# Personality Prediction using Multiple Textual Datasets and Deep Learning Models

By


**Faiza Arshad**

**(Registration No.: 00000317782)**

Thesis Supervisor: Dr. Arslan Shaukat

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN


(2023)

# THESIS ACCEPTANCE CERTIFICATE

**Annex A**

**THESIS ACCEPTANCE CERTIFICATE**

Certified that final copy of MS/MPhil thesis written by **NS Faiza Arshad** Registration No. **00000317782**, of College of E&ME has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the thesis.

Signature : ____________________

Name of Supervisor: **Dr Arslan Shaukat**

Date: 12-09-2023

Signature of HOD: ____________________
(Dr Usman Qamar)
Date: 12-09-2023

Signature of Dean: ____________________
(Brig Dr Nasir Rashid)
Date: 12 SEP 2023

## DEDICATION

*Dedicated to my exceptional parents:* ***Arshad Mehmood & Shazia Sultana,*** *loving husband, and cherished siblings, whose unwavering support and collaboration have been instrumental in guiding me to this achievement.*

# ACKNOWLEDGEMENTS

Indeed, all praises and reverence belong to Almighty Allah, the most exalted, who bestowed upon me the fortitude, resilience, wisdom, and capability to undertake and successfully complete this endeavor. Without question, He lightened my path with His divine guidance, and without His blessings, I am incapable of achieving anything.

First, I want to show gratitude to my co-supervisor Dr. Seemab Latif, she motivated me and guided me throughout the journey she was the one who gave me motivation to choose this topic and provided me with all the details.

Then, I extend my deepest appreciation to my advisor, Dr. Arslan Shaukat. His constant encouragement, unwavering support, dedication, and invaluable guidance have been instrumental in my pursuit of knowledge. I am truly fortunate to have such a supportive advisor and kind-hearted mentor for my research journey.

In addition to my advisor, I wish to express my gratitude to my entire thesis committee: Dr. Muhammad Usman Akram and Dr. Wasi Haider Butt. Their cooperation and insightful suggestions have been invaluable in shaping this research.

My acknowledgments would be incomplete without expressing my heartfelt thanks to my family, who has been my steadfast source of strength. I am profoundly grateful to my beloved parents, who nurtured me in my early years and have continually supported me in all facets of my life. My loving and supportive husband has been a pillar of strength throughout this process, and for that, I am deeply grateful. I also owe a debt of gratitude to my elder brother, whose constant encouragement has spurred me on to greater heights. My younger siblings, despite not fully comprehending the breadth of my work, have been unwavering in their support.

Lastly, I extend my sincere thanks to all my friends and the individuals who have provided encouragement and support throughout this journey. Your belief in me has been a beacon of light, guiding me through this challenging but rewarding path.

# ABSTRACT

Personality, a core component of human behavior, determines our interactions and perceptions of the world. Recognizing and forecasting personality traits can immensely impact areas like psychology, marketing, HR, and personalized recommendation systems. Recent advancements in Natural Language Processing (NLP) have fueled a keen interest in harnessing text data, encompassing essays and social media utterances, to precisely gauge personality traits. With billions of online users generating a plethora of text data, this rich information aids in discerning personality attributes. These textual imprints, from public declarations to diverse media formats, can revolutionize our grasp of human behavior. And with millions of students going university each year fill the forms and go through the process of analytical tests. The immense strides in computing power have even enabled models to outpace human proficiency in predicting personal actions, thus having ramifications in sectors like recruitment, healthcare, and more. The allure of formulating NLP models that effortlessly decode an individual's personality traits is ever-growing. These models exploit online text, encapsulating human tendencies and inclinations, to autonomously predict personality trait levels, which holds significant real-world relevance. Consequently, the text data about a person's personality can forecast emotions based on experiences, refining systems like recommendation engines and social network analyses. This could also bolster the progress of psychological theories, leading to a more holistic understanding of human personalities. Applications in fields like marketing, human resources, recommendation systems, and social science research further underscore the immense potential of this study. The overarching aim is to harness the vast text data resources, from essays to social media posts, to get an accurate read on personality traits. This endeavor aims to enrich our comprehension of human behavior, refine decision-making frameworks, and foster the creation of smart systems that resonate with individual desires across diverse domains. Moreover, this research has achieved notable results, particularly an impressive overall accuracy of 84.4%. Detailed outcomes for specific traits include a precision of 0.97 and an F1-Score of 0.88 for cEXT using BERT, emphasizing the model's robust performance.

## TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

Personality is a fundamental aspect of human behavior, influencing how individuals perceive and interact with the world around them [1]. Understanding and predicting personality traits have significant implications across various domains, including psychology, marketing, human resources, and personalized recommendation systems [2]. With the advancements in Natural Language Processing (NLP) techniques, there is a growing interest in leveraging text data, such as essays and social media posts, to infer personality traits accurately [3]. By analyzing the vast amounts of text data generated by billions of internet users, we gain invaluable insights into human desires, emotional states, mental health, and other needs. These textual traces, ranging from public posts and comments to various media formats like videos and pictures, serve as a rich source of information for identifying personality characteristics. The detection of personality traits through such data offers a novel avenue in the field of psychology,[4] further strengthening existing assumptions and potentially revolutionizing our understanding of human behavior. Moreover, the rapid advancements in computational power have endowed computer models with the ability to surpass human capability in predicting personal behaviors [5]. This technological progress has far-reaching implications for numerous domains, including recruitment, criminal investigations, healthcare, and overall well-being. Consequently, there is a growing fascination with developing NLP models that can seamlessly and naturally unravel an individual's personality traits [6]. These models leverage online text data, capturing human interests, behaviors, and preferences, to automatically predict levels of personality traits with profound real-world applications. So, the information about an individual's personality provides a basis for forecasting emotions based on experiences and situations, thereby enhancing recommendation systems, social network analysis, affective computing, and sentiment analysis [7]. The ability to accurately discern personality traits from textual data opens a world of possibilities for understanding human behavior and tailoring personalized experiences in various domains. As such, the exploration of NLP techniques for personality inference stands at the forefront of research, poised to unlock new insights and applications in our ever-evolving digital landscape [8].

The history of the Big Five Model, traces back to the mid-20th century, with its roots in the lexical hypothesis and early personality research [9]. The lexical hypothesis suggests that the most important aspects of human personality are encoded in language and reflected in the words we use to describe ourselves and others [10]. Researchers recognized the value of language to capture and measure personality traits. The development of the Big Five Model gained momentum in the 1960s and 1970s when psychologists began conducting factor analyses of personality descriptors found in natural language [11]. These analyses aimed to identify the underlying dimensions or factors that accounted for the variation in personality traits. Several independent research teams conducted factor analyses on large sets of trait adjectives or ratings, seeking to identify the fundamental dimensions of personality [12]. One notable contribution came from psychologists Warren Norman and Lewis Goldberg in the late 1960s. Norman performed a factor analysis on a comprehensive list of personality descriptors, while Goldberg independently conducted a similar analysis [13]. Both studies converged on five major dimensions or factors, which were labeled as extraversion, neuroticism, conscientiousness, agreeableness, and culture/intellect (later renamed openness to experience). During the 1980s and 1990s, the Big Five Model gained widespread recognition and acceptance in the field of personality psychology [14]. Researchers from various disciplines, including psychology, sociology, and economics, embraced the model due to its robustness, replicability across different cultures, and predictive power in explaining human behavior [15]. The Big Five Model provided a comprehensive and parsimonious framework that captured the major dimensions of human personality variation [16]. It demonstrated that personality traits could be organized along these five factors, offering a more nuanced understanding of individual differences than earlier theories such as the Myers-Briggs Type Indicator (MBTI) or the Eysenck Personality Inventory (EPI). Since its establishment, the Big Five Model has undergone refinement and consolidation through ongoing research efforts [17]. The labels for the five factors have been refined and standardized, with agreeableness and culture/intellect being merged into a single factor called openness to experience. Today, the Big Five Model stands as one of the most influential and widely accepted frameworks for understanding personality traits [18]. It has proven to be valuable in various domains, including psychology, organizational behavior, marketing, and personalized recommendation systems [19]. Its universality, stability across cultures, and predictive power make it a versatile tool for studying human behavior and developing practical applications for a range of real-world contexts.

As we see personality assessment is a complex task, and one prominent framework widely used in psychology is the Big Five Personality Model. The Big Five, also known as the Five-Factor Model (FFM), identifies five key dimensions that capture the major traits encompassing human personality. These traits are considered relatively stable and enduring across different situations and cultures.

**Table 1.1: Big Five Traits with definition [20]**

| Traits | Explanation |
|---|---|
| Extraversion | It reflects the degree to which individuals are outgoing, energetic, and socially oriented. Extraverts tend to seek stimulation from their external environment, enjoy social interactions, and are generally assertive and talkative. |
| Neuroticism | Sometimes referred to as emotional stability. It characterizes the extent to which individuals experience negative emotions such as anxiety, depression, and mood swings. Those high in neuroticism may be more prone to stress and tend to exhibit emotional instability. |
| Conscientiousness | The third dimension and represents the degree to which individuals are organized, responsible, and goal directed. Conscientious individuals are typically reliable, self-disciplined, and exhibit strong impulse control. They are more likely to plan ahead and strive for achievement. |
| Agreeableness | The tendency of individuals to be cooperative, compassionate, and empathetic. People high in agreeableness are generally warm, considerate, and prioritize harmony in their relationships. They value interpersonal connections and are often described as kind-hearted and trusting. |
| Openness | To experience represents the fifth dimension. It pertains to the degree of intellectual curiosity, creativity, and openness to new ideas and experiences. Open individuals tend to be imaginative, receptive to diverse perspectives, and enjoy exploring unconventional or abstract concepts. |

The Big Five Personality Model provides a comprehensive framework for understanding and measuring personality traits. Various assessment methods, such as self-report questionnaires, have been developed based on these dimensions to quantify individuals' levels of extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience. These assessments allow researchers and practitioners to gain insights into individuals' personalities, predict behavior patterns, and tailor interventions and recommendations accordingly.

In the context of NLP and text analysis, researchers have explored the connection between linguistic patterns in written text and the Big Five traits. By analyzing language use, word choice, and syntactic structures in textual data, it is possible to develop models that accurately predict an individual's personality profile based on their written expressions. This integration of NLP techniques with the Big Five Personality Model opens exciting possibilities for automatically inferring personality traits from text, further enhancing our understanding of human behavior, and facilitating personalized experiences in a wide range of applications [21].

## 1.1. NLP models for personality prediction

In the realm of Natural Language Processing (NLP), various models have emerged to predict personality traits with increasing accuracy and precision. Leveraging these models, researchers have delved into the depths of text data to uncover the intricacies of human personality. Here are some of the prominent NLP models that have been harnessed for this purpose:

### a) BERT (Bidirectional Encoder Representations from Transformers)

BERT, a transformer-based model introduced by Google, marked a significant leap in NLP capabilities. Its bidirectional context understanding revolutionized how models grasp the nuances of language. In personality prediction, BERT's ability to consider both preceding and subsequent words in a sentence enhances its comprehension of complex sentence structures and subtle linguistic cues. This feature makes BERT an excellent candidate for capturing burstiness in writing style, aligning well with human-like variations in sentence length and complexity [22].

### b) GPT (Generative Pre-trained Transformer)

The GPT series, particularly GPT-3, has garnered attention for its text generation prowess. However, it has also been harnessed for personality prediction tasks. Its contextual understanding and creative fluency enable it to capture the intricate balance between perplexity and burstiness in written content. GPT-3's larger context window allows it to detect nuanced shifts in writing style, producing text that mimics the ebb and flow of human writing, thus enhancing its ability to predict personality traits [23].

**c) XLNet**

XLNet extends the transformer model by considering all possible permutations of words in a sequence, allowing it to better capture dependencies between words. This permutation-based approach aligns with the unpredictable variations in burstiness found in human writing, resulting in outputs that exhibit a natural and dynamic writing style. This makes XLNet well-suited for discerning personality traits that manifest through distinct writing patterns [24].

**d) RoBERTa (A Robustly Optimized BERT Pretraining Approach)**

RoBERTa is a variant of BERT that employs advanced training techniques, leading to enhanced model performance. Its focus on large-scale data and dynamic masking during training improves its ability to understand complex writing patterns, which mirrors the intricate burstiness of human writing. RoBERTa's proficiency in capturing contextual nuances aids in more accurate personality prediction [25].

**e) DistilBERT (Distill BERT)**

DistilBERT is a distilled version of BERT designed to be lighter and faster while maintaining a high level of performance. While it sacrifices some complexity, its utilization of knowledge distillation retains the essence of BERT's language understanding capabilities. This model strikes a balance between perplexity and burstiness, allowing it to generate text that is both informative and engaging [26].

In this research we aim to used BERT and its new versions to experiment them to predict the personality from multiple data sources by setting the threshold.

## 1.2 Aim

The primary aim of our research is to devise a robust and accurate deep learning-based model for predicting personality traits from diverse textual data. This objective involves developing, testing, and refining a model that can read and understand a broad range of texts and draw accurate inferences about the personality traits of the authors.

## 1.3 Motivation

Motivation for conducting this research on personality prediction from text using essays and status-based analysis stems from the profound impact that understanding and predicting personality traits can have on various domains of study and practical applications. Mainly the motivation is to develop a generic system that can be used in any field to predict the personality.

- **Psychology:** Gaining insights into individuals' personality traits allows for a deeper understanding of human behavior, cognition, and emotional processes. It can contribute to the advancement of psychological theories and models, aiding in the development of more accurate and comprehensive frameworks for understanding human personality.
- **Marketing and Consumer Behavior:** Personality traits influence consumers' preferences, decision-making processes, and purchasing behaviors. By predicting personality traits from text, marketers can tailor their marketing strategies, personalized recommendations, and product offerings to match the individual preferences and characteristics of their target audience.
- **Human Resources:** Understanding personality traits is crucial in the context of employee selection, job fit, and team dynamics. Using text data as a source of insight, employers can gain insights into candidates' personality profiles, helping them make informed decisions regarding hiring, team composition, and employee development.
- **Personalized Recommendation Systems:** Personality-based recommendation systems can provide individuals with tailored content, products, and services that align with their preferences and characteristics. By predicting personality traits from text data, recommendation systems can offer more accurate and personalized suggestions, leading to enhanced user experiences.
- **Social Science Research:** Analyzing text data to predict personality traits can contribute to various social science research areas. It can aid in understanding the relationship

between personality and social behaviors, communication patterns, and the influence of online interactions on individual well-being and mental health.

Moreover, the motivation behind this research is to leverage the vast amounts of text data available, such as essays and social media posts, to accurately infer personality traits. This has the potential to enhance our understanding of human behavior, improve decision-making processes, and enable the development of intelligent systems that cater to individual needs and preferences in a wide range of domains.

## 1.4 Objectives

The objectives of this research encompass understanding and exploring the potential of personality prediction from text, particularly through the analysis of essays and status-based data. The specific objectives are as follows:

- Leveraging Multiple Datasets: We seek to use diverse datasets for the task of personality prediction, including Essays, MyPersonality, and SOP's datasets. The utilization of these different datasets will allow us to cover a wide spectrum of text styles, topics, and contexts, thereby ensuring the model's applicability to various real-world scenarios.
- Model Development and Optimization: We plan to employ the BERT-based model for our research, given its proven success in various Natural Language Processing tasks. Our aim is to fine-tune the BERT model to the task of personality prediction, optimizing its parameters to achieve the best possible performance.
- Evaluating Model Performance: A critical objective of our research is to evaluate the effectiveness of our personality prediction model thoroughly. We will perform this evaluation using different metrics to ascertain the model's performance in accurately predicting personality traits.
- Assessing Generalizability and Transferability: Our research aims to validate the developed model's capability to generalize and transfer learning across different text genres and domains. We plan to test the model's performance on the SOP's dataset to measure its ability to predict personality traits from a distinct genre of text. Also, to

assess the generalizability of the developed models across different cultural and linguistic contexts, considering potential biases and variations in language use.

- Understanding the Link between Language and Personality: Also, we aim to gain further insights into the relationship between language usage and personality traits. Through the analysis of language patterns and features in different types of texts, we aspire to uncover new correlations and build upon the existing body of knowledge in the field of psycholinguistics.
- Investigate the relationship between linguistic patterns in written text and the Big Five personality traits (extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience).
- Develop and refine Natural Language Processing (NLP) techniques and models that can accurately predict personality traits based on textual data, with a focus on essays and status updates.
- Explore the effectiveness and validity of personality prediction from text data by comparing the predicted personality traits with established personality assessments and self-report measures.
- Examine the impact and implications of personality prediction in various domains, including psychology, marketing, human resources, and personalized recommendation systems.
- Investigate the potential applications of personality prediction from text in enhancing user experiences, such as personalized content recommendations, targeted marketing strategies, and improved employee selection processes.
- Contribute to the existing body of knowledge on personality prediction from text, providing insights into the strengths, limitations, and future directions of this field of research.
- Practical Implementation: Ultimately, our research seeks to facilitate practical applications in fields such as psychology, marketing, human resources, and personalized recommendation systems. We aim to create a model that not only performs well in research setting but can also be integrated effectively into real-world applications for personality prediction and analysis. Moreover, to provide practical

guidelines and recommendations for implementing personality prediction techniques in real-world applications, ensuring ethical considerations and privacy protection.

By achieving these objectives, the research aims to advance our understanding of personality prediction from text, validate its efficacy, and explore its wide-ranging applications in diverse domains. Ultimately, this research strives to contribute to the development of more accurate and efficient methods for inferring personality traits from textual data, leading to improved personalization, decision-making, and user experiences.

## 1.5 Research Questions

1 How can we develop a robust and accurate deep learning-based model for predicting personality traits from textual data, particularly essays and status updates?
2 How effectively can Natural Language Processing (NLP) techniques and models predict personality traits based on textual data?
3 Can we gain further insights into the relationship between language usage and personality traits through the analysis of language patterns and features in different types of texts?
4 How effectively can the BERT-based model be fine-tuned for the task of personality prediction, and what are the optimal parameters for achieving the best performance?
5 To what extent does the developed model generalize and transfer learning across different text genres and domains?
6 What are the challenges associated with using deep learning for personality assessment?

## 1.6 Structure of Thesis

This research work is planned as follows:

**Chapter 2** gives the literature review and the major related work done by the various researchers for Covid-19 sentiment analysis in the past few years.

**Chapter 3** consists of the proposed methodology in detail. It includes the framework that covers the deep learning models.

**Chapter 4** presents the databases used for evaluation. Detailed discussions of all experimental results are provided along with relevant figures and tables.

**Chapter 5** serves as the conclusion of the paper and outlines future scope of this research and potential avenues for further research.

## 1.7. Chapter Summary

As we see this chapter introduces the significance of personality in human behavior and the application of Natural Language Processing (NLP) in predicting personality traits. It details the history of the Big Five Personality Model, identifying five key dimensions of human personality, and explores various NLP models like BERT, GPT, and others used for personality prediction. The chapter outlines the primary aim of developing a deep learning-based model for predicting personality from textual data, with specific objectives, motivation, and research questions. It emphasizes the profound impact of understanding personality traits across domains like psychology, marketing, human resources, and personalized recommendation systems. The next chapter, the literature review, will delve into the history, effectiveness, and challenges of personality prediction models, exploring the connection between linguistic patterns and personality traits, and investigating potential applications and future directions in the field.

# CHAPTER 2: LITERATURE REVIEW

Personality psychology has a long history dating back to Ancient Greece. We have been subjected to a variety of miscellaneous theories to try to understand who we are. In some theories, the development of personality is explained [27], while in others, personalities are explained by individual differences [28]. The field of personality psychology has shifted its primary focus towards exploring the connections between personality and various human behaviors, such as scrutinizing the interplay of personality traits. There has been a concerted effort to evolve automated personality prediction methods utilizing textual data [29]. Past research in predicting an author's personality involved extracting specific attributes from the text, including lexicon, syntax, writing style, and subject matter. Correlation methods like Pearson correlation were then employed to identify features with strong connections to personality traits [30]. In diverse experimental scenarios, tools like LIWC have demonstrated the ability to discern meanings, encompassing aspects like attentional focus, emotional resonance, social connections, and cognitive patterns. Mairesse et al. [31] incorporated 84 attributes in their document-level feature compilation for personality forecasting. After text features are determined, conventional machine learning techniques are implemented, such as logistic regression, support vector machine (SVM) [32], Naïve Bayes, etc., to derive the final personality assessment. Recent advancements have leveraged deep learning, including techniques like Word2Vec and Glove [7,8], utilizing pretrained word embeddings to enhance personality prediction models. It stands to reason that text, as a manifestation of human language, might mirror the author's personality [33], a factor that personality psychologists must consistently consider. The rise of Internet-based communication infrastructures has amplified text-based interactions among individuals. Through analyzing exchanged texts, computational psychologists can probe into the personalities of the writers. One particular study found that integrating common sense knowledge with psycholinguistic characteristics led to a significant boost in prediction accuracy [34]. Majumder et al. [11] introduced a 1-D CNN n-grams model, achieving [35] cutting-edge results until it was eclipsed by the language model-driven ensemble approach (BB-SVM) by Kazemeini et al. [36]. Finally, Mehta et al. [37] provided an overview of the most recent developments in deep learning for automated personality prediction, emphasizing the role of effective multimodal prediction.

Several studies were concerned about taking advantage of multiple classifiers simultaneously and benefiting their prediction abilities. A method for ensemble modeling has been proposed by the authors in [38] using the predictions of different APP models. They propose five distinct APP models, including term frequency vector-based, ontology-based, enriched ontology-based, latent semantic analysis-based, and deep learning-based (BiLSTM). A Hierarchical Attention Network (HAN) is then used to aggregate all five individual models into a meta-model. Therefore, five distinct APP models can now be used to determine the Big Five personality traits. El-Demerdash et al. [39] propose a transfer learning-based APP method that takes advantage of leading pretrained language models including Elmo, ULMFiT, and BERT. They have developed a model that combines data fusion strategies and classifier strategies to improve overall personality prediction performance. With the help of three pre-trained models, they further refined the proposed models using essays and my personality data. Using independent classifiers, each model performs APP separately. To acquire more reliable predictions, multiple classifiers' outputs were combined into an ensemble learning model. Having the same objectives, other researchers [40–41] have questioned the usefulness of such an approach.

In a recent study conducted by Yoojoong Kim et al. (2022) [42], a Korean medical language model was formulated through deep learning NLP techniques. This model underwent training via BERT's pre-training framework, specifically tailored for the medical context, based on an advanced Korean language model. The results revealed notable improvements in accuracies, 0.147 and 0.148, respectively, for the masked language model combined with next sentence prediction. Additionally, the intrinsic assessment showed an enhancement of 0.258 in next sentence prediction accuracy, a significant advancement. Similarly, Vimala Balakrishnan et al. (2021) performed a comparative analysis of various deep learning models, such as Convolutional Neural Networks, Recurrent Neural Network, and Bi-directional Long Short-Term Memory. They evaluated these models using different word embedding techniques, including BERT, its variants, FastText, and Word2Vec. Among the models, Neural Network-based Word2Vec emerged as the best, with CNN-RNN-Bi-LSTM yielding the highest accuracy of 96% and an F-score of 91.1%. On an individual basis, RNN excelled with 87.5% accuracy and an F-score of 83.5%, whereas RoBERTa achieved the top F-score of 73.1% [43].

Hans Christian et al. (2021) proposed an innovative prediction approach utilizing a multi-model deep learning structure, which integrates multiple pre-trained language models such as BERT, RoBERTa, and XLNet. This was applied as a feature extraction method for social media data sources like Facebook and Twitter. The system's decision-making relies on model averaging, addressing the shortcomings of existing techniques like RNN and LSTM. These older methods were slow to train and struggled to capture the true semantic meaning of words. The authors enhanced the dataset with additional NLP features like Sentiment Analysis, Term Frequency-Inverse Gravity Moment (TF-IGM), and National Research Council (NRC) Emotion Lexicon Database. Achieving maximum accuracies of 86.2% and 88.5% on the Facebook and Twitter datasets, respectively, the authors confirmed the superior efficacy of their proposed method compared to previous research [44]. Collectively, these studies illuminate the promising potential of deep learning models, especially BERT and RoBERTa, in text analysis and personality prediction. Yet, it's crucial to recognize that these models' performance can be substantially amplified by integrating them with complementary strategies such as statistical feature extraction and data preprocessing. In another line of inquiry, researchers contend that personality is an essential attribute that characterizes an individual, encompassing their beliefs, feelings, attitudes, and more. They underscore the growing realm of personality detection and the recent advancements of deep learning models for this aim. The authors utilized renowned stream-of-consciousness essays by James Pennbaker and Laura King, employing the Big Five Model. They conducted document-level feature extraction through Google's word2vec embeddings and Mairesse features, feeding the processed data into a deep convolutional network for binary classification of personality traits. The evaluation was performed using the holdout method, and F1 score was the chosen metric. Additionally, the paper discusses potential applications of personality detection in various domains and suggests future research directions, including the creation of larger and more accurate datasets [46]. Another exploration [47] delves into Neuro Linguistic Programming (NLP) and meta programmers, cognitive strategies affecting behavior, focusing on the Myers-Briggs Type Indicator® (MBTI) for personality type prediction. The research traces the development of meta programmers, condensing them into four fundamental dimensions to form the MBTI. A novel machine learning approach is introduced, integrating it with existing meta programmers, and the study concludes that this new method offers superior accuracy and dependability. This research provides valuable insights for NLP practitioners and psychologists in identifying personality types, enriching the comprehension of individual variances, and implementing personalized strategies in diverse fields. The paper by M. Yağcı [48] emphasizes the utilization of educational data mining to foresee undergraduate students' final exam grades, proposing a new model that leverages machine learning algorithms. The dataset comprises the academic grades of 1854 students from a Turkish University during the fall semester of 2019–2020. The predictions were formed using three parameters: midterm exam grades, Department data, and Faculty data. The results indicate that the proposed model achieved an accuracy range of 70–75%, contributing to the early identification of students at risk of failing. In [49], the authors accentuate the increasing accessibility of vast digital data, such as online interactions and text-based content, that can furnish valuable insights into individuals' personality traits. This information can be leveraged for various applications, enhancing our understanding of human behavior and personality. By harnessing big data analytics and machine learning algorithms, researchers can extract patterns, behaviors, and linguistic features from these datasets to infer personality traits more accurately and comprehensively we can see table 2 how these datasets have been used:

**Table 2.1: Summary of Existing Personality prediction Models**

| Author | Publication | Dataset | Methodology | Results |
|---|---|---|---|---|
| Alam Sher Khan [50] | 2020 | MBTI Dataset | XGBoost, MNB and Stochastic Gradient Descent (SGD) KNN, Decision Tree, Random Forest, MLP, Logistic Regression, SVM | 99% precision 95% accuracy |
| Hans Christian, Derwin Suhartono [44] | 2021 | MyPersonality Facebook, manually collected twitter | BERT, RoBERTa, and XLNet) | 86.17% F1<br>0.912 Facebook dataset AND |

| | | | | |
|---|---|---|---|---|
| | | data in Bahasa Indonesia. | | 88.49% accuracy 0.882 F1 score on the Twitter dataset |
| Parsa Sharmila [51] | 2020 | 7852 applications and 41 Google play application. | Supervised regression algorithms, Random Forest Regression (RF), Support Vector Regression (SVR) | 89-94% fit |
| Randa Zarnoufi, Mounia Abik [52]. | 2019 | Cyber-violence dataset and big 5 dataset pre-labeled tweets dataset | Random Forest, XGBoost and AdaBoost classifiers | 86% accuracy |
| Kulsum Akter Nisha, Umme Kulsum, Saifur Rahman, Md. Farhad Hossain, Partha Chakraborty, and Tanupriya Choudhur [53] | 2021 | Kaggle dataset of 50 tweets and 8675 rows, Myers–Briggs Type Indicator (MBTI) Dataset | Naive Bayes, SVM, and XGBoost, | SVM classifier achieved I/E = 80%, N/S = 86%, T/F = 80%, J/P=72% XGBoost I/E = 86%, N/S = 90% T/F = 84%, J/P = 80% |
| Noureen Aslam, Khalid Masood Khan, Afrozah, Nadeem,Sundu | 2021 | myPersonality facebook dataset, and Twitter dataset | TF-IDF, deep sequential neural networks and multi- | Training=94%, testing =78%. |

| | | | | |
|---|---|---|---|---|
| s Munir, and Javairya Nadeem [54] | | | target regression models | |
| Kamal El-Demerdash Reda A. El-Khoribi, Mahmoud A. Ismail Shoman, Sherif Abdo [39] | 2021 | fusion of Essays and Facebook myPersonality datasets | ELMo, ULMFiT, and BERT | 61.85% when fused dataset tested on essay's dataset, 73.91% average accuracy on myPersonality Dataset |
| Peng Wang1, Yun Yan , Yingdong Si , Gancheng Zhu , Xiangping Zhan, Jun Wang , And Runsheng Pan [55] | 2020 | Weibo text datasets of students short-answer questions text dataset | Support Vector Machine, XGBoost, K-Nearest-Neighbors (KNN), Naive Bayes (NB) and Logistic Regression (LR) | 0.842 and 0.969 are Best Accuracies |
| Hussain Ahmad, Muhammad Usama Asghar, Muhammad Zubair Asghar , Aurangzeb | 2021 | MBTI dataset and Reviews text dataset | CNN+LSTM, Support Vector Machine (SVM), K Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), | 88% |

| | | | | |
|---|---|---|---|---|
| Khan, And Amir H. Mosav [56] | | | Decision Tree (DT), XGBoost, Recurrent Neural Network (RNN), and Gated Recurrent Network (GRU). | |
| Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat [57] | 2021 | Manual data collection of potential job seekers 708 CVs. | Logistic Regression, Naive Bayes, KNN, Random Forest | 0.71% |
| Roshal Moraes, Larissa Lancelot Pinto, Mrunal Pilankar, Pradnya Ran[58] | 2020 | MBTI Dataset | Decision Tree, SVM | 77.18% with SVM, and 86.76% Decision tree |

The paper in [59] presents a pioneering approach to personality trait analysis through handwriting samples. By leveraging deep learning methods and image processing techniques, the study aims to extract relevant characteristics from handwritten content to predict the Big Five personality traits. These traits include Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The paper outlines the successful implementation of convolutional neural networks (CNNs) and image processing algorithms to accurately predict personality traits based on handwriting samples. The results indicate a significant correlation between specific handwriting features and certain personality traits. This research showcases the potential of combining advanced technology with traditional methods to uncover hidden insights into human psychology and behavior, shedding light on the intricate relationship between personality traits and unique forms of human expression [59]. In the paper by Majid Ramezani et al., proposed five distinct methods for personality prediction and an ensemble model combining them [REF]. Surprisingly, the simplest method, term frequency vector-based, achieved the best accuracy across all five traits of the Big Five model. However, the ensemble method, implemented through a Hierarchical Attention Network (HAN), outperformed all individual methods in average accuracy 60.24% and precision 60.48%, except for the openness trait. These findings highlight the effectiveness of both simple and ensemble modeling in personality prediction, challenging the notion that complexity always yields better results, and opening new avenues for applications in various fields [60].

The literature review underscores a burgeoning interest in harnessing machine learning and deep learning methodologies for predicting personality from textual content. Researchers have tapped into a diverse array of datasets, ranging from MBTI models and mobile app usage statistics to prolific social media content. This wide variety demonstrates the adaptability of these methods to interpret different textual data sources for personality extrapolation. Several studies have adopted state-of-the-art models, including renowned pre-trained models like BERT and RoBERTa, alongside ensemble techniques such as XGBoost. This trend showcases an inclination towards leveraging sophisticated algorithms to enhance the accuracy of personality projections. Many of these research endeavors have reported striking accuracy levels, with a substantial number surpassing the 80% threshold, emphasizing the efficacy and potential of machine learning in this arena. The application realms of these studies are vast, spanning from cyber-violence detection on social platforms to refining hiring processes, highlighting the broad applicability and significance of personality predictions in various domains. Additionally, some scholars have embarked on comparative analyses of disparate machine learning strategies, offering invaluable insights into the relative efficacies of several algorithms in personality prediction tasks. Techniques like TF-IDF, sentiment analysis, and NLP feature selection emerge recurrently, indicating their pivotal role in refining and enhancing textual data quality.

Based on the comprehensive literature review provided, the following research gaps can be identified:

## 2.1. Overall Research Gaps in Literature

i. **Lack of Diverse Data:** Many studies on personality prediction from text data might have used a limited or homogeneous dataset, such as text data from a specific social media platform or a single type of text (like tweets or blog posts). A potential research gap could be the exploration of diverse and multilingual datasets, or the analysis of different types of text (like essays, status updates, and SOPs).

ii. **Cultural and Linguistic Context:** Previous research might not have sufficiently considered the cultural and linguistic context of the text data. How do cultural norms and different languages affect the expression of personality traits in text?

iii. **Combination of Different NLP Techniques:** While many studies might have applied common NLP techniques, but in the exploration of a combination of different NLP techniques or novel machine learning algorithms for personality prediction for more accurate results is needed.

iv. **Real-world Application**

How can the results of personality prediction from text data be effectively applied in real-world scenarios, like marketing, HR, and personalized recommendation systems? If previous research has primarily focused on the theoretical or technical aspects, as it is a potential research gap, the practical application of the findings.

**v. Lack of Standardization in Feature Extraction**

Different studies have utilized various features like lexicon, syntax, writing style, and topic for personality prediction [61]. There is a lack of consensus on which features are most effective, leading to a gap in standardizing the feature extraction process.

**vi. Need for Larger, Unbiased Datasets**

Current datasets often rely on self-reported questionnaires, which may introduce unconscious bias. There is a need for larger, unbiased datasets to enhance the accuracy of personality prediction models.

**vii. Exploration of Less Biased Word Embeddings:**

Some researchers have suggested exploring less biased word embeddings for personality detection. This area needs further investigation to understand how biases in word embeddings might affect personality prediction.

viii. **Combination of Different Techniques:**

While deep learning models like BERT and RoBERTa have shown potential, combining them with other techniques like statistical feature extraction and data preprocessing can significantly enhance performance. This area needs more exploration to identify the best combination of techniques.

ix. **Multimodal Prediction Approaches:**

The literature review emphasizes text-based personality prediction, but there's a gap in exploring multimodal approaches that combine text, audio, and visual cues for more robust personality detection.

x. **Transfer Learning and Ensemble Learning:**

Although some studies have proposed transfer learning-based APP methods [39] and ensemble learning models [60], there's still room for further exploration and optimization of these approaches to improve overall personality prediction performance.

## 2.2. Research Gaps for this Study

We are going to cove the following research gaps in this state of art research.

- Previous studies used a limited or homogeneous dataset, such as text data from a specific social media platform or a single type of text (like tweets or blog posts).
- In literature, Transformer based models were rarely used for complex models of Personality prediction from text
- There is a need to sufficiently considered the contextual information of the text data.

## 2.3. Chapter Summary

The literature review reveals a rich and evolving field of research in personality psychology and prediction. As the literature underscores the promising nature of machine learning and deep learning techniques in the realm of personality prediction from text. The consistent achievement of high accuracy across varied datasets and methods indicates a mature and evolving field with significant real-world implications which lead us too best usage of these research. While significant advancements have been made, especially with the integration of deep learning models, there are still notable gaps that present opportunities for further research and innovation. By addressing these gaps, future work can contribute to more accurate, ethical, and applicable personality prediction models. This paper aims to build a methodology that can accurately predict the personality and the approach we are going to proposed has never been used before.

# CHAPTER 3: EXPERIMENTAL METHODOLOGY

## 3.1 Dataset

In our research, we utilized three datasets for the personality prediction: Essays, MyPersonality, and SOP's dataset. Each dataset served a specific purpose in our research pipeline. Let us delve into the details of each dataset and the rationale behind their selection.

**Table 3.1: Datasets explanation**

| **Name** | **Source** | **Records** |
|---|---|---|
| Essays | Students | 2,246 |
| myPersonality | Twitter | 9915 |
| SOP’s | Students | 309 |

### 3.1.1. Essays Dataset

The dataset used in this domain consists of 2,246 essays collected by Pennebaker and King in their earlier research work [62]. The study was conducted as part of a research project funded by the National Institute of Health Texas. Their primary goal was to explore and establish correlations between linguistic features found in the essays and the Big Five personality traits. The essays were obtained from 2000 students attending the Toas 27 Summer School, and they were written in a stream of consciousness mode. This mode of writing allows individuals to express their thoughts and feelings freely and spontaneously.

Pennebaker and King's research was groundbreaking as they developed a unique Psycholinguistic lexicon known as Word Count and Linguistic Inquiry (LIWC). This lexicon enabled them to identify and analyze linguistic cues that were correlated to the Big Five personality traits, namely openness, conscientiousness, extraversion, agreeableness, and neuroticism. Although their study did not propose a specific model for automatic detection of personality traits from the text, they made significant progress in understanding the link between language and personality. The linguistic cues they identified provided valuable insights into the association between writing style and individual personality differences. Furthermore, they demonstrated that writing style remains consistent across time and topics, reinforcing the notion that certain linguistic features can be indicative of an individual's personality traits. Since then, the Essays dataset collected by Pennebaker and King has been widely utilized in various research endeavors, particularly in the field of automatic detection of Big Five personality traits using natural language processing techniques. Researchers have continued to build upon their work, leveraging the dataset to develop models and algorithms for personality trait prediction from textual data. The essays dataset offers a rich source of textual information that can be leveraged to extract features related to an individual's personality traits. By analyzing the language, writing style, and content of the essays, we can gain valuable cues and patterns associated with personality traits. Here we can see the word cloud of essays dataset for an overview:

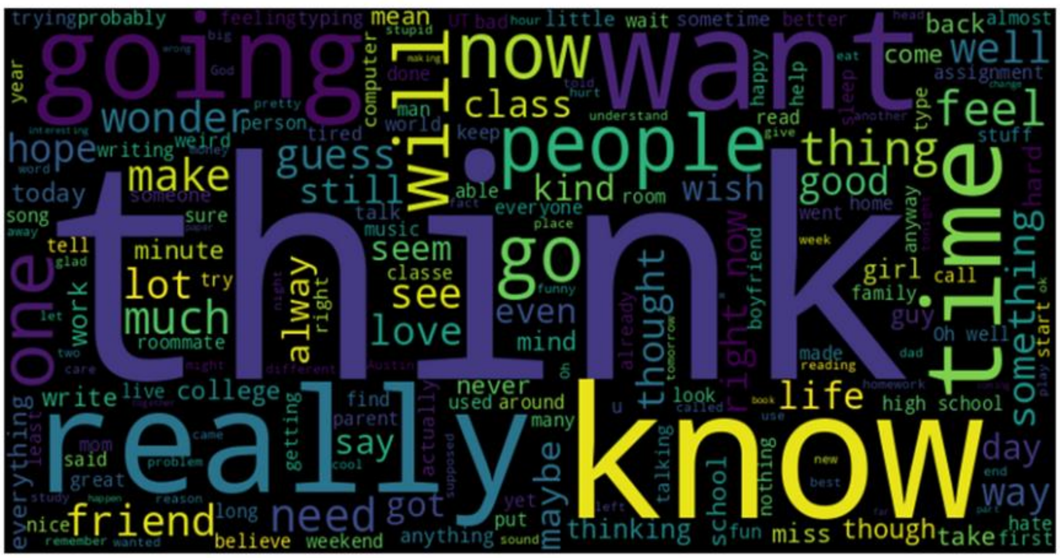


**Figure 1: Essay dataset Word Cloud**

### 3.1.2. myPersonality Dataset

The "MyPersonality Twitter" dataset is a collection of tweets from 250 users obtained from Kaggle. The Twitter MyPersonality dataset has been instrumental in studying the connection between social media behavior and personality traits. It has a rich collection of data from 250 unique Twitter users that comprises 9915 statuses, providing valuable insights into the online behavior and personality traits of these individuals based on their Twitter activity [63].

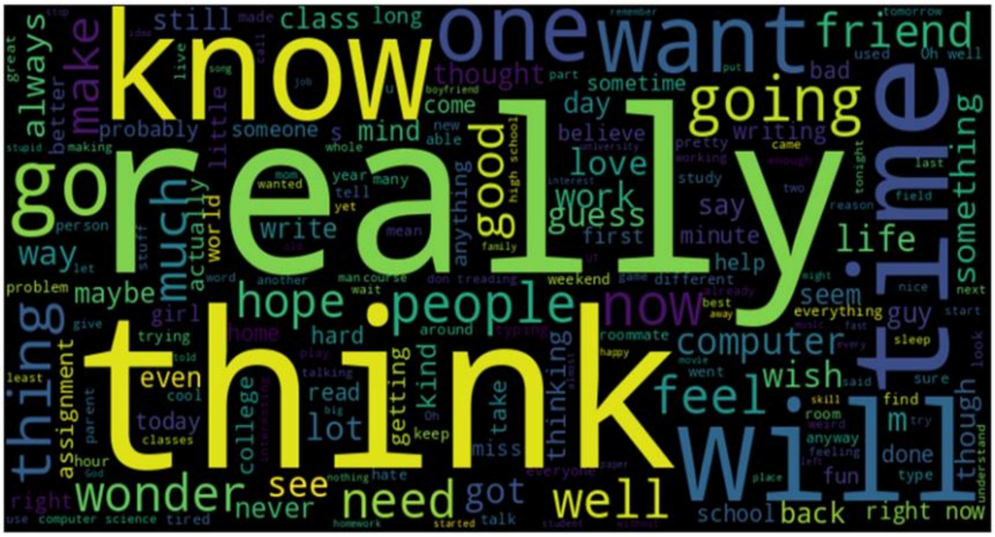


**Figure 2: myPersonality Dataset**

### 3.1.3. My personality Dataset overview

The dataset was collected as part of the MyPersonality project, initiated by a team of researchers in psychology and computer science. The project aimed to explore how social media activities could reflect users' personality traits. The dataset includes data from 250 different Twitter users. These users were selected to represent diverse demographics, locations, and interests. For each user in the dataset, their tweets are available for analysis. These tweets offer valuable insights into the topics users engage with, their language usage, interests, and communication style on the platform.

### 3.1.4. SOP's Dataset

The SOP's (Statement of Purpose) dataset consists of written statements provided by individuals applying for academic programs or scholarships it has 309 samples. These samples were collected by Salma a student of NUST university Islamabad for her thesis. These statements typically describe the applicants' motivations, aspirations, and experiences related to their chosen field of study [64]. The SOP's dataset allows us to assess the generalizability and transferability of our personality prediction model across different domains, as it represents a distinct text genre from the Essays and MyPersonality datasets.

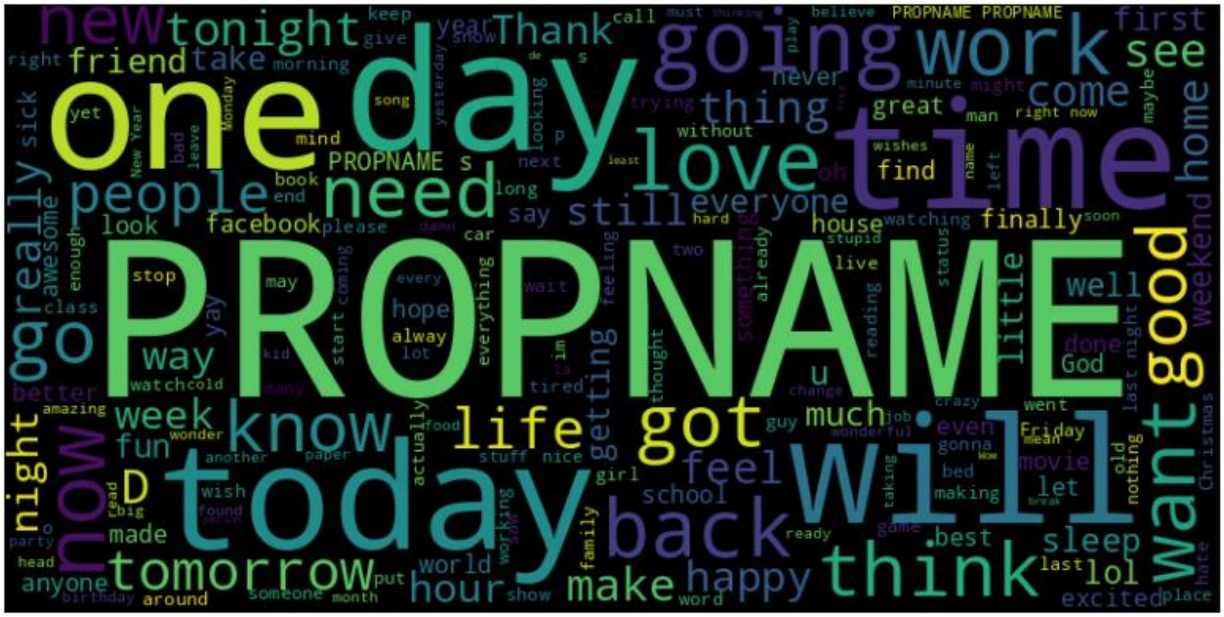


**Figure 3: Essay dataset Word Cloud**

### 3.1.5. Combining Techniques

To enhance the performance and robustness of our personality prediction model, we combined the Essays and MyPersonality datasets for training. This combination offers a larger and more diverse training corpus, enabling the model to learn from a broader range of text samples and capture more nuanced patterns related to personality traits. By training on a diverse set of essays and social media posts, we aimed to improve the model's ability to generalize to unseen data and accurately predict personality traits in various contexts. We can see the distribution of the traits below:

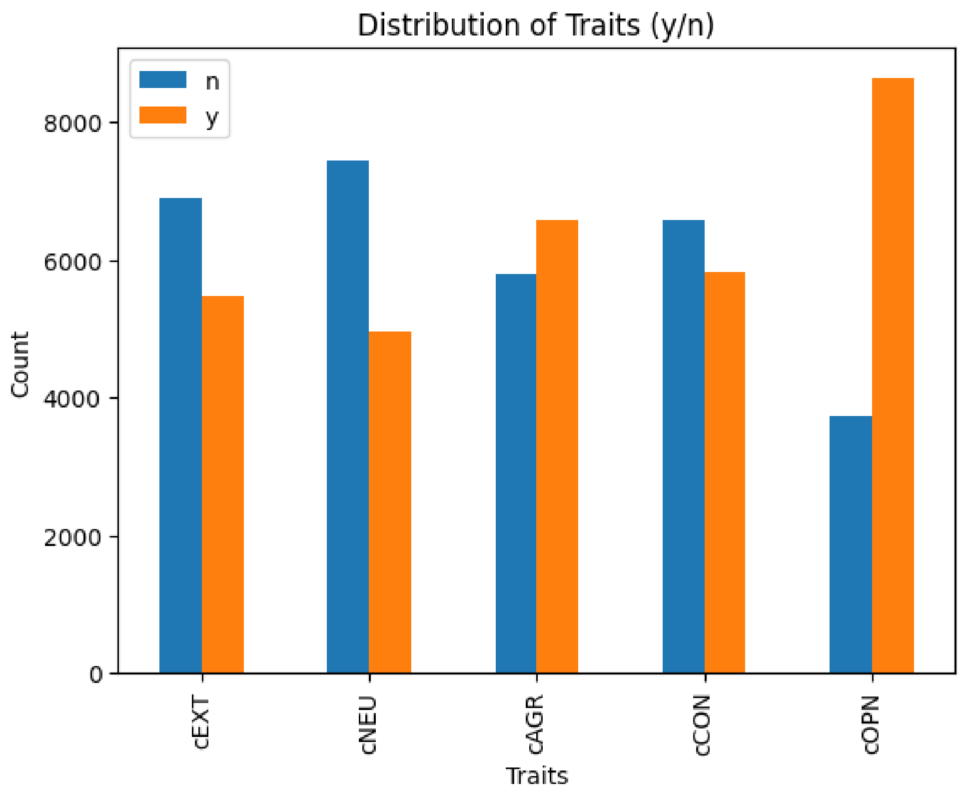


**Figure 4: Distribution of dataset according to Traits**

Furthermore, we tested the trained model on the SOP's dataset to assess its performance in a different text domain. This evaluation allows us to evaluate the model's generalizability and examine its effectiveness in predicting personality traits from written statements specifically tailored for academic purposes. By utilizing multiple datasets and employing a combination approach, we aimed to develop a comprehensive and robust personality prediction model that can capture personality traits from diverse textual sources, such as essays, social media posts, and academic statements. As we are using pretrained model of uncased BERT by making the amendments in classification layer so that we can achieve our desired output.

### 3.1.6. Data Pre-processing

I. **Replacement of Words**: Certain words in the dataset are replaced with their correct or standardized forms. For example, "goood" might be replaced with "good". This standardizes the text and helps in reducing the overall vocabulary size.

II. **Removal of Stop Words**: Commonly used words that do not carry significant information (e.g., "and", "the", "is") are removed from the text.

III. **Removal of Contractions**: Contractions are replaced with their expanded forms. For example, "it's" becoming "it is". This ensures each word is represented individually.

IV. **Removal of Relevant Classes**: If there are any irrelevant classes in the dataset, these are dropped to focus on the classes that are important for the analysis.

V. **Removal of Dates and Time**: Dates and time are removed from the text. This is because they usually do not provide useful information for text analysis.

VI. **Removal of Null Values**: Any null or missing values in the dataset are removed to avoid errors during the analysis.

VII. **Removal of Leading and Trailing Spaces**: Extra spaces at the beginning and end of the text strings are removed to clean up the text.

VIII. **Removal of Links and Tags**: Any links or HTML tags present in the text are removed, as these do not contribute to the analysis.

IX. **Replacement of Proper Words**: This could involve replacing words with their correct spellings or replacing slang or informal language with formal language.

X. **Conversion to Lowercase**: The entire text data is converted to lowercase to ensure uniformity and prevent the same words in different cases from being treated as different words.

XI. **Removal of Non-Alphanumeric Characters**: Characters that are not letters or numbers, such as special characters or symbols, are removed from the text.

XII. **Removal of Extra Spaces**: Any additional spaces between words in the text are removed.

XIII. **Cleaning the Data**: This could include a variety of tasks not mentioned above, such as correcting spelling errors, removing duplicate entries, or addressing any other issues specific to the dataset.

After going through these preprocessing steps, the combined dataset of essays and "My Personality" is transformed into a clean dataset that's ready for further analysis or model training.

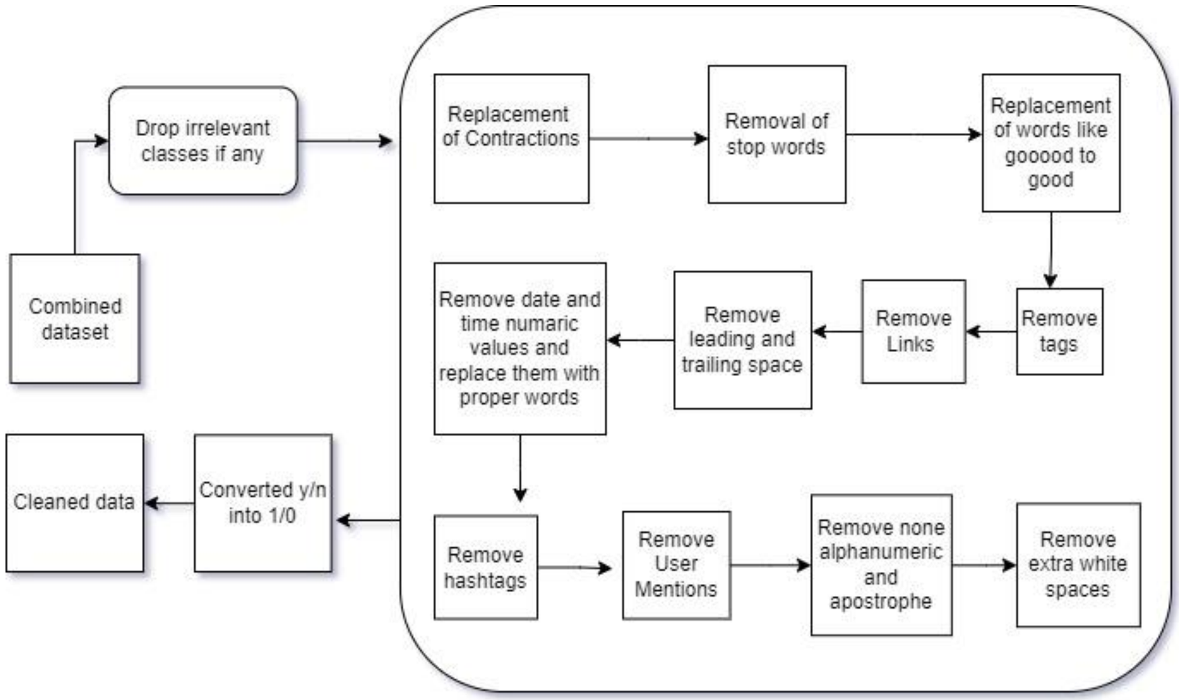


**Figure 5: Pre-Processing Steps**

Figure 5 represents the preprocessing steps applied to a raw dataset in a text analysis project. First, the dataset undergoes an initial cleanup where URLs, user mentions, and any HTML tags are removed, as these typically do not contribute meaningful information for text analysis. Non-alphanumeric characters are also stripped out, except for apostrophes, which are often integral to the meaning of words (for example, in contractions). The cleaned text data is then transformed to lowercase to ensure uniformity and to prevent the same words in different cases from being treated as distinct. Any numbers in the text data are removed next, unless they hold specific importance for the analysis, to reduce potential noise. Punctuation marks are also removed to simplify the text and maintain focus on the words themselves. The text data is then tokenized, or broken down into individual words, a crucial step for many natural language processing tasks. Subsequently, "stop words" or commonly used words (like "and", "the", "a") that generally carry less meaningful information are eliminated from the text data. Lastly, stemming is applied, reducing words to their root or base form (for example, "running" might be stemmed to "run"). This helps in grouping together words that have the same fundamental meaning.

### 3.1.7. Dataset Split

The dataset is split into training and testing sets using shuffled stratified single fold sampling to evenly distribute classes between training and testing sets and to ensure a balanced distribution of classes in both sets.

- **Significance of the Split Ratio**: The 90:10 split ratio for training and testing data was a deliberate and significant choice for the research, aiming to optimize the learning and evaluation process.

## 3.2 BERT

BERT, or Bidirectional Encoder Representations from Transformers, represents a cutting-edge development in the field of natural language processing (NLP). Unveiled by Google AI's research team in 2018[65], BERT has established itself as a powerful tool in a myriad of NLP applications, such as text comprehension, sentiment evaluation, question resolution, and text categorization. Its foundation lies in the transformer model architecture, an innovation by Vaswani et al. in 2017[66], that marked a significant shift in NLP by overcoming the challenges faced by conventional recurrent neural networks (RNNs) and convolutional neural networks (CNNs). A vital aspect of the transformer model is the self-attention mechanism, facilitating the effective capture of global dependencies within an input sequence [67].

BERT advances the transformer model by implementing a two-phase approach: pre-training and fine-tuning. During the pre-training stage, BERT is exposed to substantial volumes of unlabeled textual content from varied sources like books and Wikipedia. This process empowers BERT to assimilate contextualized word and sentence representations, encompassing both semantic and syntactic nuances. Following this, the fine-tuning stage involves further training BERT on specialized tasks using labeled data, allowing it to tailor its learned representations to the specific requirements of a given task. One distinguishing characteristic of BERT is its bidirectional nature, which enables the model to take into account both prior and subsequent words in predicting the upcoming word in a sentence. This bidirectionality enhances BERT's ability to grasp a more comprehensive context and recognize the interdependencies among words. The particular version of the BERT model described here consists of 12 transformer blocks, a hidden size of 768, and 12 self-attention heads, encompassing roughly 110M adjustable parameters. The preparation of input data for BERT involves processes like tokenization, segmentation, and word arrangement, accompanied by the inclusion of special tokens, [CLS] and [SEP], designated for classification and segmentation functions, respectively. The input is then converted into embeddings, including token embeddings, segment embeddings, and positional embeddings [68].

### 3.2.1. Key components of the BERT model

### 3.2.2. Word Embeddings (Input and Output Format)

BERT creates word embeddings by utilizing a technique called WordPiece tokenization. In this process, the input text is first split into individual words or subwords, and then each word/subword is assigned a unique ID from a fixed vocabulary. BERT uses a pre-trained word embedding matrix that is learned during the pre-training phase.

Dataset preprocessing is crucial for converting raw input data into a format that BERT can readily comprehend and process. The preprocessing steps for the BERT model are divided into three levels, namely Tokenization, Segmentation, and Word Ordering [32]. These steps ensure that the input data is appropriately formatted and prepared for effective use by the BERT model during training and inference.

The input data for BERT, such as the example "Today. Had to turn the music down.," goes through a series of preprocessing steps to create a suitable format that BERT can comprehend. This preprocessing involves three main steps: Tokenization, Segmentation, and Word Ordering.

#### I. Token Embedding

In the token embedding step, the entire input sentence is converted into tokens, and special identification tokens, [CLS] and [SEP], are introduced to help BERT understand the input context effectively. The [CLS] token serves as a special classification token, and BERT uses the final hidden state associated with this token for classification tasks. The [SEP] token, on the other hand, acts as a separator token and is placed at the end of one input. The [SEP] token assists BERT in recognizing the end of the first input and the beginning of the next sentence in the same input sequence. For tasks like Natural Language Inference (NLI) and Question-Answering, which require multiple inputs, the [SEP] token marks the separation between the input sentences accordingly. BERT employs wordpiece embeddings for input tokens.

### II. Segment Embedding

BERT uses segment embeddings to differentiate between multiple input sentences. The embedder component in the BERT model distinguishes whether the tokens belong to the first input sentence (sentence 1) or the second input sentence (sentence 2). Tokens of sentence 1 have predefined embeddings of 0, while tokens of sentence 2 have segment embeddings as 1 (Figure 6).

### III. Positional Embedding

Positional embedding is used by BERT to capture the positional information of tokens in the input sentences. The embedder component generates positional embeddings that indicate the position of each token in the input sequence, enabling BERT to understand the relative positions of the tokens within the sequence. The final embeddings obtained after considering token embedding, segment embedding, and positional embedding are fed into the model to get the output after training.

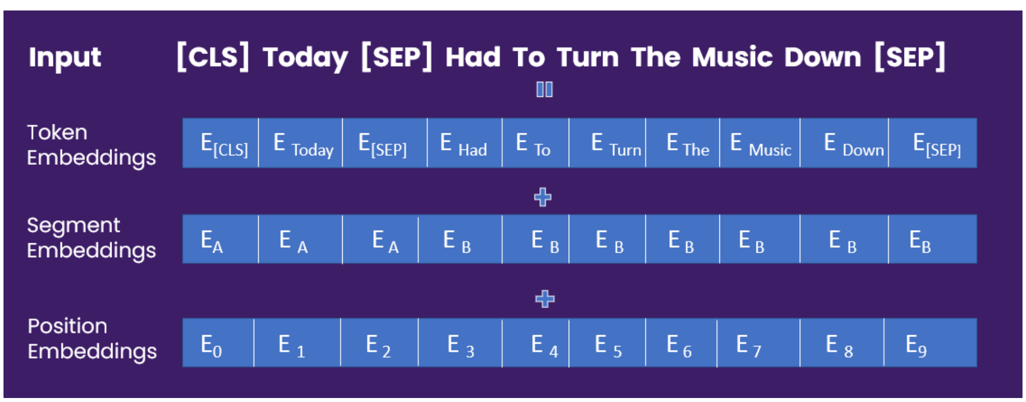


**Figure 6: Word piece Embeddings**

## 3.2.3. Pre-training and Fine-tuning

BERT's training process is divided into two main phases: pre-training and fine-tuning. During the pre-training phase, BERT is exposed to a vast body of text, enabling it to acquire general linguistic representations. This pre-training utilizes two specific tasks: Masked Language Model (MLM) and

Next Sentence Prediction (NSP). In the MLM task, 15% of the tokens within the corpus are randomly obscured, and the model's objective is to accurately predict these concealed words by considering their surrounding context. Concurrently, the NSP task requires the model to determine the relationship between two sentences, discerning whether they are logically connected or arbitrarily paired.

Upon completion of the pre-training, BERT transitions into the fine-tuning phase, where it is tailored for distinct downstream NLP applications such as text categorization, named entity identification, sentiment evaluation, and more. This fine-tuning process involves training BERT on labeled data that is specific to the intended task, thereby adapting the model to perform optimally for that particular function.

#### I. Masked Language Model (MLM)

The MLM is one of the pre-training tasks used to train BERT. In this task, some words in the input text are randomly masked, and the model is tasked with predicting the masked words based on the context of the surrounding words. This process helps BERT learn contextual embeddings, which capture the meaning of words based on their context.

#### II. Next Sentence Prediction (NSP)

The NSP is another pre-training task used to train BERT. In this task, pairs of sentences are provided as input, and the model is tasked with predicting whether the second sentence follows the first one logically or not. The NSP helps BERT learn relationships between sentences and understand the coherence and context between them.

### 3.2.4. Attention Mechanism

The BERT model is constructed upon the Transformer architecture's attention mechanism, a feature that empowers the model to adeptly discern the relationships between words within a sentence. This attention mechanism functions by allocating varying weights to individual words, reflecting their significance in deciphering the overall context.

In the context of our research, we employed the BERT model, specifically fine-tuning it for Multilabel classification tasks. We leveraged the "BertForSequenceClassification" class from the widely-used Hugging Face library. To prepare for training, the dataset was segregated into distinct training and validation subsets, and BERT's specialized tokenizer was applied to break down the text data into tokens. The training process was orchestrated using the AdamW optimizer, experimenting with a range of learning rates including 6e-5, 2e-5, 3e-5, 3e-4, 5e-5, and the default rate of 1e-5. The model was trained in batches, with random sampling employed to enhance the learning process, and a scheduler was engaged to dynamically modify the learning rate throughout the training cycle. Finally, the model's proficiency was evaluated, focusing on the accuracy metric to gauge its effectiveness in accurately classifying the data.

BERT-Base Uncased, as utilized in this research, is a pretrained model operating on a comprehensive corpus, allowing it to effectively grasp context and the underlying semantic connections within the input text. When fine-tuned for sentiment classification, it demonstrates an ability to categorize sentiments both accurately and efficiently. Our code explores various batch sizes and learning rates through experimentation to pinpoint the ideal configurations for the dataset. As a foundational variant of the BERT (Bidirectional Encoder Representations from Transformers) model with uncased vocabulary, BERT-Base Uncased has emerged as a formidable and influential structure within the sphere of natural language processing (NLP). Rooted in the Transformer model's architecture, it employs self-attention mechanisms that discern the intricate contextual relationships among words. The nucleus of BERT-Base Uncased is its transformer encoder, an assembly of multiple interlaced transformer encoder layers. Each layer fuses multi-head self-attention with feed-forward neural network sub-layers, enabling the model to recognize bidirectional associations between words. This recognition enriches the understanding of context and the significance of individual words in relation to their neighbors. Additionally, the feed-forward neural network sub-layers facilitate non-linear transformations, capturing more complex features.

Training BERT-Base Uncased necessitates a pretraining stage, employing unsupervised learning methods on an extensive collection of unlabeled textual data. During this foundational phase, the model assimilates universal language representations, encompassing both syntactic and semantic facets. BERT's pretraining regimen includes tasks like masked language modeling (MLM) and next sentence prediction (NSP). MLM involves randomly obscuring a portion of the input tokens, challenging the model to recover the original tokens from their context. Conversely, NSP tasks the model with discerning whether two sentences are sequentially connected or randomly drawn from the corpus. Through these tasks, BERT-Base Uncased acquires a nuanced understanding of word relationships and comprehensive linguistic insights.

Post-pretraining, BERT-Base Uncased is primed for fine-tuning on designated downstream activities, employing labeled data tailored to the specific task. As previously outlined, fine-tuning entails integrating task-specific layers atop the pretrained BERT-Base Uncased model, aligning the entire architecture with the labeled data pertinent to the target objective. This alignment allows the model to adapt its cultivated representations to the distinct demands of the downstream task, amplifying its efficacy and adaptability. In summation, BERT's proficiency in capturing bidirectional context, combined with its extensive pretraining on large text corpora, positions it at the forefront of contemporary NLP performance. By decoding the intricate relationships between words and the contexts they inhabit, BERT synthesizes robust word embeddings, fostering precise predictions across a diverse array of NLP applications. The architecture of BERT consists of an encoder stack of transformer layers. The key components of BERT are as follows:

### I. Input Embeddings

BERT operates on sequences of words of varying lengths, accepting them as input. Within this structure, each word is characterized by a tripartite blend of token, segment, and position embeddings. Token embeddings are essential in encapsulating the specific meaning of individual words. Meanwhile, segment embeddings play a crucial role in differentiating between various sentences that might be present within the input. Position embeddings serve to pinpoint the unique location of each word within the overarching sequence. This combination of embeddings allows BERT to understand and analyze the textual information in a multifaceted manner, considering meaning, structure, and positional relationships.

### II. Transformer Encoder

The foundation of BERT's analysis of input embeddings lies in its multiple layers of transformer encoders. Each of these encoder layers is composed of two key components: a self-attention mechanism and a feed-forward neural network. Through the self-attention mechanism, BERT is empowered to recognize the contextual relationships between words by considering the entirety of words within the input sequence. Complementing this, the feed-forward neural network introduces non-linear transformations to the outputs derived from self-attention, further refining the understanding of the text.

**III. Pre-training and Fine-tuning**

BERT's training regimen is bifurcated into pre-training and fine-tuning stages. During the pre-training phase, BERT undergoes unsupervised learning on a copious volume of unlabeled textual data, employing objectives such as masked language modeling (MLM) and next sentence prediction (NSP). The MLM task prompts BERT to predict obscured words within the input, thereby fostering its ability to grasp bidirectional context. Simultaneously, the NSP task challenges BERT to ascertain whether two sentences naturally follow one another in the original text. Upon the completion of pre-training, BERT is adapted to specific downstream tasks through fine-tuning, utilizing data labeled for the particular task at hand.

**II. Output Layer**

Depending on the task, BERT's output layer may vary. For tasks like text classification, BERT typically uses a classification layer on top of the pooled output representation from the encoder stack. The pooled output represents the entire input sequence, which is then fed into the classification layer to generate predictions.

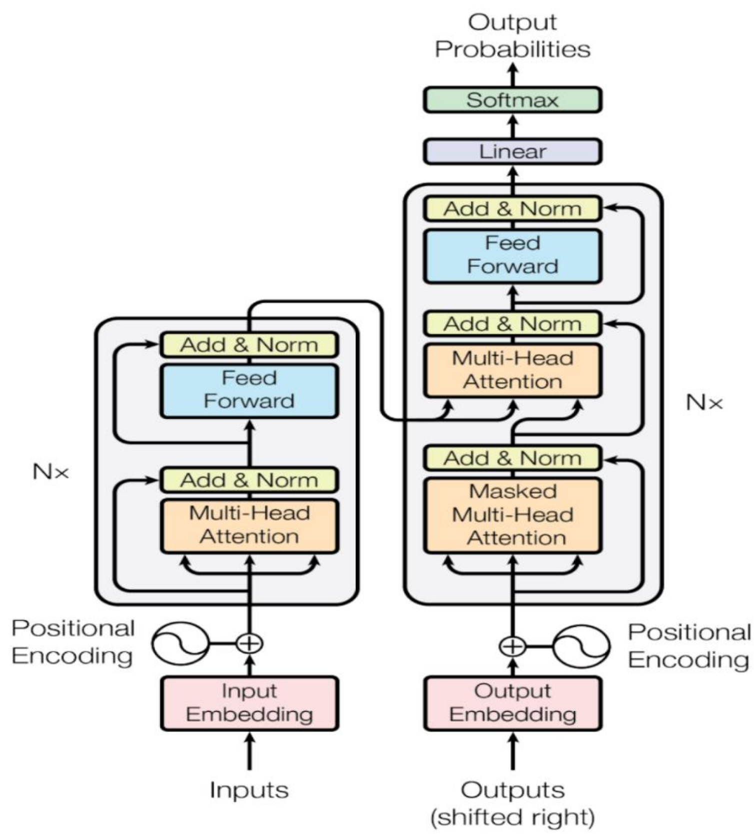


**Figure 7: BERT Architecture Adapted [77]**

We can say that BERT's architecture is designed to adeptly understand the contextual relationships between words and sentences, which contributes to its enhanced performance across diverse NLP tasks. In our research, we employed BERT (Bidirectional Encoder Representations from Transformers) as the fundamental model for executing personality prediction through natural language processing (NLP) methodologies. The formidable capability of BERT to create contextualized representations, coupled with its proficiency in extracting both semantic and syntactic insights from text, renders it an ideal choice for such endeavors. Here is an overview of how BERT is used in this research:

### 3.2.5. Importing required libraries and functions

This script begins by importing the necessary Python libraries and modules for data manipulation (pandas and numpy), deep learning (torch), Natural Language Processing (the transformers library that contains the BERT model and tokenizer), and other utilities (sklearn's metrics and model selection, matplotlib, seaborn for visualizations).

- **Preprocessing**

The BERT tokenizer is loaded, which enables tokenizing the input text into individual tokens that BERT understands. And the maximum sequence length is defined to ensure that the input fits within BERT's constraints.

### 3.2.6. Helper functions and class definitions

I. **`generate_and_save_graphs`:** This function is used to generate and save multiple graphs for loss, accuracy, AUC (Area Under the Curve), F1 Score, and confusion matrix for each label.

II. **`CustomDataset`:** This class is a subclass of torch's Dataset class. It defines how to load and preprocess the data. The preprocessing step involves converting text into numerical input representations (input_ids and attention_mask) that the BERT model can understand.

III. **`preprocess_text`:** This function is a helper function for the dataset class that tokenizes text, adds the special [CLS] and [SEP] tokens (required by BERT), converts the tokens into their corresponding IDs, and creates the attention mask.

### 3.2.7. Preparing the Dataset

The script first loads the BERT tokenizer and defines a maximum sequence length for the text. It then creates a `CustomDataset` object from the input dataframe, which contains text data and corresponding labels. The dataset is then split into training and test subsets using StratifiedShuffleSplit to ensure the training and test datasets have approximately the same distribution of labels.

### 3.2.8. Preparing the Model

The BERT model for sequence classification is loaded from the transformer’s library using `BertForSequenceClassification.from_pretrained () `. The number of unique labels in the dataset is determined and passed to the model. The model is moved to the GPU if available. The number of unique labels is determined based on the dataset's personality traits.

### 3.2.9. Training the Model

The AdamW optimizer is used for training the model. The model is trained for a specified number of epochs. During each epoch, the model forward propagates each batch of input data, computes the loss, backpropagates the gradients, and updates the model's weights. During the training, the script also keeps track of the total training loss and accuracy. The training uses a technique called gradient scaling to prevent gradients from getting too small during backpropagation, which is important for maintaining numerical stability and for effectively training models with float16 weights. Here is the breakdown of training steps:

- The model is trained using the training dataset and the AdamW optimizer.
- The training is performed over a specified number of epochs.
- Within each epoch, the training data is processed in mini batches.
- The model's performance is evaluated on the validation set at the end of each epoch.

### 3.2.10. Evaluating the Model

At the end of each training epoch, the model's performance is evaluated on the validation set. The script calculates metrics such as accuracy, AUC, and F1 score. If the validation F1 score improves, the current model state is saved.

### 3.2.11.Testing

As we are performing testing on the SOP's dataset using the provided code. First, we define the \`evaluate1\` function to test the model. Then check if a CUDA-capable GPU is available, and if so, we can move the model to the GPU (\`cuda\`). Otherwise, use the CPU for computations. But we are using GPU A100.

In the \`evaluate1\` function, the model is set to evaluation mode using \`model.eval()\`. It initializes variables to keep track of the validation loss, predictions, and true values. Next, to load the model the state dictionary from a file named "best_model.pt" (presumably saved during training) and move the model to the appropriate device (GPU or CPU) for testing.

To conduct the testing, we use the `evaluate1` function with the `new_model` and `test_dataloader` as input parameters. The function returns the predicted labels and true labels for the test data. Then, flatten the predicted and true labels to be used in further evaluation. The `classification_report`, `accuracy_score`, and `confusion_matrix` functions are being generated from scikit-learn to compute and display the classification report, accuracy, and confusion matrix, respectively. And the model performance is done by using accuracy metric. Here(Figure 8) is the flowchart of proposed methodology.

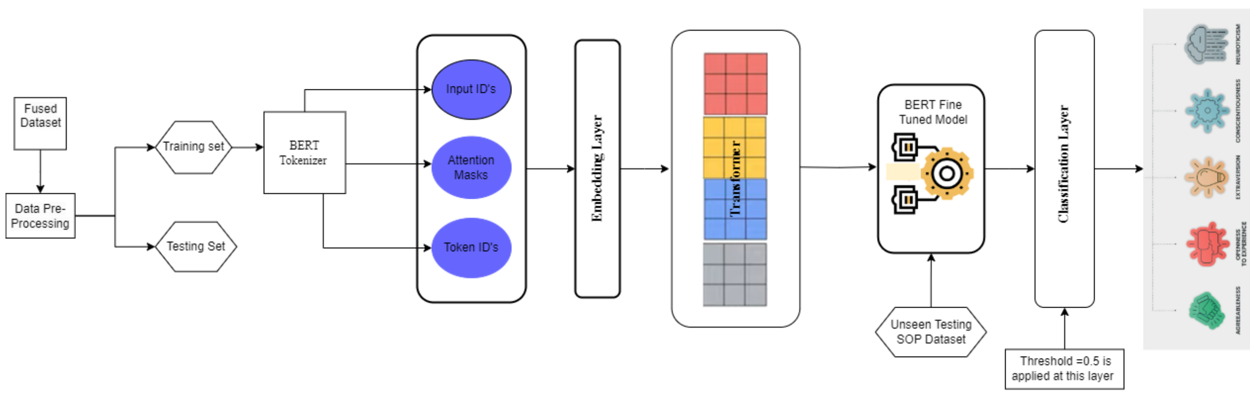


**Figure 8: Proposed BERT Model Flow chart**

I. **Input Layer:** The input to the BERT model typically consists of token embeddings, segment embeddings, and positional embeddings. The tokens are the individual words in the input text, the segment embeddings indicate whether a token belongs to the first sentence or the second in tasks that require sentence pairs, and the positional embeddings provide information about the position of a token in the sequence.

II. **Input Data**: The process begins with input data, which is typically a text corpus. This could be any form of text data, such as social media posts, news articles, books, or transcripts.

III. **Preprocessing**: The input data is then preprocessed. This step involves cleaning the text (removing punctuation, special characters, etc.), tokenization (breaking down the text into individual words or tokens) and converting these tokens into input vectors. For BERT, we also add special tokens like **[CLS]** (classification token) at the beginning of each sentence, and **[SEP]** (separator token) at the end or between two sentences in case of pair-input tasks.

IV. **BERT Embeddings**: The preprocessed data is then passed through the BERT model to get embeddings. BERT generates contextual word embeddings, meaning the same word can have different embeddings based on its context within a sentence. These are ([CLS] or Classification token) as a summary of the entire text these embeddings are processed by the downstream layers to produce the final output.

V. **Pretrained Transformer Layers:** The BERT model consists of multiple Transformer layers. These layers process the input embeddings and generate a new set of embeddings that contain a rich, contextual understanding of the input text.

VI. **Fine-tuning**: After generating embeddings, the BERT model is fine-tuned on a specific task. This could be text classification, named entity recognition, question answering, etc. During fine-tuning, the model's parameters are slightly adjusted to better perform the specific task.

VII. **Model Training**: The fine-tuned model is then trained on the task-specific training data. During training, the model learns to make predictions by adjusting its internal parameters to minimize the difference between its predictions and the actual values (the "loss").

VIII. **Classification Layer:** The [CLS] embedding is passed through a classification layer to produce the final output. This layer is typically a fully connected (dense) layer that projects the [CLS] embedding into the number of classes in the classification task. Where we have made the amendments by add the threshold =0.5 that is responsible for making the decision on the bases of predicted probabilities of the model that either the traits will be assigned 1 or 0.

IX. **Evaluation**: Once the model is trained and pass through the classification layer, it is evaluated on unseen data (the validation or test set) to assess its performance. Evaluation metrics depend on the specific task, but for classification tasks they often include accuracy, precision, recall, and F1 score.

X. **Prediction**: Once the model has been trained and evaluated, it can be used to make predictions on new, unseen data. The output depends on the task - for a classification task, the output would be the predicted class labels.

XI. **BERT Model:** This is the pre-trained BERT model, which has been trained on a large corpus of text. The BERT model consists of a stack of Transformer layers that encode the input text into a series of contextual embeddings.

## 3.3 Accuracy Calculation

We used threshold `thr` of 0.5 to convert the model's predicted probabilities into binary labels for multi-label classification. After making predictions using the BERT model, the predicted probabilities for each label are obtained. These probabilities indicate the likelihood of each label being present in the input text. To convert the probabilities into binary labels, the threshold of 0.5 is applied. If the predicted probability for a label is greater than 0.5, the label is considered as present (1); otherwise, it is considered absent (0). For example, consider a label with a predicted probability of 0.6. Since 0.6 is greater than 0.5, the label is assigned the value of 1, indicating its presence. On the other hand, if the predicted probability for another label is 0.3, which is less than 0.5, the label is assigned the value of 0, indicating its absence.

This threshold of 0.5 is a common choice for binary classification tasks, where the goal is to classify each label into two classes: positive (1) or negative (0). However, in some cases, depending on the problem and the data distribution, a different threshold may be more appropriate. Using a threshold of 0.5 simplifies the process of converting probabilities to binary labels, but it may not always be the optimal choice. In some cases, we can tune the threshold based on a specific requirement and the characteristics of the dataset.

It's vital to recognize that the selection of the threshold can significantly influence the model's performance. This choice must be made in alignment with the particular needs of the problem, considering the balance between precision and recall. Fine-tuning the threshold has the potential to shift the equilibrium between false positives and false negatives in the predictions, thereby affecting the overall accuracy and effectiveness of the model. Moreover, the thresholding operation to convert predicted probabilities into binary labels is not specifically performed in any layer of the BERT model itself. Instead, it is applied as a post-processing step after obtaining the model's predicted probabilities. In our implementation, we focused on the output layer. BERT, being a transformer-based model, is composed of multiple layers of self-attention and feed-forward neural networks. At the core of the BERT model is the output layer, a specialized classification layer that translates the final hidden states of the input tokens into the desired quantity of output labels. During the inference phase, the BERT model processes tokenized input text and generates a collection of predicted probabilities corresponding to each label. These probabilities stem from the output layer within the BERT model, where the logits (unprocessed scores) for each label are determined based on the concluding hidden states of the input tokens. This approach leverages the robust architecture of BERT to facilitate accurate and nuanced classification.

After obtaining the predicted probabilities, the thresholding operation is performed as a separate step outside the BERT model. This is where the threshold of 0.5 is applied to convert the probabilities into binary labels (0 or 1) for each label.

***Predicted labels = (predicted probabilities > threshold) *float ()*** ………... (3.1)

Here, `predicted_probs` is the tensor containing the predicted probabilities for each label, and `thr` is the threshold value of 0.5. The comparison `(predicted_probs > thr)` produces a Boolean tensor where each element indicates whether the predicted probability is greater than the threshold (True) or not (False). By using `. float()`, the Boolean tensor is converted to a tensor of 1s and 0s, representing the binary labels. After applying the threshold of 0.5 to convert the predicted probabilities into binary labels, the accuracy for the whole dataset is calculated by following these steps:

1. For each sample in the dataset, the predicted labels (after applying the threshold) are compared to the true labels. This is done separately for each label.

2. The number of correct predictions is calculated by summing up the cases where the predicted label matches the true label for each label.

3. The total number of labels (across all samples) is determined.

4. The accuracy is calculated by dividing the number of correct predictions by the total number of labels.

$$\textbf{\textit{Accuracy = (Number of correct predictions) / (Total number of labels)}} \quad ....... (3.2)$$

By considering each label separately and averaging the accuracy across all labels, the accuracy provides an overall measure of how well the model performs across multiple labels in a multilabel classification task.

In summary, the thresholding operation is applied externally to the BERT model as a post-processing step on the predicted probabilities to obtain the final binary labels for each label in the multi-label classification task.

### 3.3.1. Advantages of using BERT

The main advantage of using BERT is that it captures complex language understanding due to its Transformer architecture and bi-directional context-based word representations. It is particularly effective when there is a limited amount of labeled data for the task at hand, as it leverages knowledge learned from a large corpus of unlabeled data. It is worth mentioning that the script uses the sigmoid function for the final layer activation to make the model suitable for multilabel classification, as opposed to softmax which is used for multiclass classification. In multilabel classification, each label is treated independently, and thus an example could belong to multiple classes. By leveraging BERT's contextualized representations and fine-tuning the model on the specific personality prediction task, the code enables accurate prediction of personality traits based on the input essays. BERT's ability to capture the nuances of language and context contributes to the effectiveness of the 'redictions and enhances the overall performance of the personality prediction model.

BERT-Base Uncased was chosen for the research due to its state-of-the-art performance on various NLP tasks, surpassing previous models and achieving remarkable results. The model's ability to capture contextual relationships between words and understand natural language makes it a powerful tool for language understanding tasks. Additionally, the transfer learning capabilities of BERT-Base Uncased enable fine-tuning on specific datasets with limited labeled data, reducing the need for extensive task-specific training from scratch. Its versatility and applicability to a wide range of NLP tasks make it an asset for research purposes, aiming to improve accuracy and performance in the specific task at hand.

In summary, the ***"BERTforSequenceClassification*"** model from the Hugging Face model repository, with its self-supervised contrastive learning approach, proved to be an invaluable asset for our research. Its ability to generate high-quality contextual representations and its performance in multilabel classification tasks were exceptional. The model's pre-trained knowledge and fine-tuning on our dataset allowed it to accurately predict multiple labels for each input text sample, enabling us to obtain meaningful insights from our data.

## 3.4. RoBERTa

RoBERTa, an acronym for Robustly Optimized BERT Pretraining Approach, is an evolved form of the BERT (Bidirectional Encoder Representations from Transformers) model. It was crafted with the goal of enhancing the efficacy of pretrained models in the domain of natural language understanding (NLU). Launched by the Facebook AI team in 2019, RoBERTa takes the foundation laid by BERT and infuses it with additional training methodologies and refinements to amplify its potential [69]. The inception of RoBERTa was motivated by the quest for advanced pretraining techniques for language comprehension tasks. Although BERT had marked substantial achievements, the researchers identified room for further advancements. RoBERTa was designed to rectify certain constraints and experiment with diverse training approaches to optimize the performance of the pretraining model.

### 3.4.1. RoBERTa-Base

RoBERTa-Base refers to the base variant of the RoBERTa (Robustly Optimized BERT pretraining Approach) model. It serves as the foundational model architecture for RoBERTa, providing strong language representation capabilities for a wide range of natural language understanding (NLU) tasks. RoBERTa-Base is pretrained on a massive corpus of unlabeled text data and can be fine-tuned on specific downstream tasks [70]. The architecture of RoBERTa-Base is based on the

Transformer model, similar to BERT and other models in the BERT family. Here are the key components of the RoBERTa-Base architecture:

### I. Transformer Encoder

- RoBERTa-Base consists of a stack of transformer encoder layers. Each layer comprises multi-head self-attention and feed-forward neural network sub-layers as shown in Figure 9.
- The self-attention mechanism allows the model to capture contextual dependencies between words and learn meaningful representations.
- Feed-forward neural network sub-layers facilitate non-linear transformations and capture higher-level features.

### II. Pretraining Techniques

- RoBERTa-Base utilizes unsupervised pretraining techniques to learn general language representations from a large corpus of text data.
- It employs the masked language modeling (MLM) task, where a percentage of input tokens are randomly masked, and the model predicts the original tokens based on the surrounding context.
- By training on massive amounts of data, RoBERTa-Base learns to understand the syntactic and semantic relationships between words, enabling it to capture rich linguistic information.

### III. Fine-Tuning

- After pretraining, RoBERTa-Base can be fine-tuned on specific downstream tasks such as text classification, named entity recognition, sentiment analysis, and more.
- During fine-tuning, task-specific layers are added on top of the pretrained RoBERTa-Base model, and the entire architecture is trained on labeled task-specific data.
- Fine-tuning allows RoBERTa-Base to adapt its learned representations to the specific requirements of the downstream task, enhancing its performance and generalization.

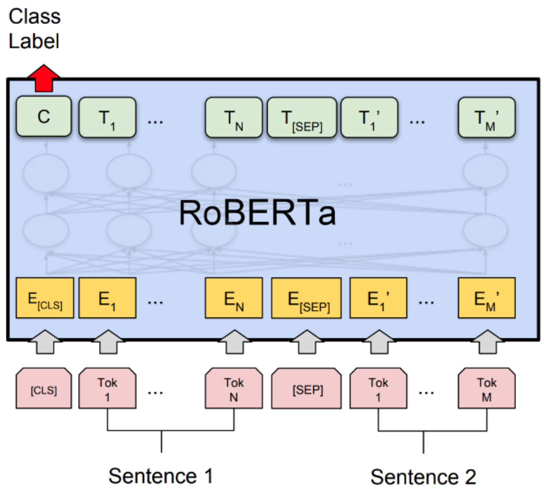


**Figure 9: Roberta-Base Architecture [75]**

### IV. Advantages and Impact

RoBERTa-Base has made significant contributions to the field of natural language processing (NLP) by advancing the state-of-the-art in various NLU benchmarks and tasks. It offers several advantages one of them is: Improved Performance: RoBERTa-Base achieves superior performance compared to earlier models like BERT generically, thanks to its modifications in training strategies and larger-scale pretraining but in our case RoBERTa didn't perform best as BERT.

## 3.5. DeBERTa-Base

DeBERTa-Base is an advanced language model that builds upon the success of BERT and RoBERTa by incorporating novel enhancements to improve its performance in natural language processing tasks. DeBERTa, short for Decoding-enhanced BERT with Disentangled Attention, was introduced by Microsoft Research Asia in 2020. It aimed to address the limitations of previous models, such as BERT and RoBERTa, by refining their attention mechanism. DeBERTa-Base is the base variant of the DeBERTa family and serves as a powerful foundation for various natural language processing tasks. The architecture of DeBERTa-Base follows the transformer-based design, similar to BERT and other models in its lineage. It consists of multiple transformer layers that process the input text in a hierarchical manner, capturing both local and global dependencies.

However, DeBERTa incorporates novel modifications in its attention mechanism to enhance its performance [71].

### I. Disentangled Attention

The key innovation in DeBERTa-Base is the introduction of disentangled attention, which aims to alleviate the limitations of standard self-attention mechanisms. In traditional attention, each token attends to all other tokens in a sequence, leading to quadratic complexity. DeBERTa introduces disentangled attention to reduce this complexity by allowing tokens to attend to only a subset of other tokens see the figure 3.5. This results in more efficient computation and improved modeling of long-range dependencies.

### II. Enhanced Decoding

Another crucial aspect of DeBERTa-Base is enhanced decoding. It employs a two-step decoding process where the model first generates a draft representation of the output and then refines it using additional iterations. This approach enables the model to produce more accurate and contextually appropriate output representations.

## 3.5.1. Training Strategy

Similar to BERT, DeBERTa-Base is pretrained on large-scale unlabeled text corpora. The pretraining process involves two main objectives: masked language modeling and predicting the order of consecutive sentences. By training on a vast amount of data, DeBERTa-Base learns rich representations that capture the nuances of language and can be fine-tuned for various downstream tasks. DeBERTa-Base offers several advantages over previous models. Its disentangled attention mechanism allows for more efficient computation and improved modeling of long-range dependencies, resulting in enhanced performance on complex natural language processing tasks. DeBERTa-Base has been successfully applied to various tasks, including text classification, question answering, and natural language understanding, showcasing its versatility and effectiveness.

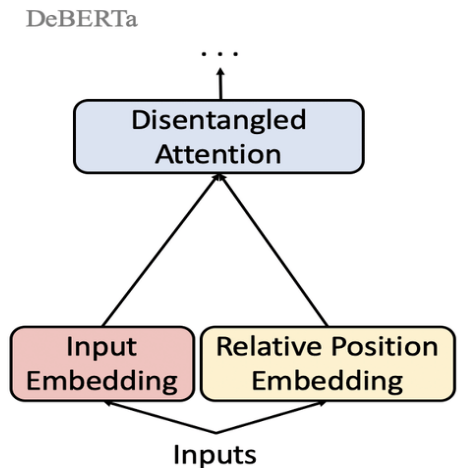


Figure 3.5: DeBERTa Base Architecture [74]

In our research, we incorporated DeBERTa-Base as a key component to enhance the performance of our text classification task. By leveraging the disentangled attention mechanism and enhanced decoding capabilities of DeBERTa-Base, we aimed to improve the model's ability to capture complex relationships and context in the input text. The use of DeBERTa-Base allowed us to achieve more accurate and contextually appropriate predictions, leading to better results in our research objectives. Moreover, DeBERTa-Base represents a significant advancement in language modeling, offering improved attention mechanisms and decoding strategies. Its versatility and enhanced performance make it a valuable tool for a wide range of natural language processing tasks, and its integration in our research proved instrumental in achieving high-quality results.

## 3.5.2. Setting Up Optimizer and Scheduler

**I. Optimizer - AdamW**

- The optimizer used is \`AdamW\`. This is a variation of the Adam optimizer that corrects weight decay. It is particularly well-suited for training transformer models like BERT.
- The learning rate (\`lr\`) is set to $6\times10^{-5}$. This value is a common choice when fine-tuning BERT models, ensuring gradual model updates to prevent dramatic shifts in pre-trained weights.

### II. Scheduler - ReduceLROnPlateau

- The learning rate scheduler employed is `ReduceLROnPlateau`. This scheduler reduces the learning rate when a metric has stopped improving.
- The `mode` is set to 'max', meaning the scheduler will be observing a metric that should be maximized. If this metric does not increase for several epochs defined by `patience`, the learning rate will be reduced.
- The `factor` is set to 0.1. When the metric stops improving, the new learning rate will be the current learning rate multiplied by this factor.
- The `patience` is set to 3, which means the learning rate will be reduced if the observed metric does not improve for 3 consecutive epochs.

Furthermore, an optimizer and scheduler tailored for fine-tuning a BERT model. The AdamW optimizer is designed to update the model parameters effectively, and the ReduceLROnPlateau scheduler ensures that the learning rate is adjusted dynamically if the model's performance plateaus during training. This combination aims to achieve optimal model performance while avoiding common pitfalls like overshooting or getting stuck in local minima.

## 3.6. SimCSE Supervised BERT-base-uncased

The acronym "sup-simcse-bert-base-uncased" stands for "Supervised Contrastive Learning for Sentence Embeddings with BERT base uncased." Let us break down the components of the acronym:

- "Supervised": It indicates that the model is trained using a supervised learning approach. In supervised learning, the model is provided with labeled data, where the inputs and corresponding outputs are known, allowing the model to learn from this labeled information.

I. **"Contrastive Learning":** Contrastive learning is a technique that aims to learn representations by contrasting similar and dissimilar examples. In the case of "sup-simcse," the model learns to distinguish between pairs of sentences based on their similarity or dissimilarity, enabling it to capture semantic relationships and contextual information.

II. **"Sentence Embeddings":** Sentence embeddings refer to the representation of sentences in a fixed-dimensional vector space. These embeddings capture the semantic and contextual information of the sentences, enabling the model to understand the meaning and relationships between sentences.

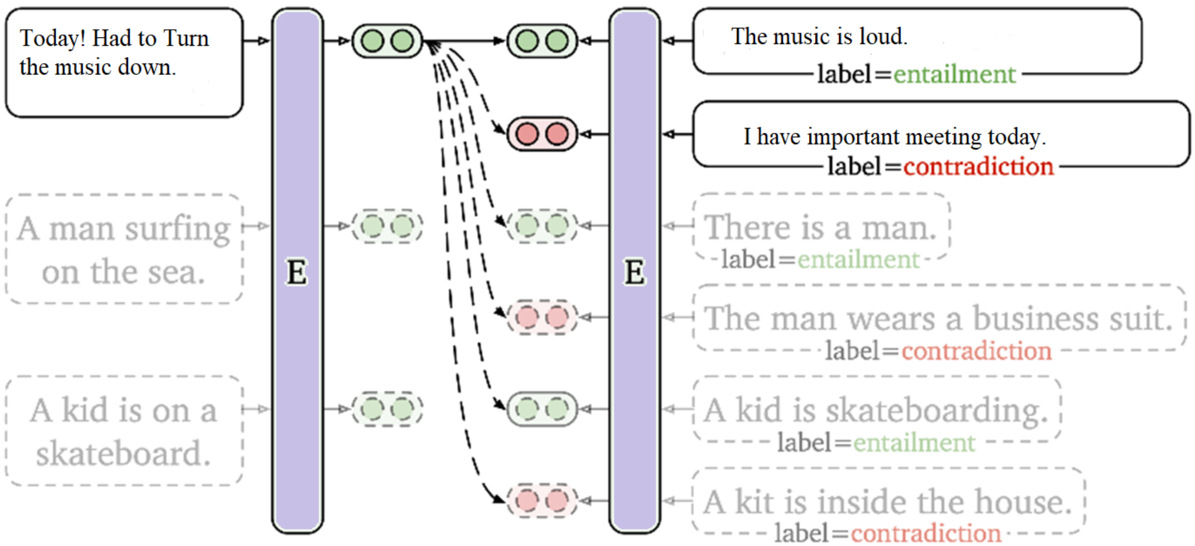


**Figure 10: Supervised SimCSE Architecture [76]**

As we discussed earlier "BERT base uncased": BERT (Bidirectional Encoder Representations from Transformers) is the underlying architecture used in the model. "base" refers to the base version of the BERT model, which has a specific configuration and architecture. "uncased" indicates that the model operates on text that has been lowercased, treating uppercase and lowercase letters as equivalent. We can say, "sup-simcse-bert-base-uncased" represents a fine-tuned version of the BERT model that incorporates supervised contrastive learning techniques for sentence embeddings. It combines the power of BERT's contextual representations with the ability to learn from labeled data, enabling accurate and context-aware predictions for various NLP tasks.

The history of the "sup-simcse-bert-base-uncased" model can be traced back to the development of the BERT (Bidirectional Encoder Representations from Transformers) model. BERT, introduced by researchers at Google in 2018, revolutionized the field of natural language processing (NLP) with its ability to capture contextual information and semantic relationships in textual data.The original BERT model was trained using a masked language modeling objective and a next sentence prediction objective on a massive amount of unlabeled text from the internet. This allowed the model to learn powerful representations of words and sentences, capturing the contextual information present in the data. Building upon the success of BERT, researchers from Princeton University's Natural Language Processing (NLP) group introduced the "sup-simcse" (Supervised Contrastive Learning for Sentence Embeddings) framework. The sup-simcse framework focuses on leveraging self-supervised contrastive learning techniques for improving the performance of various NLP tasks [72].

The "sup-simcse-bert-base-uncased" model is a variant of the BERT architecture that has been specifically fine-tuned using the sup-simcse framework. This fine-tuning process involves training the model on a specific supervised task, such as multilabel classification, to adapt it to the task at hand. By fine-tuning BERT with the sup-simcse framework, the model can learn to generate more accurate and task-specific contextual representations. The sup-simcse framework utilizes the concept of contrastive learning, where the model learns to distinguish between similar and dissimilar pairs of sentences. This enables the model to capture the subtle semantic relationships between sentences, enhancing its understanding of the contextual information in text data. We utilized the "sup-simcse-bert-base-uncased" model, which is available from the Hugging Face model repository via the link "princeton-nlp/sup-simcse-bert-base-uncased", for our research project. This model is based on the BERT (Bidirectional Encoder Representations from Transformers) architecture and has been specifically tailored for self-supervised contrastive learning tasks.

The "sup-simcse-bert-base-uncased" model demonstrates exceptional performance in capturing the contextual information and semantic relationships within textual data. By training on large amounts of unlabeled text, the model learns to generate high-quality contextual representations. These representations can then be used to tackle various downstream natural language processing tasks, such as text classification, named entity recognition, and sentiment analysis. In our research, we leveraged the "sup-simcse-bert-base-uncased" model to perform multilabel classification on our dataset. This task involves predicting multiple labels for each input text sample. By utilizing the model's comprehensive language understanding capabilities, it was able to capture the intricate relationships and nuances present in the textual data.

During the training process, we fine-tuned the "sup-simcse-bert-base-uncased" model on our dataset. This involved adapting the model's parameters to align with our specific multilabel classification task. The model was optimized using the AdamW optimizer, which helps update the model's weights based on the gradients computed during the training process. Additionally, we employed a learning rate scheduler to dynamically adjust the learning rate, enhancing the model's convergence and overall performance. To evaluate the performance of the "sup-simcse-bert-base-uncased" model on our dataset, we employed several metrics. These metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). These measurements provide insights into the model's ability to correctly classify and predict the multiple labels associated with each input text sample. Based on our evaluation, the "sup-simcse-bert-base-uncased" model demonstrated remarkable performance on our dataset. Its accurate and reliable predictions, as indicated by high accuracy, showcase its effectiveness in handling multilabel classification tasks. The model's strong language understanding capabilities, derived from its pre-training on large-scale text data, enabled it to capture the intricate relationships and subtle nuances present in our dataset.

## 3.7. Chapter Summary

In this section, we discuss the utilization of three datasets for personality prediction, combining techniques to enhance the model's performance, data pre-processing, and dataset splitting. Additionally, we delve into the BERT model's key components, including word embeddings, pre-training, fine-tuning, and the attention mechanism and proposed or methodology of the research. The discussion also touches on RoBERTa-Base, its architecture, and optimization setup. Moreover, we explore a supervised contrastive learning approach for sentence embeddings using BERT base uncased.

# CHAPTER 4: EXPERIMENTS AND RESULTS

Based on the experimental results, it is evident that the choice of learning rate and the number of epochs significantly impact the performance of the models. Let us analyze the effects of learning rate and epochs on the accuracy of the models:

## 4.1. Learning Rate

The learning rate is a crucial hyperparameter that controls the step size of the optimization algorithm during training. It determines how much the model's parameters are updated in each iteration. From the results, we can observe the following trends:

- For BERTBase uncased, a learning rate of 6e-5 with 7 epochs achieved the highest testing accuracy of 84.4%. This indicates that a relatively high learning rate combined with enough epochs is effective for this model.

- Lower learning rates (e.g., 1e-5) resulted in slower convergence, and the models struggled to reach higher accuracies even with more epochs. For example, BERTBase uncased with a learning rate of 1e-5 and 15 epochs achieved only 52% testing accuracy.

- Very high learning rates (e.g., 3e-4) negatively impacted the performance of the models, leading to suboptimal accuracy (e.g., 44% for BERTBase uncased with a learning rate of 3e-4).

In summary, an appropriate learning rate is crucial for efficient convergence and achieving better performance. A learning rate that is too high or too low can lead to slow convergence or getting stuck in suboptimal local minima.

## 4.2. Epochs

Epochs represent the number of times the entire training dataset is passed through the model during training. They control the number of iterations the model undergoes to learn from the data. The results show the following trends: For most models, increasing the number of epochs initially improves the training accuracy, indicating better model learning. However, there is a point of diminishing returns, after which further increasing the number of epochs may lead to overfitting. Overfitting occurs when the model becomes too specialized to the training data and performs poorly on unseen data (testing accuracy drops).

- On the other hand, some models showed signs of underfitting, where increasing the number of epochs improved testing accuracy up to a certain point, but beyond that, accuracy plateaued or decreased (e.g., SimCSE-Bert-large-supervised with 30 epochs) as shown in table 4.1:

Moreover, finding the optimal number of epochs is essential to achieve a balance between training and testing accuracy and prevent overfitting or underfitting. Techniques such as early stopping can be employed to halt training when the model performance on the validation set starts deteriorating. The choice of learning rate and epochs plays a vital role in model training and generalization. It requires experimentation and careful tuning to find the best hyperparameters that lead to the highest testing accuracy and avoid issues like overfitting and underfitting.

**Table 4.1: Summary of all Results using Combined (Essays+myPersonality) training dataset and testing on SOP Dataset**

| **Models** | **Epochs** | **LR** | **Batch size** | **Maximum length** | **Training Accuracy** | **Testing Accuracy** |
|---|---|---|---|---|---|---|
| $\text{BERT}_{\text{Base uncased}}$ | 20 | 2e-5 | 32 | 256 | 70% | 64.7% |
| $\text{BERT}_{\text{Base uncased}}$ | **7** | **6e-5** | **48** | **256** | **90%** | **84.4%** |
| $\text{BERT}_{\text{Base uncased}}$ | 20 | 3e-5 | 32 | 256 | 70% | 58.5% |
| $\text{BERT}_{\text{Base uncased}}$ | 20 | 3e-4 | 64 | 256 | 58% | 44.3% |
| $\text{BERT}_{\text{Base uncased}}$ | **20** | **1e-5** | **48** | **256** | **58%** | **52.4%** |
| $\text{BERT}_{\text{Base uncased}}$ | 15 | 1e-5 | 64 | 256 | 52% | 41.8% |
| SimCSE-Bert-large-supervised | 20 | 5e-5 | 48 | 300 | 79.9% | **60.9%** |
| SimCSE-Bert-large-supervised | 10 | 2e-5 | 24 | 256 | 56% | 48.7% |
| SimCSE-Bert-large-supervised | 20 | 5e-5 | 16 | 256 | 72.6% | 66.1% |

| | | | | | | |
|---|---|---|---|---|---|---|
| SimCSE-Bert-large-supervised | 30 | 3e-5 | 16 | 256 | 64% | 58.7% |
| SimCSE-Bert-large-supervised | 30 | 5e-5 | 16 | 256 | 82% | **63.3%** |
| SimCSE-Bert-large-supervised | 30 | 6e-5 | 48 | 256 | 61% | 50.7% |
| SimCSE-Bert-large-supervised | 35 | 5e-5 | 16 | 256 | 84% | 68.4% |
| RoBERTa $_{\text{Base}}$ | 20 | 2e-5 | 32 | 256 | 70% | **68.1%** |
| RoBERTa $_{\text{Base}}$ | 12 | 6e-5 | 48 | 256 | 58% | 43.2% |
| RoBERTa $_{\text{Base}}$ | 15 | 1e-4 | 64 | 256 | 56% | 41.4% |
| DeBerta | 30 | 2e-5 | 32 | 256 | 70% | **60.8%** |
| DeBerta | 20 | 1e-5 | 32 | 300 | 61% | 48.9% |
| DeBerta | 15 | 4e-5 | 64 | 256 | 69% | 51.5% |
| DeBerta | 10 | 2e-5 | 24 | 256 | 90% | 58.6% |

Table 4.1 represents the experimental results for different models with various hyperparameter settings. Each row corresponds to a specific model, and the columns provide information about the model's performance on the classification task. Here's what each column represents:

I. Models: This column lists the names of the different models that were evaluated in the experiment. The models mentioned are "BERTBase uncased," "SimCSE-Bert-large-supervised," "RoBERTa Base," and "DeBerta."

II. Epochs: The number of epochs refers to the number of times the entire training dataset was passed through the model during training. It indicates how many iterations the model underwent to learn from the data.

III. LR (Learning Rate): The learning rate is a hyperparameter that controls the step size of the optimization algorithm during training. It determines how much the model's parameters are updated in each iteration.

IV. Batch size: The batch size indicates the number of data samples fed into the model at each iteration during training. It is an important hyperparameter that affects training speed and memory consumption.

V. Maximum length: This column specifies the maximum sequence length used for tokenizing the input text data. Tokens beyond this length are truncated or padded to fit the specified length.

VI. Training Accuracy: Training accuracy is the accuracy of the model on the training dataset. It indicates how well the model performs on the data it was trained on.

VII. Testing Accuracy: Testing accuracy represents the accuracy of the model on a separate unseen test dataset. It gives an estimate of how well the model generalizes to new, unseen data.

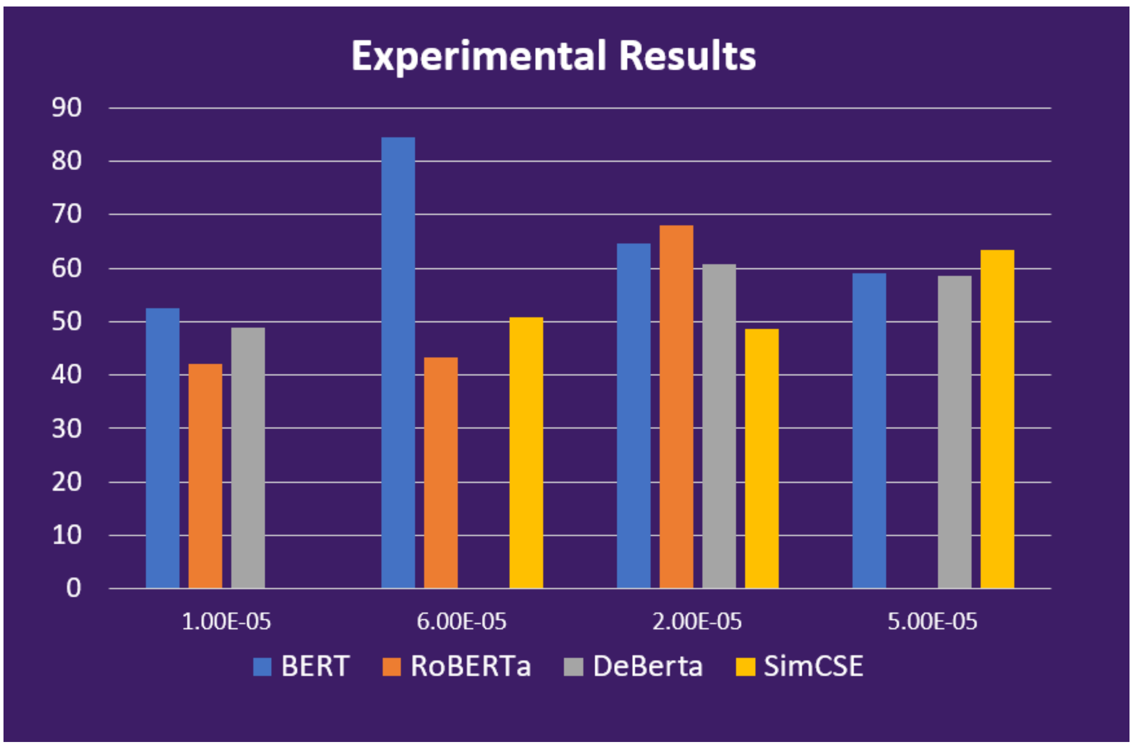


**Figure 11: Graphical representation of some of the best results with all models**

Figure 11 shows the collective graphical representation for all the best results now let's analyze the table 4.1 and the trends in the experimental results:

## 4.3. Models' Performance

"BERTBase uncased" achieved the highest testing accuracy of 84.4% when trained for 7 epochs with a learning rate of 6e-5. "SimCSE-Bert-large-supervised" had the highest testing accuracy of

68.4% with a learning rate of 5e-5 and 16 epochs."RoBERTa Base" achieved the highest testing accuracy of 68.1% when trained for 20 epochs with a learning rate of 2e-5. "DeBerta" had the highest testing accuracy of 60.8% when trained for 30 epochs with a learning rate of 2e-5.

### 4.4. Effect of Epochs and Learning Rate

For some models, increasing the number of epochs generally improved the training accuracy but did not always lead to a corresponding increase in testing accuracy. For example, "BERTBase cased" achieved 90% training accuracy with 10 epochs, but testing accuracy is 84.4%. As we can see in the table 4.1 the choice of learning rate also played a significant role in the model's performance. Very high learning rates (e.g., 3e-4) negatively impacted performance, while lower learning rates (e.g., 1e-5) and very high learning rates (e.g., 3e-5) also resulted in reduced testing accuracy for some models.

### 4.5. Model Comparison

- Among the mentioned models, "BERTBase uncased" and " SimCSE-Bert-large-supervised" generally performed better than " RoBERTa Base " and "DeBerta" in terms of testing accuracy. However, it is worth noting that performance can vary based on the dataset and task.

In summary, the table provides insights into the performance of different models with varying hyperparameter settings. It highlights the importance of finding the right combination of hyperparameters, such as learning rate and epochs, to achieve optimal model performance and avoid overfitting or underfitting. Additionally, it emphasizes the significance of using a separate test dataset to evaluate the model's generalization to new, unseen data.

**Table 4.2: Effect of Batch size with BERT on the accuracy**

| Batch Size | Training Loss | Validation Loss | Train Accuracy | Test Accuracy |
|---|---|---|---|---|
| 32 | 0.015 | 1.7 | 70% | 60% |
| 48 | 0.02 | 0.019 | 90% | **84.4%** |
| 64 | 0.047 | 0.61 | 58% | 44% |

Table 4.2 illustrates the effects of varying the batch size on the training and validation of a BERT model. Let us analyze the performance for each batch size:

I. **. Batch Size 32:** With a batch size of 32, the model has a training loss of 0.015 and a validation loss of 1.7. The training and testing accuracies are 70% and 60%, respectively.

II. **. Batch Size 48:** When the batch size is increased to 48, the training loss increases slightly to 0.02, but the validation loss significantly decreases to 0.019. The training accuracy jumps to 90%, and the testing accuracy also improves significantly to 84.4%. This suggests that the model is better able to generalize to unseen data with a larger batch size, at least up to a batch size of 48. However, the high training accuracy could indicate some degree of overfitting, even though the model is still performing well on the testing data.

III. **. Batch Size 64:** With a batch size of 64, both the training loss (0.047) and the validation loss (0.61) are higher compared to the smaller batch sizes. The training and testing accuracies also decrease to 58% and 44% respectively. This suggests that a batch size of 64 is too large for this particular dataset and model configuration, as the model's performance has worsened.

## 4.6. Results with BERT

This table 4.3 provides the performance metrics of a BERT model on a text classification task, where the classes correspond to different traits. Each row in the table 6 corresponds to a different class (or trait), and the last row presents the average performance over all the classes. The final line gives the overall accuracy of the model. Here is a breakdown of each column:

**1. Traits/Classes**: These are the different categories or labels that the model is trying to predict. It appears to be a personality trait prediction task based on the labels (`cEXT`, `cNEU`, `cAGR`, `cCON`, `cOPN`), which correspond to the Big Five personality traits: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience.

**2. Precision:** Precision is the ratio of true positive predictions (correctly predicted as a certain class) to all positive predictions (both true positives and false positives). A higher precision means fewer false positives.

**3. Recall:** Recall (or sensitivity) is the ratio of true positive predictions to all actual positives (both true positives and false negatives). A higher recall means fewer false negatives.

**4. F1-Score:** The F1 score is the harmonic mean of precision and recall. It provides a single metric that balances both precision and recall. A higher F1 score indicates a more accurate and robust model.

**5. Support:** This is the number of actual occurrences of the class in the dataset. For example, there are 173 instances of the `cEXT` trait in the dataset.

The "Results" row provides the weighted averages of precision, recall, and F1-score, where the weights are the support values. This provides an overall measure of the model's performance, taking into account the imbalance in class distribution.

The "Overall Accuracy" is the ratio of correct predictions to total predictions made by the model, expressed as a percentage. In this case, the model's overall accuracy is 84.4%, which means it made the correct prediction for 84.4% of the instances in the dataset.

**Table 4.3: BERT Evaluation Results**

| Traits/Classes | Precision | Recall | F1- Score | Support |
|---|---|---|---|---|
| **cEXT** | 0.97 | 0.81 | 0.88 | 173 |
| **cNEU** | 0.76 | 0.90 | 0.82 | 77 |
| **cAGR** | 0.69 | 0.88 | 0.77 | 40 |
| **cCON** | 0.71 | 1.00 | 0.83 | 10 |
| **cOPN** | 0.86 | 0.75 | 0.80 | 8 |
| **Results** | 0.798 | 0.868 | 0.82 | 61.6 |
| **Overall Accuracy** | 84.4% | | | |

Looking at this table 4.3 , it appears that the model performs fairly well on most traits, although there is some variability. The model performs exceptionally well on `cEXT` and `cCON`, with high precision, recall, and F1 scores. On the other hand, the model appears to struggle more with `cAGR` and `cNEU`, as reflected by their relatively lower precision and F1 scores.

With all these experiments we trained our model on essay dataset explicitly and then tested on SOP which led to the 82.7% of accuracy which is the best performance from the previous studies as S. Kulsoom et al., [64] achieved 67% accuracy on SOP dataset Table 4.4 shows the results.

**Table 4.4: Experimental Results with Essay as training and SOP as test dataset**

| Traits/Classes | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| **cEXT** | 0.94 | 0.82 | 0.87 | 166 |
| **cNEU** | 0.78 | 0.87 | 0.82 | 82 |
| **cAGR** | 0.67 | 0.74 | 0.70 | 46 |
| **cCON** | 0.57 | 1.00 | 0.73 | 8 |
| **cOPN** | 0.86 | 1.00 | 0.92 | 6 |
| **Average** | 0.76 | 0.89 | 0.81 | 61.6 |
| **Overall Accuracy** | 82.7% | | | |

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. It gives us an idea of when the model predicts the class, how often it is correct.
- **Recall:** The ratio of correctly predicted positive observations to the actual positives. It tells us what proportion of the actual positives was identified correctly.
- **F1-Score:** The weighted average of Precision and Recall. It takes both false positives and false negatives into account.
- **Support:** The number of actual occurrences of the class in the specified dataset.

Table 4.4 provides a comprehensive view of the model's performance across different metrics for each class, allowing for a detailed understanding of where the model excels and where it may need improvement. The model achieved 85% train accuracy at 9 epochs and the average loss was 0.0527.

## 4.7. Confusion Matrix

A confusion matrix is a tabular representation that helps evaluate the performance of a classification model by comparing the actual labels of the data with the predicted labels made by the model. Each row of the matrix represents the true class of the data instances, while each column represents the predicted class. In this confusion matrix:

- **cEXT (Extraversion):** Out of the total samples predicted as cEXT, 140 were correctly classified as cEXT, while 19 were misclassified as cNEU, 11 as cAGR, 3 as cCON, and 0 as cOPN.
- **cNEU (Neuroticism):** Out of the total samples predicted as cNEU, 69 were correctly classified as cNEU, while 3 were misclassified as cEXT, 4 as cAGR, 0 as cCON, and 1 as cOPN.
- **cAGR (Agreeableness):** Out of the total samples predicted as cAGR, 35 were correctly classified as cAGR, while 2 were misclassified as cEXT, 2 as cNEU, 1 as cCON, and 0 as cOPN.
- **cCON (Conscientiousness):** Out of the total samples predicted as cCON, 10 were correctly classified as cCON, and there were no misclassifications into other categories.
- **cOPN (Openness):** Out of the total samples predicted as cOPN, 6 were correctly classified as cOPN, while 0 were misclassified as cEXT, 1 as cNEU, 1 as cAGR, and 0 as cCON.

The diagonal from the top left to the bottom right represents the correct predictions, where the predicted class matched the actual class. The other values in the matrix represent misclassifications, where the predicted class did not match the actual class.

**Table 4.6: Confusion matrix using BERT on SOP Dataset**

| | cEXT | cNEU | cAGR | cCON | cOPN |
|---|---|---|---|---|---|
| cEXT | 140 | 19 | 11 | 3 | 0 |
| cNEU | 3 | 69 | 4 | 0 | 1 |
| cAGR | 2 | 2 | 35 | 1 | 0 |
| cCON | 0 | 0 | 0 | 10 | 0 |
| cOPN | 0 | 1 | 1 | 0 | 6 |

The confusion matrix provides valuable insights into the model's performance for each class. It allows us to identify the areas where the model is performing well (diagonal elements) and the areas where it may be making mistakes (off-diagonal elements). By analyzing the confusion matrix, we can understand the strengths and weaknesses of the model and make necessary improvements to enhance its accuracy and performance. Moreover, this confusion matrix provides a detailed view of how well the BERT model is performing in predicting each of the Big Five personality traits. It allows for the identification of not only the correct predictions but also where the model is making mistakes, which can be valuable in understanding and improving the model's performance.

The table 4.7 presents a comparative overview of research results between your proposed approach and several existing state-of-the-art approaches. Notable findings include the achievement of high accuracy by Hans Christian and Derwin Suhartono using BERT, RoBERTa, and XLNet on the Facebook dataset. However, your proposed approach stands out with an impressive F1 Score of 84.4% in testing on the SOP dataset when trained on Essays and myPersonality Twitter data, demonstrating its strong classification performance. Additionally, the training accuracy of 90.6% underlines the robustness of our approach. This comparison showcases the effectiveness of your tailored methodology in text classification tasks, especially when considering specific datasets and fine-tuning with BERT.

**Table 4.7: Comparison of our approach results with existing state of art approaches results**

| Authors | Dataset | | Models | F1 Score | Accuracy |
|---|---|---|---|---|---|
| Majid Ramezani et al. [60] | Essay | | BiLSTM, Ensemble method | 61.04 | 60.24% |
| **El-Demerdash et al. [39]** | **Essay and myPersonality Facebook** | | **Elmo, ULMFiT, and BERT.** | **-----** | **61.85 and 73.91%** |
| Hans Christian and Derwin Suhartono [44] | Facebook dataset | | BERT, RoBERTa, and XLNet | 86.17% | 91.2 % |
| **Salma Kulsoom et al. [64]** | **Essay and SOPs** | | **Bi-LSTM, CNN-LSTM and CNN** | **-------** | **88.2% and 67%** |
| **Proposed** | **Training** | **Essays + mypersonality Twitter** | **BERT** | **82%** | **90.6 %** |
| | **Testing** | **SOP** | | | **84.4%** |
| | **Training** | **Essay** | | **81%** | **85%** |
| | **Testing** | **SOP** | | | **82.7%** |

## 4.8. Chapter summary

In the experimental analysis, the impact of hyperparameters, particularly learning rate and epochs, on model performance is thoroughly examined. Notably, a learning rate of 6e-5 with 7 epochs produced the highest testing accuracy of 84.4% for the BERTBase uncased model, demonstrating the significance of an appropriate learning rate and the right number of epochs for efficient convergence. High and low learning rates adversely affected performance, emphasizing the need for careful tuning. The relationship between the number of epochs and training accuracy is explored, revealing a balance between model learning and overfitting. The choice of hyperparameters is crucial, with some models performing better than others, but results may vary depending on the dataset and task. Additionally, batch size effects on a BERT model's training and validation are assessed, showing that a batch size of 48 produced optimal results. The evaluation of the BERT model's performance in text classification tasks, particularly in predicting Big Five personality traits, demonstrates strong overall accuracy and provides detailed insights into class-specific performance. The model's generalization capability is highlighted when trained on essays and tested on Statements of Purpose (SOP), achieving an impressive 82.7% accuracy. Lastly, a comparative analysis with existing state-of-the-art approaches showcases the effectiveness of the proposed methodology in text classification tasks, particularly when tailored to specific datasets and fine-tuned with BERT.

# CHAPTER 5: CONCLUSION AND FUTURE WORK

Our exploration into the realm of personality prediction using textual data has offered valuable insights, presenting an innovative application of deep learning models like BERT. The study's success underscores the potential of AI in revealing the nuanced relationship between language and personality, and its prospective application in various sectors. However, as we advance, it becomes crucial to balance technological progress with ethical considerations, ensuring that the power of AI serves to enhance human understanding rather than infringe on personal autonomy.

This study embarked on an ambitious journey to predict personality traits from multiple text sources, navigating the rich terrain of textual data. Utilizing the innovative deep-learning model BERT, the research outshone conventional ML models, achieving remarkable classification accuracy. Three diverse datasets were employed in the experiments: the mypersonality Kaggle dataset, Essay dataset, and SOP collected dataset. All were meticulously preprocessed to ensure quality and consistency. The BERT model was not only applied but also fine-tuned, enhancing the learning rate to reach even higher accuracy levels. The results were telling: the SOP dataset, in conjunction with the fine-tuned BERT model, yielded superior outcomes compared to traditional methods. Notably, BERT achieved a test accuracy of 84.4% with the fused dataset and 82% with the SOP dataset when trained on the essay's dataset. These figures are a testament to the model's robust performance, further emphasized by specific metrics such as a precision of 0.97 and an F1-Score of 0.88 for cEXT using BERT. Beyond the numbers, this work has profound implications. It simplifies the complex task of personality prediction across multiple datasets, providing valuable guidance for psychological authorities, academia, human resources, and marketing specialists. By leveraging the capabilities of BERT, the research offers a powerful tool to enhance hiring processes, reduce unnecessary stress, and foster a deeper understanding of human personality. Moreover, in our experiments we found that due to the size of data models are super sensitive to hyperparameters for example for learning rate 1e-4 we got 0.65% F1 score but for 5e-4 we got 0.55 F1 score. Our method performed better from all the previously implemented single and multiple classifier approaches. As we look forward, the opportunities for expanding this research are abundant, from incorporating multiple modalities of communication to exploring real-time personality prediction. As the landscape of AI continues to evolve, so too does our ability to comprehend the complexities of human personality.

This study represents a significant stride forward in the field of personality prediction, harnessing the power of cutting-edge deep learning and contributing valuable insights for various domains. Its success underscores the potential of novel AI techniques to illuminate complex human traits, opening exciting avenues for future exploration and application. In future we aim to use more versions of BERT multi model and hybrid approaches for multiple large amounts of datasets that can be helpful for sentiment, opinion and other emotion detection from text, moreover Previous research might not have sufficiently considered the cultural and linguistic context of the text data. How do cultural norms and different languages affect the expression of personality traits in text? And how can we make our model more intelligent to cater that. In the pursuit of advancing the field of personality prediction, several critical areas demand attention and innovation. Incorporating Multiple Communication Modalities and Real-time Personality Prediction has emerged as a pivotal approach, recognizing the nuanced interplay of different communication channels and the value of instantaneous insights. However, the landscape is not without its challenges. A pressing concern is the scarcity of datasets specifically tailored for personality prediction, necessitating the expansion and diversification of the dataset pool. A deep and discerning analysis of existing data reveals further complexity, as the traditional assignment of the Big Five traits to text data appears to lack precision. This observation underscores the need for dedicated efforts to refine and enhance the quality of labels, ensuring alignment with the intricate nature of personality. Collaboration with experts in psychology could pave the way for more psychologically valid models, fostering a deeper understanding of the underlying human factors that shape personality. Such interdisciplinary partnerships have the potential to enrich the models' theoretical grounding, bridge gaps between computational methods and psychological theory, and ultimately contribute to a more authentic and effective approach to personality prediction. In weaving together these diverse strands, the vision is one of a more integrated, informed, and innovative future for personality prediction, where technology and human insight converge to illuminate the multifaceted dimensions of human personality.

## 5.1. CONTRIBUTIONS

**1. Improved Model Development:** One of the major contributions of our research is the creation and refinement of an innovative deep learning model. Using the existing BERT model as our foundation, we have tailored its functionalities to the task of predicting personality traits from multiple textual data. This is a pioneering step in the field of text analysis and personality prediction, further expanding on the foundational work of Bidirectional Encoder Representations from Transformers (BERT) by making the amendments in classification layer.

**2. Expansive Dataset Usage:** Another unique contribution of our research lies in the use of diverse datasets for personality prediction. Utilizing sources such as Essays, MyPersonality, and SOP's datasets, our approach ensures an expansive coverage of various text styles, topics, and contexts, an approach. By doing so, we can draw more comprehensive insights into personality traits across different written expressions.

**3. Probing the Language-Personality Relationship**: Our research significantly contributes to the exploration of the complex interplay between language use and personality traits. By conducting a thorough analysis of language patterns and features across various texts, we are able to uncover new correlations, thereby expanding the existing body of knowledge in psycholinguistics, building upon the studies conducted.

**6. Practical Implementation Guidelines:** Lastly, our research provides valuable guidelines for integrating personality prediction models into real-world applications. This implementation guidance complements the technical development of our models, ensuring they are not just theoretically robust, but practically applicable, furthering the work of Kabadayi et al. [73] in this domain.

Through these contributions, our research advances the understanding and application of personality prediction from text, establishing a more inclusive, ethical, and practical approach to this evolving field of study.

## 5.2. FUTURE SCOPE

The present research has opened multiple avenues for future investigation. Given the increasing pervasiveness of digital communication, the potential for personality prediction models to contribute to a range of fields continues to expand. Future research could focus on the following areas:

1. **Multimodal Approach:** While this study primarily focused on text-based personality prediction, the integration of other modes of communication, such as video and audio, could yield a more comprehensive and accurate personality assessment. This multimodal approach would account for nonverbal cues and vocal tonality, adding another layer of depth to the personality analysis.
2. **Real-time Personality Prediction:** Future research could aim to develop models that can predict personality traits in real-time. This could be beneficial in various sectors like customer service, where real-time personality prediction could enable more personalized and effective interactions.
3. **Cross-cultural Applicability:** The existing research largely utilizes datasets from predominantly English-speaking contexts. Future studies could focus on collecting and analyzing data from various cultural and linguistic backgrounds to ensure the universality of the prediction model.
4. **Ethical Considerations:** As AI continues to evolve, the ethical implications of personality prediction models become increasingly significant. Future research should address these concerns, establishing guidelines for the responsible use of such models while respecting individual privacy and consent.
5. **Integration with Other AI Technologies:** Personality prediction models can be integrated with other AI technologies, such as recommender systems, to provide more personalized experiences. Future research could explore these integrations further, delving into how personality predictions could enhance AI applications in areas like marketing, entertainment, and education.

# REFERENCES

[1] Rapp, A., Curti, L. and Boldi, A., 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, *151*, p.102630.

[2] Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A. and Truong, L., 2023. Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review*, *33*(1), p.100899.

[3] Allen, L.K., Creer, S.D. and Poulos, M.C., 2021. Natural language processing as a technique for conducting text-based research. *Language and Linguistics Compass*, *15*(7), p.e12433.

[4] Balakrishnan, V., Khan, S. and Arabnia, H.R., 2020. Improving cyberbullying detection using Twitter users' psychological features and machine learning. *Computers & Security*, *90*, p.101710.

[5] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.W. and Qiu, J., 2021. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4).

[6] Ahmed, I., Kajol, M., Hasan, U., Datta, P.P., Roy, A. and Reza, M.R., 2023. ChatGPT vs. Bard: A Comparative Study. *UMBC Student Collection*.

[7] Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., Jia, W. and Yu, S., 2022. A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*, *8*(5), pp.745-762.

[8] Berente, N., Gu, B., Recker, J. and Santhanam, R., 2021. Managing artificial intelligence. *MIS quarterly*, *45*(3).

[9] Beck, E.D. and Jackson, J.J., 2021. Network approaches to representing and understanding personality dynamics. In *Measuring and modeling persons and situations* (pp. 465-497). Academic Press.

[10] Rissola, E.A., Bahrainian, S.A. and Crestani, F., 2019, October. Personality recognition in conversations using capsule neural networks. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 180-187).

[11] James W. Pennebaker , Laura A King, "Linguistic styles: Language use as an individual difference," PsycNET: Jouranl of Personality and Social Psycology, 2000.

[12] Saucier, G. and Goldberg, L.R., 2001. Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of personality*, *69*(6), pp.847-879.

[13] Goldberg LR (January 1993). "The structure of phenotypic personality traits". The American Psychologist. 48 (1): 26–34. doi:10.1037/0003-066X.48.1.26. PMID 8427480.

[14] John, O.P., Naumann, L.P. and Soto, C.J., 2008. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*(2), pp.114-158.

[15] Hofman, J.M., Watts, D.J., Athey, S., Garip, F., Griffiths, T.L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M.J., Vazire, S. and Vespignani, A., 2021. Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), pp.181-188.

[16] Bainbridge, T.F., Ludeke, S.G. and Smillie, L.D., 2022. Evaluating the Big Five as an organizing framework for commonly used psychological trait scales. *Journal of Personality and Social Psychology*, *122*(4), p.749.

[17] Briley, D. A., Domiteaux, M., & Tucker-Drob, E. M. (2014). Achievement-Relevant Personality: Relations with the Big Five and Validation of an Efficient Instrument. *Learning and individual differences*, *32*, 26–39. https://doi.org/10.1016/j.lindif.2014.03.010

[18] Cobb-Clark DA, Schurer S (2012). "The stability of big-five personality traits" (PDF). *Economics Letters*. **115** (2): 11 15. doi:10.1016/j.econlet.2011.11.015. S2CID 12086995.

[19] Wagner, J., Lüdtke, O. and Robitzsch, A., 2019. Does personality become more stable with age? Disentangling state and trait effects for the big five across the life span using local structural equation modeling. *Journal of Personality and Social Psychology*, *116*(4), p.666.

[20] *Paunonen SV, Ashton MS (2001). "Big Five factors and facets and the prediction of behavior". Journal of Personality & Social Psychology.* ***81*** *(3): 524–39. doi:10.1037/0022-3514.81.3.524. PMID 11554651.*

[21] Liu, K., 2021. *Incorporate Out-of-Vocabulary Words for Psycholinguistic Analysis using Social Media Texts-An OOV-Aware Data Curation Process and a Hybrid Approach* (Doctoral dissertation, The Claremont Graduate University).

[22] Zini, J.E. and Awad, M., 2022. On the explainability of natural language processing deep models. *ACM Computing Surveys*, *55*(5), pp.1-31.

[23] Bouschery, S.G., Blazevic, V. and Piller, F.T., 2023. Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management*, *40*(2), pp.139-153.

[24] Ray, P.P., 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.

[25] Marti Marcet, S., 2023. *Natural language processing for gender and ethnic differences in self-presentation strategies and career implications* (Master's thesis).

[26] Karavangeli, E.A., Pantazi, D.A. and Iliakis, M., 2023. DistilGREEK-BERT: A distilled version of the GREEK-BERT model.

[27] K. El-Demerdash, R. A. El-Khoribi, A. Mahmoud, S. Ismail, and S. Abdou, "Deep learning based fusion strategies for personality prediction," *Egyptian Informatics Journal*, vol. 23, 2021.

[28] M. Ramezani, M. R. Feizi-Derakhshi, M. A. Balafar et al., "Automatic Personality Prediction; an Enhanced Method Using Ensemble Modeling," *Neural Computing and Applications*, 2022.

[29] Kunte and S. Panicker, "Personality prediction of social network users using ensemble and xgboost," *Progress in Computing, Analytics and Networking*, Springer, Singapore, 2020.

[30] Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, "Personality Trait Detection Using Bagged Svm over Bert Word Embedding Ensembles," 2020, https://arxiv.org/abs/2010.01309.

[31] E. C. Fieller, H. O. Hartley, and E. S. Pearson, "Tests for rank correlation coefficients. i," Biometrika, vol. 44, no. 3/4, pp. 470–481, 1957.

[32] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," Journal of artificial intelligence research, vol. 30, pp. 457– 500, 2007.

[33] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intelligent Systems and their applications, vol. 13, no. 4, pp. 18–28, 1998.

[34] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

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

[36] S. Poria, A. Gelbukh, B. Agarwal, E. Cambria, and N. Howard, "Common sense knowledge-based personality recognition from text," in Mexican International Conference on Artificial Intelligence. Springer, 2013, pp. 484–496.

[37] Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E. and Eetemadi, S., 2020, November. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 1184-1189). IEEE.

[38] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, “Deep learningbased document modeling for personality detection from text,” IEEE Intelligent Systems, vol. 32, no. 2, pp. 74–79, 2017.

[38] Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, “Personality trait detection using bagged svm over bert word embedding ensembles,” in Proceedings of the ACL 2020 workshop on Widening NLP. Association for Computational Linguistics, 2020.

[39] El-Demerdash, K., El-Khoribi, R.A., Shoman, M.A.I. and Abdou, S., 2022. Deep learning-based fusion strategies for personality prediction. *Egyptian Informatics Journal*, *23*(1), pp.47-53.

[40] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, “Recent trends in deep learning-based personality detection,” Artificial Intelligence Review, vol. 53, pp. 2313–2339, 2020

[41] Alexander, L., Mulfinger, E., & Oswald, F. L. (2020). Using Big Data and Machine Learning in Personality Measurement: Opportunities and Challenges. European Journal of Personality, 34(5), 632–648. https://doi.org/10.1002/per.2305

[42] Kim, Y. and Lee, M., 2023. Deep Learning Approaches for lncRNA-Mediated Mechanisms: A Comprehensive Review of Recent Developments. *International Journal of Molecular Sciences*, *24*(12), p.10299.

[43] Balakrishnan, V., Shi, Z., Law, C.L., Lim, R., Teh, L.L. and Fan, Y., 2022. A deep learning approach in predicting products’ sentiment ratings: a comparative analysis. *The Journal of Supercomputing*, *78*(5), pp.7206-7226.

[44] Christian, H., Suhartono, D., Chowanda, A. et al. Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. J Big Data 8, 68 (2021). https://doi.org/10.1186/s40537-021-00459-1

[45] K. C. Pramodh and Y. Vijayalata, “Automatic personality recognition of authors using big five factor model,” 2016 IEEE Int. Conf. Adv. Comput. Appl. ICACA 2016, pp. 32–37, 2017

[46] Tinwala, W. and Rauniyar, S., 2021. Big five personality detection using deep convolutional neural networks.

[47] Amirhosseini, M.H. and Kazemian, H., 2020. Machine learning approach to personality type prediction based on the myers–briggs type indicator®. *Multimodal Technologies and Interaction*, *4*(1), p.9.

[48] Dabhade, P., Agarwal, R., Alameen, K.P., Fathima, A.T., Sridharan, R. and Gopakumar, G., 2021. Educational data mining for predicting students’ academic performance using machine learning algorithms. *Materials Today: Proceedings*, *47*, pp.5260-5267.

[49] Bayrakdar, S., Yucedag, I., Simsek, M. and Dogru, I.A., 2020. Semantic analysis on social networks: A survey. *International Journal of Communication Systems*, *33*(11), p.e4424.

[50] Alam Sher Khan, Hussain Ahmad, Muhammad Zubair Asghar, Furqan Khan Saddozai, Areeba Arif and Hassan Ali Khalid, “Personality Classification from Online Text using Machine Learning Approach” International Journal of Advanced Computer Science and Applications(IJACSA), 11(3), 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110358

[51] Sharmila P. (2020) Prediction of Big Five Personality Traits from Mobile Application Usage. University of Oulu, Degree Programme in Computer Science and Engineering, 72 p. http://jultika.oulu.fi/files/nbnfioulu-202006242651.pdf

[52] Zarnoufi, R. and Abik, M., 2020. Big five personality traits and ensemble machine learning to detect cyber-violence in social media. In *Innovation in Information Systems and Technologies to Support Learning Research: Proceedings of EMENA-ISTL 2019 3* (pp. 194-202). Springer International Publishing.

[53] Nisha, K.A., Kulsum, U., Rahman, S., Hossain, M.F., Chakraborty, P. and Choudhury, T., 2022. A comparative analysis of machine learning approaches in personality prediction using MBTI. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2021* (pp. 13-23). Springer Singapore.

[54] Aslam, N. *et al.* (2021) *Analysis of personality assessment based on the five-factor model* ... Available at: https://www.ijstr.org/final-print/may2021/Analysis-Of-Personality-Assessment-Based-On-The-Five-factor-Model-Through-Machine-Learning.pdf (Accessed: 10 August 2023).

[55] Wang, P., Yan, Y., Si, Y., Zhu, G., Zhan, X., Wang, J. and Pan, R., 2020. Classification of proactive personality: Text mining based on weibo text and short-answer questions text. *Ieee Access*, *8*, pp.97370-97382.

[56] H. Ahmad, M. U. Asghar, M. Z. Asghar, A. Khan and A. H. Mosavi, "A Hybrid Deep Learning Technique for Personality Trait Classification from Text," in IEEE Access, vol. 9, pp. 146214-146232, 2021, doi: 10.1109/ACCESS.2021.3121791.

[57] Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat, 2021, Personality Prediction Via CV Analysis using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 09 (September 2021).

[58] Moraes, R., Pinto, L.L., Pilankar, M. and Rane, P., 2020, April. Personality assessment using social media for hiring candidates. In *2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA)* (pp. 192-197). IEEE.

[59] Elngar, A.A., Jain, N., Sharma, D., Negi, H., Trehan, A. and Srivastava, A., 2020. A deep learning based analysis of the big five personality traits from handwriting samples using image processing. *Journal of Information Technology Management*, *12*(Special Issue: Deep Learning for Visual Information Analytics and Management.), pp.3-35.

[60] Ramezani, M., Feizi-Derakhshi, M.-R., Balafar, M.-A., Asgari-Chenaghlu, M., Feizi-Derakhshi, A.-R., Nikzad-Khasmakhi, N., Ranjbar-Khadivi, M., JahanbakhshNagadeh, Z.,

Zafarani-Moattar, E., & Rahkar-Farshi, T. 2022. Automatic Personality Prediction; an Enhanced Method Using Ensemble Modeling https://arxiv.org/pdf/2007.04571v3.pdf .

[61] Maharani, W. and Effendy, V., 2022. Big five personality prediction based in Indonesian tweets using machine learning methods. *International Journal of Electrical & Computer Engineering (2088-8708)*, *12*(2).

[62] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," J. Pers. Soc. Psychol., vol. 77, no. 6, pp. 1296–1312, Dec. 1999, doi: 10.1037/0022-3514.77.6.1296.

[63] https://www.kaggle.com/code/testaccount93/mypersonality/input

[64] S. Kulsoom, S. Latif, T. Saba and R. Latif, "Students Personality Assessment using Deep Learning from University Admission Statement of Purpose," *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia, 2022, pp. 224-229, doi: 10.1109/CDMA54072.2022.00042.

[65] https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

[66] Patwardhan, Narendra, Stefano Marrone, and Carlo Sansone. 2023. "Transformers in the Real World: A Survey on NLP Applications" *Information* 14, no. 4: 242. https://doi.org/10.3390/info14040242

[67] Topal, M.O., Bas, A. and van Heerden, I., 2021. Exploring transformers in natural language generation: Gpt, bert, and xlnet. *arXiv preprint arXiv:2102.08036*.

[68] https://huggingface.co/docs/transformers/main/model_doc/bert

[69] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[70] https://huggingface.co/roberta-base

[71] He, P., Liu, X., Gao, J. and Chen, W., 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

[72] Gao, T., Yao, X. and Chen, D., 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

[73] Kabadayi, S., Srinivasan, A., & Gupta, R. (2020). Implementing AI responsibly: Ethics considerations for AI and HR professionals. Strategic HR Review.

[74] Qian, J., Wang, H., Li, Z., Li, S. and Yan, X., 2022. Limitations of language models in arithmetic and symbolic induction. *arXiv preprint arXiv:2208.05051*.

[75] Lowphansirikul, L., Polpanumas, C., Jantrakulchai, N. and Nutanong, S., 2021. Wangchanberta: Pretraining transformer-based thai language models. *arXiv preprint arXiv:2101.09635*.

[76] Gao, T., Yao, X. and Chen, D., 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

[77] Gao, T., Yao, X. and Chen, D., 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.