, and J. Zhao. Handling cold-start problem in review spam detection by jointly embedding texts and behaviors. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 366–376, 2017.

[280] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857, 2018.

[281] Z. Wang, Y. Zhang, and T. Qian. Fake review detection on yelp. 2017.

[282] W. Wei and X. Wan. Learning to identify ambiguous and misleading news headlines. *arXiv preprint arXiv:1705.06031*, 2017.

[283] J. Weng and B.-S. Lee. Event detection in twitter. *Icwsm*, 11(2011):401–408, 2011.

[284] C. Whitehouse, T. Weyde, P. Madhyastha, and N. Komninos. Evaluation of fake news detection with knowledge-enhanced language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1425–1429, 2022.

[285] D. Widyantoro and Y. Wibisono. Modeling credibility assessment and explanation for tweets based on sentiment analysis. *Journal of Theoretical and Applied Information Technology*, 70(3), 2014.

[286] Y. A. Winatmoko, A. A. Septiandri, and A. P. Sutiono. Aspect and opinion term extraction for hotel reviews using transfer learning and auxiliary labels. *arXiv preprint arXiv:1909.11879*, 2019.

[287] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang. A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Systems*, 148:66–73, 2018.

[288] L. Wu, J. Li, X. Hu, and H. Liu. *Gleaning Wisdom from the Past: Early Detection of Emerging Rumors in Social Media*, pages 99–107.

[289] L. Wu and H. Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*, pages 637–645, 2018.

[290] L. Wu, F. Morstatter, X. Hu, and H. Liu. Mining misinformation in social media. In *Big data in complex and social networks*, pages 135–162. Chapman and Hall/CRC, 2016.

[291] D. Xenos, P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis, and I. Androutsopoulos. Aueb-absa at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In *SemEval@ NAACL-HLT*, pages 312–317, 2016.

[292] N. Xu, G. Chen, and W. Mao. Mnrd: A merged neural model for rumor detection in social media. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018.

[293] Y. Xu, C. Wang, Z. Dan, S. Sun, and F. Dong. Deep recurrent neural network and data filtering for rumor detection on sina weibo. *Symmetry*, 11(11):1408, 2019.

[294] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu. Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PloS one*, 15(9):e0239441, 2020.

[295] F. Yang, A. Mukherjee, and E. Dragut. Satirical news detection and analysis using attention mechanism and linguistic features. *arXiv preprint arXiv:1709.01189*, 2017.

[296] K.-C. Yang and F. Menczer. Large language models can rate news outlet credibility. *arXiv preprint arXiv:2304.00228*, 2023.

[297] K.-C. Yang, T. Niven, and H.-Y. Kao. Fake news detection as natural language inference. *arXiv preprint arXiv:1907.07347*, 2019.

[298] Q. Yang, H. Alamro, S. Albaradei, A. Salhi, X. Lv, C. Ma, M. Alshehri, I. Jaber, F. Tifratene, W. Wang, et al. Senwave: Monitoring the global sentiments under the covid-19 pandemic. *arXiv preprint arXiv:2006.10842*, 2020.

[299] Z. Yang, C. Wang, F. Zhang, Y. Zhang, and H. Zhang. Emerging rumor identification for social media with hot topic detection. In *2015 12th Web Information System and Application Conference (WISA)*, pages 53–58. IEEE, 2015.

[300] A. Yenkikar, N. Babu, S. Sangve, and M. Bali. Oacd-sa: Online aspect category detection for sentiment analysis using unsupervised learning. In *2019 IEEE Pune Section International Conference (PuneCon)*, pages 1–7. IEEE, 2019.

[301] H. Yin, S. Yang, and J. Li. Detecting topic and sentiment dynamics due to covid-19 pandemic using social media. In *International Conference on Advanced Data Mining and Applications*, pages 610–623. Springer, 2020.

[302] K. H. Yoo and U. Gretzel. Comparison of deceptive and truthful travel reviews. In *ENTER*, pages 37–47, 2009.

[303] F. Yu, Q. Liu, S. Wu, L. Wang, T. Tan, et al. A convolutional approach for misinformation identification. 2017.

[304] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. *arXiv preprint arXiv:1909.04465*, 2019.

[305] S. Zannettou, S. Chatzis, K. Papadamou, and M. Sirivianos. The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 63–69. IEEE, 2018.

[306] L. Zeng, K. Starbird, and E. S. Spiro. # unconfirmed: Classifying rumor stance in crisis-related social media messages. In *Tenth International AAAI Conference on Web and Social Media*, 2016.

[307] Z.-Y. Zeng, J.-J. Lin, M.-S. Chen, M.-H. Chen, Y.-Q. Lan, and J.-L. Liu. A review structure based ensemble model for deceptive review spam. *Information*, 10(7):243, 2019.

[308] C. Zhang, A. Gupta, C. Kauten, A. V. Deokar, and X. Qin. Detecting fake news for reducing misinformation risks using analytics approaches. *European Journal of Operational Research*, 279(3):1036–1052, 2019.

[309] Q. Zhang, A. Lipani, S. Liang, and E. Yilmaz. Reply-aided detection of misinformation via bayesian deep learning. In *The World Wide Web Conference*, pages 2333–2343, 2019.

[310] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[311] X. Zhang, J. Zhao, and Y. LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.

[312] Z. Zhang, Z. Zhang, and H. Li. Predictors of the authenticity of internet health rumours. *Health Information & Libraries Journal*, 32(3):195–205, 2015.

[313] Z. Zhao, P. Resnick, and Q. Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, page 1395–1405, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.

[314] H.-T. Zheng, J.-Y. Chen, X. Yao, A. K. Sangaiah, Y. Jiang, and C.-Z. Zhao. Clickbait convolutional neural network. *Symmetry*, 10(5):138, 2018.

[315] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani. Fake news early detection: A theory-driven model. *arXiv preprint arXiv:1904.11679*, 2019.

[316] X. Zhou, X. Tao, J. Yong, and Z. Yang. Sentiment analysis on tweets for social events. In *Proceedings of the 2013 IEEE 17th international conference on computer supported cooperative work in design (CSCWD)*, pages 557–562. IEEE, 2013.

[317] Y. Zhou. Clickbait detection in tweets using self-attentive network. *arXiv preprint arXiv:1710.05364*, 2017.

[318] Y. Zhu, H. Wang, Y. Wang, Y. Li, Y. Yuan, and J. Qiang. Clickbait detection via large language models. *arXiv preprint arXiv:2306.09597*, 2023.

[319] L. Ziora. The sentiment analysis as a tool of business analytics in contemporary organizations. *Studia Ekonomiczne*, 281:234–241, 2016.

[320] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2):1–36, 2018.

[321] N. A. Zuhroh and N. A. Rakhmawati. Clickbait detection: A literature review of the methods used. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 6(1):1–10, 2020.