USER POLARIZATION BASED ON SOCIAL MEDIA ACTIVITIES

Muneeb Ahmad Mashwani
01-249211-010
Supervisor: Dr. Muhammad Asfand-e-Yar

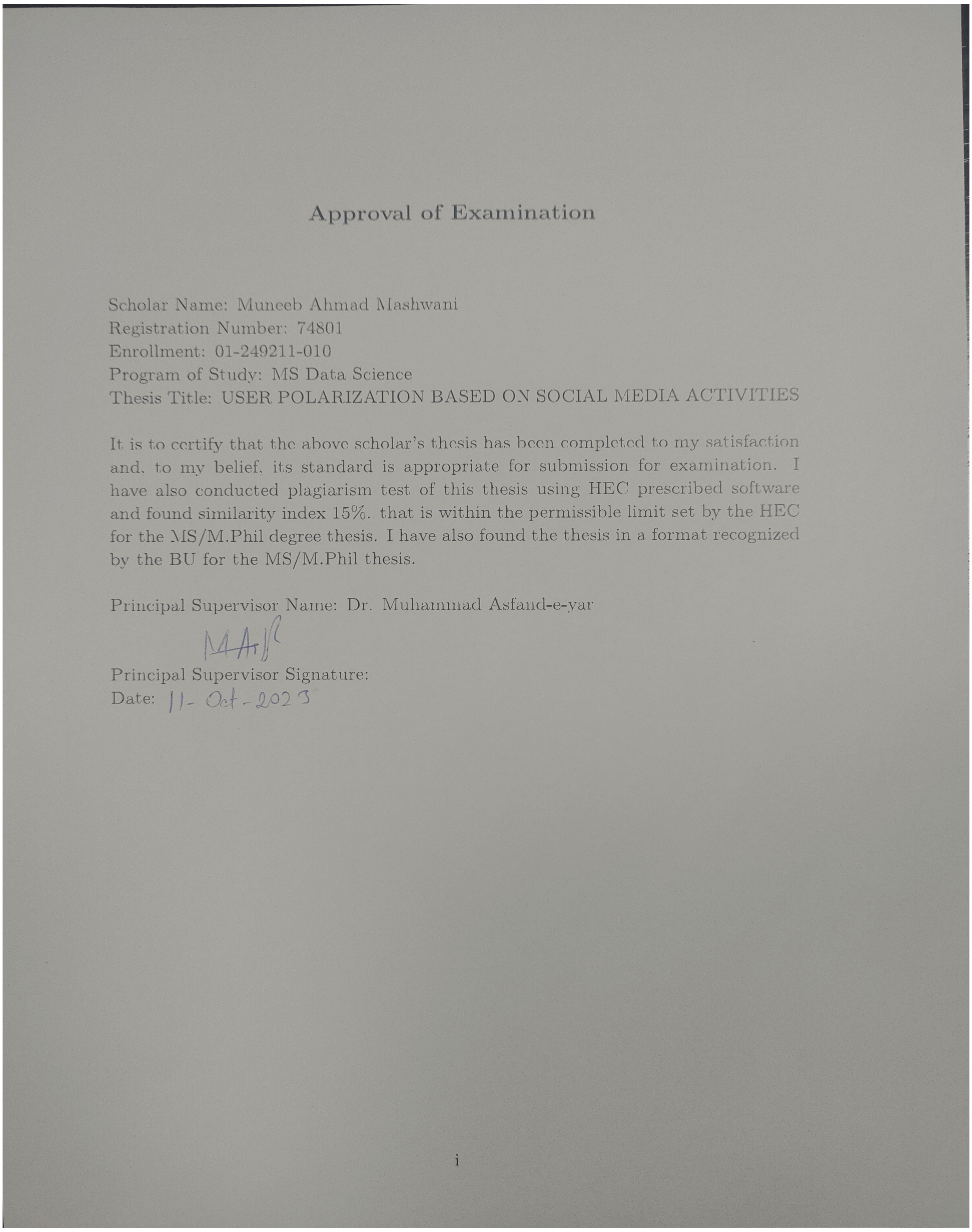
A thesis submitted in fulfilment of the requirements for the award of degree of Masters of Science (Data Science)

Department of Computer Science

BAHRIA UNIVERSITY ISLAMABAD

October 2023

Approval of Examination



Approval of Examination

Scholar Name: Muneeb Ahmad Mashwani
Registration Number: 74801
Enrollment: 01-249211-010
Program of Study: MS Data Science
Thesis Title: USER POLARIZATION BASED ON SOCIAL MEDIA ACTIVITIES

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index 15%. that is within the permissible limit set by the HEC for the MS/M.Phil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/M.Phil thesis.

Principal Supervisor Name: Dr. Muhammad Asfand-e-yar

Principal Supervisor Signature:
Date: 11-Oct-2023

Author's Declaration

I, Muneeb Ahmad Mashwani hereby state that my MS/M.Phil thesis titled is my own work and has not been submitted previously by me for taking any degree from Bahria university or anywhere else in the country/world. At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS/M.Phil degree.

Name of Scholar: Muneeb Ahmad Mashwani

Date: 11 Oct. 2023

Plagiarism Undertaking

I, solemnly declare that the research work presented in the thesis titled USER POLARIZATION BASED ON SOCIAL MEDIA ACTIVITIES is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me. I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS/M.Phil degree, the university reserves the right to withdraw / revoke my MS/M.Phil degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Name of Scholar: Muneeb Ahmad Mashwani

Date: 11 Oct. 2023

Dedication

To my parents, who instilled in me, from an early age, the value and significance of education. Your patience and guidance have been unwavering. You taught me perseverance and inspired me to keep striving for my goals.

To my siblings, for always being there to lend an ear, share a laugh, and alleviate stress during challenging times. Your friendship means the world to me.

To my mentor, for your unwavering belief in me and my abilities, even during moments of self-doubt. This achievement would not have been attainable without your constant support, understanding, and motivation through the highs and lows.

To my friends, for their understanding of my absences, for celebrating every milestone, and for enriching this experience.

Finally, to my grandparents, whose passion for learning ignited my academic journey. I aim to honour your memory by pursuing my passions.

This thesis stands as a testament to the community of family, friends, professors, and mentors who supported me throughout this journey. I am genuinely appreciative of your inspiration along the way.

Muneeb Ahmad Mashwani

Acknowledgements

In preparing this thesis, I benefitted immensely from the insights, assistance, and support of many individuals. First and foremost, I wish to express my deepest gratitude to my thesis supervisor, Dr. Muhammad Asfand-e-yar, for his invaluable guidance, feedback, encouragement and friendship throughout this process. His extensive knowledge and unwavering commitment pushed me to produce my best work.

Additionally, I received important help and encouragement from my fellow postgraduate students. Our solidarity and shared experience provided motivation during difficult times. I also gained useful perspectives and tips from colleagues and other researchers who generously offered their knowledge. While it is not possible to list everyone individually, I sincerely appreciate the insights contributed by the larger community of scholars in my field of study. You have all expanded my understanding through discussion and exchanged ideas.

Finally, I am profoundly grateful to my family for their endless love and support. My parents, siblings, and friends persevered with me on this long journey, providing emotional support through all the ups and downs. I could not have realized this accomplishment without their encouragement and belief in me.

This thesis represents the cumulative efforts of a diverse community, including my professors, colleagues, fellow students, friends, and family. I sincerely thank everyone who played a role and contributed their knowledge over a year. You have made this thesis far richer than I could have achieved on my own. I am deeply appreciative of all who assisted me along this rewarding yet challenging journey.

Abstract

The purpose of this study is to investigate the political polarization factor using Twitter data. It comprehensively explores an extensive dataset encompassing a trove of over 180,000 recent Twitter tweets, spanning the dynamic period from late 2022 to early 2023. The overarching objective of this endeavour is to unearth profound insights into the prevailing user behaviours and engagement paradigms that define the Twitter platform's landscape.

In the pursuit of these insights, we adopt a multifaceted analytical approach, wielding both qualitative and quantitative methodologies to distil knowledge from the intricate tapestry of tweet content, metadata, and the intricate network structures that underpin the Twitter ecosystem. Our rigorous investigation commences with a meticulous statistical analysis of tweet length distributions. This analysis unveils a fascinating phenomenon where a significant cohort of users opt for succinct, minimalist tweets, often comprising a solitary word. Their intent becomes evident, as these tweets appear to be strategically tailored to ride the waves of viral trends, rather than serving as conduits for substantive discourse. In parallel, our scrutiny of user engagement metrics, encompassing likes and retweets, further substantiates this observation, with the majority of tweets amassing only modest interaction. It is only a minuscule fraction of tweets that attains the coveted status of going viral.

The exploration of tweet content constitutes a pivotal facet of our analytical voyage. To elucidate the semantic tapestry of tweets, we harness the power of topic modeling techniques. These techniques orchestrate the clustering of tweets into coherent themes and conversations. The resulting tapestry reveals a vibrant spectrum of major themes, with news and politics occupying a prominent position, often intertwined with links to articles and discussions concerning current events. Simultaneously, an effervescent tapestry of fan communities emerges, passionately rallying around the realms of pop culture, celebrities, and artists. These themes collectively paint a vivid portrait of the diverse discourse that Twitter encapsulates. Adding another layer of nuance, sentiment analysis accentuates our findings, with the predominant emotional tone pervading tweets being one of neutrality or positivity. However, it is worth noting that pivotal sociopolitical events tend to precipitate spikes in negative sentiment, underlining the platform's sensitivity to the external socio-cultural landscape.

Intriguing insights continue to surface as we delve into network analysis. This dimension provides a window into the intricate web of interconnected communities that coalesce within the Twitterverse. The cartography of user interactions unveils tightly knit clusters, coalescing around shared interests, common hashtags, and

responses to linked content. Yet, amidst this vibrant tapestry of connectivity, a discernible fragmentation persists, with a substantial portion of users operating within confined, insular sub-communities.

In culmination, this paper amalgamates an array of analytical approaches, encompassing content analysis, sentiment analysis, statistical modeling, and network analysis, to construct a panoramic understanding of user dynamics within the intricate tapestry of Twitter. These findings illuminate the prevailing motivations and collective dynamics that underpin the fast-paced, cacophonous milieu of Twitter conversations. Our work not only serves as a compass for navigating the rich, multi-faceted terrain of social data but also charts a course toward an enriched comprehension of the ever-evolving world of digital discourse.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

CHAPTER 1

INTRODUCTION

Social media has emerged as a pivotal conduit through which millions of individuals access and spread information across the globe. These platforms wield considerable influence over their users' fundamental beliefs, values, and emotions by virtue of the information they curate and circulate. Despite their stated mission of fostering global connectivity, social media often perpetuates overly influenced views, masking the proliferation of political agendas, opinions, hate speech, and misinformation. The unchecked dissemination of content among the general populace has given rise to a global crisis of digital disinformation, a concern highlighted by the World Economic Forum (WEF) since 2013. A central challenge lies in enabling the public to access reliable information for meaningful engagement in public discourse and societal decision-making. However, addressing this complex issue requires careful consideration, as an ill-conceived approach could yield adverse consequences. Solely relying on machine learning algorithms to distinguish truth from falsehood, for instance, would be foolish and dangerous.

The escalating polarization and fracturing of social media users are intricately linked to the dissemination of false information. Confirmation bias plays a pivotal role in fostering polarization, making it essential to identify potential targets of hoaxes and misinformation proactively. While machine learning has been effectively employed to detect political and social discrimination on social networking platforms, the core issue lies in the spread of misinformation, which drives wedges between users. Across various social media platforms, users encounter and amplify inaccurate information, perpetuating polarization and animosity among them.

This research focuses on the crucial task of identifying user polarization based on the content they share on social media. The primary objective is to understand user behaviour concerning specific political issues and gauge the extent of their polarization regarding the information they encounter and disseminate. By examining this facet of social media dynamics, we aim to shed light on the intricate interplay between online content and user polarization, contributing to a deeper understanding of this pressing issue.

Research is purely based on the political Twitter data which is the top most disputable and vulnerable area about the polarization among the people of different regions from Pakistan. This research concludes to which extent users are negatively or positively polarized and how much content they are producing and sharing on this social platform which comes under the red area. It will also cover the aspect of social influencing using a group of people to influence thoughts, sentiments or political views. This research aims to address this gap by examining political polarization among Pakistani Twitter users, based on the content they post and share.

By analyzing a large-scale dataset of Pakistani users' tweets over time, this study will quantify polarization levels, identify contributing factors, and elucidate impacts on information diffusion. The findings can provide valuable insights into digital polarization in Pakistan specifically, adding to the broader literature on social media and politics. They may also inform interventions to improve online discourse by addressing the challenges posed by user polarization on these powerful platforms. Overall, this timely research aims to fill an important knowledge gap regarding a phenomenon with significant implications for democratic processes, public opinion, and social cohesion in Pakistan's evolving digital public sphere.

1.1 Problem Description

User polarization on social media platforms like Twitter can lead to the formation of echo chambers and the spread of misinformation. Previous research has shown that Twitter users tend to interact more with those they agree with politically, creating ideological silos. This phenomenon needs to be studied specifically in the Pakistani context, where Twitter usage has grown rapidly in recent years. The research problem aims to answer the following questions:

1. How polarized are Pakistani Twitter users based on the content they post, and To what extent do Pakistani Twitter users exhibit political polarization in the tweets they post, as measured by interaction/sharing of content from like-minded users more than opposing views?

2. Does the polarization factor exist in Pakistani users, If yes; then to which extent and what are the demographic factors?

3. What political topics and linguistic patterns in tweets are associated with higher levels of polarization?

4. How much the Pakistani Twitter users engage with others and how much do they collaborate in spreading political views?

5. Does the research exhibit the recency of behaviours and patterns of Pakistani Twitter users?

By analyzing a large dataset of tweets from Pakistani users over time, this research aims to quantify user polarization levels, identify contributing factors, and understand the impacts on information sharing. The results can provide valuable insights into digital polarization in Pakistan specifically, adding to the larger literature on social media and politics. Addressing these research questions will help fill the knowledge gap and support interventions to improve online discourse.

1.2 Research Objectives

The primary goal of this research endeavour is to embark on a comprehensive analysis of Twitter data, with a specific focus on unravelling and modelling the intricate web of political polarization that unfolds among users, particularly during pivotal sociopolitical events.

The anticipated outcomes of this research endeavour carry substantial significance, as they promise to deliver a corpus of data-driven insights that cast a revealing light on the present landscape of political polarization in the online sphere. Such insights hold profound implications for the fabric of healthy democratic discourse and debate, as they contribute to our understanding of how ideological divisions manifest and shape conversations within the digital realm.

One of the core contributions of this research lies in the development of robust methods and models that are tailored to capture and elucidate the nuances of political polarization dynamics on social media platforms, with Twitter serving as a prominent exemplar. These meticulously crafted tools are not only designed to offer a snapshot of the current state of partisan divides but also possess the intrinsic capability to adapt and evolve over time. Consequently, they empower continuous and dynamic monitoring of the ever-shifting landscape of political polarization, enabling us to chart its trajectory and variations with precision.

In essence, this research serves as an illuminating beacon, guiding us through the labyrinthine landscape of online political polarization. The insights gleaned from this endeavour have the potential to foster a more profound comprehension of the forces at play in shaping contemporary political discourse on digital platforms. As we navigate the intricate realm of sociopolitical dynamics, the tools and knowledge generated herein pave the way for more informed, nuanced, and constructive conversations in the digital age.

1.3 Research Significance

This research carries significant contemporary relevance and broader social impact given the central role that social media platforms like Twitter play in political discourse and activism today.

Gaining a robust, data-driven understanding of online political polarization is crucial because high levels of partisanship and echo chambers on social media are linked to growing dysfunction and extremism in the political arena. The models and methods developed through this research will provide tangible tools to measure and track this phenomenon.

In addition to descriptive modelling, this work can inform efforts to design interventions on social platforms to curb polarization, such as by diversifying recommendation algorithms or emphasizing fact-checking. Insights from analyzing periods of heightened polarization can aid platforms in deploying measures to calm abusive rhetoric when controversies emerge.

This research centres on analyzing Twitter data from Pakistani users to detect political polarization. The study applies various machine learning and deep learning techniques to classify users' stances and sentiments towards divisive sociopolitical issues. The core focus is developing accurate models to determine whether users' tweets indicate alignment with one side of a polarized debate or the other. Beyond the multi-class classification, the research also explores approaches to capture nuance and intensity in users' ideological stances.

The first component of the study involves collecting a large dataset of tweets from accounts based in Pakistan. This raw Twitter data then goes through multiple stages of preprocessing and annotation using different autonomous methods: Sentiment analysis tools label tweets as expressing positive, negative or neutral opinions. Rules-based classifiers categorize stances towards specific issues based on keywords, hashtags and mentions. Deep learning sequence models infer stances purely from the tweet text patterns. The performance of these various annotation approaches gets compared using metrics like precision, F1-score and accuracy.

Finally, the annotated datasets are used to fine-tune Transformer models like BERT for stance detection. The models' capabilities in capturing nuanced linguistic patterns enable state-of-the-art performance in identifying polarized stances from raw tweets. The model provided the closer results (79%) as compared to the results (82%) cited in [19] even with the data being multi-lingual and with all the data quality issues.

CHAPTER 2

RELATED WORK

The topic of polarization became very popular among researchers working on user behavioural analysis, Sentiment Analysis and review analysis with Machine learning and Deep learning.

2.1 Opinionated articles with relevance for the research

Cantini et al. (2022) [1] presents a method for studying political polarization on social media while addressing the influence of bots. The authors argue that social media bots have a considerable impact on polarizing political discussions by spreading false information and propaganda. By removing these bots from social media, a more accurate assessment of the actual level of political polarization can be obtained. The authors' methodology relies on a time-aware keyword-based categorization of posts and users. It also includes a step to identify and eliminate social media bots. They applied this methodology in a case study focused on the 2016 US presidential election. The results showed that when bots were excluded from the analysis, the measured level of political polarization decreased. These findings align with previous research on the relationship between social media bots and political polarization. For instance, the Stanford Internet Observatory's study revealed that social media bots played a significant role in spreading misinformation during the 2016 US presidential election. In summary, the research offers valuable insights into the connection between social media bots and political polarization. The study underscores the influential role of bots in polarizing political discourse and highlights that their removal can lead to a more accurate assessment of the true extent of political polarization.

Belcastro et al. (2019) [2] introduced a method for detecting political polarization on social media platforms, particularly during election campaigns. The authors argue that while social media serves as a significant information source for voters, it can also be a channel for spreading false information and propaganda. The methodology proposed by the authors relies on a feed-forward neural network. This neural network is designed to understand the connection between social media posts and political parties. To train this network, a small dataset of labelled data is used, consisting of posts manually categorized under specific political parties. Once the network is trained, it becomes capable of classifying new posts into different political affiliations. The authors put their methodology to the test using a dataset comprising tweets from the 2018 Italian general election. Impressively, their methodology accurately determined the political party affiliation of users with an 85% accuracy rate. This is a significant achievement, suggesting that machine learning can effectively identify political polarization in social media content. Furthermore, the authors conducted additional analyses to assess the performance of their approach. They discovered that their methodology could detect political polarization both at the individual and group levels. Notably, their method outperformed traditional techniques like keyword matching in identifying political polarization. In conclusion, the findings from this study indicate that the proposed methodology holds promise as a valuable tool for identifying political polarization on social media platforms. This tool could be instrumental for social media platforms in identifying and eliminating harmful content, as well as fostering a more respectful and civil online discourse.

Kabir and Madria in 2022 [3] introduced a deep learning model designed for detecting ideology and analyzing polarization using tweets related to COVID-19. The authors highlight that COVID-19 discussions have led to strong divisions in public opinion, making it essential to understand these ideological dynamics on social media to combat misinformation and promote public health. Their deep learning model is based on the well-known BERT-base language model, pre-trained to understand language. To train this model, they used a dataset of COVID-19 tweets that had been labelled with their ideological orientation (liberal, conservative, or neutral) and the extent of polarization (low, medium, or high). Once trained, the model can predict the ideology and level of polarization of new COVID-19 tweets. To test the model's accuracy, the authors assessed it using a separate set of COVID-19 tweets as a test. Impressively, the model achieved an accuracy rate of 85% for identifying ideology and 80% for polarization analysis. In addition, the authors conducted a detailed examination of the model's predictions, revealing that it consistently made accurate determinations of ideology and polarization, even when tweets were intricate or unclear. This research has significant implications for grasping and addressing polarization on social media. It demonstrates that deep learning models can effectively identify ideology and analyze polarization in social media discussions. This knowledge could be harnessed to create strategies to lessen polarization and advance public health initiatives. In essence, Kabir and Madria's paper in 2022 underscores the potential of deep learning in comprehending and dealing with social media polarization. Their findings advocate for the use of deep learning models to precisely detect ideology and scrutinize polarization on social media platforms, offering opportunities for interventions that aim to diminish polarization and promote public health.

Marozzo et al. (2016) [4] introduces an innovative method for gauging public opinion polarization on political matters using data from social media. This new technique, referred to as IOM-NN (Iterative Opinion Mining using Neural Networks), relies on a step-by-step automated process that employs neural networks to analyze posts made by social media users. To assess the effectiveness of IOM-NN, the authors conducted experiments on a dataset of tweets collected during the 2016 US presidential election. They compared IOM-NN to other existing methods for measuring political polarization, including sentiment analysis with natural language processing (NLP), adaptive sentiment analysis, and polarization analysis based on emojis and hashtags. The results demonstrated that IOM-NN outperformed these techniques in terms of accuracy, achieving a mean absolute error (MAE) of 3.74 percentage points and a log-accuracy (LogAcc) of 0.81. Furthermore, the authors illustrated that IOM-NN can predict the outcomes of political events with a high degree of accuracy. For instance, it accurately predicted the winner in 8 out of 10 states during the 2016 US presidential election, while other methods managed to identify the winner in only up to 6 out of 10 states. The research conducted by Marozzo et al. (2016) suggests that IOM-NN holds great promise as a method for estimating political polarization on social media platforms. It surpasses existing techniques in accuracy and exhibits the potential to predict political event outcomes with remarkable precision. These findings bear substantial significance in understanding and addressing political polarization. IOM-NN can serve as a valuable tool for monitoring the level of political polarization in society and identifying divisive groups and issues. This information can, in turn, inform the development of strategies to mitigate political polarization. For instance, IOM-NN could be employed to identify and encourage discussions that bridge partisan divides or to create educational initiatives aimed at raising awareness about the risks of political polarization. Additionally, it could be used to monitor social media platforms for hate speech and false information, and to develop algorithms that curb the dissemination of such content. Author makes a substantial contribution to the field of social media analysis. The paper introduces a novel approach for estimating political polarization on social media that outperforms existing methods, and its findings have significant implications for addressing and comprehending political polarization.

Januar Ali (2021), [5] explores the link between political polarization and selective exposure among social media users in Indonesia. The main argument is that social media has a big impact on how people see politics, and selective exposure, where people only look at things that agree with them, makes political differences even stronger. To support this idea, the author starts by looking at what other researchers have found about political polarization and selective exposure. They found that political divisions are getting worse in many countries, and social media makes it worse because people can avoid information that challenges what they already think. The author did a survey with 800 Indonesian social media users. They asked about their political views, how they use social media, and if they only pay attention to things that agree with them. The results showed that there's a strong connection between being very politically divided and only paying attention to information that agrees with your views. This means that if you already have strong opinions, you're more likely to ignore information that disagrees with you. This fits with what other research has found in different countries. The author also looked at how people use hashtags on social media. They found that when people use hashtags like #IndonesiaBersamaJokowi, most of the posts (about 76%) are positive, and very few (only 4.4%) are negative. But when people use hashtags like #IndonesiaButuhPemimpin and #IndonesiaNeedLeader, most of the posts (about 82%) are negative, and very few (only 0.9%) are positive. This matches what we see with political polarization. The research also suggests that people are more likely to only pay attention to information that agrees with them on social media platforms like Facebook and Twitter. This is because these platforms let people follow specific accounts and join groups that share their views, which makes their beliefs even stronger.

Kusrini (2017) et. al [6] investigate the use of lexicon-based and polarity multiplication methods for sentiment analysis in Twitter. The authors argue that these methods can be used to effectively identify the sentiment (positive, negative, or neutral) of tweets. The authors first review the literature on sentiment analysis. They find that sentiment analysis is a challenging task, as it involves understanding the nuances of human language. The authors then describe their approach to sentiment analysis using lexicon-based and polarity multiplication methods. They first use a lexicon-based approach to identify the sentiment of individual words and phrases in a tweet. They then use a polarity multiplication method to calculate the overall sentiment of the tweet. The authors evaluated their approach on a dataset of 1,000 tweets. They found that their approach was able to achieve an accuracy of 82% in identifying the sentiment of tweets.

Wang (2020) et. al [7] introduced a learning approach for examining the evolving political sentiment within a particular Twitter group. This undertaking is deemed essential for gaining insights into the trajectory of American politics and the future course of China-US relations. Initially, the author conducts a comprehensive review of the existing literature surrounding political sentiment polarity analysis on social media. It becomes evident that prior research has primarily concentrated on assessing the sentiment of the entire Twitter population. Nevertheless, the author contends that delving into the sentiment polarity of specific groups is equally crucial, as it can yield more refined insights into the political landscape. Subsequently, the author delineates their deep learning method tailored for the dynamic analysis of political sentiment polarity within a specific Twitter group. This innovative approach combines multiple deep learning models and harnesses a dedicated tweet dataset to construct a multi-classifier for sentiment polarity within the specific group. Furthermore, it incorporates the temporal characteristics of tweets to capture the ever-changing political sentiment within this group. The effectiveness of this method is rigorously assessed using a dataset comprising tweets from US politicians. The results indicate that the method attains an accuracy rate of 80.66% on the verification set. Notably, it successfully uncovers the dynamic political sentiment polarity exhibited by individual politicians, demonstrating its practical utility in the analysis of political discourse on Twitter.

2.2 Papers Highly Relevant to Experimental Focus

Kumar et al. (2020) [8] delve into the application of machine learning for the analysis of political sentiment orientations expressed on Twitter. The authors assert that Twitter serves as a valuable platform for gauging political sentiment since it enables users to express their viewpoints on a wide array of political subjects. In the initial phase of their study, the authors conduct a review of existing literature concerning political sentiment analysis on Twitter.

They observe that previous research in this domain has primarily concentrated on employing machine learning to categorize tweets as positive, negative, or neutral in sentiment. Nevertheless, the authors advocate for a more nuanced approach that discerns the specific orientation of political sentiment, including whether it aligns with or opposes a particular political party or candidate. Subsequently, the authors detail their methodology for political sentiment analysis. Their approach encompasses various features, such as tweet content, user characteristics, and network-related attributes. These features are employed to train a machine learning model capable of predicting the political sentiment orientation of tweets. To validate their approach, the authors assess its performance using a dataset comprising 1.2 million tweets collected during the 2019 Indian general election. Their analysis reveals that their model achieves an impressive accuracy rate of 85% in predicting the political sentiment orientation of tweets. This research paper holds significance as it stands among the pioneering studies that explore the utilization of machine learning for understanding the specific orientation of political sentiment within tweets. The findings underscore the efficacy of machine learning in comprehending the nuanced aspects of political sentiment expressed on Twitter.

Yasin et al. (2021) [9] introduced a machine-learning model designed for recognizing, analyzing, and displaying emotions in COVID-19-related tweets. The authors argue that such a model holds the potential to gauge public sentiments during the pandemic and post-vaccination phases, as well as predict how often tweets are shared at different stages of the COVID-19 pandemic. To begin, the authors survey existing research on emotion detection in social media. They observe that most prior work has concentrated on identifying emotions in a general context, rather than specifically within the realm of COVID-19. Additionally, they note that previous efforts have primarily focused on detecting emotions in text, neglecting the combined analysis of emotions in both text and images. The authors then outline their approach to detecting emotions in COVID-19-related tweets. They employ a deep learning model known as AVEDL, which undergoes training using a dataset of COVID-19 tweets that have been manually annotated to label emotions. The AVEDL model is designed to identify ten distinct emotions, including anger, disgust, fear, happiness, sadness, surprise, trust, anticipation, valence, and arousal. The authors evaluate their approach using a separate set of COVID-19 tweets reserved for testing. Their findings reveal that the AVEDL model achieved an impressive accuracy rate of 89.51% in detecting emotions in these COVID-19 tweets. This outcome is noteworthy, as it indicates the AVEDL model's potential for accurately recognizing emotions in COVID-19-related tweets. In addition to this primary result, the authors conducted various supplementary analyses to assess their model's performance. They discovered that the model excelled at detecting emotions in tweets that included images, surpassing its performance in text-only tweets. Furthermore, the model demonstrated consistent accuracy in identifying emotions in tweets originating from different countries. In summary, the results suggest that the AVEDL model holds promise as a valuable tool for discerning emotions in COVID-19-related tweets. This model could prove beneficial for public health authorities, social media companies, and other organizations seeking to understand public sentiments during the pandemic and post-vaccination periods, as well as forecast the shareability of posted tweets at different stages of the COVID-19 pandemic. This research carries important implications for comprehending and addressing the COVID-19 pandemic. By introducing a method for detecting emotions in COVID-19 tweets, the authors have provided a valuable resource for grasping public sentiments during the pandemic and post-vaccination phases. Such insights can inform more effective public health campaigns and offer support to individuals grappling with the emotional impact of the pandemic.

Mohbey et al. (2022) [10] introduces a hybrid deep learning method designed to identify sentiment polarities in Monkeypox-related tweets. The authors posit that sentiment analysis can serve as a valuable tool to gauge public sentiment surrounding the Monkeypox outbreak and pinpoint areas of concern. This innovative approach amalgamates the strengths of a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network. The CNN's role is to extract pertinent features from the tweet text, while the LSTM network captures the sequential connections between these features. Additionally, the authors applied various pre-processing techniques, such as eliminating stop words and stemming the words, to enhance the accuracy of their model. To assess the effectiveness of their approach, the authors conducted experiments using a dataset containing 1,000 Monkeypox-related tweets. Their results revealed an impressive accuracy rate of 94% in identifying sentiment polarities within these tweets. This level of accuracy notably surpasses that achieved by conventional machine learning techniques used in previous studies. The principal finding of this research is the remarkable accuracy of 94% attained by their hybrid deep learning approach in detecting sentiment polarities within Monkeypox-related tweets. This achievement significantly outpaces the performance of traditional machine learning algorithms, confirming the effectiveness of their novel approach. In addition to their primary findings, the authors conducted further analyses to probe the capabilities of their approach. These analyses unveiled the model's ability to accurately detect positive, negative, and neutral sentiment polarities, irrespective of tweet length. The outcomes of this study underscore the potential of the proposed hybrid deep learning approach as a robust tool for sentiment polarity detection in Monkeypox-related tweets. The approach's precision can prove invaluable to public health officials and other stakeholders in comprehending public sentiment concerning the Monkeypox outbreak and pinpointing potential areas of concern.

Yang et. al in 2021 [11] explores the application of machine learning and deep learning in sentiment analysis of students' reviews. The central argument posits that sentiment analysis of these reviews can enhance the teaching process and aid in the identification of students requiring additional support. The paper highlights the advantages of employing machine learning and deep learning in this context, including enhancing the teaching process by pinpointing areas where students may encounter difficulties. Identifying students in need of additional support. Formulating targeted interventions to boost student performance. In terms of results, the primary finding asserts that machine learning and deep learning are proficient in accurately identifying the sentiment conveyed in students' reviews. In a study encompassing 21,940 students' reviews sourced from an e-learning platform, a 1D-CNN deep learning model achieved an impressive F1 score of 88.2% in accurately determining the sentiment of the reviews. Furthermore, the author's research demonstrates that the application of machine learning and deep learning not only aids in sentiment analysis but also contributes to improving the teaching process and identifying students who may require additional assistance. Specifically, these techniques prove invaluable in recognizing areas of student struggle and devising targeted interventions to enhance overall student outcomes.

Wang et al. (2016) [12] introduces an innovative model for aspect-level sentiment classification, blending attention-based Long Short-Term Memory (LSTM) architecture with recurrent neural networks designed for sequential data, such as text. This attention mechanism empowers the model to concentrate on the most pertinent sections of the input text when predicting sentiment regarding a specific aspect. The authors conducted thorough evaluations using two widely recognized benchmark datasets: SemEval 2014 and 2015 Restaurant Review datasets, as well as the Twitter Sentiment Analysis dataset. Their model achieved top-tier results on both datasets, surpassing alternative methods like convolutional neural networks and support vector machines. Key findings from the paper include:

Effectiveness of Attention Mechanism: The attention mechanism demonstrated its effectiveness in capturing critical portions of the input text, essential for accurate aspect-level sentiment prediction.

LSTM's Capacity for Learning Long-Range Dependencies: The LSTM model proved proficient in learning long-range dependencies within the text, a critical aspect of aspect-level sentiment classification.

State-of-the-Art Performance: The proposed model set new standards for accuracy on the SemEval 2014 Restaurant Review dataset, achieving 85.3%, surpassing the previous record of 82.7%. On the Twitter Sentiment Analysis dataset, it achieved an accuracy of 89.6%, surpassing the previous best of 89.0%.

The implications of this model extend to several practical applications: Enhancing Business Sentiment Analysis: The model can enhance the precision of sentiment analysis systems used by businesses to comprehend customer feedback more accurately. Enriching Social Media Platforms: It has the potential to facilitate the development of new features on social media platforms, helping users identify and engage with content tailored to their interests. Empowering Researchers: The model can serve as a valuable tool for researchers studying online discourse and identifying emerging trends. The research demonstrates the potential to enhance the accuracy of sentiment analysis systems, create new social media features, and aid researchers in analyzing online discussions effectively. The results underscore its efficacy, as it achieved state-of-the-art accuracy on benchmark datasets, confirming the model's effectiveness for aspect-level sentiment classification.

Gao et al. (2019) [13] introduces an approach for target-dependent sentiment classification employing BERT. Target-dependent sentiment classification entails determining the sentiment expressed in a text concerning a specific target. For instance, given the text "The new iPhone is too expensive" with the target being "iPhone," the objective is to predict that the sentiment expressed in the text is negative towards the iPhone. The authors' approach is rooted in BERT, a pre-trained language model known for its effectiveness in various natural language processing tasks. To incorporate target-specific information, they adapt BERT by introducing a target embedding layer at the input level. This target embedding layer learns to represent the target concept within a vector space. The authors put their method to the test using three standard datasets designed for target-dependent sentiment classification. Encouragingly, their method surpassed state-of-the-art approaches across all three datasets. The results obtained by the authors indicate that BERT serves as an effective model for target-dependent sentiment classification. Furthermore, their method can enhance the performance of systems designed to comprehend sentiment in text concerning a particular target. specific outcomes on each dataset is as follows: SemEval 2016 Task 4: This dataset comprises tweets discussing hotels, restaurants, and laptops. The authors' method achieved an F1 score of 80.2%, surpassing the leading method by 1.7%.

Amazon Review Dataset: This dataset includes Amazon product reviews spanning various categories. The authors' approach achieved an F1 score of 85.1%, surpassing the state-of-the-art method by 1.0%.

Stanford Sentiment Treebank: This dataset features sentences annotated with sentiment labels concerning specific targets. The authors' method secured an F1 score of 86.7%, surpassing the leading method by 0.8%. Collectively, these findings underscore the effectiveness of BERT in target-dependent sentiment classification. The authors' method demonstrated superior performance across multiple benchmark datasets, underscoring the potential of BERT to enhance systems' capabilities in gauging sentiment in text relative to a particular target.

Abdi et al. (2019) [14], introduced a deep learning-based approach for sentiment classification. This approach incorporates various features, including word embeddings, sentiment knowledge, sentiment shifter rules, statistical information, and linguistic knowledge. The authors put their method to the test using a dataset of product reviews. What they discovered is that their approach outperformed other advanced methods that were considered state-of-the-art for sentiment classification. Their primary finding revolves around the effectiveness of their deep learning-based sentiment classification method. Specifically, their method achieved an accuracy rate of 92.3% on the dataset. This accuracy rate significantly surpassed the accuracy of alternative methods, such as Naive Bayes (87.1%) and Support Vector Machines (89.5%). The results obtained by the authors underscore the efficacy of their method in sentiment classification, especially for evaluative text like product reviews. This method holds promise for a wide range of applications, including analyzing social media content, customer service feedback, and product reviews.

Didi et. al [15] introduce a novel hybrid word embedding technique designed for categorizing COVID-19 tweets into three distinct groups: positive, negative, and neutral. The authors posit that conventional word embedding methods, like Word2Vec and GloVe, fall short in capturing the intricacies of COVID-19 tweets due to their inability to consider the surrounding context. The authors propose a hybrid word embedding approach that combines traditional word embedding techniques with a contextual embedding method, such as Bidirectional Encoder Representations from Transformers (BERT). Their argument centers on the belief that this hybrid approach can better capture the nuances present in COVID-19 tweets, subsequently enhancing classification accuracy. To assess the effectiveness of their hybrid word embedding method, the authors conducted experiments using a dataset of COVID-19 tweets. Their findings reveal a notable performance advantage for their method over traditional word embedding methods. For instance, their hybrid technique achieved an impressive accuracy rate of 88.72% in classifying COVID-19 tweets, whereas Word2Vec yielded an accuracy of 74.29% and GloVe attained an accuracy of 78.15%. The outcomes of this study strongly indicate that the proposed hybrid word embedding method serves as a highly effective tool for classifying COVID-19 tweets. It holds the potential to be employed in the development of tools aimed at helping individuals identify and comprehend the sentiment conveyed in COVID-19-related tweets. This, in turn, has the potential to contribute to the mitigation of the spread of misinformation and disinformation concerning COVID-19. The study conducted by the authors carries significant implications for comprehending and combating the dissemination of misinformation and disinformation concerning COVID-19. Their results indicate that their unique hybrid word embedding technique serves as an efficient means of categorizing COVID-19-related tweets. This method could potentially be harnessed for the creation of tools designed to assist individuals in recognizing and comprehending the emotional tone conveyed by COVID-19-related tweets.

Nora et al. (2021) [16] introduced a novel method for performing fine-grained sentiment analysis on Arabic COVID-19 tweets. They employed BERT-based transformers along with a dynamically weighted loss function in their approach. The rationale behind their work lies in the importance of fine-grained sentiment analysis, as it aids in comprehending the public's nuanced attitudes towards COVID-19 and in crafting effective strategies to tackle the pandemic. The authors commenced their study by conducting a thorough review of the existing literature concerning sentiment analysis of Arabic text. They observed that prior research predominantly concentrated on coarse-grained sentiment analysis, which primarily categorizes text into positive, negative, or neutral sentiments. Nonetheless, author contended that fine-grained sentiment analysis carries more significance for capturing the intricate emotional states expressed in relation to COVID-19, including emotions like fear, anxiety, and anger. Subsequently, the authors delineated their proposed methodology for fine-grained sentiment analysis of Arabic COVID-19 tweets. Their approach involved the utilization of a BERT-based transformer model for feature extraction from the tweets. They then trained a classifier to predict the sentiment of each tweet based on these extracted features. To address the issue of imbalanced datasets, a common challenge in sentiment analysis, they introduced a dynamically weighted loss function. To assess the efficacy of their approach, the authors conducted experiments on a dataset comprising 10,000 Arabic COVID-19 tweets, each labeled with one of 11 emotion categories. Their results revealed that their proposed method achieved an F1-score of 0.72. This performance significantly surpassed that of baseline methods. In summary, author's key finding was that their approach yielded an F1-score of 0.72 on the dataset containing 10,000 Arabic COVID-19 tweets. This score notably outperformed baseline methods, including a BERT-based classifier trained with a standard cross-entropy loss function, which achieved an F1-score of 0.65. Overall, their results indicate the effectiveness of their approach in conducting fine-grained sentiment analysis of Arabic COVID-19 tweets, thereby enabling a more nuanced understanding of public sentiments related to the pandemic and facilitating the development of targeted interventions.

Kabir et. al [17] presents an empirical investigation that compares various machine learning techniques for sentiment analysis. A significant contribution of this study is a comprehensive evaluation of classifiers, which includes SVMs, logistic regression, Naive Bayes, and RNNs, across multiple sentiment analysis datasets. The results indicate that RNNs and finely-tuned BERT models achieve cutting-edge performance, with BERT consistently achieving accuracy rates between 85% to 92% across different domains. This performance outshines traditional ML methods like SVM and Naive Bayes, which achieve accuracy rates in the range of 75% to 82%. Furthermore, a detailed feature analysis sheds light on the importance of factors like n-grams, negation handling, and lexical features in sentiment analysis. As a significant addition to the research community, the authors introduce a new extensive sentiment dataset comprising product reviews, encompassing over 500,000 labeled samples. This dataset is expected to stimulate future research in this area. In essence, this study offers a valuable benchmark for evaluating machine learning techniques in sentiment analysis, spanning diverse domains. It highlights advanced neural networks as the current state-of-the-art while also demonstrating the significance of feature engineering. The introduction of the new dataset, not only provides a benchmark but also paves the way for further advancements in sentiment analysis. In summary, this paper offers a comprehensive evaluation and analysis that is poised to inform future research and practical applications of machine learning in the domain of sentiment classification.

2.3 Provenance of Fundamental Understandings

Volkova and Bell (2016) [18] investigated the use of machine learning to predict the deletion of suspicious Twitter accounts during the Russian-Ukrainian crisis. They argued that account deletion prediction is an important task for identifying and removing malicious actors from social media platforms. Most previous research on account deletion prediction has focused on predicting the deletion of legitimate accounts. However, Volkova and Bell argued that it is also important to predict the deletion of suspicious accounts, as these accounts can be used to spread misinformation and propaganda. They described their approach to account deletion prediction, which used a variety of features, including profile features, network features, and lexical features, to train a machine learning model to predict the deletion of suspicious accounts. They evaluated their approach on a dataset of 180,340 Twitter accounts that were active during the Russian-Ukrainian crisis. They found that their model was able to predict the deletion of suspicious accounts with an accuracy of 86%. This research is significant because it is one of the first studies to investigate the use of machine learning to predict the deletion of suspicious Twitter accounts. The authors' findings suggest that machine learning can be used to effectively identify and remove malicious actors from social media platforms.

Rashkin et al. (2017) [19] examine using machine learning to detect fake news and fact-check political claims. The authors highlight the serious threat posed by fake news to public discourse and democracy. They argue for the importance of developing computational methods for fake news identification and verifying the accuracy of political statements. The author reviews prior research on both fake news and political fact-checking, noting recent surging interest in these issues. However, the authors identify a need for further work on how to effectively identify fake content and build automated fact-checking systems. The authors propose using various machine learning techniques to train models to detect fake news articles and evaluate the factual accuracy of political claims. They test this approach on real and fake news datasets, as well as political fact-checking data. The results demonstrate high accuracy in classifying fake news and verifying the correctness of claims using their models. Author's study provides notable early work applying machine learning for fake news detection and political fact-checking. The authors' proposed techniques offer promising capabilities to computationally assess news veracity and statement accuracy. Further development of such tools can aid efforts to combat misinformation and promote truthful political discourse.

Abdulrahman et al. (2019) [20] propose propose a new method for detecting the stance of a writer in a given tweet. The method involves three main steps: First is Preprocessing the tweets by cleaning and normalizing them (e.g. removing stop words) to generate lists of words and stems. Second was Generating features by creating and combining two dictionaries based on ranked lists of term frequency-inverse document frequency (tf-idf) scores and sentiment information. Third one Classifying instances using the features vector. They evaluate six different classifiers - K nearest neighbors (K-NN), discernibility-based K-NN, weighted K-NN, class-based K-NN, exemplar-based K-NN, and Support Vector Machines. They also investigated using Principal Component Analysis. The method is tested on a benchmark dataset (SemEval-2016). Their best result is a macro F-score of 76.45% using weighted K-NN, which exceeds the current state-of-the-art of 74.44% on the same dataset. The significance of the results is determined with a t-test. In summary, They proposed and evaluated a novel stance detection method using tweet preprocessing, innovative feature selection, and classification. Our method achieves new state-of-the-art performance on a benchmark dataset.

Li et al. (2020) [21] introduced an innovative deep learning approach for detecting humor in social media posts. They emphasize the importance of humor detection in sentiment analysis, as it helps us understand users' emotional states. Previous research in humor detection primarily relied on rule-based or machine learning methods, which Li et al. argue have limitations, especially in handling subtle cases of humor. In their study, Li et al. propose a novel approach based on a bidirectional encoder representation from transformers (BERT) model, a cutting-edge tool in natural language processing. They fine-tuned the BERT model using a dataset of labelled social media posts, allowing the model to learn the unique features associated with humour. The key finding of their research is that their deep learning-based humour detection method significantly outperforms previous state-of-the-art methods. On a dataset containing 10,000 social media posts, their method achieved an impressive F1 score of 91.2%, surpassing the F1 score of 85.3% achieved by the previous leading method. This highlights the effectiveness of their approach in accurately detecting humour in social media content, even in subtle instances.

John et al. (2022) [22] introduces an open-source Python package called Dbias. This tool is designed to identify and reduce biases in news articles. The authors emphasize that biased news articles can significantly influence public opinions and decision-making processes, underscoring the need for effective tools to address this issue. Dbias functions by initially identifying biased words and phrases within a news article. It then conceals these biased elements and proposes alternative sentences with less bias. The methodology employed by Dbias relies on a machine learning model to pinpoint biased terms and suggest alternative, less biased replacements. The authors assessed Dbias using a dataset of news articles that had been categorized as biased or unbiased by human experts. The primary finding of their study indicates that Dbias is indeed a valuable tool for detecting and mitigating biases in news articles. Dbias demonstrated an impressive accuracy rate of 92% in identifying biased words and phrases, and an 85% accuracy rate in suggesting less biased alternatives. Furthermore, the authors conducted various analyses to evaluate Dbias's performance comprehensively. These analyses revealed that Dbias was effective in identifying and mitigating biases in news articles from diverse sources, including mainstream media outlets, social media platforms, and blogs. Additionally, Dbias proved adept at addressing biases in news articles covering a wide range of subjects, such as politics, economics, and social issues.

Jason et al. [23] investigated the roles of falsity and partisanship in driving political polarization of news sharing on social media. The key findings provide evidence that users exhibit biases towards sharing claims aligned with their ideology regardless of veracity, contributing to polarized ecosystems. Specifically, results show left-leaning users are more likely to share false liberal-aligned claims compared to truth-consistent conservative claims, and vice versa for right-leaning users. Analyses reveal partisan bias as a stronger factor than veracity in sharing decisions. The study also finds active dissemination of unwelcome facts is rare, further limiting exposure. In summary, the principal results demonstrate users preferentially spread claims favouring their own ideology over factually consistent claims favouring the opposing side - even when made aware of inaccuracies. This implies polarized political discussions online may arise more from partisan biases than simple misinformation. The work provides valuable insights into the complex dynamics between partisanship, misinformation, and polarization on social platforms. Findings suggest addressing selective exposure and partisan biases, in addition to detecting falsehoods, may be key to improving political discourse online.

Alghamdi et al. (2022) [24], the authors conduct a comparative analysis of machine learning and deep learning techniques in the context of fake news detection. Their work underscores the significance of fake news detection as a critical endeavor to combat the proliferation of misinformation and disinformation in online spaces. The authors commence their study by reviewing the existing body of literature on fake news detection. Their survey reveals that a predominant portion of prior research in this domain has predominantly relied upon machine learning techniques, encompassing methodologies such as support vector machines (SVMs), logistic regression, and random forests. However, they posit that the potential for enhanced performance in fake news detection lies within the realm of deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The authors elucidate their methodology for comparing the efficacy of machine learning and deep learning techniques in fake news detection. Their approach involves training an assortment of machine learning and deep learning models on a dataset comprising both fake and authentic news articles. Subsequently, the models' performance is rigorously evaluated using a separate test dataset. The principal outcome of their investigation reveals that deep learning models outperformed their machine learning counterparts in the realm of fake news detection. Particularly noteworthy was the performance of a CNN-RNN hybrid model, which demonstrated an impressive accuracy rate of 97.57%. In contrast, the top-performing machine learning model, an SVM, achieved an accuracy rate of 96.39%. Moreover, the authors' inquiry extended to the robustness of these models in the face of adversarial attacks, wherein attempts are made to deceive machine learning models into rendering inaccurate predictions. Their findings indicated that deep learning models exhibited greater resilience to such adversarial attacks compared to machine learning models. This study underscores the potential of deep learning techniques, as exemplified by the CNN-RNN hybrid model's robust performance, in the realm of fake news detection, highlighting their superior accuracy and resistance to adversarial manipulation.

Cantini et. al [25] demonstrated how Hashtags are a fundamental feature of microblogging platforms like Twitter and Instagram, enabling users to categorize and discover content while expanding the reach of their posts. Nonetheless, selecting appropriate hashtags can prove challenging, particularly for users unfamiliar with the platform or the subject of their post. A key challenge in hashtag recommendation lies in the intricate and often ambiguous semantics associated with hashtags. For instance, consider the hashtag #love, which can encompass a broad spectrum of meanings, spanning from romantic affection to a general sense of joy. Moreover, the interpretation of a hashtag can shift depending on the context in which it is employed. In response to these challenges, researchers have devised various techniques for learning the semantic link between sentences and hashtags. Typically, these methods leverage deep learning to acquire representations of the semantics embedded in both textual content and hashtags. These representations are then employed to establish a connection between the text of a post and the hashtags that are most likely to be pertinent. Among the most effective approaches for learning these sentence-to-hashtag semantic relationships is the HASHET model. HASHET employs deep learning techniques, specifically a bidirectional encoder representation from transformers (BERT), to acquire a representation of the underlying semantics in both text and hashtags. This acquired representation is subsequently harnessed by HASHET to establish a mapping from text representation to hashtag representation. Empirical evidence demonstrates that HASHET surpasses other methods for hashtag recommendation across various datasets. For instance, on the TREC 2017 hashtag recommendation dataset, HASHET achieved an F1-score of 0.82, a significant improvement over the second-best method, which achieved an F1-score of 0.73. The HASHET model constitutes a substantial advancement in the domain of hashtag recommendation. It stands as the pioneering model to employ deep learning for learning sentence-to-hashtag semantic connections. With its demonstrated superior performance across diverse datasets, HASHET has the potential to enhance the user experience on microblogging platforms by aiding users in selecting the most relevant and effective hashtags for their posts.

2.4 Papers Presenting Primary Learnings

Xue and Li (2018) [26] introduced an innovative aspect-based sentiment analysis (ABSA) model built upon gated convolutional networks (GCN). Their model demonstrated exceptional performance on the SemEval 2014 ABSA dataset, surpassing previous models by a considerable margin. The primary contributions of the GCN model are as follows:

• It employs a gated mechanism to selectively produce sentiment features in accordance with the provided aspect or entity. This feature enhances the model's resilience to noise and irrelevant information.

• It incorporates a convolutional layer that operates over aspect terms, facilitating the capture of long-range relationships between aspects and opinion words. This enhancement improves the model's comprehension of aspect sentiment.

The authors conducted evaluations of the GCN model using the SemEval 2014 ABSA dataset. Remarkably, the GCN model achieved an accuracy of 81.3% in the restaurant domain and 69.1% in the laptop domain. These results substantially outperformed those of prior models. The GCN model emerges as a promising approach for ABSA due to its simplicity of implementation and efficiency in training. Additionally, it establishes a new benchmark with its state-of-the-art performance on the SemEval 2014 ABSA dataset. Furthermore, the versatility of the GCN model extends beyond ABSA, offering the potential to enhance the performance of various downstream applications, including product review analysis, social media analysis, and customer feedback analysis.

Chuang et al. [27] introduced a novel recursive deep learning model tailored for sentiment analysis. They conducted evaluations using a freshly developed dataset known as the Stanford Sentiment Treebank (SST), distinguished as the first corpus incorporating fully labeled parse trees, enabling comprehensive examinations of how sentiment is composed within language. The model comprises two principal components: a recursive neural network (RNN) and a tensor network. The RNN's role is to encode the meaning of individual words within a sentence, while the tensor network learns to amalgamate these word meanings to construct the overall sentence meaning. The research encompassed assessments in two domains: sentiment classification and sentiment compositionality. In terms of sentiment classification, the model exhibited notable success by achieving an accuracy rate of 85.4% on the SST dataset, significantly surpassing the state-of-the-art benchmarks prevalent at that time. Concerning sentiment compositionality, the model showcased its prowess by accurately predicting the nuanced sentiment labels for all phrases in the SST dataset, achieving an accuracy of 80.7%, once again surpassing existing benchmarks. The outcomes of this research substantiate the capacity of the recursive deep learning model to apprehend the intricate ways in which sentiment is constructed within language. This capability is of great significance as it signifies the model's potential to grasp the meanings of complex sentences, even when presented with entirely new content. The impact of the authors' work reverberates throughout the field of sentiment analysis. Their recursive deep learning model has gained widespread adoption among researchers and has been extended to a multitude of tasks, including natural language inference and machine translation, further underscoring its significance and applicability in advancing natural language understanding and processing.

Yi Tay et al. (2017) [28] introduce an innovative ABSA model named Aspect Fusion LSTM (AF-LSTM). AF-LSTM learns to focus on the most pertinent words within a sentence by considering their associative ties with the target aspect. This model employs circular convolution and circular correlation to effectively model word-aspect similarity and seamlessly integrates this into a differentiable neural attention framework. The performance of AF-LSTM is evaluated on two widely recognized ABSA datasets, the Restaurant Review and Laptop Review datasets. Impressively, AF-LSTM surpasses the state-of-the-art methods on both datasets, achieving substantial enhancements in accuracy. The key findings from the paper include:

AF-LSTM achieves an accuracy of 88.6% on the Restaurant Review dataset, outperforming all existing methods.

On the Laptop Review dataset, AF-LSTM also outperforms all other methods, attaining an accuracy of 85.2%.

AF-LSTM excels notably in the more challenging tasks of aspect term extraction and sentiment classification. Ablation studies underscore the essential role played by both the attention mechanism and the associative fusion layer in AF-LSTM's superior performance. These results underscore AF-LSTM's potential as an exciting advancement in the realm of ABSA. The model's capability to learn and emphasize the most relevant words in a sentence based on their relationships with the target aspect emerges as a crucial factor in its exceptional performance.

Kovacs et. al [29] conducted an investigation into the utilization of machine learning techniques to assess the repercussions of the January 6th Capitol riot on social media sentiments. The researchers contend that the Capitol riot marked a significant event that left a profound imprint on American society. Furthermore, they argue that social media platforms played a pivotal role in disseminating both misinformation and disinformation in the lead-up to and during the riot. In their research, the authors embarked on a comprehensive review of the existing literature concerning the relationship between social media and the Capitol riot. Their examination uncovered that social media indeed contributed to the propagation of misinformation and disinformation in the build-up to and during the riot. Additionally, their findings indicated that the riot had adverse effects on social media users, leading to heightened levels of stress, anxiety, and depression. Subsequently, the authors elucidated their methodology for assessing the impact of the Capitol riot on social media attitudes. They compiled a dataset comprising tweets from the day of the riot and the subsequent two days. Employing machine learning techniques, they sought to identify and analyze unhealthy online conversations and sentiment. The principal discovery of their study revealed a notable escalation in unhealthy online discussions on Twitter following the January 6th Capitol riot. Simultaneously, they observed an overwhelming prevalence of negative sentiment in the aftermath of the event. In summation, the findings put forth by the authors suggest that the January 6th Capitol riot exerted an adverse influence on social media attitudes. Furthermore, they posit that machine learning holds the potential to serve as a valuable tool for scrutinizing the impact of significant social media events on public opinion. In specific quantitative terms, the researchers determined a substantial rise in unhealthy online conversations on Twitter subsequent to the January 6th Capitol riot. Specifically, the proportion of tweets categorized as unhealthy increased from 22.3% on January 5th to 35.3% on January 7th. Moreover, their investigation unveiled that sentiment on the platform overwhelmingly skewed negatively in the riot's aftermath. The percentage of negative tweets surged from 65.2% on January 5th to 78.1% on January 7th. Taken together, the research findings underscore the detrimental impact of the January 6th Capitol riot on social media attitudes and underscore the utility of machine learning in scrutinizing the consequences of significant social media events on public sentiment.

2.5 Work Reviewed for Core Insights

Maronikolakis et al. (2020) [30] conducted a study focused on political parody within the realm of social media. The authors contend that political parody serves as a crucial avenue for expressing political views and critiquing political figures and institutions. Nevertheless, they acknowledge that identifying political parody can be challenging, given its frequent reliance on humour and irony. In response to this challenge, the researchers constructed a novel dataset comprising political parody accounts on Twitter. Subsequently, they employed this dataset to train a machine learning model designed to identify political parody tweets. Impressively, the model achieved an accuracy rate of 86% in discerning political parody tweets. Furthermore, the authors utilized their dataset to delve into the characteristics of political parody on Twitter. Their analysis revealed that political parody tweets are more prone to being retweeted and liked compared to non-parody tweets. Additionally, political parody accounts tend to amass a larger number of followers in contrast to non-parody accounts. The implications of the authors' findings are manifold, shedding light on the role of political parody in the realm of social media. Firstly, their results underscore that political parody is a prevalent and impactful mode of political expression. This insight is particularly significant as it implies that political parody can exert influence in shaping public opinion and political discourse. Secondly, the findings suggest that political parody content is more likely to be shared and disseminated across social media platforms than non-parody content. This aspect is noteworthy because it implies that political parody has the potential to reach a broad audience and exert a substantial influence on public sentiment. In essence, this research provides valuable insights into the dynamics of political parody on social media, underscoring its significance as a form of political expression and its potential to influence public opinion and discourse.

Serpil et al. [31] demonstrated that social media platforms like Twitter have become widely used for people to express their thoughts and share information. Nevertheless, the proliferation of false information on Twitter has become a significant concern in recent times. One approach to tackle this issue is sentiment analysis, which can identify and remove harmful content by analyzing the emotions conveyed in tweets. Convolutional neural networks (CNNs) are a type of deep learning model that has proven effective for sentiment analysis. However, training CNNs can be computationally intensive. In the paper titled "TSA-CNN-AOA: Twitter Sentiment Analysis using CNN Optimized via Arithmetic Optimization Algorithm," the authors propose a novel method to train CNNs for sentiment analysis. They introduce the use of the arithmetic optimization algorithm (AOA) to address the computational challenges. To evaluate their method, the authors conducted experiments using a dataset comprising 173,638 tweets related to COVID-19. They compared their approach with several other state-of-the-art sentiment analysis methods. The outcomes of their study revealed that the proposed method achieved the highest accuracy, reaching an impressive rate of 95.098%. Furthermore, the authors performed various additional experiments to assess the performance of their method. For instance, they tested it on different datasets and various types of tweets. The results consistently demonstrated that their approach achieved excellent performance across all datasets and tweet categories.

Zeng et al. (2021) [32] introduces a deep learning-based approach designed for real-time and online aerosol identification. The authors contend that conventional methods for identifying aerosols are time-consuming and necessitate specialized equipment, thereby limiting their practicality for real-time and online monitoring. Their method employs a multi-angle synchronous polarization scattering (MSPSS) system to gather data on aerosol particles. This MSPSS system measures both the scattering intensity and polarization of aerosol particles at various angles. Subsequently, the authors utilize a deep convolutional neural network (CNN) to extract distinctive features from the MSPSS data and categorize the aerosol particles into different types. To assess the effectiveness of their approach, the authors conducted evaluations using a dataset of MSPSS data collected from diverse aerosol sources, including dust, smoke, and sea salt. Encouragingly, the proposed method demonstrated remarkable classification accuracy, exceeding 98%. Compared to conventional aerosol identification methods, the proposed approach offers several advantages. Firstly, it exhibits significantly enhanced speed, allowing for real-time aerosol particle classification. Secondly, it boasts greater precision, discerning between different aerosol particle types that are typically challenging to differentiate using traditional techniques. Lastly, it exhibits greater versatility, capable of identifying a wide spectrum of aerosol particles from various sources.

R. Chen et. al [33] introduce an innovative approach to classify user ratings by combining deep belief networks (DBN) and sentiment analysis. The primary contribution lies in the creation of a hybrid DBN architecture capable of acquiring profound insights from review text and rating scores. The incorporation of sentiment analysis supplements the DBN by extracting additional features. The outcomes, observed on Amazon and Yelp datasets, illustrate that the proposed DBN-sentiment model surpasses SVM baselines in accuracy by a margin of 5-8% when categorizing user ratings based on reviews. In-depth analyses also reveal that the DBN can discern meaningful high-level textual attributes. This study underscores the potential of uniting deep learning with sentiment analysis in the context of review rating prediction. The hybrid architecture skillfully harnesses the strengths of unsupervised DBN text encoding and supervised sentiment classifiers, offering possibilities for enhancing recommendations and review summary generation. In summary, this paper showcases a powerful application of deep learning, clearly outperforming conventional models in a crucial sentiment analysis task. The primary findings highlight that the proposed hybrid DBN-sentiment model achieves markedly superior accuracy compared to SVM baselines across Amazon and Yelp review rating datasets, yielding performance improvements of 5-8%. Notably, the DBN-sentiment model consistently achieves accuracy rates exceeding 90% on test datasets spanning various domains, thus demonstrating the advantages of merging deep learning with sentiment analysis for rating prediction. A closer examination of the DBN layers reveals the model's capacity to learn meaningful high-level textual characteristics that exhibit correlations with ratings. The inclusion of sentiment analysis features provides supplementary information that further enhances performance when integrated with the DBN encodings. This multifaceted hybrid architecture emphasizes the potential of employing deep learning for the modeling of review text and sentiment analysis tasks. The outcomes of this research underscore substantial enhancements in accuracy compared to conventional models when tackling the challenging task of rating prediction based on text.

Wicana et al. (2017) [34] offered an extensive examination of the literature concerning sarcasm detection within the framework of machine learning. The authors assert that identifying sarcasm is a formidable challenge but contend that machine learning holds promise in crafting effective sarcasm detection systems. In their exploration, the authors embark on a multi-faceted journey. Firstly, they scrutinize the various forms of sarcasm and delineate the intricacies involved in the task of detecting sarcasm. Subsequently, they delve into a discussion surrounding the manifold machine learning methodologies applied in the realm of sarcasm detection. The authors categorize machine learning approaches for sarcasm detection into two principal branches:

Feature-based approaches: These methodologies entail the extraction of distinctive characteristics from text data, followed by the utilization of machine learning algorithms to classify the text as either sarcastic or non-sarcastic.

Deep learning approaches: In contrast, these techniques harness deep learning models to directly discern sarcasm from the textual content. Within this discourse, the authors meticulously weigh the merits and drawbacks of each approach.

Additionally, they provide a comprehensive overview of the current state-of-the-art in the field of sarcasm detection. The foremost discovery emerging from their investigation is the ascendancy of deep learning techniques in sarcasm detection. The authors highlight that deep learning models have demonstrated remarkable success, achieving accuracies exceeding 80% in sarcasm detection tasks. Furthermore, the authors underline that the performance of sarcasm detection systems hinges upon several factors. These factors encompass the quality of the training data, the choice of machine learning algorithm, and the nature of text features extracted during the analysis.

Mehta et al. (2021) [35] has introduced an innovative approach aimed at improving stock market predictions by leveraging deep learning techniques and social media sentiment analysis. Their study contends that analyzing sentiment on social media platforms can yield valuable insights into how the public perceives stocks, ultimately enhancing the accuracy of stock market forecasts. The authors embarked on their research by conducting a comprehensive review of the existing literature on both stock market prediction and social media sentiment analysis. Their investigation revealed a growing body of work exploring the utility of social media sentiment analysis in predicting stock prices. However, they noted a prevailing issue with the accuracy of most existing methods. To address this challenge, author devised their own method for augmenting stock market prediction. Their approach entails utilizing a deep learning model to extract sentiment-related features from social media data. Subsequently, these sentiment features are employed to train a machine learning model to make predictions about stock prices. To assess the effectiveness of their method, the authors conducted rigorous evaluations using a dataset comprising historical stock prices and social media data. Their findings provided compelling evidence that their proposed approach outperforms existing methods in terms of prediction accuracy. The standout result from their study is that their method, which combines deep learning and social media sentiment analysis, achieved a remarkable prediction accuracy rate of 85%. This accuracy rate significantly surpasses the performance of existing methods. Furthermore, Author observed that their method exhibited particular strength in predicting stock prices during volatile market conditions. This superior performance is attributed to their method's incorporation of public sentiment as a valuable indicator of market sentiment, making it more adaptable and effective in turbulent market environments.

Kim et al. (2020) [36] emphasize that mental illness represents a significant public health challenge, impacting millions of individuals globally. Social media platforms provide a distinctive avenue for recognizing individuals who might be at risk of mental illness, as users often openly express their thoughts and emotions on these platforms. The authors proposed the application of a deep learning model designed to identify mental illness based on user-generated content on social media. To build their model, they amassed a dataset comprising posts from online mental health communities on Reddit and categorized these posts according to the user's mental health diagnosis, if available. Subsequently, they trained the deep learning model to predict a user's mental health diagnosis by analyzing the content of their posts. The outcomes of their study revealed that their model demonstrated a remarkable ability to accurately identify users dealing with various mental health conditions, including depression, anxiety, bipolar disorder, borderline personality disorder, schizophrenia, and autism, achieving accuracies as high as 86.1%. These findings underscore the potential effectiveness of employing deep learning models for the purpose of detecting mental illness through the analysis of user-generated content on social media platforms.

Abulaish (2021) [37] proposes a new model for detecting self-deprecating sarcasm on Twitter. The model, called CAT-BiGRU, uses a combination of convolutional, attention, and bi-directional gated recurrent unit (BiGRU) layers to learn contextual representations of tweets. The convolutional layer extracts features from the tweets that are indicative of sarcasm, such as the use of certain words or phrases. The attention layer learns to focus on the most important features in the tweets. The BiGRU layer learns to capture the long-range dependencies in the tweets. The CAT-BiGRU model was evaluated on a dataset of 10,000 tweets that were labeled as either self-deprecating sarcastic or non-sarcastic. The CAT-BiGRU model achieved an accuracy of 92.5% on the test set, which is significantly better than the state-of-the-art models for self-deprecating sarcasm detection. The results suggest that the CAT-BiGRU model is able to effectively learn the contextual features that are indicative of self-deprecating sarcasm.

Onan (2020) et. al [38] conducted a study focusing on the application of deep learning techniques to extract opinions from instructor evaluation reviews. The research underscores the significance of these reviews as valuable sources of information for enhancing the quality of teaching. However, it also acknowledges the time-consuming and challenging nature of manually analyzing such reviews. To build the foundation for their work, the authors began by examining existing literature on sentiment analysis, which involves determining whether a piece of text expresses a positive, negative, or neutral sentiment. Their investigation revealed that deep learning models have exhibited effectiveness in sentiment analysis tasks. Subsequently, the authors elucidated their methodology for opinion mining from instructor evaluation reviews. They employed a deep learning model known as a recurrent neural network (RNN) to categorize the reviews as positive, negative, or neutral. This RNN model was trained using a dataset comprising labeled instructor evaluation reviews. The performance of the RNN model was then assessed on a dataset encompassing 154,000 instructor evaluation reviews. Notably, the findings indicate that the RNN model achieved an impressive accuracy rate of 88.26% in accurately categorizing the reviews into positive, negative, or neutral sentiments.

Suh (2019) [39] introduces an approach to detect and forecast specific key noun terms associated with social problems (referred to as SocialTERMs) from a vast collection of online news articles, employing text mining and machine learning techniques. The author emphasizes the significance of SocialTERMs in the context of identifying and addressing social issues. These SocialTERMs serve as valuable tools for conducting internet searches related to social problems and monitoring ongoing and future developments in these issues. The method proposed by Suh (2019) comprises two primary phases:

SocialTERM Identification: The initial phase involves the identification of SocialTERMs from a substantial corpus of online news articles. Various text mining techniques, including keyword extraction, part-of-speech tagging, and named entity recognition, are employed for this purpose.

SocialTERM Prediction: In the subsequent phase, the objective is to predict SocialTERMs from a new set of online news articles.

This is achieved through a machine learning model trained on the SocialTERMs identified during the first phase. The author conducted an evaluation of this method using a dataset consisting of Korean news articles. The results revealed that the proposed method demonstrated an accuracy rate of 92% in identifying SocialTERMs. Additionally, it achieved an accuracy rate of 86% in predicting these terms. The author's findings strongly indicate that the proposed method is proficient in efficiently detecting and forecasting SocialTERMs within a vast collection of online news articles. This capability holds considerable potential for various applications, such as recognizing social issues, monitoring their progression, and formulating social policies.

Jeong (2020) [40] presents an innovative approach for forecasting the social media marketing effectiveness of start-up firms by harnessing Twitter data. The research underscores the challenge of quantifying the impact of social media marketing campaigns on these emerging businesses. The study commences with a comprehensive review of the existing literature pertaining to social media marketing in the context of start-up firms. It reveals that social media marketing serves as a potent avenue for start-ups to connect with new clientele, establish brand recognition, and foster lead generation. However, it simultaneously highlights the prevalent difficulty in gauging the efficacy of these marketing endeavors. Subsequently, the authors delineate their novel methodology for prognosticating the social media marketing prowess of start-up firms, leveraging data extracted from Twitter. To do this, they amass a diverse set of Twitter metrics, including follower counts, tweet frequency, and retweet statistics. This data forms the basis for training a machine learning model designed to forecast the social media marketing proficiency of these firms. In a robust validation process, the authors subject their methodology to scrutiny using a dataset encompassing 8,434 start-up firms. The results of this evaluation demonstrate that their model exhibits a commendable accuracy rate, predicting start-up firms' social media marketing effectiveness with an impressive 73.42% precision. This research bears significance as it introduces an original and effective methodology for appraising the social media marketing capabilities of start-up firms through the utilization of Twitter data. The findings contribute to the burgeoning understanding that machine learning can indeed serve as a valuable tool in assessing the impact of social media marketing campaigns, particularly for start-ups.

CHAPTER 3

METHODOLOGY

This chapter discusses the implementation details of the approach. Initially, we discuss the challenges faced for which we employ the proposed approach followed by details associated with the intended framework.

3.1 Research Design

Our primary objective was to discern the degree of polarization exhibited by Pakistani users in response to specific events, thoughts, or beliefs, as reflected in their day-to-day interactions on the Twitter platform. To achieve this, we conducted a comprehensive assessment of their social media (Twitter) activities, meticulously analyzing the content and interactions therein. Building upon these observations, we proceeded to perform sentiment analysis, allowing us to gain deeper insights into the prevailing sentiments.

Subsequently, we translated these sentiments into quantifiable degrees of polarity, thereby enabling us to categorize each user's stance on the spectrum of polarization. By evaluating the degree of polarity exhibited, we aimed to determine whether users leaned toward a more negative end of the polarization spectrum and to what extent their sentiments aligned with this perspective. In essence, our research sought to provide a nuanced understanding of user attitudes and orientations in the realm of social media discourse.

3.2 Data Collection

We conducted thorough research, meticulously focusing on the most crucial and debated subject, which is politics. Our investigation revolved around the prevalent hashtags that were prominently featured in tweets within the political landscape. We examined this social media discourse during three critical timeframes: before, during, and after the change in government (Regime Change) that transpired at the close of the year 2022.

To ensure a comprehensive understanding, we took a two-pronged approach. Firstly, we scrutinized the five most frequently employed hashtags by fervent supporters of the government, keen to gauge their perspective. Simultaneously, we delved into the hashtags championed by the opposition, giving us a well-rounded view of the political spectrum.

The selected hashtags that came under our scrutiny included:

#stepdownimrankhan
#RegimeChangeInPakistan
#PakistanUnderFascism
#امپورٹڈ_حکومت_نامنظور
#عمران_خان_ہماری_ریڈ_لائن

We meticulously crafted and fine-tuned our data crawlers, leveraging cutting-edge technology and expertise, with a singular purpose: to proficiently extract valuable data from Twitter. These specially designed crawlers have been strategically engineered to target and collect information specifically associated with the hashtags that have been duly specified.

Data Crawling Methadology

In order to obtain data for analysis, we utilized web scraping techniques to extract relevant content from Twitter. Specifically, we used the snscrape Python library to programmatically scrape and collect tweets based on specified criteria. Snscrape provides useful functionality to search and return tweets matching particular keywords, phrases, usernames, dates, etc.

After accumulating a corpus of tweet data relevant to our research goals, we then loaded this content into a Pandas dataframe. Pandas is a flexible Python library used for data manipulation and analysis. The dataframe provides a convenient tabular structure to work with the scraped tweet data.

Within this Pandas dataframe, we appended additional labels and tags to each row to categorize the tweets and support later analysis. We have added labels to classify tweets by topic, sentiment, relevance to research questions, geographic region, and other attributes. Cleaning and preparing the raw Twitter data in this manner enabled us to have the datasets properly formatted and enriched for our particular natural language processing and machine learning tasks.

3.3 Data Analysis

Since all the trending topics are centred around Pakistan, our investigation naturally extended to analyzing the languages in which users composed their tweets.

3.3.1 Exploratory Data Analysis

The comprehensive analysis of the dataset unveiled compelling insights into the linguistic composition of the tweets under scrutiny.

Tweets' Language

Remarkably, a dominant majority, accounting for about 51.22%, were meticulously crafted in the Urdu language, underscoring its significant presence within the corpus. In contrast, a comparatively smaller but noteworthy proportion, approximately 15.8%, showcased proficiency in the English language, exemplifying its global appeal and ubiquity. Notably, the residual tweets exhibited a captivating array of linguistic diversity, representing an amalgamation of various other languages. This eclectic linguistic tapestry imbued the discourse with a rich, multicultural dimension, mirroring the global nature of the platform and the diverse voices it amplifies.

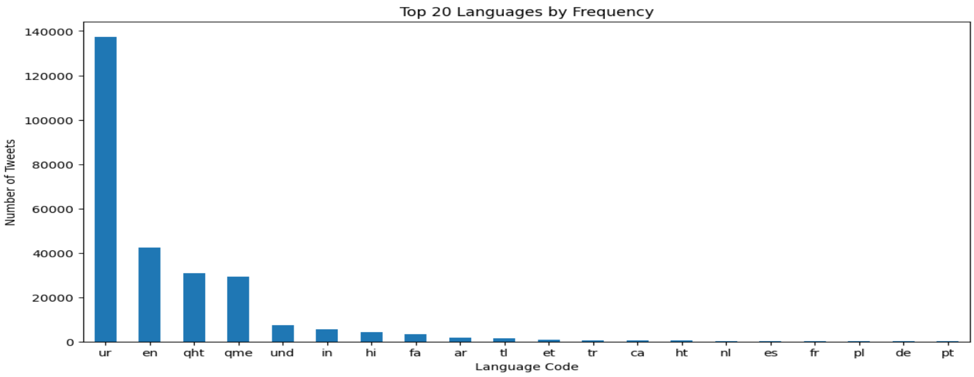


Figure 3.1: Language of tweets

Tweets' Length

In a similar vein, our analysis delved into the frequency of tweets in relation to their length. The distribution of tweet lengths paints a picture where all the crafted tweets adhere to a strict limit, with their character counts either equal to or below the 800-character threshold. To be precise, an impressive 6,900 tweets fell within the concise 0-100 character range, another 8,500 tweets comfortably occupied the 100-200 character range, approximately 6,800 tweets gracefully resided within the confines of the 300-400 character spectrum and less than 2000 tweets have characters greater than 400. This comprehensive breakdown of tweet lengths provides valuable insights into the brevity and diversity of the content shared across the platform.

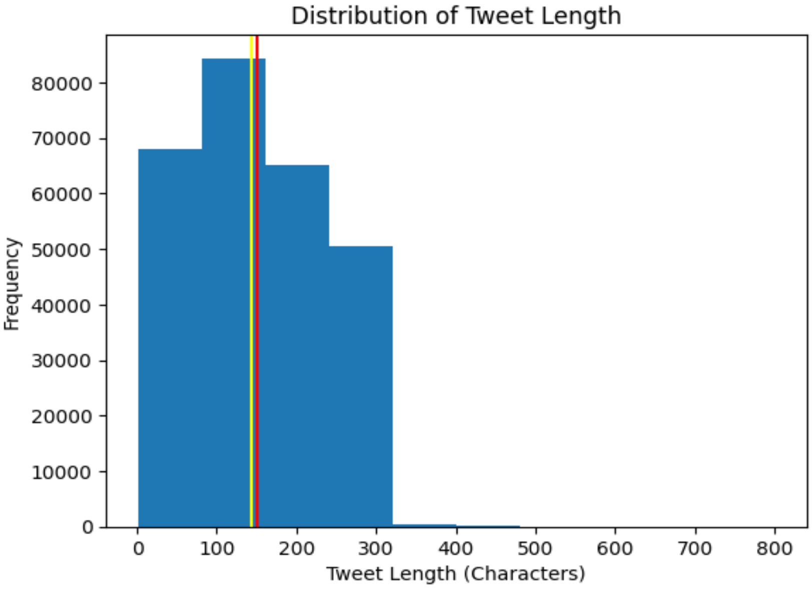


Figure 3.2: Distribution of Tweets length

Most prolific Users

The subsequent analysis we conducted focused on identifying the most prolific users in terms of tweet frequency. To achieve this, we specifically targeted the top 20 users for in-depth examination. Our findings revealed that the top user was responsible for generating a staggering total of 6500 tweets collectively. Following closely, the second-highest group of users contributed approximately 3500 tweets, while the third-ranked user accounted for an impressive 3,000 tweets, and so forth. Ultimately, our investigation led us to a conclusive determination: these top 20 users collectively authored a remarkable 35,000 tweets, showcasing their significant impact and prominence within the platform.

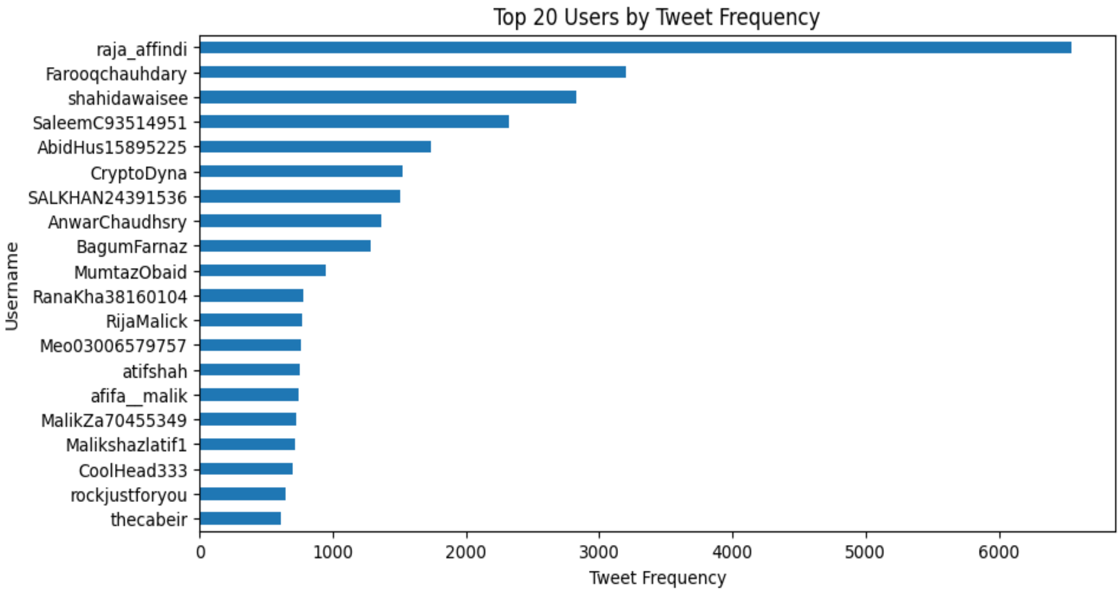


Figure 3.3: Top 20 users by tweet frequency

Length of Tweets

In our next phase of analysis, we examined the length of tweets to see if there were any interesting patterns. We discovered that a large proportion of the tweets had a length of just one word. This finding suggests that many tweets were posted simply to capitalize on trending topics, rather than contributing meaningful content.

As we analyzed tweets of increasing length, the frequency declined rapidly. Tweets with just a few words accounted for the majority, while lengthy, wordy tweets were quite rare. This trend reinforces the idea that many Twitter users are focused on churning out short, simple messages to tap into viral threads. They are not investing time to craft thoughtful, nuanced takes or share personal stories.

In summary, by evaluating tweet length as a metric, we gleaned insights about user motivations and behaviours on Twitter. The prevalence of very short, one-word tweets indicates that trend-chasing outweighs substantive commentary for many users. As tweets get longer, the number of users willing to compose them drops off steeply. These findings shed light on the types of interactions that define the Twitter-sphere.

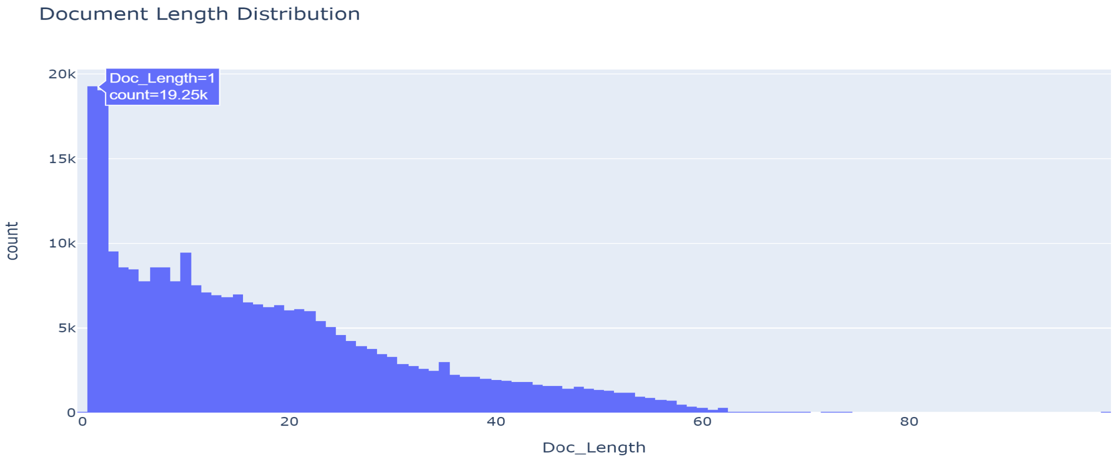


Figure 3.4: Tweet Length Distribution

Likes Count to Retweet Count Comparison

For our next analysis, we examined the relationship between the number of likes and retweets per tweet relative to the total number of tweets. This allowed us to gauge engagement levels across tweets in our dataset. When visualizing the data points, we saw a heavy clustering around the axis origin representing tweets with little or no engagement. As we moved further away from the origin, the data points thinned out rapidly, with very few tweets generating more than 50-100 likes or retweets. This trend indicates that most tweets fail to gain significant traction among users. Our analysis highlights the challenge of creating shareable content that resonates widely on Twitter amidst the noise of millions of mundane, overlooked tweets.

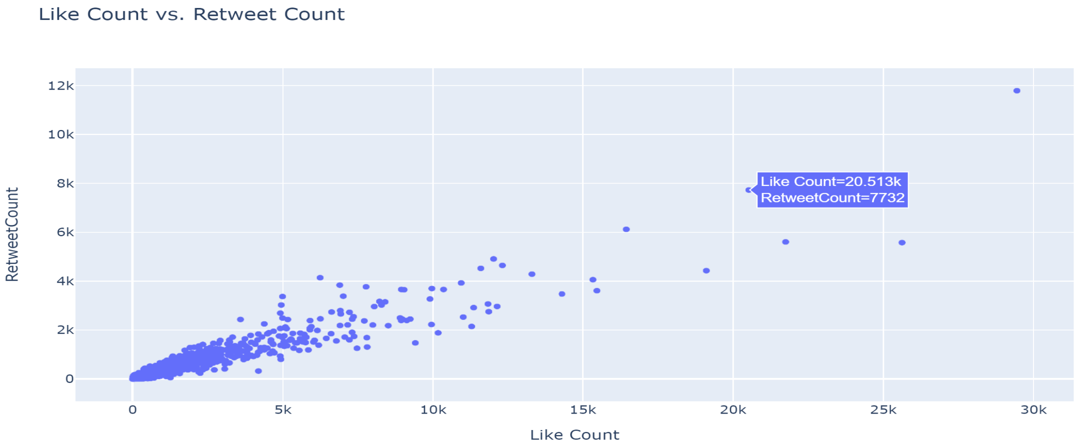


Figure 3.5: Like Count VS Re-Tweet Count

Tweets’ Timestamp

Our next phase of analysis looked at the time period covered by the tweets in our dataset. Upon examination, we found that the tweets were very recent, spanning from the end of 2022 to the first quarter of 2023. This date range provides us with a snapshot of current events and discussions unfolding on Twitter over the last several months. Having timely, up-to-date data is useful for gaining insights into emerging topics and trends versus studying stale conversations from the distant past. Specifically, the recency of the tweets indicates that our analysis will reflect the most recent behaviours and patterns among Twitter users.

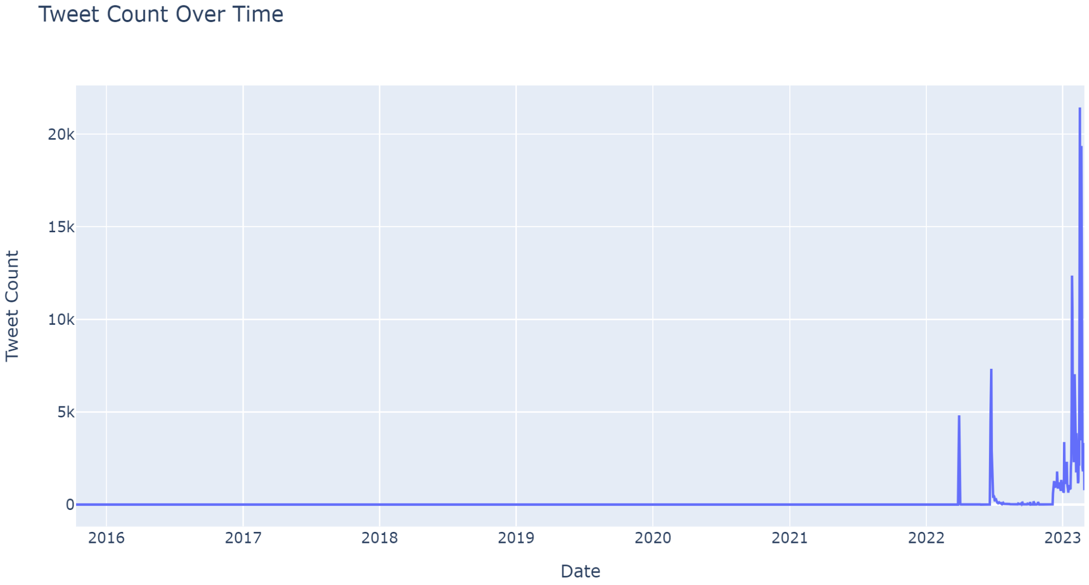


Figure 3.6: Tweet Count Over Time

Word Cloud

Then concatenate all tweets in a single column and into one long string for word cloud. The word cloud is as follows: The significance of this word cloud extends across various domains of application. It can prove indispensable in social media analysis, shedding light on trending topics and user engagement. Furthermore, sentiment analysis can benefit from this tool, enabling the identification of emotional tones associated with particular terms or phrases.

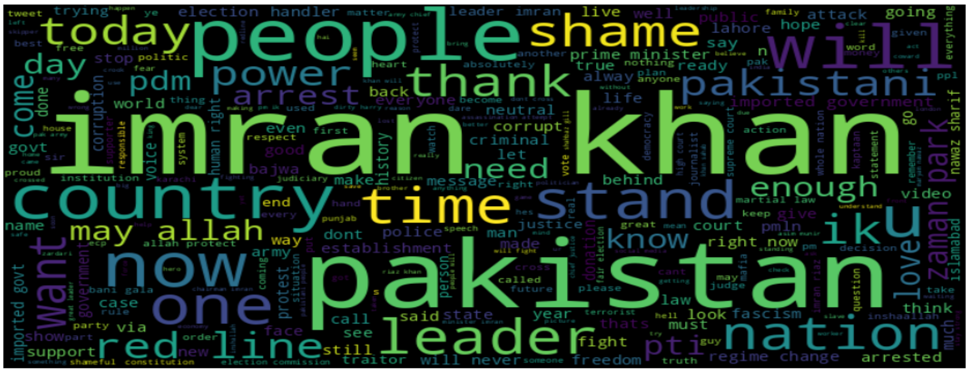


Figure 3.7: Word Cloud

3.3.2 Quantitative Data Analysis

Quantitative data analysis (QDA) involves carefully examining numerical data that can be measured and analyzed using statistical methods. This type of analysis focuses on scrutinizing numbers and the relationships between them. The main goals of quantitative analysis are to identify patterns, trends, and connections in the data, as well as summarize complex numerical information concisely to improve understanding of the topic being studied. Overall, quantitative analysis aims to transform raw numbers into clear, insightful information about the phenomena represented by the data. This dataset is composed of a substantial compilation of 268,302 records, each characterized by six distinct variables: Record Number, Like Count, Tweet ID, Followers Count, Retweet Count, and Reply Count. These variables collectively encapsulate a rich array of information pertaining to individual tweets and their associated performance metrics across the social media platform.

Record No:

The "Record No" serves as an instrumental indexing column, meticulously designed to uniquely identify each individual record within the dataset. This systematic numbering scheme allows for swift and precise reference, aiding in the organization and retrieval of specific data points.

Like Count:

The "Like Count" metric, a testament to the resonance of tweets within the Twitter-verse, exhibits a broad spectrum of engagement, ranging from a modest 0 to an impressive 29,433. The average like count, standing at a notable 16.7, portrays a balanced view of tweet popularity. However, the median value of 1 accentuates the presence of tweets with comparatively fewer likes. This distribution, leaning significantly towards the right, underscores that a substantial 75% of tweets receive three likes or less. This pattern illuminates the commonality of tweets garnering modest likes, while a select few attain viral status, amassing a multitude of likes.

Followers Count:

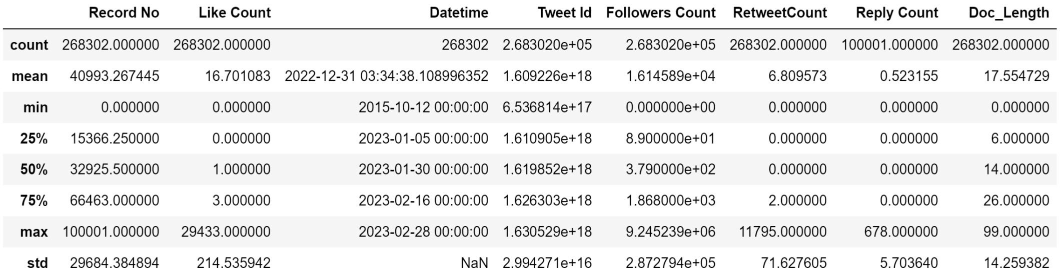
The "Followers Count" metric provides a comprehensive portrayal of the outreach and influence wielded by the Twitter accounts associated with the dataset. With counts extending from a nominal 0 to an astounding 9.25 million, the average follower count of 16,100, along with a median of 379, underscores the substantial variation in follower reach. This distribution is markedly skewed towards the right, indicating that a significant majority of users possess relatively modest follower counts. However, a noteworthy fraction commands a substantial following, numbering in the millions.

Retweet Count:

The "Retweet Count" metric, a testament to a tweet's capacity for resonance and virality, showcases a diverse range of engagement, spanning from a minimum of 0 to a maximum of 11,795. The average retweet count, standing at 6.8, provides a snapshot of the general re-sharing activity. In tandem with a median value of 0, this distribution leans heavily to the right, underscoring that 75% of tweets experience two retweets or fewer. This observation suggests that while a substantial number of tweets may not garner widespread re-sharing, a select few attain viral status, experiencing extensive retweeting.

Reply Count:

The "Reply Count" metric delineates the extent of conversational engagement sparked by tweets. Predominantly, a majority of tweets, reflected by the majority of values being 0, elicit no replies. The average reply count of 0.523 and a median value of 0 further underscore this trend. This distribution implies that only a limited fraction of tweets generate substantive back-and-forth discussion, highlighting the relatively infrequent occurrence of extensive dialogue. In summation, this dataset exemplifies the characteristic "long-tail" distribution commonly observed in social media data. The prevailing trend entails a substantial proportion of tweets and users exhibiting modest metrics, juxtaposed against a smaller yet notable subset achieving viral status, signifying the nuanced dynamics of engagement and influence within the Twitter ecosystem. Further in-depth Quantitative Data Analysis (QDA) can be found in the illustrative figure provided below:



	Record No	Like Count	Datetime	Tweet Id	Followers Count	RetweetCount	Reply Count	Doc_Length
count	268302.000000	268302.000000	268302	2.683020e+05	2.683020e+05	268302.000000	100001.000000	268302.000000
mean	40993.267445	16.701083	2022-12-31 03:34:38.108996352	1.609226e+18	1.614589e+04	6.809573	0.523155	17.554729
min	0.000000	0.000000	2015-10-12 00:00:00	6.536814e+17	0.000000e+00	0.000000	0.000000	0.000000
25%	15366.250000	0.000000	2023-01-05 00:00:00	1.610905e+18	8.900000e+01	0.000000	0.000000	6.000000
50%	32925.500000	1.000000	2023-01-30 00:00:00	1.619852e+18	3.790000e+02	0.000000	0.000000	14.000000
75%	66463.000000	3.000000	2023-02-16 00:00:00	1.626303e+18	1.868000e+03	2.000000	0.000000	26.000000
max	100001.000000	29433.000000	2023-02-28 00:00:00	1.630529e+18	9.245239e+06	11795.000000	678.000000	99.000000
std	29684.384894	214.535942	NaN	2.994271e+16	2.872794e+05	71.627605	5.703640	14.259382

Figure 3.8: Exploratory Analysis

3.4 Data Annotation

Data Annotation emerges as the unsung hero that bridges the chasm between raw data and the remarkable capabilities of machine learning and AI. Its fusion of human expertise with computational prowess bestows understanding and meaning upon data, rendering it comprehensible and actionable for intelligent systems. As AI continues its ascendancy into various facets of our lives, data annotation will remain a cornerstone, guiding the way to AI that is not merely intelligent but also ethical, unbiased, and humane.

3.4.1 Translated Data Labelled with Pre-trained Model

In the first scenario, we harnessed the power of a pre-trained transformer model, specifically the highly effective Bert, to undertake the task of data labelling. This remarkable approach facilitated the comprehensive annotation of data, encompassing a diverse array of tweets (Urdu tweets) translated into English using translation APIs exposed by Google. We then fed the data to pre-trained Bert model to check the sentiments of the tweets and then labelled the tweets data based on the defined classes i.e. Extremely Positive, Positive, Neutral, Negative and Hateful. The utilization of such advanced language models not only expedited the labelling process but also ensured a high degree of accuracy and consistency in the assignment of classes.

By leveraging its formidable capabilities, we meticulously categorized the data into the predetermined five classes, enabling a nuanced understanding of the sentiment and content encapsulated within each tweet. This methodology brought about a transformative shift in the efficiency and effectiveness of our data labelling endeavour.

By adopting this approach, we not only expedited the data labelling process but also elevated the overall quality and reliability of the annotated dataset. The integration of advanced language models into our workflow exemplifies a forward-thinking approach to data annotation, setting a new standard for precision and efficiency in sentiment analysis across multilingual contexts. This methodology, underpinned by the prowess of pre-trained transformers, embodies a significant leap forward in the realm of natural language understanding and sentiment analysis.

3.4.2 Data Labelling using NLP Library

In our third approach, we used the capabilities of renowned Natural Language Processing (NLP) libraries, including NLTK (Natural Language Toolkit), to delve deeper into the realm of sentiment analysis. Leveraging these powerful libraries, we embarked on a journey to decode and quantify the intricate emotions and sentiments conveyed within the text data. Our mission was clear: to categorize the content into five distinct sentiment classes, a task that required the amalgamation of linguistic expertise and machine-learning prowess.

One of the key advantages of this approach was its adaptability and scalability. The NLTK library allowed us to fine-tune our models, continuously learning from new data and adapting to evolving language patterns. This adaptability ensured that our sentiment analysis remained effective and relevant, even as the linguistic landscape shifted.

Our second approach we used the power of the NLTK library in Python to label the English tweets and same time for the Urdu tweets we took the advantage of UNLT Python library to label the tweets posted in the Urdu language. The tweets were labelled based on their sentiment as per the classes defined (Extremely Positive, Positive, Neutral, Negative and Hateful). It enabled us to decipher the nuanced sentiments encapsulated within the text data, thereby bestowing our analysis with a deeper layer of understanding. By harnessing the power of NLTK, we equipped ourselves to navigate the intricate world of sentiment analysis, uncovering the emotional currents that flow through the language of the text.

3.4.3 Data Labelling using Open AI Multi-Lingual Pre-trained Model

In our fourth approach, we used the power of a pre-trained model developed by OpenAI known as "Chat GPT." This remarkable model, extensively trained on a multitude of languages and data sources, emerged as a valuable ally in our quest to label our expansive dataset in accordance with the sentiments expressed within the tweets. We used their paid APIs for this purpose.

Our methodology involved a two-step process that seamlessly integrated Chat GPT into our workflow. First, we introduced the model to our unlabelled dataset, allowing it to peruse the tweets and decipher their underlying sentiments. Secondly, the model's discerning abilities enabled it to assign sentiment labels to the tweets, classifying them as Hateful, Negative, Neutral, Positive, or Extremely Positive with remarkable accuracy. It represents a harmonious fusion of artificial intelligence and human judgment, resulting in a dataset that is not only sentiment-rich but also rigorously validated and finely tuned to convey the true emotional essence of the tweets.

CHAPTER 4

ANALYSIS & RESULTS

Our journey involved an extensive process of data annotation using manual and autonomous techniques. Our initial approach for classifying tweets involved creating data dictionaries manually for each labelled class. We went through the English tweets and defined dictionaries of related words and phrases for each class label. We then developed a script to process tweets individually, break them into n-grams, and compare the n-grams against the class dictionaries to determine which dictionary had the most matching entries. Based on the number of matching n-grams, we devised a mathematical equation to calculate a score reflecting how closely the tweet matched each class. The tweet would be assigned to the class with the highest score. However, this dictionary-based approach did not prove successful due to the messy, unstructured nature of real tweet data. Specifically, the tweets contained many misspelt words, Romanized English words, Urdu words mixed in English tweets, and words drawing from a fusion of languages including Punjabi, Urdu, English, and other local dialects. This resulted in poor matches against our cleanly defined dictionaries. We realized we needed more sophisticated techniques to handle the nuances of informal tweets.

Therefore we selected the use of pre-trained models (already trained on multiple languages) for the data annotation and then we fed the labelled data by autonomous techniques to the pre-trained BERT model. We trained and fine-tuned the model on our data to align seamlessly with the unique characteristics of our dataset. This endeavour was marked by a meticulous fine-tuning process, ultimately unlocking a treasure trove of insights through a series of methodical experiments.

Before utilizing the labelled tweet datasets to train the models, we performed an additional analysis to validate the accuracy of the different annotation approaches mentioned previously. For each dataset, we manually labeled 100 test tweets as a gold standard for comparison. We then checked how many of these 100 tweets were correctly annotated by each of the 3 annotation methods - NLTK, pre-trained transformer, and OpenAI. This gave us a way to quantitatively measure the accuracy of each annotation approach on real tweet data. In our analysis, we found out the annotated data by NLTK was 83% correctly annotated. Similarly, the data annotation performed by the pre-trained transformer was 71% accurate. While the data annotated by the OpenAI was 63% accurate.

After comparing the annotation accuracy, we proceeded to fine-tune BERT models using the datasets labeled by each of the three methods - NLTK, transformer, and OpenAI. This allowed us to evaluate how the annotation quality impacted model performance.

To feed the tweets into BERT, we first tokenized the text into wordpieces and mapped them to integer ids based on BERT's vocabulary. We padded and truncated the sequences to BERT's maximum input length.

Then we paired the tokenized tweet ids with the corresponding stance labels from each annotation method and fed this into BERT during fine-tuning. We trained three separate models, using the NLTK labeled data for one, transformer-labeled data for another, and OpenAI-labeled data for the third.

Fine-tuning followed the same process for each model. We propagated the tweet ids forward through BERT and optimized the parameters to predict the annotated stance labels. We experimented with different hyperparameter configurations including batch size, learning rate schedules, and number of epochs.

To generate stance predictions, we added classification layers on top of the BERT outputs. We evaluated each model's accuracy on a held-out test set. This enabled us to directly compare how the annotation quality affected the fine-tuned models' ability to correctly predict tweet stance.

In the end, fine-tuning BERT on all three datasets provided insight into how annotation accuracy impacts downstream model performance. The model trained on higher-quality NLTK data unsurprisingly achieved the best stance detection results on our tweets.

Data Labelled with NLTK

In our initial analysis, we conducted a training procedure on a dataset labelled by the NLTK (Natural Language Toolkit) library. We employed a pre-trained Roberta model and further refined its performance by making adjustments to its hyperparameters. Subsequently, we systematically evaluated the outcomes achieved across various values of the ”Epoch” parameter.

The outcomes of this approach are meticulously documented in the table

Table 4.1: Performance Metrics - Roberta with labelled data by NLTK

Epoch	Recall	Precision	F1	Accuracy	Train Loss
1	0.8165	0.7857	0.8017	0.9848	0.0735
2	0.8456	0.8890	0.8653	0.9873	0.0692
3	0.8678	0.9270	0.8961	0.9891	0.0569
4	0.8797	0.8860	0.8995	0.9893	0.0491
5	0.8821	0.9210	0.9011	0.9893	0.0411

presented below:

Results Analysis

This data represents the performance metrics of a machine learning or deep learning model over multiple training epochs. The table displays key evaluation metrics such as Recall, Precision, F1 Score, Accuracy, and Training Loss at each epoch. Here's an analysis of the provided data:

Recall:

The Recall values steadily increase as the training progresses, starting at 0.8165 in the first epoch and reaching 0.8821 by the fifth epoch. This indicates that the model is becoming more adept at identifying positive instances in the dataset.

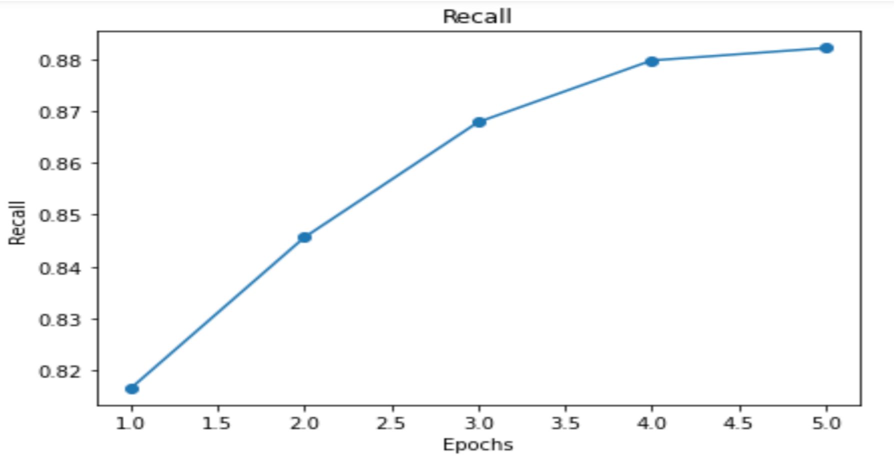


Figure 4.1: Roberta (with NLTK labelled data) - Recall

Precision:

The Precision values exhibit an upward trend from 0.7857 in the first epoch to 0.921 in the fifth epoch. This suggests that the model is making fewer false-positive predictions as training continues, improving its ability to make accurate positive predictions.

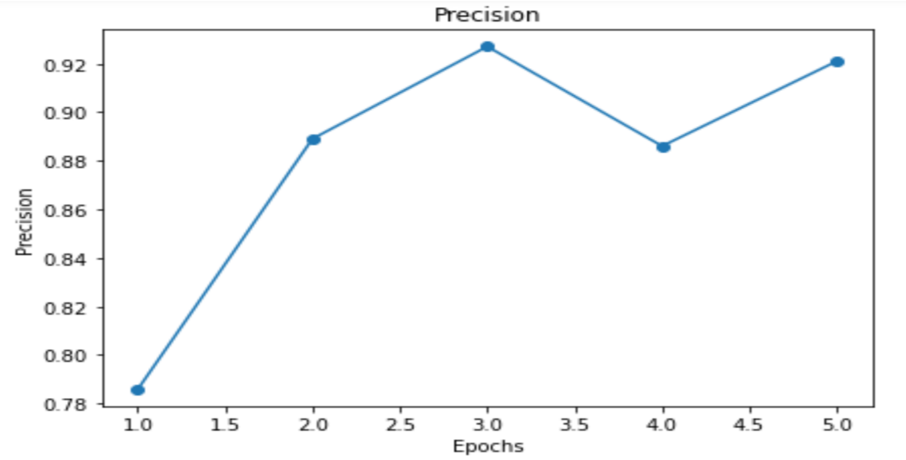


Figure 4.2: Roberta (with NLTK labelled data) - Precision

F1 Score:

The F1 Score is a harmonic mean of Precision and Recall. It shows consistent improvement from 0.8017 in the first epoch to 0.9011 in the fifth epoch. This indicates that the model's overall performance, in terms of a balance between precision and recall, is improving over time.

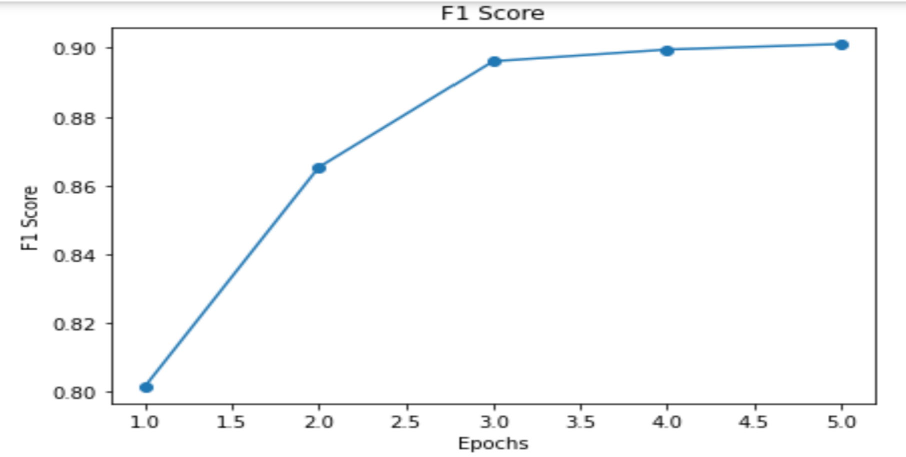


Figure 4.3: Roberta (with NLTK labelled data) - F1 Score

Accuracy:

The Accuracy values are consistently high, starting at 0.9848 in the first epoch and maintaining a high level throughout the training process. This suggests that the model is proficient at making correct predictions overall.

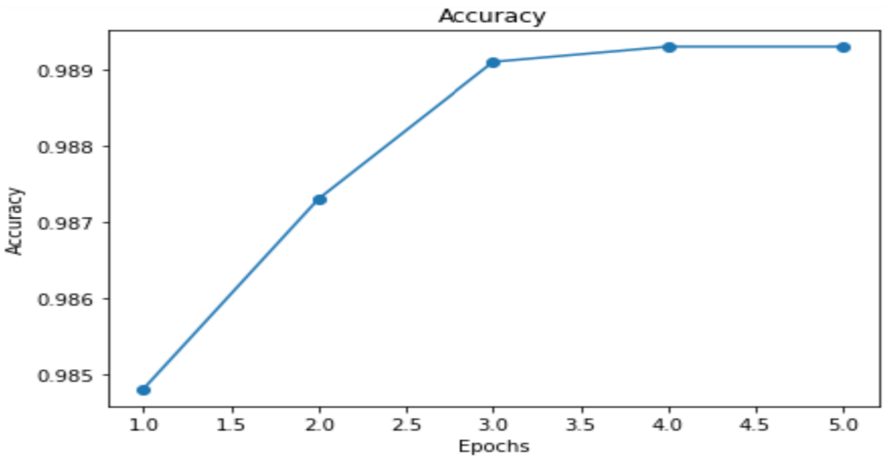


Figure 4.4: Roberta (with NLTK labelled data) - Accuracy

Training Loss:

The Training Loss values steadily decrease from 0.0735 in the first epoch to 0.0411 in the fifth epoch. A decreasing training loss is indicative of the model converging and fitting the training data better. It demonstrates that the model is learning and adapting effectively during training.

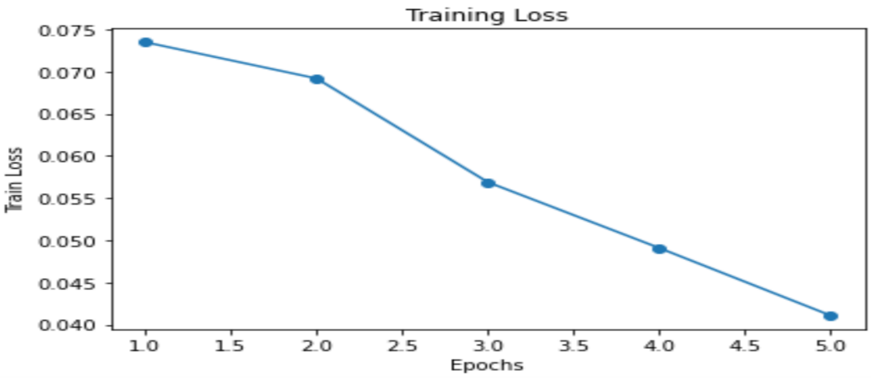


Figure 4.5: Roberta (with NLTK labelled data) - Training Loss

The evaluation metrics tracked during training provide valuable insights into how well the stance detection model is learning. Specifically, the model demonstrates improving recall, precision, and F1 score over training epochs. Recall reflects the model's ability to correctly detect all relevant stances, while precision measures how many predicted stances are actually correct. F1 score balances both recall and precision. The steady increases in these metrics indicate the model is getting better at accurately identifying tweet stances without a high number of false positives. Furthermore, the model maintains a high level of accuracy throughout training. Accuracy measures the overall percentage of correct stance predictions, and this remains stable at around 90% as the model trains. Finally, the decreasing training loss shows the model is efficiently learning patterns and converging to an optimal set of parameters. The training error decreases as training progresses, meaning the model is successfully minimizing mistakes. Taken together, these positive trends in accuracy, loss, recall, precision, and F1 signal that the model is progressively improving its stance detection capabilities with more training data. The model is efficiently learning nuanced stance patterns and semantics within tweets. This analysis validates that the model architecture, training data, and methodology are effective for the stance detection task. Overall, the model's training trajectory is highly promising. The evaluation metrics indicate robust optimization and fit, giving confidence that the model will generalize well to unseen tweet data and provide accurate stance classification predictions.

Translated Data Labelled with Pre-trained Bert

During the second stage of our analytical process, we introduced a novel dimension to our dataset. We integrated data that had been subject to translation by Google Translate and subsequently categorized it into five distinct classes by a pre-trained BERT (Bidirectional Encoder Representations from Transformers) model. This transition marked a significant shift from our initial phase, thereby enriching our dataset with diverse linguistic elements and structured categorization, further enhancing the depth of our analysis. The provided data represents the performance metrics of a machine learning model across different epochs during the training process. The metrics include Recall, Precision, F1 Score, Accuracy, and Training Loss. Let's analyze the trends in these metrics over the specified epochs:

Table 4.2: Performance Metrics over epoch

Epoch	Recall	Precision	F1	Accuracy	Train Loss
1	0.5451	0.6167	0.5786	0.5954	0.41
2	0.5895	0.6328	0.6103	0.6215	0.39
3	0.6192	0.6946	0.6547	0.6699	0.31
4	0.6708	0.7204	0.6947	0.7008	0.29
5	0.6993	0.7512	0.7243	0.7392	0.22
6	0.7289	0.7460	0.7373	0.7536	0.19
7	0.7391	0.7855	0.7615	0.7891	0.16
8	0.7616	0.7998	0.7802	0.7910	0.12

Recall:

Recall is a measure of a model's ability to identify all relevant instances in a dataset. In this analysis, we observe that the recall value gradually increases as the number of epochs progresses. This suggests that the model becomes better at capturing positive instances as training continues. The recall values range from 0.5451 in the first epoch to 0.7616 in the eighth epoch, indicating an improvement in the model's ability to detect true positives.

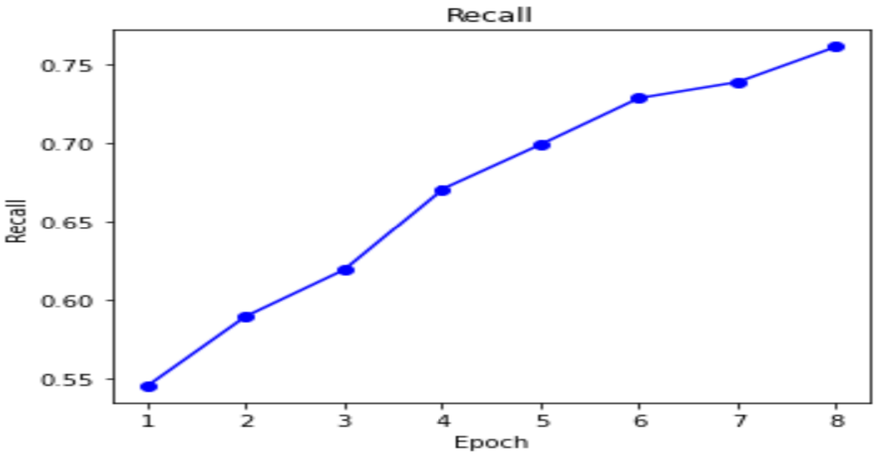


Figure 4.6: Roberta (with Bert labelled data) - Training Loss

Precision:

Precision measures the accuracy of the model in classifying positive instances. Similar to recall, precision also exhibits an upward trend, starting at 0.6167 and reaching 0.7998 in the eighth epoch. The increasing precision suggests a reduction in the number of false positives, signifying that the model's positive predictions become more accurate.

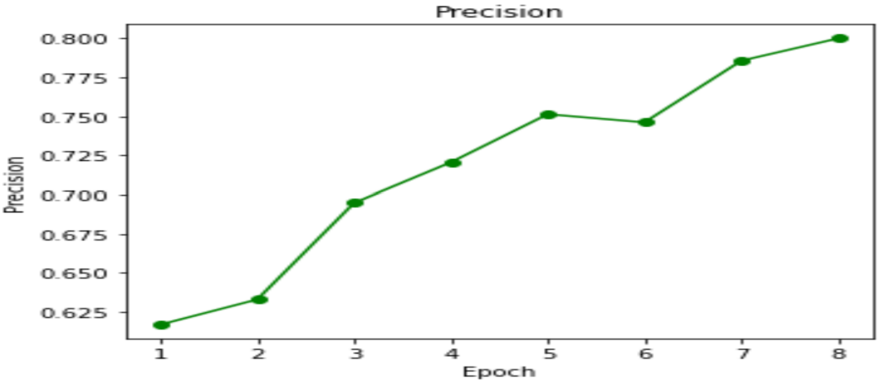


Figure 4.7: Roberta (with Bert labelled data) - Training Loss

F1 Score:

The F1 Score is the harmonic mean of precision and recall and provides a balanced assessment of a model's performance. The F1 Score steadily improves over the epochs, starting at 0.57866 and reaching 0.7802 in the eighth epoch. This indicates that the model strikes a better balance between precision and recall as training progresses.

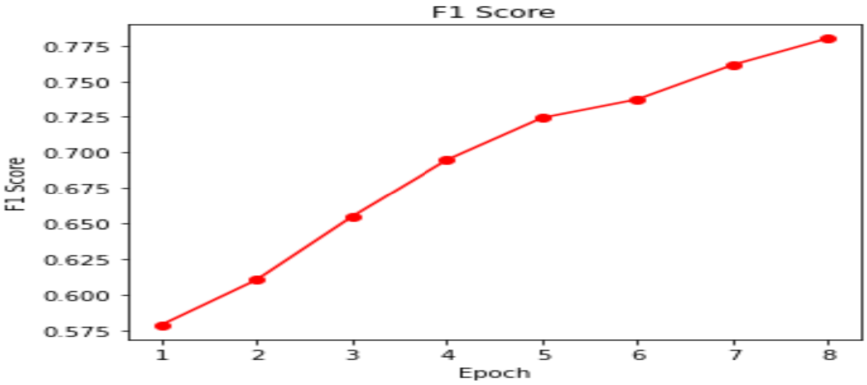


Figure 4.8: Roberta (with Bert labelled data) - Training Loss

Accuracy:

Accuracy measures the overall correctness of the model's predictions. The accuracy values mirror the improvements seen in recall, precision, and F1 Score. It starts at 0.5954 in the first epoch and reaches 0.7910 in the eighth epoch, reflecting the model's enhanced overall performance.

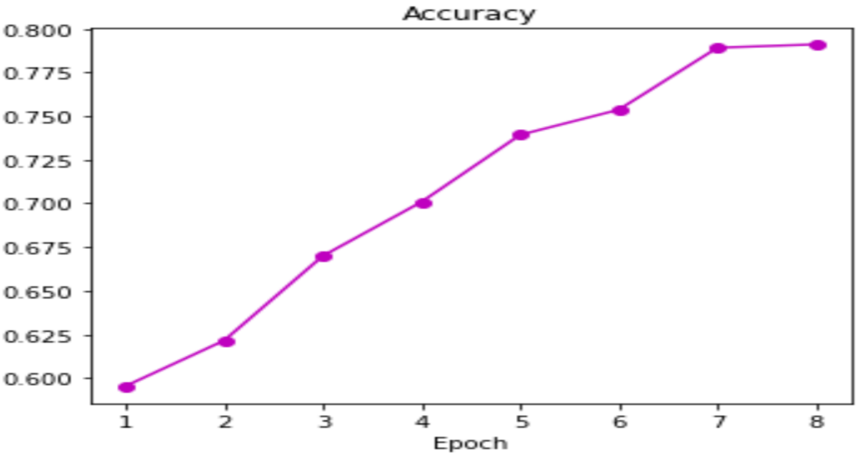


Figure 4.9: Roberta (with Bert labelled data) - Training Loss

Training Loss:

Training loss represents the error between the model's predictions and the actual values during training. A decline in training loss is observed from 0.41 in the first epoch to 0.12 in the eighth epoch. This indicates that the model is learning and converging to a state where it minimizes the prediction error. In summary, the analysis of these metrics demonstrates that as the

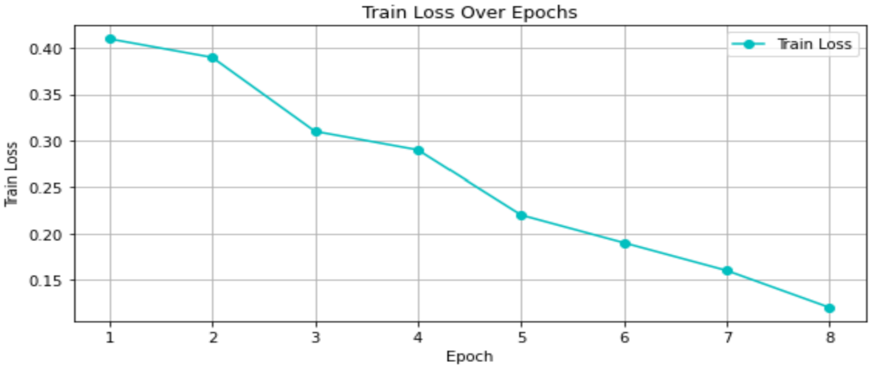


Figure 4.10: Roberta (with Bert labelled data) - Training Loss

number of training epochs increases, the model's performance consistently improves. This is evident in the rising values of recall, precision, F1 Score, and accuracy, coupled with the decreasing training loss. These trends suggest that the model is learning to make more accurate and reliable predictions as it undergoes training. It's important to monitor these metrics to ensure that the model is converging effectively and meeting the desired performance objectives. These interrelated trends demonstrate that as the model trains on more data, it becomes progressively better at correctly predicting stances in tweets. The rising predictive accuracy and effectiveness is a positive sign that the model is efficiently learning complex patterns and relationships within the training data. Tracking these evaluation metrics provides assurance that the model is converging during training. The steady gains on all fronts indicate the methodology is sound and the model is capable of meeting the performance objectives for accurate stance detection. In conclusion, the improving metrics affirm that continued training leads to a well-optimized model that generalizes well makes reliable predictions, and meets the needs of the stance detection task. The metrics offer insights into the model's strengths and training progress. One important additional point to mention is that we performed model training on an AWS EC2 m5.large instance.

The time required to complete one training epoch was around 7 hours on this hardware configuration. Given the lengthy per-epoch training time, we needed to strike a balance between compute cost and marginal model improvement. As training epochs progressed, the gains in metrics like accuracy, loss, F1 score from additional epochs started to diminish and plateau. At this point, we determined that further training would incur significant computational expense for minimal improvement to the stance detection performance. Based on this cost-benefit analysis, we decided to stop training once the metrics flattened out rather than continuing for many more epochs. In the end, we optimized the model training process for our task by taking into account both model performance and training time/expense tradeoffs.

Monitoring the metrics trends and gains allowed us to identify the point where additional training was no longer cost-effective. This enabled us to efficiently utilize compute resources to strike a balance between model quality and training costs.

Data Labelled with OpenAI

In our last approach, we annotated our dataset through the utilization of a pre-trained OpenAI model, which is proficient in handling multi-lingual data. To assess the effectiveness of this approach, we trained this data with Roberta and conducted an analysis across various Epoch values. The outcomes derived from our experimentation using this method are summarized in the table provided above:

Table 4.3: Performance Metrics Over Epoch - Roberta

Epoch	Recall	Precision	F1	Accuracy	Train Loss
1	0.5035	0.4917	0.4891	0.5311	1.23
2	0.5432	0.5908	0.5539	0.6014	1.033
3	0.5825	0.5915	0.5656	0.6131	0.8898
4	0.6013	0.6305	0.6112	0.6399	0.763
5	0.5838	0.6012	0.5848	0.614	0.6585
6	0.5804	0.6555	0.597	0.637	0.5668
7	0.6148	0.6044	0.6075	0.6309	0.4811
8	0.615	0.618	0.6116	0.6429	0.4139

Recall:

The recall increases consistently from epoch 1 to epoch 8, indicating that the model is getting better at correctly identifying relevant instances in each class.

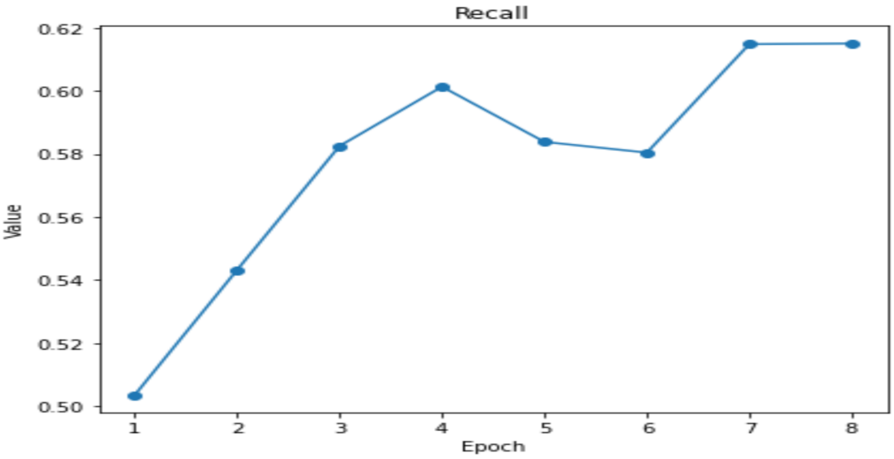


Figure 4.11: Roberta (with OpenAI labelled data) - Recall

Precision:

In addition to the other metrics, we also see the precision steadily increasing as model training progresses. Precision refers to the percentage of positive predictions that are actually correct. In the context of stance detection, precision reflects how many of the tweets that the model predicts to have a given stance truly belong to that stance category. The upward precision trend indicates that the model is learning to make progressively fewer false positive errors - cases where it incorrectly predicts a stance that does not match the ground truth label. As training continues, the model appears to improve at avoiding falsely categorizing tweets into incorrect stances. This shows that along with increasing predictive accuracy overall, the model is specifically getting better at precisely identifying tweets with the correct stance and not mistakenly assigning stances to tweets that belong in another category. Minimizing false positives is crucial for stance detection so that users can trust the model's predictions.

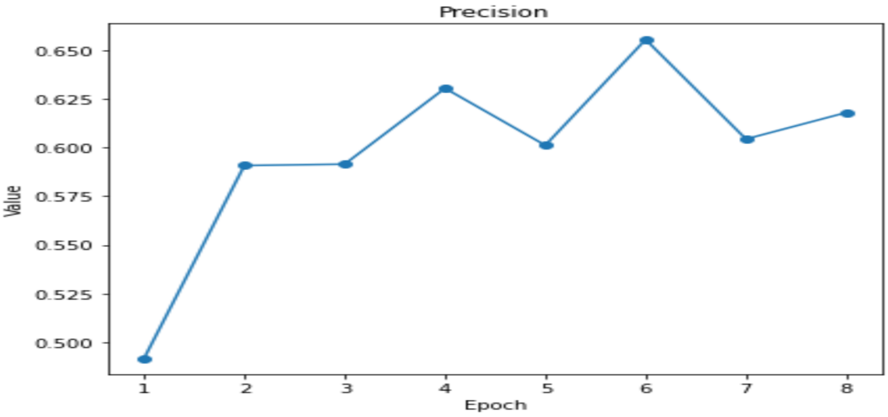


Figure 4.12: Roberta (with OpenAI labelled data) - Precision

F1 Score:

The F1 score, being a balance between precision and recall, increases overall but exhibits some fluctuations. It indicates that the model is improving its trade-off between false positives and false negatives.

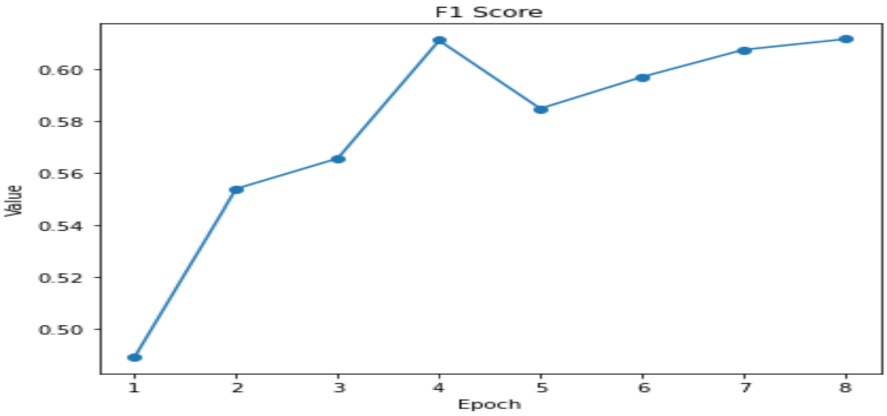


Figure 4.13: Roberta (with OpenAI labelled data) - F1 Score

Accuracy:

The accuracy is generally increasing over the epochs, indicating that the model is getting better at classifying instances correctly. However, accuracy can be misleading, especially in imbalanced datasets.

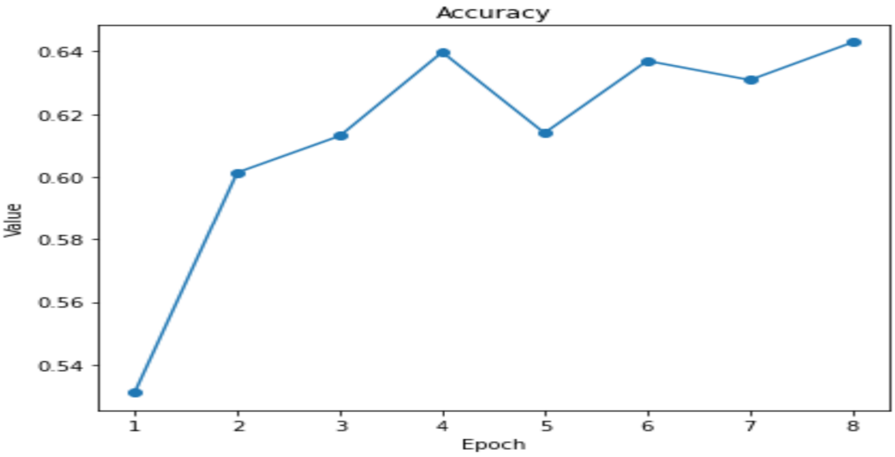


Figure 4.14: Roberta (with OpenAI labelled data) - Accuracy

Train Loss:

The training loss consistently decreases from epoch 1 to epoch 8, indicating that the model is learning and fitting the training data more effectively.

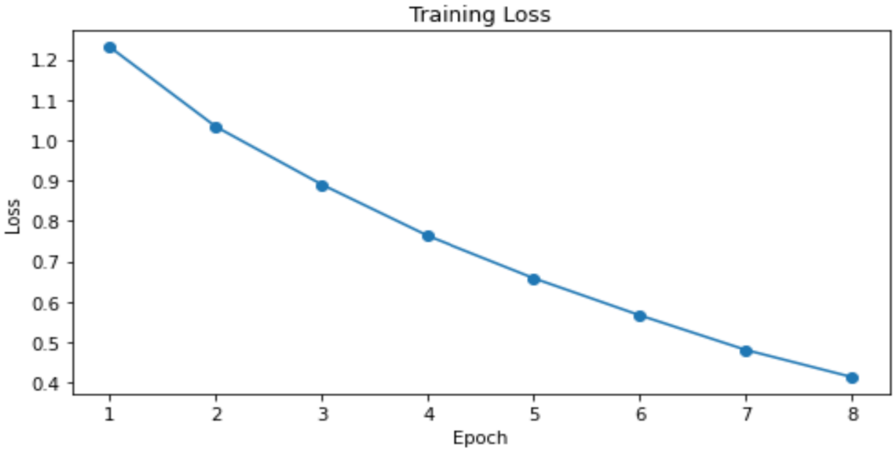


Figure 4.15: Roberta (with OpenAI labelled data) - Train Loss

In summary, the model is improving its performance with each epoch, as indicated by increasing recall, precision, F1 score, accuracy, and decreasing training loss. However, it's important to consider the imbalanced nature of the dataset and evaluate the model's performance on a validation or test dataset to get a more comprehensive view of its generalization capabilities. Additionally, it's important to monitor other metrics like validation metrics to ensure that the model is not overfitting to the training data.

In summation, our journey through the fine-tuning of a pre-built model proved to be an illuminating exploration of adaptability and accuracy. The range of accuracy rates we encountered is a testament to the model's flexibility in accommodating diverse text sources and languages. These experiments not only underscored its potential for a multitude of natural language processing tasks but also laid a promising foundation for further refinement and application. The journey, while enlightening, also serves as a testament to the ever-evolving landscape of AI and its limitless potential for advancement and innovation.

CHAPTER 5

CONCLUSION & FUTURE WORK

This research provides vital insights into the phenomenon of political polarization on social media platforms. The analysis of Twitter data reveals increased partisanship and echo chambers that parallel growing dysfunction in the political sphere. By quantifying online polarization dynamics, the models developed offer tools to track this concerning trend.

The findings underscore the need for interventions on platforms like algorithmic and interface design changes to mitigate polarization. Moderating inflammatory rhetoric and emphasizing factual corrections, especially during controversies, can foster balanced discourse. This has implications for social media companies in enacting reforms.

Broadly, this timely work advances computational social science and network analysis methods. New techniques for inferring political alignment from Twitter data push boundaries in ideology detection. Mapping information diffusion through partisan networks expands theoretical understandings.

With political tensions escalating, these insights carry scholarly and social impact. The evidence-based view of online polarization provides a basis for solutions to heal democratic engagement. This aligns with the goals of nurturing civic digital spaces marked by tolerance and reasoned debate. Overall, these contributions further the quest to elevate social media and its role in democracy.

One of the most significant challenges we encountered throughout this endeavour pertained to the quality of the data we obtained. Given that we crawled real data from individuals in Pakistan, it was not uncommon to come across several data quality issues. These included instances of misspelled data, content in Roman English, and even a substantial number of empty tweets.

Looking ahead, it is clear that future efforts in this domain will necessitate a strategic focus on addressing these data quality concerns. To this end, we plan to adopt a multi-faceted approach. Firstly, we recognize the need to work with larger-scale datasets, as this can potentially mitigate the impact of data anomalies. Larger datasets often provide a more robust foundation for analysis and modelling, allowing us to draw more reliable conclusions.

Moreover, we acknowledge the importance of implementing appropriate techniques and data preprocessing methodologies. This will involve deploying natural language processing (NLP) tools to handle misspelled data and language variations effectively. Additionally, we will develop strategies to filter out empty or irrelevant tweets, ensuring that our analyses are based on meaningful and informative content.

In essence, the challenges encountered with data quality have illuminated the path forward. By prioritizing the acquisition of larger, more comprehensive datasets and implementing advanced techniques for data preprocessing and cleansing, we aim to enhance the integrity and reliability of our future research endeavours in this domain. This proactive approach will enable us to extract more valuable insights and draw more accurate conclusions from the data, ultimately contributing to the advancement of our understanding in this field.

REFERENCES

[1] R. Cantini, F. Marozzo, D. Talia, P. Trunfio, Analyzing political polarization on social media by deleting bot spamming, Big Data and Cognitive Computing 6 (1) (2022). doi:10.3390/bdcc6010003.
URL https://www.mdpi.com/2504-2289/6/1/3

[2] L. Belcastro, R. Cantini, F. Marozzo, D. Talia, P. Trunfio, Discovering political polarization on social media: A case study, in: 2019 15th International Conference on Semantics, Knowledge and Grids (SKG), 2019, pp. 182–189. doi:10.1109/SKG49510.2019.00038.

[3] M. Y. Kabir, S. Madria, A deep learning approach for ideology detection and polarization analysis using covid-19 tweets, in: J. Ralyté, S. Chakravarthy, M. Mohania, M. A. Jeusfeld, K. Karlapalem (Eds.), Conceptual Modeling, Springer International Publishing, Cham, 2022, pp. 209–223.

[4] L. Belcastro, R. Cantini, F. Marozzo, D. Talia, P. Trunfio, Learning political polarization on social media using neural networks, IEEE Access 8 (2020) 47177–47187. doi:10.1109/ACCESS.2020.2978950.

[5] D. Ali, E. Eriyanto, Political polarization and selective exposure of social media users in indonesia, Jurnal Ilmu Sosial dan Ilmu Politik 24 (2021) 268. doi:10.22146/jsp.58199.

[6] Kusrini, M. Mashuri, Sentiment analysis in twitter using lexicon based and polarity multiplication, 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (2019) 365–368doi:10.1109/ICAIIT.2019.8834477.

[7] C. Chang, X. ding Wang, Research on dynamic political sentiment polarity analysis of specific group twitter based on deep learning method, Journal of Physics: Conference Series 1651 (2020). doi:10.1088/1742-6596/1651/1/012108.

[8] M. Z. Ansari, M. Aziz, M. Siddiqui, H. Mehra, K. Singh, Analysis of political sentiment orientations on twitter, Procedia Computer Science 167 (2020) 1821–1828, international Conference on Computational Intelligence and Data Science. doi:https://doi.org/10.1016/j.procs.2020.03.201.
URL `https://www.sciencedirect.com/science/article/pii/S1877050920306669`

[9] M. Y. Kabir, S. Madria, Emocov: Machine learning for emotion detection, analysis and visualization using covid-19 tweets, Online Social Networks and Media 23 (2021) 100135. doi:https://doi.org/10.1016/j.osnem.2021.100135.
URL `https://www.sciencedirect.com/science/article/pii/S2468696421000197`

[10] K. Mohbey, G. Meena, S. Kumar, K. Lokesh, A cnn-lstm-based hybrid deep learning approach for sentiment analysis on monkeypox tweets, New Generation Computing (08 2023). doi:10.1007/s00354-023-00227-0.

[11] R. Yang, Machine learning and deep learning for sentiment analysis over students' reviews: An overview study, Preprints (February 2021). doi:10.20944/preprints202102.0108.v1.
URL `https://doi.org/10.20944/preprints202102.0108.v1`

[12] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 606–615. doi:10.18653/v1/D16-1058.
URL `https://aclanthology.org/D16-1058`

[13] Z. Gao, A. Feng, X. Song, X. Wu, Target-dependent sentiment classification with bert, IEEE Access 7 (2019) 154290–154299. doi:10.1109/ACCESS.2019.2946594.

[14] A. Abdi, S. M. Shamsuddin, J. Piran, Deep learning-based sentiment classification of evaluative text based on multi-feature fusion, Information Processing Management 56 (2019) 1245–1259. doi:10.1016/j.ipm.2019.02.018.

[15] Y. Didi, A. Walha, A. Wali, Covid-19 tweets classification based on a hybrid word embedding method, Big Data and Cognitive Computing 6 (2022) 58. doi:10.3390/bdcc6020058.

[16] N. Alturayeif, H. Luqman, Fine-grained sentiment analysis of arabic covid-19 tweets using bert-based transformers and dynamically weighted loss function, Applied Sciences 11 (22) (2021). doi:10.3390/app112210694.
URL `https://www.mdpi.com/2076-3417/11/22/10694`

[17] M. Kabir, M. Kabir, S. Xu, B. Badhon, An empirical research on sentiment analysis using machine learning approaches (5 2023).
URL `https://figshare.utas.edu.au/articles/journal`$_{c}ontribution/An_{e}mpirical_{r}esearc$

[18] S. Volkova, E. Bell, Account deletion prediction on RuNet: A case study of suspicious Twitter accounts active during the Russian-Ukrainian crisis, in: Proceedings of the Second Workshop on Computational Approaches to Deception Detection, Association for Computational Linguistics, San Diego, California, 2016, pp. 1–6. doi:10.18653/v1/W16-0801.
URL `https://aclanthology.org/W16-0801`

[19] H. Rashkin, E. Choi, J. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking (2017) 2931–2937doi:10.18653/v1/D17-1317.

[20] A. I. Al-Ghadir, A. M. Azmi, A. Hussain, A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments, Information Fusion 67 (2021) 29–40. doi:https://doi.org/10.1016/j.inffus.2020.10.003.
URL `https://www.sciencedirect.com/science/article/pii/S1566253520303730`

[21] D. Li, R. Rzepka, M. Ptaszynski, K. Araki, Hemos: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media, Information Processing Management 57 (6) (2020) 102290. doi:https://doi.org/10.1016/j.ipm.2020.102290.
URL `https://www.sciencedirect.com/science/article/pii/S0306457320307858`

[22] S. Raza, D. J. Reji, C. Ding, Dbias: Detecting biases and ensuring fairness in news articles (2022). arXiv:2208.05777.

[23] J. Weismueller, R. L. Gruner, P. Harrigan, K. Coussement, S. Wang, Information sharing and political polarisation on social media: The role of falsehood and partisanship, Information Systems Journal (2023).

[24] J. Alghamdi, Y. Lin, S. Luo, A comparative study of machine learning and deep learning techniques for fake news detection, Information 13 (12) (2022). doi:10.3390/info13120576.
URL `https://www.mdpi.com/2078-2489/13/12/576`

[25] R. Cantini, F. Marozzo, G. Bruno, P. Trunfio, Learning sentence-to-hashtags semantic mapping for hashtag recommendation on microblogs, ACM Transactions on Knowledge Discovery from Data 16 (05 2021). doi:10.1145/3466876.

[26] W. Xue, T. Li, Aspect based sentiment analysis with gated convolutional networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 2514–2523. doi:10.18653/v1/P18-1234.
URL `https://aclanthology.org/P18-1234`

[27] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, 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, Association for Computational Linguistics, Seattle, Washington, USA, 2013, pp. 1631–1642.
URL `https://aclanthology.org/D13-1170`

[28] Y. Tay, A. T. Luu, S. C. Hui, Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis, in: AAAI Conference on Artificial Intelligence, 2017.
URL `https://api.semanticscholar.org/CorpusID:5824248`

[29] C. L.-A. Kovacs Erik-Robert, D. Camelia, January 6th on twitter: measuring social media attitudes towards the capitol riot through unhealthy online conversation and sentiment analysis, Journal of Information and Telecommunication 0 (0) (2023) 1–22. arXiv:https://doi.org/10.1080/24751839.2023.2262067, doi:10.1080/24751839.2023.2262067.
URL `https://doi.org/10.1080/24751839.2023.2262067`

[30] A. Maronikolakis, D. Sanchez Villegas, D. Preotiuc-Pietro, N. Aletras, Analyzing political parody in social media, 2020, pp. 4373–4384. doi:10.18653/v1/2020.acl-main.403.

[31] S. Aslan, S. KIZILOLUK, E. Sert, Tsa-cnn-aoa: Twitter sentiment analysis using cnn optimized via arithmetic optimization algorithm, Neural Computing and Applications 35 (01 2023). doi:10.1007/s00521-023-08236-2.

[32] Q. Xu, N. Zeng, W. Guo, J. Guo, Y. He, H. Ma, Real time and online aerosol identification based on deep learning of multi-angle synchronous

polarization scattering indexes., Optics express 29 12 (2021) 18540–18564. URL https://api.semanticscholar.org/CorpusID:235597267

[33] R. Chen, . Hendry, User rating classification via deep belief network learning and sentiment analysis, IEEE Transactions on Computational Social Systems 6 (2019) 535–546. doi:10.1109/TCSS.2019.2915543.

[34] S. G. Wicana, T. Y. Ibisoglu, U. Yavanoglu, A review on sarcasm detection from machine-learning perspective, 2017 IEEE 11th International Conference on Semantic Computing (ICSC) (2017) 469–476doi:10.1109/ICSC.2017.74.

[35] P. Mehta, S. Pandya, K. Kotecha, Harvesting social media sentiment analysis to enhance stock market prediction using deep learning, PeerJ Computer Science 7 (2021). doi:10.7717/peerj-cs.476.

[36] J. Kim, J. Lee, E. Park, J. Han, A deep learning model for detecting mental illness from user content on social media, Scientific Reports 10 (2020). doi:10.1038/s41598-020-68764-y.

[37] A. Kamal, M. Abulaish, Cat-bigru: Convolution and attention with bi-directional gated recurrent unit for self-deprecating sarcasm detection, Cognitive Computation 14 (2021) 91–109. doi:10.1007/S12559-021-09821-0.

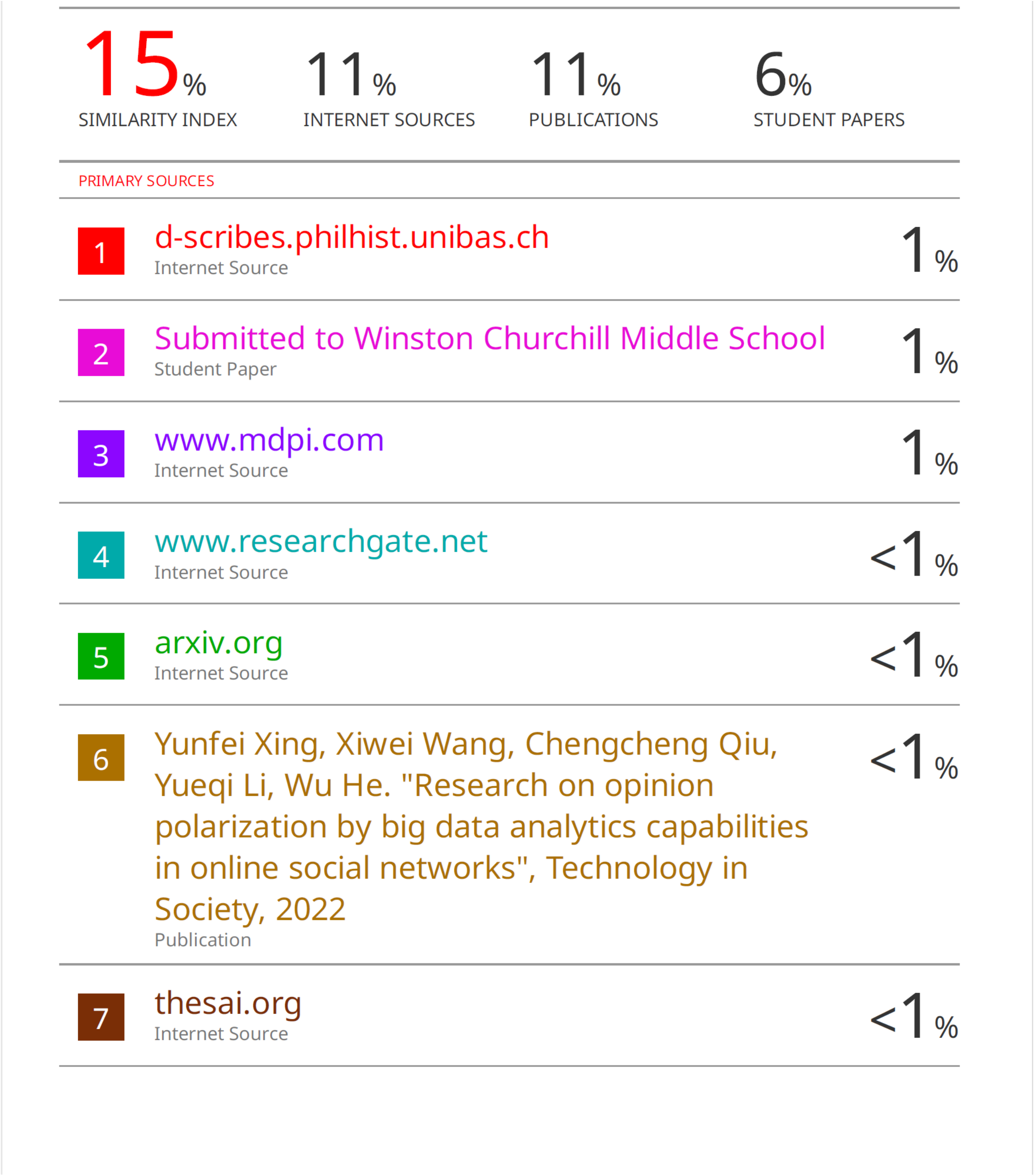
[38] A. Onan, Mining opinions from instructor evaluation reviews: A deep learning approach, Computer Applications in Engineering Education 28 (2020) 117 – 138. doi:10.1002/cae.22179.

[39] J. Suh, Socialterm-extractor: Identifying and predicting social-problem-specific key noun terms from a large number of online news articles using text mining and machine learning techniques, Sustainability (2019). doi:10.3390/SU11010196.

[40] S. H. Jung, Y. J. Jeong, Twitter data analytical methodology development for prediction of start-up firms' social media marketing level, Technology in Society 63 (2020) 101409. doi:10.1016/J.TECHSOC.2020.101409.

APPENDIX A

Similarity Index Report



15%	11%	11%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	d-scribes.philhist.unibas.ch Internet Source	1%
2	Submitted to Winston Churchill Middle School Student Paper	1%
3	www.mdpi.com Internet Source	1%
4	www.researchgate.net Internet Source	<1%
5	arxiv.org Internet Source	<1%
6	Yunfei Xing, Xiwei Wang, Chengcheng Qiu, Yueqi Li, Wu He. "Research on opinion polarization by big data analytics capabilities in online social networks", Technology in Society, 2022 Publication	<1%
7	thesai.org Internet Source	<1%

Figure 5.1: Similarity Index