PERSONALITY PREDICTION BASED ON SOCIAL MEDIA POSTS USING DEEP LEARNING

BY

Nazmul Hosain Nahid
ID: 192-15-2881

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Nasima Islam Bithi
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Md. Sabab Zufiker
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

This Project titled **"PERSONALITY PREDICTION BASED ON SOCIAL MEDIA POST USING DEEP LEARNING"**, submitted by Nazmul Hosain Nahid, ID No: 192-15-2881 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26th January 2024.

BOARD OF EXAMINERS

Chairman

Dr. Md. Ismail Jabiullah(MIJ)
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Raja Tariqul Hasan Tusher(THT)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Md. Abbas Ali Khan (AAK)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Md. Zulfiker Mahmud (ZM)
Associate Professor, External Member
Department of Computer Science and Engineering
Jagannath University

DECLARATION

I, therefore, declare that this undertaking has been finished by us under the supervision of **Ms. Nasima Islam Bithi, Lecturer**, Department of CSE, Daffodil International University. I further declare that neither an application or an educational grant has been made anywhere for this project or any part of it.

Supervised by:

For.

Ms. Nasima Islam Bithi
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Sabab Zufiker
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Nazmul

Nazmul Hosain Nahid
ID: 192-15-2881
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final year project successfully.

I am really grateful and wish my profound indebtedness to **Ms. Nasima Islam Bithi**, Lecturer, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of "Deep Learning" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to Professor **Dr. Sheak Rashed Haider Noori,** Head**,** Department of CSE, for his kind help to finish my project and also to other faculty members and the staff of CSE department of Daffodil International University.

I would like to thank all of my classmates at Daffodil International University who participated in the discussion while also attending the course.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

This study analyzes the complicated field of personality prediction through the use of the diverse data embedded in social media posts. Our research uses many different areas of machine learning and deep learning algorithms to study the complicated connection between individual characteristics and digital expressions. The algorithms that were selected include CNN, BiLSTM, LSTM, LR, Linear SVC, DT, RF, and Multinomial Naive Bayes (MNB) deep learning and machine learning architectures. Analyzing these algorithms provides small variations in approach, with every algorithm presenting different points of view on the prediction task. With an accuracy of 81.36%, Linear SVC was the clear winner, closely followed by Logistic Regression at 80.25%. .. The outcomes of this study not only improve the accuracy of personality prediction from social media posts but also determine a basis for future study actions. The combination of deep learning and machine learning to understand the specifics of human behavior on digital platforms has significant promise for a variety of uses, including mental health monitoring and customized advertising methods. The knowledge achieved from this study prepares the way for the responsible and significant application of predictive algorithms in gaining a knowledge of human personalities in the online environment, as technological advances keep changing our digital connections.

Keywords: *Personality, Social Media, Deep Learning, Machine Learning, Linear SVC, CNN, Bi-LSTM*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

A unique field of research in machine learning and natural language processing is personality prediction from social media posts. Social media platforms function as huge archives of content created by users, offering a unique view into people's ideas, feelings, and modes of communication. This novel method looks for hidden patterns in textual data to identify and predict characteristics using the skills of deep learning algorithms.

Recent developments on human behavior have been made possible by the rise in the frequency of online contacts in recent years. Standard methods for evaluating personality usually depend on self-reporting or organized surveys, which have potential limitations and biases. Making use of the huge quantity of data included in social media posts enables a more dynamic and culturally rich knowledge of people's personalities.

Deep learning algorithms that are especially successful at capturing the nuances of language and context include The opposite direction of long-term memory (BiLSTM), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). These models are particularly good at processing sequential data and extracting hierarchical characteristics, which helps them identify small trends that traditional methods might miss.

Besides simply theoretical curiosity, there is more to the motivation for personality prediction from social media posts. Applications are found in many different fields, such as social research, mental health monitoring, individualized suggestions, and targeted marketing. Through the analysis of subtle language and affective signals included in social media content, deep learning models can provide new insights that help create a more customized and responsive digital environment.

But there are many obstacles in the way, from privacy-related moral problems to the fundamental difficulties in simulating human behavior. It is critical that we carefully discuss these obstacles as we begin the investigation of personality prediction with deep learning, guaranteeing ethical and open processes. This journey has an opportunity to shape customized interactions in the digital age and improve our understanding of individual changes.

1.2 Motivation

Deep learning algorithms are being used to predict personality based on social media posts because of their huge capacity for unlocking the complex web of human behavior that is contained in digital expressions. Social networking sites now function as real-time journals, recording users' feelings, ideas, and responses. Conventional personality evaluations frequently depend on reported information, which can be unpredictable and insufficient for reflecting the dynamic character of personalities. Deep learning algorithms provide a unique chance of understanding textual material by utilizing techniques like BiLSTM, LSTM, and CNN. These algorithms may identify hidden characteristics by analyzing the language patterns, sentiment, and context of social media posts. This allows for a more thorough understanding of persons. The personality prediction has applications in a wide range of fields, including social research, mental health monitoring, and marketing strategy and content recommendation customization. This methodology not simply improves the precision of personality evaluations but also fosters a more customized and flexible digital landscape. It is essential that we approach this project with ethical considerations, protecting user privacy and making sure predictive models are used responsibly as we discuss this terrain. The ultimate driving force is the ability to use cutting-edge technology to interpret social media and provide deep understanding into the wide range of human personalities that influence our online interactions.

1.3 Rationale of the Study

Understanding digital human behavior has the potential to be transformative, which is why deep learning algorithms are being utilized for investigating personality prediction based on social media posts. Social media platforms have developed into rich sources of unfinished, real-time expressions that are free of restriction, offering a never-before-seen chance to gain an in-depth knowledge of people. Conventional personality evaluations frequently incorporate biases and limitations by using self-reports and organized surveys.

Textual data analysis is made dynamic and context-aware by deep learning techniques like CNN, BiLSTM, and LSTM. These machine learning techniques attempt to identify patterns predictive of personality characteristics by examining language subtleties, emotional undertones, and contextual indications present in social media posts. The study's justification is based on its potential uses in an array of fields, such as personalized content recommendations, mental health monitoring, and targeted marketing. Recognizing and predicting personalities from social media posts helps create more personalized and responsive interactions in the digital world as society grows more digitally connected. This study has the potential to improve the accuracy and flexibility of personality tests, leading to a better understanding of people's online identities.

1.4 Research Question

1. What extent can personality qualities from social media posts be identified by deep learning algorithms?

2. How can multiple deep learning architectures, such CNN, LSTM, and BiLSTM, affect the accuracy of personality prediction?

3. How much can sentiment analysis and language usage in social media posts support the efficacy of personality prediction models?

4. How do imbalances in the dataset's personality characteristic distribution impact the way deep learning algorithms work?

5. How does the amount of the dataset affect personality prediction models' generalizability and dependability?

6. Is it possible to improve personality characteristic predictions by taking into account contextual data like social media user interactions and temporal trends?

1.5 Expected output

By applying deep learning algorithms to predict personality based on social media posts, realistic and understandable knowledge of people's personality traits are what is expected. Model projections for every personality dimension, including characteristics like "ENFJ," "INFP," and "ISTJ," among others, would make up the output. With the help of their digital expressions, these predictions hope to offer a thorough comprehension of individuals' behavioral patterns and emotional preferences. For every personality feature, the output might contain probability scores or categorical predictions, providing an advanced understanding of the chance that particular traits would be displayed. The results' interpretability is essential for improving the model's transparency and helping people who use it understand the foundation for predictions. In addition, charts and graphs that show the connections between projected personality characteristics and social media content may be included in the expected result. Based on the projected personality profiles of individuals, this comprehensive output enables applications in personalized suggestions, mental health monitoring, and targeted marketing, resulting in a more responsive and customized digital experience.

1.6 Project Management and Finance

In the case of personality prediction based on social media posts utilizing deep learning algorithms, project management and finance include resource allocation, strategic planning, and financial considerations to guarantee a fruitful and long-lasting research project. A complete plan will specify the activities, deadlines, and checkpoints for gathering data, preparing it, developing a model, and evaluating it in project management. Effective resource allocation is also required, including the use of human resources for data annotation and analysis as well as computer power for model training. Creating a budget is necessary for the project's computational resources, possible use of cloud services, and the cost of labor. To guarantee that the project stays within budgetary restrictions and produces results of the highest quality, financial planning is necessary. Successful leadership requires consistent tracking and reporting of project progress. Furthermore, risk analysis and backup plans are essential for handling unforeseen problems that could affect schedules or resource usage. The character analysis project will proceed smoothly, stay in line with organizational objectives, and use deep learning algorithms to gain valuable insights from social media data if project management and finance are successfully integrated.

TABLE 1.1: PROJECT MANAGEMENT TABLE

Work	**Time**
Data Collection	1 month
Papers and Articles Review	3 month
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 month
Total	8 month

1.7 Report Layout

- Introduction
- Background
- Research Methodology
- Experimental Result and Discussion
- Impact on Society, Environment
- Summary, Conclusion, Future Research
- Reference

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

The starting point processes that lay the groundwork for successful research take place in the preliminary stages of a deep learning algorithm-based personality prediction study based on social media posts. To start, the project's goals, objectives, and research questions must be established in order to direct the inquiry. An essential part of the preliminary stages is data collecting and selection, where researchers carefully select a dataset from resources like Kaggle, making sure it complies with the objectives of the study and ethical guidelines. Next, preprocessing activities including text normalization, feature extraction, and data cleaning are performed to get the social media information ready for model training. This is the selection of suitable deep learning algorithms, taking into account the goals of the research and the characteristics of the data, such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM). In the initial phase of preparation, the computer system must also be configured, any necessary permissions for data usage must be obtained, and a clear project timeline must be created. The foundation for further activities in model training, assessment, and interpretation is laid during this phase, guaranteeing a methodical and well-structured approach to personality prediction from social media posts.

2.2 Related Works

The literature review for this study will introduce previous variations of papers by various scholars. To predict a mentality based on social media comment using deep learning it is necessary. as a lot of study has been done in this field. I studied a few study papers in order to determine the methods and approaches they used:

Valanarasu, R.et al[1] researched presents a novel hybrid machine learning approach for predicting human personality based on social media digital footprints. By utilizing dynamic multi-context information from various platforms such as Facebook, Twitter, and YouTube, the suggested model addresses the shortcomings of earlier approaches that were attributed to different hiring views. The study shows that the hybrid approach is more accurate than existing methods in predicting people's mentalities. With a remarkable 96.7% accuracy rate, it outperforms traditional methods like Naive Bayes (88%), Random Forest (90%), and Support Vector Machines (92%).

Al Marouf, Ahmed, et al. [2] produced, predicting personality from these digital footprints becomes computationally challenging. The profile-based method is examined in this article, which emphasizes the usage of user-generated textual information from social media. Traditional linguistic aspects and social network attributes are highlighted. Based on the well-known Big Five Factor Model (BFFM), which consists of neuroticism, agreeableness, extraversion, conscientiousness, and openness-to-experience, the study reinterprets personality prediction as a personalized evaluation of these characteristics. chi-squared approach, and symmetric uncertainty evaluator, emphasizing their effects on accuracy metrics for personality prediction models as Random Forest (59.26%), Decision Trees (62.70%), and Naive Bayes (72.13%).

Zhu, Yu. [3] studies employ computational technologies and personality theory to anticipate and analyze user actions on social networking sites. They use data mining techniques, focusing especially on text data, to collect raw data from social network members. The study examines the relationship between personality traits and user behaviors on the Weibo platform using a user personality prediction method that incorporates multi-label learning based on the random walk model. It highlights the value of combining deep neural networks and data mining technologies for precise predictions.

Garg, Shruti, and Ashwani Garg.[4] used of social media (SM) interactions to predict personality traits has gained popularity in recent years, with a special emphasis on the Big Five personality dimensions (Extraversion, Consciousness, Agreeableness, and Openness to Experiences). In order to predict four important Big Five personality traits that are comparable to the Myers-Briggs Type Indicator (MBTI), six supervised machine learning (SML) algorithms were used. The study examines three feature extraction techniques—

global vector for word representation (GloVe), bag of words (BOW), and term frequency and inverse document frequency (TF-IDF)—to address the issues raised by unstructured and unbalanced social media conversations. TF-IDF has been shown to be more accurate than word2vec representation, and GloVe is recommended as a more potent feature extractor because it can preserve word spatial information.

Lima, et al. [5] focused on the developing area of data science, which offers a wealth of information about user trends and habits. This work presents a novel method for personality prediction in social media data by excluding user profiles and evaluating groupings of texts rather than individual ones, which departs from traditional approaches. Using a semi-supervised learning strategy, the prediction task is transformed into a multi-label classification issue by utilizing a Big Five model for personality traits. This effectively addresses the obstacle of annotating vast amounts of social media data. With individual classification rates varying for distinct personality qualities, the implemented system, trained with neural network methods such as Naïve Bayes, Support Vector Machine, and Multilayer Perceptron, revealed an approximate 83% accuracy in predicting personality traits from Tweets.

Al-Fallooji, et al. [6] researched aims to forecast users' personality types using the Myers-Briggs Type Indicator (MBTI) model by analyzing their SMS posts in the field of social media personality prediction. Using a dataset of 8668 records, each with fifty postings, the study primarily compares the performance accuracy of several preprocessing and data mining techniques. Interestingly, the results show that lightGBM performs better than stemming, lemmatization, and grid search optimization combined. It achieves 100.0% prediction accuracy for the four MBTI personality dimensions (Introversion-Extroversion, Intuition-Sensing, Feeling-Thinking, and Judging-Perceiving), offering businesses, educational institutions, and SMS providers promising insights to modify their online platforms based on users' predictive personality behaviors.

Tadesse, Michael M., et al. [7] Used the myPersonality project dataset, this study examines the prediction of personality traits among Facebook users based on a variety of factors and measures generated from the Big 5 model. The study looks at the relationship between various feature sets and personality qualities using four machine learning models. Notably, the XGBoost classifier achieves a maximum prediction accuracy of 74.2%, outperforming the average baseline across all feature sets. The best accuracy of 78.6% is notably obtained for predicting the extraversion trait in the individual Social Network Analysis (SNA) features set. These results provide important new understandings of how language characteristics, social network structures, and machine learning models interact to predict personality in the context of Facebook users.

Tandera, Tommy, et al. [8] studied using the Big Five Model Personality by utilizing deep learning architectures instead of older machine learning technologies. With an amazing average accuracy of 74.17%, the research outperforms earlier studies of a similar nature and shows higher prediction accuracy using a thorough analysis procedure. The adoption of deep learning methodologies highlights the effectiveness of sophisticated methods in improving the accuracy of personality prediction models, which is a significant development in the field of personality research.

Khan, Alam Sher, et al. [9] Using a resampling strategy (random over-sampling) to improve performance, this research addresses the problem of dataset skewness in previous work on personality prediction. To further investigate personality recognition from text, the study also looks at a number of pre-processing methods, including tokenization, word stemming, stop word removal, and feature selection using TF-IDF. The results indicate that a strong personality identification system should be created, with implications for hiring practices, personnel selection, and customer-focused business development. Notably, the XGBoost classifier achieved exceptional precision and accuracy, surpassing 99% for various personality traits.

Wei, Honghao, et al.[10] presented a unique framework that incorporates various ways to extract semantic representations from heterogeneous social media input, with the goal of improving personality prediction ability. The study uses real-world data from personality surveys and social media usage trends from a sizable volunteer cohort to undertake lengthy experiments. The results show that the suggested approaches outperform commonly used state-of-the-art baseline methods. Furthermore, the research broadens its scope by presenting DiPsy, a personalized chatbot that uses heterogeneous data from digital traces and conversation logs in real-world interactive settings to predict user personality.

Tay, Louis, et al.[11] growing use of text data from social media platforms for personality evaluation calls into question the validity of psychometric validation initiatives. The current body of research on personality evaluation using social media text mining (SMTM) lacks thorough validation, hence a targeted study is needed to identify previous attempts at psychometric validation and point out areas that need more research. This review of the literature tackles issues regarding the validity of SMTM techniques, stressing the significance of creating a solid reference (ground truth) and comprehending the causal relationship between personality traits and scores obtained from text data on social media. The article also examines difficulties with generalizability across various social media platforms and demographics, offering a comprehensive grasp of the fundamental validity and validation concerns essential to the advancement of machine learning techniques in personality

Imran, et al. [12] informed exchange and virtual communities, personality prediction has emerged as a major area of study. This analysis of the literature looks at how users' interactions, posts, and activities on Facebook can be used to predict personality traits using in-house technologies, namely a Python-based program. Using a dictionary-based strategy that combines WordNet, SenticNet, and Opinion Lexicon, the study outperforms existing approaches in predicting Big-5 personality traits. For additional validation, a comparative examination of machine learning classifiers is included.

Souri, et al. [13] research states that when users collaborate on social networks, they display comparable behavioral patterns. Using an application created with the Facebook API, the study focuses on personality detection through the analysis of user behavior on Facebook. By using data mining techniques, specifically the boosting-decision tree model, the authors were able to predict personality traits based only on user profiles with an impressive accuracy of 82.2%, surpassing the accuracy of previous studies in this domain. The study involved 100 volunteers who completed the NEO personality questionnaire.

El-Demerdash, Kamal, et al. [14] growing interest in affective computing and sentiment analysis for automated personality characteristic recognition from text data. Previous work has mostly concentrated on features engineering for psycholinguistic databases and linguistic styles, but current developments in natural language processing—especially transfer learning using trained language models like as Elmo, ULMFiT, and BERT—have become essential. When compared to recent results on gold standard datasets like Essays and myPersonality, this literature review highlights a novel deep learning-based model proposed for personality prediction, leveraging both data and classifier level fusion, and demonstrates its effectiveness by achieving a noteworthy accuracy enhancement of 1.25% to 3.12%.

Golbeck, Jennifer, et al. [15] highlighted the impracticality of depending solely on personality tests for personality analysis in social media domains, this paper tackles the significance of users' personalities in various interactions. The authors describe a revolutionary approach that does away with the requirement for explicit personality tests by effectively predicting individuals' personalities using publicly available information from their Twitter profiles. The study highlights the wider implications for social media design and interface design as it examines the data collection procedure, analytical approaches, and effective use of machine learning techniques in personality prediction.

Table 1: Accuracy Comparison of Existing Related Papers

SN.	Author	Applied Algorithms	Dataset	Best Accuracy

1	Valanarasu, R.et al[1]	NB, RF, SVM	social media	96.7%
2	Al Marouf, Ahmed, et al.[2]	RF, DT, NB	social media	72.13%
3	Garg, Shruti, and Ashwani Garg.[4]	DT	social media	97.6%
4	Lima, et al. [5]	NB, SVM, MP	social media	83%
5	Tadesse, Michael M., et al. [7]	XGBoost	social media	78.6%
6	Tandera, Tommy, et al. [8]	CNN	social media	74.17%
7	Souri, et al. [13]	DT	Facebook	82.2%

2.3 Comparative Analysis and Summary

The unique breakthrough in the field of medical diagnostics is the use of improved deep learning to detect colon cancer. This approach aims to improve the efficiency and accuracy of colon cancer diagnosis by comparing several deep learning models. With the help of advanced methods like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), the model is able to analyze medical imaging data and spot minute patterns that could be signs of colorectal abnormalities. In a comparison analysis, the effectiveness of several deep learning architectures is determined, showing both their advantages and disadvantages in terms of colon cancer detection accuracy. This technique seeks to outperform conventional diagnostic techniques by utilizing cutting-edge algorithms, providing a more accurate and automated answer. In the end, this advanced deep learning approach represents a major advancement in the detection of colon cancer. Researchers want to improve patient outcomes and treatment planning by validating and improving the model through a thorough comparison analysis. This will provide medical professionals a valuable tool for early and accurate colon cancer diagnosis.

2.4 Scope of the Problem

Algorithms based on deep learning have the potential to predict personality characteristics from social media posts, which has wide-ranging implications in multiple fields. Analyzing social media content for personality traits offers important insights into people's preferences, behaviors, and communication styles in the age of universal digital communication. Marketing can benefit from this research since it can be used to create targeted and customized advertisements based on expected personality profiles. Furthermore, social media post analysis in mental health monitoring could help identify and help at-risk persons early on .In the field of social research, the ability to comprehend individuals through deep learning algorithms can be helpful since it provides a distinct viewpoint on society trends and behavioral patterns. By using these findings, educational institutions can use individualized learning strategies and modify their curriculum to better suit the needs of each student. Employers can utilize personality predictions to improve team chemistry, streamline hiring procedures, and customize employee development plans in human resources. The range of information is not without difficulties, though, such as moral questions about privacy and appropriate data use. Reaching the full potential of deep learning algorithms for personality prediction based on social media posts requires navigating these obstacles.

2.5 Challenges

The algorithms used with deep learning for personality prediction based on social media posts face a number of difficulties that need to be carefully considered. To begin with, it is critical to protect privacy and use personal data in a moral way. Highly sensitive data must be handled when analyzing social media content, and strict security measures must be taken to maintain user privacy and uphold moral standards. The fluid and changing character of language on social media presents another difficulty. Slang, irony, and idioms unique to a certain context provide challenges for deep learning models, which may find it challenging to accurately represent the specifics of human speech. The ability of the model to generalize effectively over a range of user profiles may be impacted by addressing imbalances in the dataset, wherein certain characteristics may be underrepresented. A major issue with deep learning models is interpreting their predictions. Some models' "black-box" characteristics raise questions regarding interpretability and transparency since they make it difficult to understand the underlying logic behind predictions. In addition, it's essential to remove biases present in both the algorithms and the training data in order to prevent the maintenance of unjust predictions or views based on demographic characteristics. It necessitates careful planning to strike a balance between the possible hazards and difficulties and the improved personalization that could result from using social media data to construct personality prediction models. This includes highlighting ethical and open development and deployment processes.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The purpose of the research project is to extract and predict personality traits from the huge quantity of user-generated information on social media platforms by utilizing cutting-edge machine learning techniques. This topic includes the analysis of textual data to find latent personality factors, such as sentiment, language patterns, and clues from context. Latest deep learning algorithms including BiLSTM, LSTM, and CNN are used in the instrumentation for this study. The computational tools used to process and examine language material found in social media posts are these algorithms. Natural language processing (NLP) tools for extraction of features, tokenization, and data prior to treatment are also included in the instrumentation.The objective of the research is to use these tools to create accurate models that can predict personality traits with accuracy. The project aims to improve the comprehension of people's personalities through their digital expressions by utilizing deep learning. This will advance personalized applications in a variety of sectors and make an important contribution to the field of natural language processing. I classified my text data into 16 categories based on the prediction. Liner SVC achieved the highest accuracy among all models, with a value of 81.36%. The Python programming language, as well as Anaconda and related tools, were used.On a Windows 11 Pro 64-bit system with a 3.5 GHz Intel Core i5 10th generation CPU and 16 GB of RAM.

3.2 Data Collection Procedure

Thoroughly considered processes are involved in the data collecting process to provide a representative and moved dataset for personality prediction based on social media posts utilizing deep learning algorithms. The data used in this study came from the well-known dataset the website Kaggle. The Kaggle dataset has a large number of entries with 16 target attributes that represent different personality traits. It was selected based on its relevance to the research aims. Using social media information as a basis, deep learning models are developed and trained with the help of the Kaggle dataset in an effort to properly predict personality attributes. I've included: Figure 3.1 shows which number of words in text of target attribute can be seen, in Figure 3.2 is a distribution of length of the post which is collected from social media. Figure 3.3 shows a plot of percentages of the targeted attributed value , Given below:

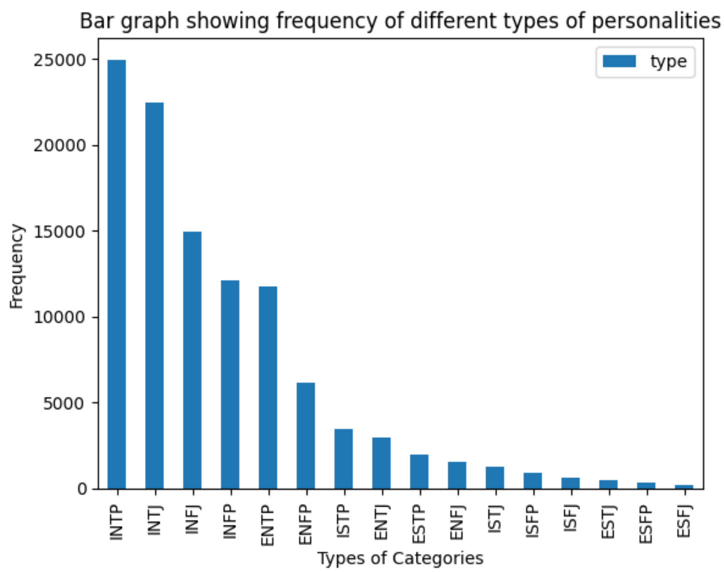


Figure 3.1: Number of words in text

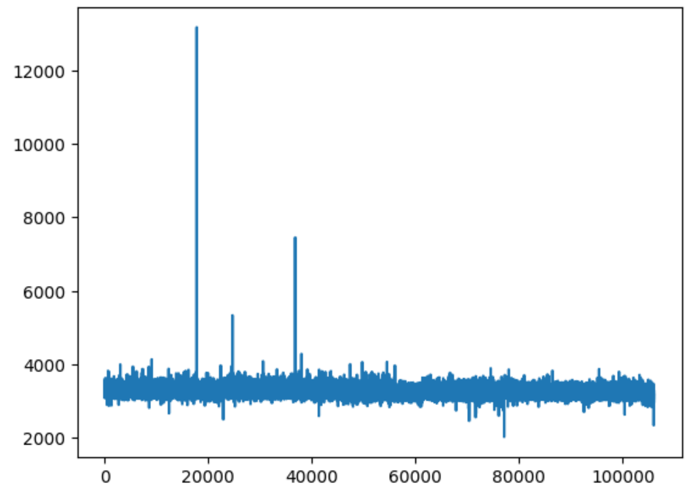


Figure 3.2: Distribution of length of the post

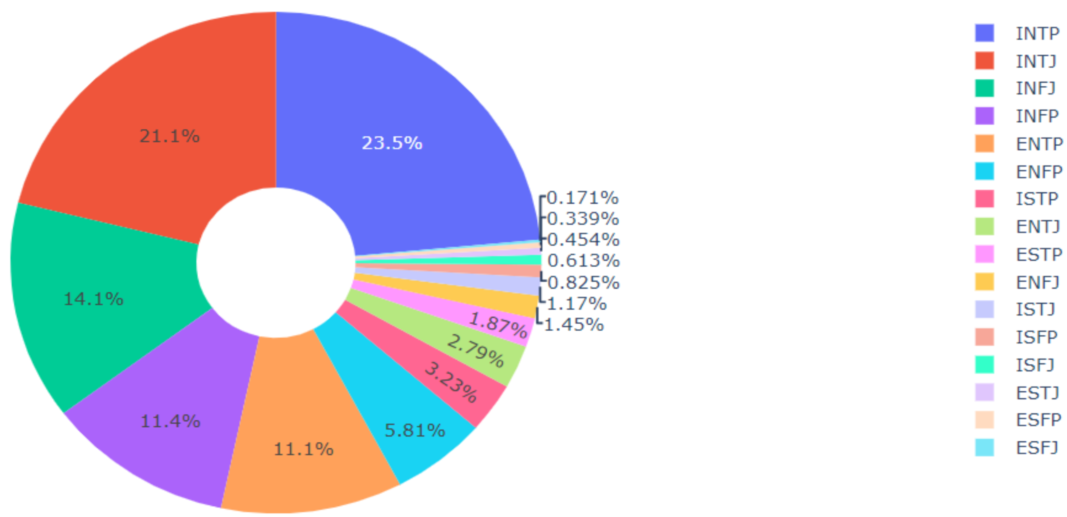


Figure 3.3: Pie Chart of Percentages of Target Attribute

3.3 Statistical Analysis

In the study of personality prediction based on social media posts utilizing deep learning algorithms, statistical analysis is essential. The dataset is first subjected to descriptive statistics, which offer insights into the distribution of personality traits and the features of content on social media. Understanding the fundamental tendencies and variation within the data is made easier by metrics like mean, median, standard deviation, and quartiles. In order to avoid biases in model training, class distribution analysis looks at the frequency of particular personality characteristics. Potential links between personality characteristics are investigated via correlation analysis, which shows patterns that help the model make sense of complex relationships. In order to ensure that each personality trait is fairly represented, statistical approaches like the oversampling and underestimating are used to correct imbalances in the dataset. Quality checks improve the dependability of later analyses by locating and fixing problems like missing numbers or outliers. Informed decisions on feature selection, data preprocessing, and model development are based on statistical analysis, which enhances the interpretability and robustness of personality prediction models that depend on social media posts and deep learning algorithms.

3.4 Proposed Methodology

In below we are following methodology for predicting Personality Based On Social Media Posts Using Deep Learning Algorithms:

Flow Chart:

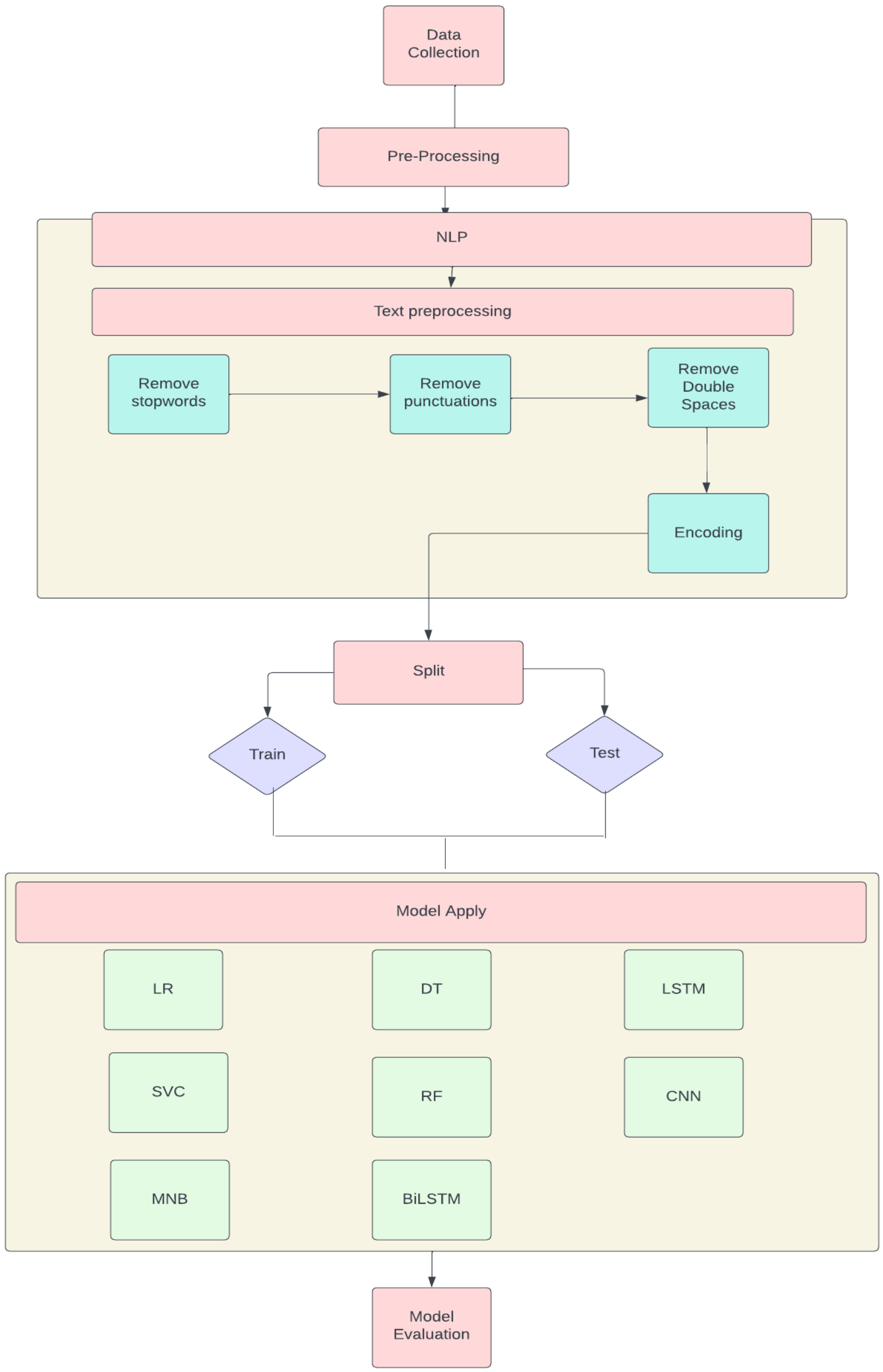


Figure 3.4: Methodology Flowchart

Data Collection:

Obtain information from a variety of sources, such as Kaggle, which offers social media posts along with personality labels. Ensure respect to data usage policies and ethical issues.

Data Labeling:

Label the dataset with personality traits ('ENFJ', 'INFP', etc.) either manually or semi-automatically. The ground truth is created in this step and used for model evaluation and training.

Text Preprocessing:

Text the normalization process, special character handling, and noise reduction are used to clean up and preprocess the textual data. Turn text into numerical representations by using tokenization and other natural language processing methods.

Data Split:

Construct training, validation, and test sets out of the dataset. To ensure that the model is trained on one subset, modified on another, and assessed on an unknown section, a typical split might be 80-10-10.

Model Training:

Choose deep learning algorithms like CNN, LSTM, or BiLSTM. Establish the model architecture's parameters, such as dropout rates, activation functions, and layer counts. Utilizing the training set, train the model and also using Machine learning algorithm like Logistic Regression (LR), Linear Support Vector Classifier (Linear SVC), Multinomial Naive Bayes (MNB), Decision Tree, Random Forest,

Logistic Regression (LR):

Logistic Regression (LR) is a core technique in personality detection based on social media posts. Textual features are analyzed by LR, and probabilities are assigned to distinct personality categories. Its ease of use and interpretability make it ideal for categorizing users and providing vital insights into online behavior and communication patterns. LR got 80.25% accuracy in this study .

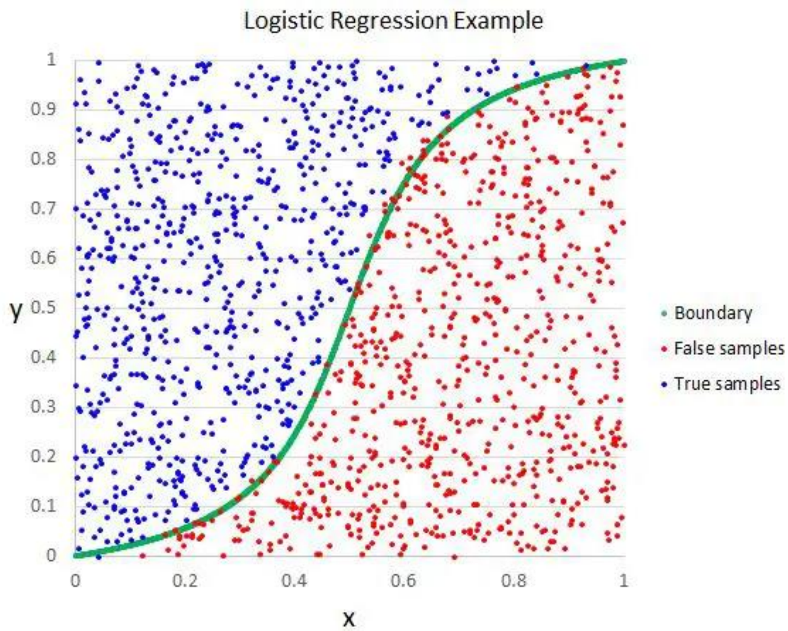


Figure 3.2 : Logistic Regression (LR) Model Architecture

Linear Support Vector Classifier (Linear SVC): The Linear Support Vector Classifier (Linear SVC) is significant in predicting personality from social media posts. Using its ability to delineate complex decision boundaries, Linear SVC efficiently categorizes users into distinct personality types. It improves the model's predictive accuracy by optimising the separation between personality classes. Linear SVC's robustness in handling high-dimensional feature spaces makes it well-suited for extracting meaningful patterns from the diverse and complex language present in social media content. Liner SVC got 81.36% accuracy in our test run.

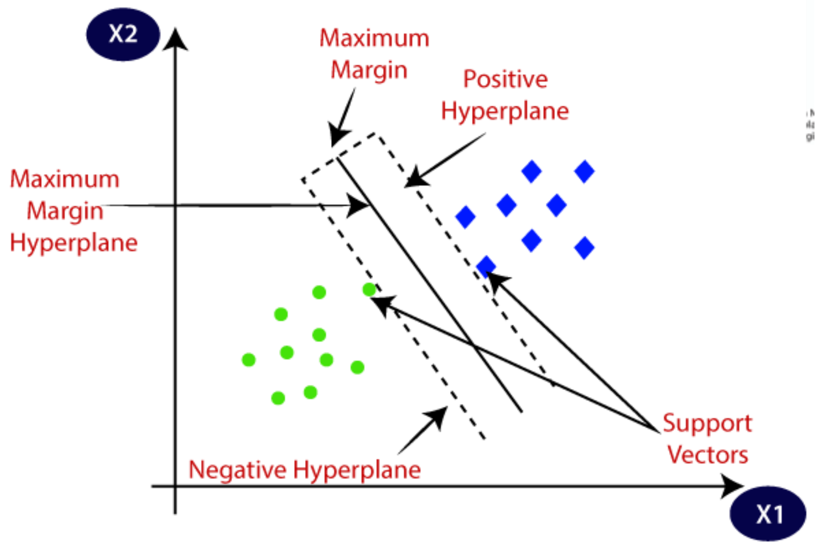


Figure 3.3 : Linear SVC Model Architecture

Multinomial Naive Bayes (MNB):

Multinomial Naive Bayes is applicable in predicting personalities from social media posts. Using its probabilistic methodology, it models the likelihood of different personality characteristics based on word frequency distributions. By assuming independence between features, it efficiently handles textual data. This algorithm is good in capturing the varied language patterns associated with various personality types, which contributes to a greater awareness of users' online expressions in the context of personality prediction studies. MNB got 57.08% accuracy in our test run.

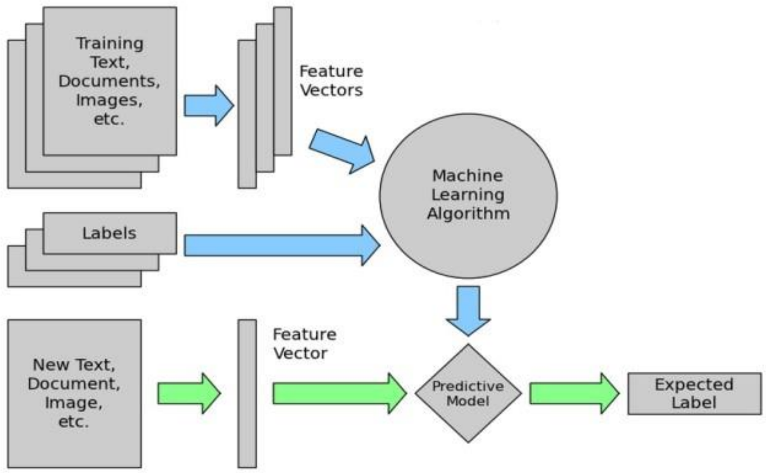


Figure 3.3 :Multinomial Naive Bayes (MNB) Model Architecture

Decision Tree:

Decision Trees offer a versatile application in personality prediction from social media posts. By infinitely separating the feature space, Decision Trees capture deep connections between textual elements and personality factors. Their understanding allows for insights into significance of features and decision paths. In this study, Decision Trees provide a clear framework for discerning patterns in user behavior, aiding the understanding of how specific linguistic cues contribute to the classification of individuals into distinct personality categories. DT got 63.34% accuracy in our test run.

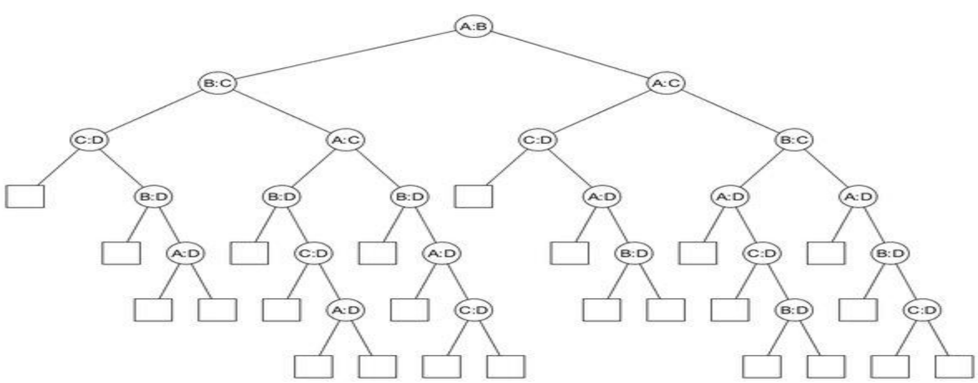


Figure 3.3 :Decision Tree (DT) Model Architecture

Random Forest:

By combining multiple Decision Trees, random forest analysis proves useful in identifying personalities from social media posts. This ensemble method improves predictive accuracy and expansion by reducing overfitting. Random Forests succeed in capturing diverse linguistic patterns and interaction between features present in social media content, providing a robust model for personality classification. The integration of individual trees' views offers an in-depth knowledge of the subtle links between textual signals and numerous personality qualities, contributing to more reliable predictions. RF got 64.92% accuracy in our test run.

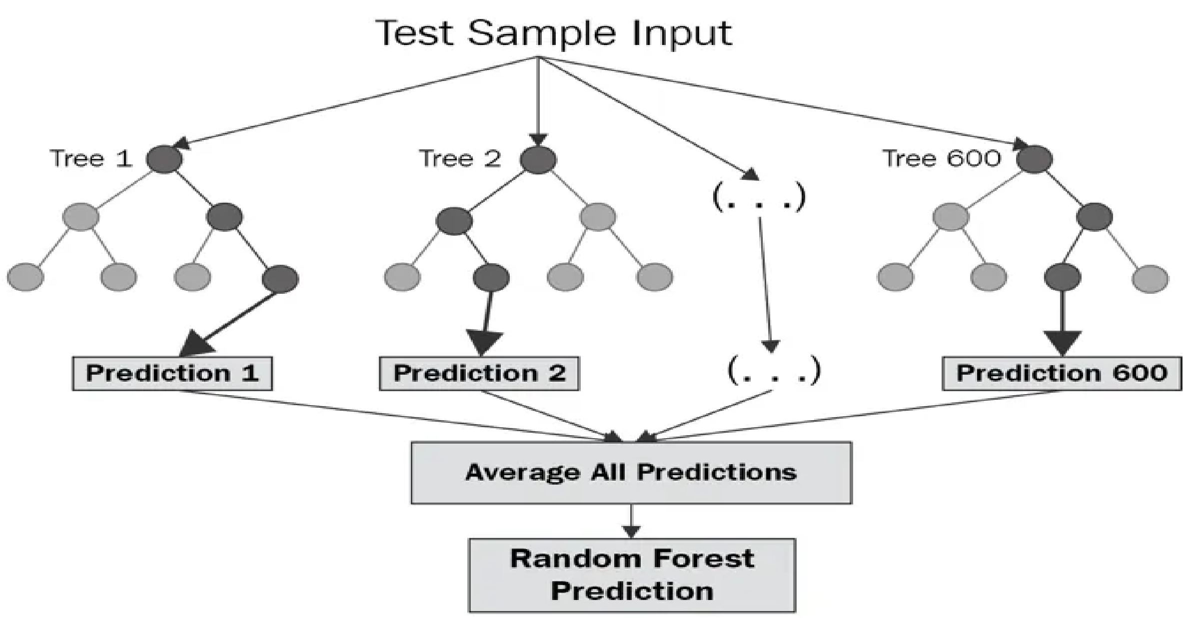


Figure 3.4 : Random Forest (RF) Model Architecture

CNN:

Convolutional Neural Networks (CNNs) show effectiveness in personality prediction from social media posts. By utilizing filters to catch local patterns in textual material, CNNs excel at learning hierarchy features. In this study, CNNs efficiently extract and interpret complex linguistic structures, distinguishing variations in online expressions suggestive of distinct personality types. The hierarchical structure of CNN architectures enables them to capture both short and long-range dependencies in social media text, enhancing the model's ability to discern intricate patterns. CNN got 54.09 % accuracy in our test run.

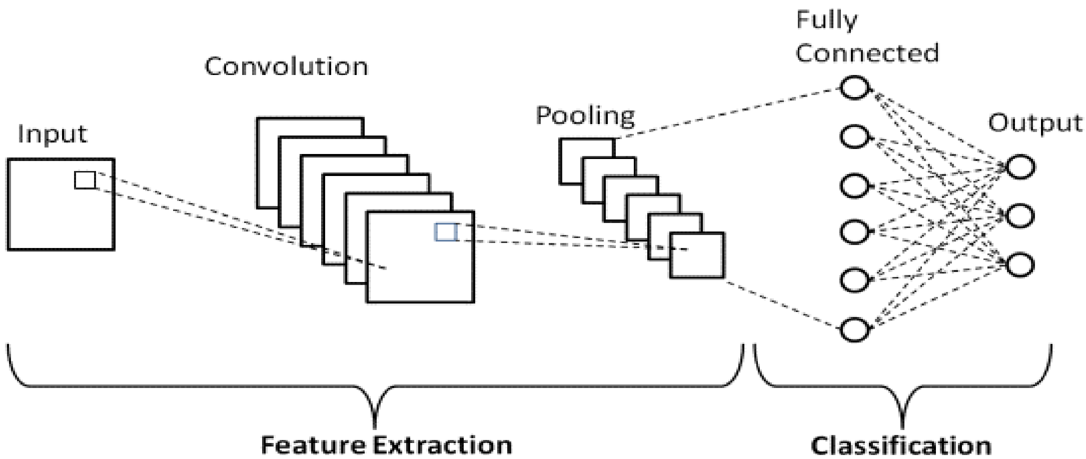


Figure 3.5 :CNN Model Architecture

LSTM:

Long Short-Term Memory networks (LSTMs) have an important role in predicting personality from social media posts. LSTMs effectively capture interdependence and context in textual input due to their progressive learning capabilities. LSTMs excel in this study in holding the temporal dynamics of linguistic expression, discerning subtle complexities, and long-term dependencies indicative of various personality qualities. Because of their ability to model sequential patterns, they are well-suited to extract useful knowledge from the sequential nature of social media messages.LSTM got 59.62 % accuracy in our test run.

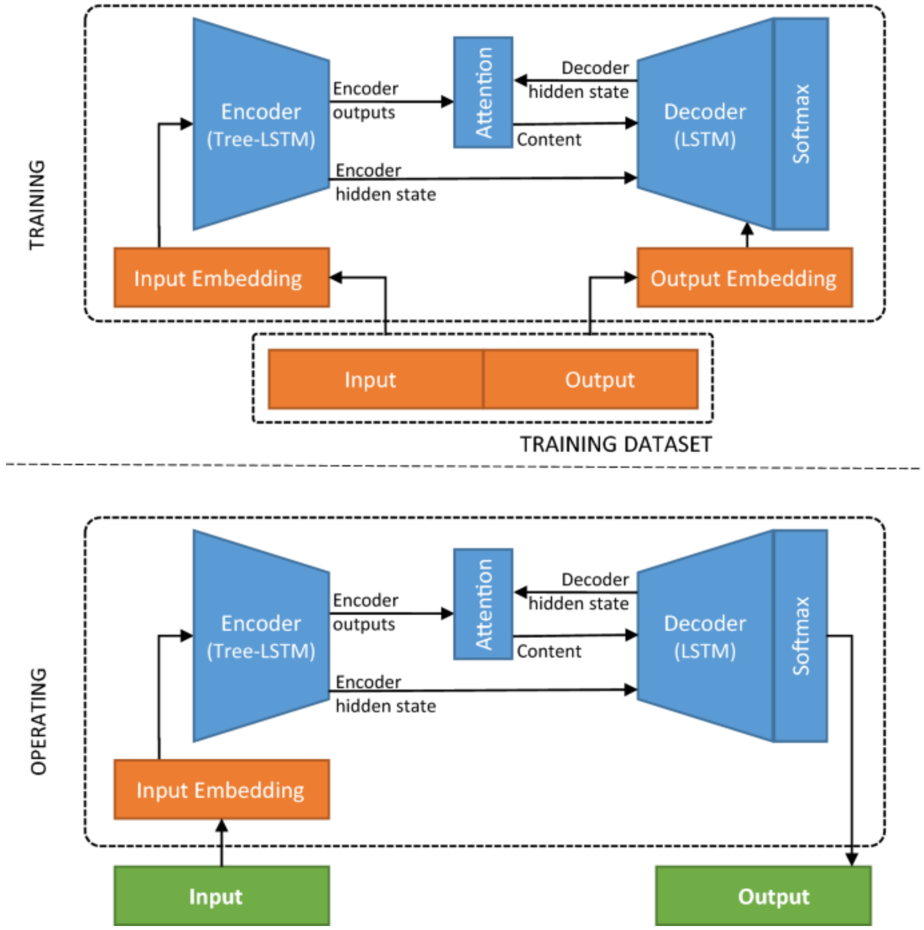
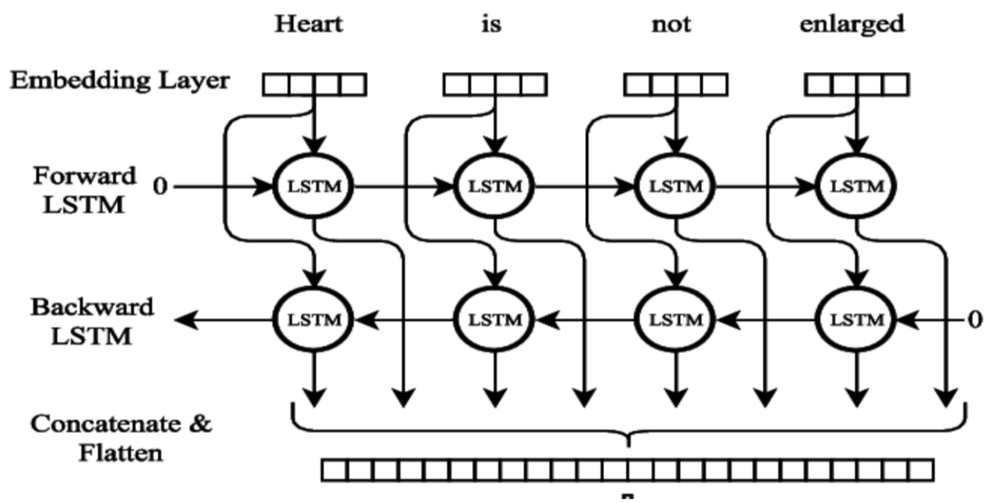


Figure 3.6 :LSTM Model Architecture

BiLSTM:

The opposite direction Long Short-Term Memory (BiLSTM) networks improve personality prediction from social media posts by taking into account both past and future context. This bidirectional method captures complex connections and temporal dynamics in textual data, improving the model's knowledge of many personality traits. BiLSTMs excel at detecting complex language patterns, providing a more complete view of the sequential nature of social media content. BiLSTMs contribute to a more complex and accurate classifying of users based on their online expressions by using information from both directions. Bi LSTM got 62.34 % accuracy in our test run.



Recognition:

Apply the trained model to forecast the validation set's personality attributes. To avoid overfitting, adjust hyperparameters in light of validation performance.

Model Evaluation:

Analyze the model's performance with measures such as F1-score, accuracy, precision, and recall on the test set. Examine the confusion matrix to determine the predictive strengths and weaknesses of particular qualities.

Output:

Provide understandable results, such as probability ratings for every personality characteristic. Provide a visual representation of the outcomes and insights into the model's thinking. Make sure the predictions are represented clearly and in an approachable manner.

The stages involved in creating a personality prediction model utilizing deep learning algorithms are described in this suggested technique, with a focus on robust evaluation, data preprocessing, and model training to ensure accurate and significant outcomes.

3.5 Implementation Requirements

Watchful consideration of a few essential parameters is necessary for the successful deployment of a deep learning-based personality prediction system based on social media posts. The most important thing is to have access to a well-annotated, diverse dataset. To ensure the model's generality, this dataset should include a diverse range of social media posts that represent different personality qualities.In terms of technology, deep learning algorithms require a powerful computing infrastructure to meet their computational demands. It can be essential to have high-performance GPUs or TPUs for effective model training. Having access to pertinent software frameworks such as the Torch or TensorFlow as is also essential for putting deep learning models into practice.Textual data must be cleaned and encoded using data pretreatment techniques, such as natural language processing libraries (NLTK or spaCy). Programming knowledge of Python and deep learning model creation frameworks are also necessary for the implementation.The implementation should be guided by ethical considerations, with a focus on responsible data usage and user privacy. Ensuring adherence to ethical norms and data protection requirements is crucial during the development and deployment phases.To sum up, a well-curated dataset, powerful computer infrastructure, programming language expertise, and adherence to ethical norms are necessary for a successful implementation. Fulfilling these requirements guarantees the creation of a strong personality prediction system capable of generating useful data from social media posts.

CHAPTER 4

Experimental results and discussion

4.1 Experimental Setup

The data is divided into training, validation, and test sets. The data is made up of 2012 photos of spinach leaves. A representative distribution of disease categories within each set is guaranteed via stratified sampling. Resizing photos, defining pixel values, and using augmentation techniques are all part of data preprocessing, which improves model durability. To enable effective training of deep learning models the computational infrastructure employs high-performance GPUs or TPUs. During model training, pre-trained architectures are refined using the spinach dataset, and performance is maximized by hyperparameter adjustment. For an accurate model assessment, assessment standards like accuracy, precision, recall, and F1 score are used. Confusion matrix production is another feature of the experimental setting that allows for the analysis of model predictions across various illness categories. Transparency and repeatability are guaranteed by the thorough documentation of the entire experimental procedure. This well-organized and methodical experimental approach is essential to obtaining useful information about how well deep learning algorithms identify regional spinach illnesses.

4.2 Experimental Results & Analysis

The Multilayer CNN model determines local spinach diseases with 87.34% accuracy, showing that it's able to capture complicated patterns. The learning process of the model reduces training loss and improves accuracy over time. The performance of the model for various disease categories is evaluated using confusion matrix standards such as precision, recall, and F1 score. Further research may focus on modifying hyperparameters or exploring ensemble techniques to improve the model's reliability and generalization capabilities in real-world agricultural contexts.Each statistic provides significant insights into a number of facets of the model's performance:

Accuracy: The accuracy of the model's predictions is determined by comparing the number of correctly classified samples to the total number of samples. Unbalanced classes give a general idea of the model's efficacy, but they may not give a complete picture.

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

Precision: Precision is concerned with the number of true positive forecasts made by the model out of all positive predictions generated by the model.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Recall: The percentage of true positive predictions created out of all actually positive samples is referred to as recall. It's also known as sensitivity or true positive rate.

$$Recall = \frac{TruePositive}{TruePositive + \ FalseNegative}$$

F1 Score: The F1 score is determined as the harmonic mean of recall and precision. Its fair evaluation metric considers recall and precision. The F1 score is useful in cases where class sizes are not equal since it accounts for both false positives and false negatives. A high F1 score indicates a good precision to recall ratio.

$$F - 1\,Score = 2 * \frac{Recall * \ Precision}{Recall + Precision}$$

In given below I am describing the result analysis part also show the training accuracy rate and confusion matrix also:

Logistic Regression

Achieving Test Accuracy of logistic regression is 80.25%. In below Figure 4:1 describing the confusion matrix of Logistic Regression:

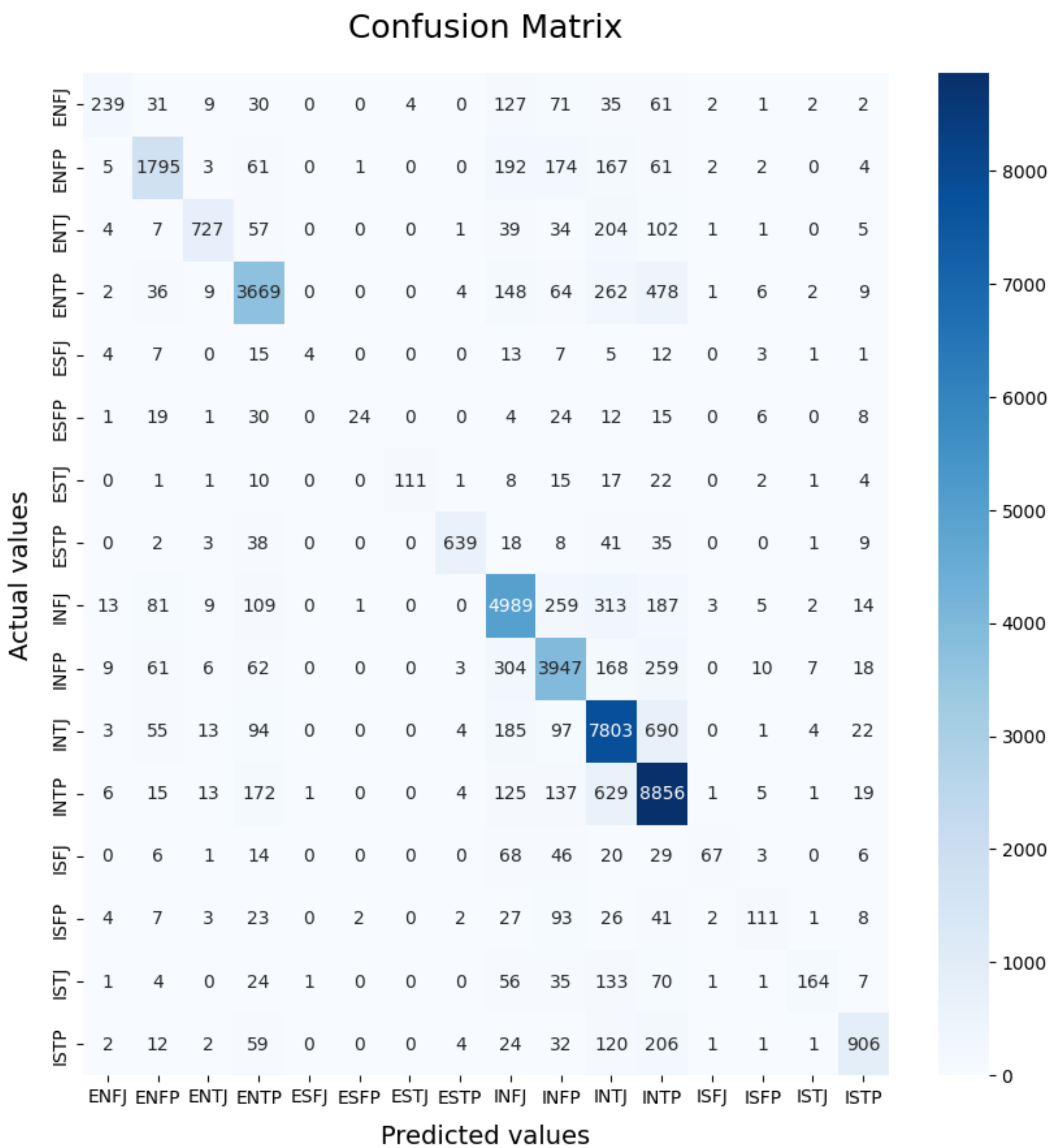


Figure 4.1: Confusion Matrix (LR)

Figure 4:1: show The model's accuracy is rated at 81% for all classes in the classification report, with an overall average accuracy of 0.83, recall of 0.63, and F1-score of 0.69. The precision, recall, and F1-score are all 0.81 on the weighted average. High precision, recall, and F1-scores have been shown by certain personality types (ESTP, ESTJ, INTJ, and INTP), while areas for improvement are present in ESFJ, ESFP, ISFJ, and ISFP. The study proposes focusing on identifying particular personality types, and possible changes include collecting more data, improving features, and looking into bias.

Linear SVC

Achieving the Test Accuracy of Linear Support Vector classifier: 81.36% In below Figure 4:2 describing the confusion matrix of Linear SVC:

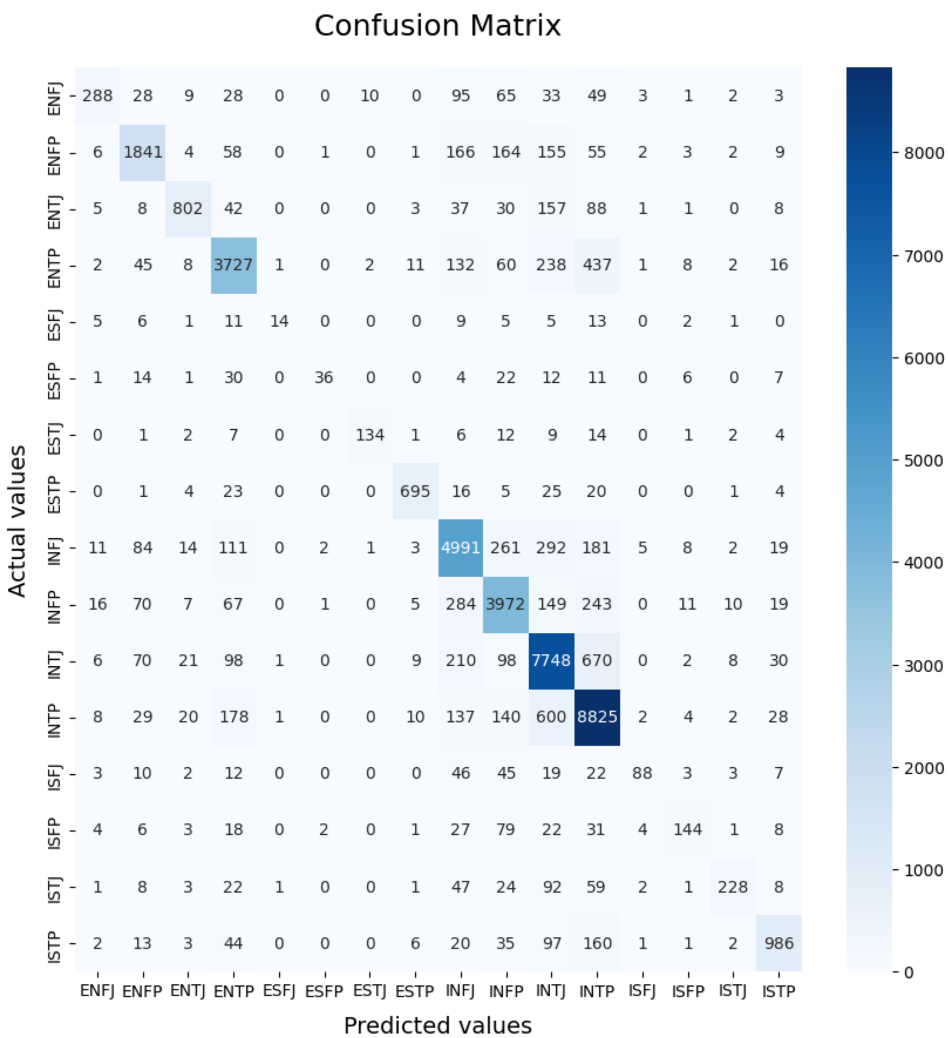


Figure 4.2: Confusion Matrix (Linear SVC)

Figure 4.2: shown Based on a test dataset of 42,427 samples, the report evaluates the performance of a classification model that predicts 16 personality types. Metrics include precision, recall, and F1-score; support indicates how many samples there are in each class. Class-level performance shows that different personality types have different precision,

recall, and F1-scores. For example, the ENFJ personality type has a precision of 0.80, a recall of 0.47, an F1-score of 0.59, and 614 sample support. The overall performance shows 81% accuracy, and weighted and macro averages offer more information. Important findings show that while the model performs well overall, there are notable differences between personality types.

Multinomial Naive Bayes

Achieving the Test Accuracy of Multinomial Naive Bayes: 57.08% In below Figure 4:3 describing the confusion matrix of Multinomial Naive Bayes:

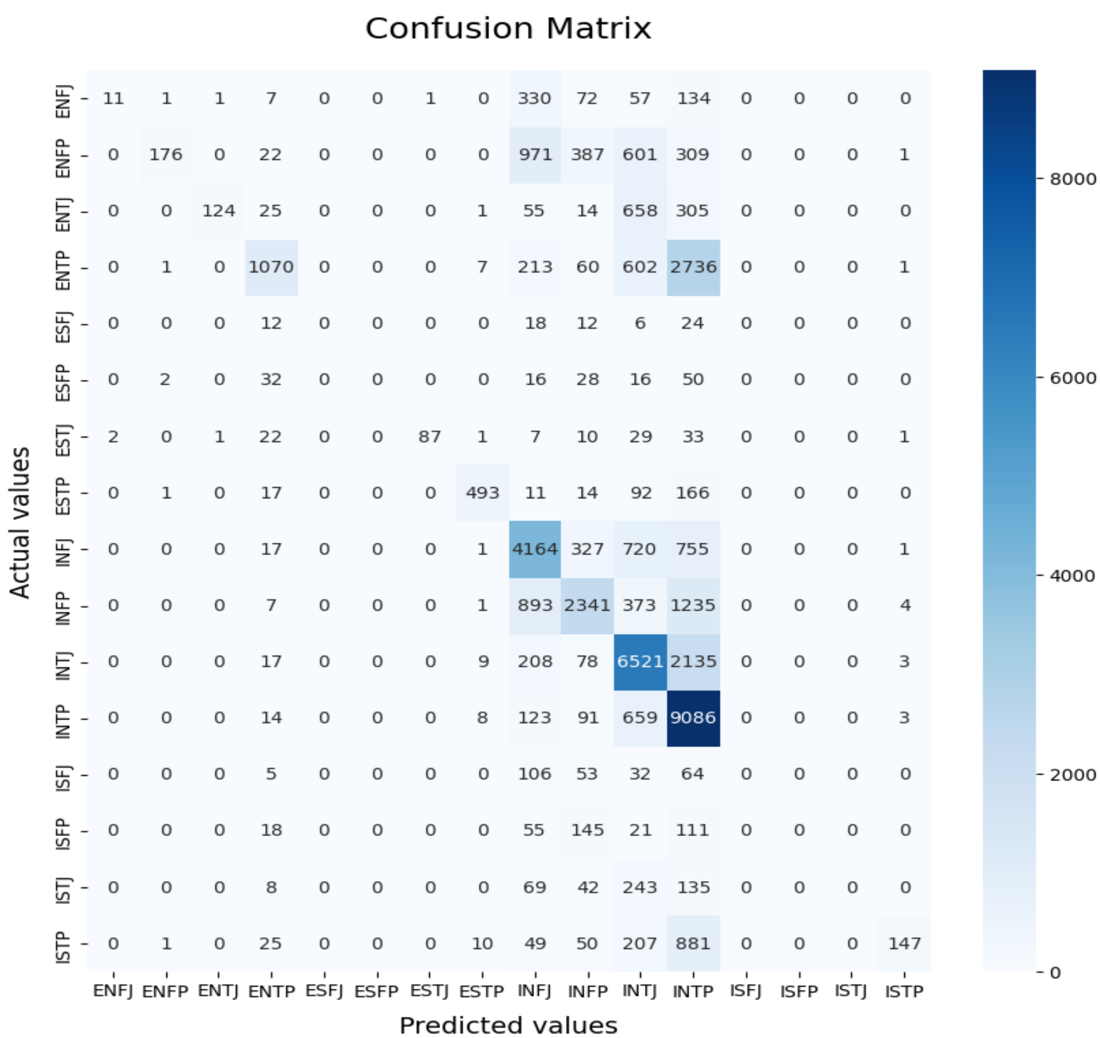


Figure 4.3: Confusion Matrix (Multinomial Naive Bayes)

Figure 4.3: showed the performance of a model that predicts sixteen personality types is evaluated in the test classification report that is provided. For every class, metrics such as precision, recall, and F1-score are provided, and the support shows how many samples there are in each class. However, there are significant variations in the model's performance between classes. The ENFJ class, for example, has a low F1-score of 0.04 given having a high precision of 0.85 and an extremely low recall of 0.02. In the same way, the precision, recall, and F1-scores of the ESFJ, ESFP, ISFJ, ISFP, and ISTJ classes are almost zero, indicating poor performance in recognizing these personality types. The macro and weighted averages show an unweighted and weighted average of precision, recall, and F1-score across all classes, respectively, and the overall accuracy is reported as 0.57.

Decision Tree Classifier

Achieving the Test Accuracy of Decision Tree Classifier is 63.34 % In below Figure 4:4 describing the confusion matrix of Decision Tree Classifier:

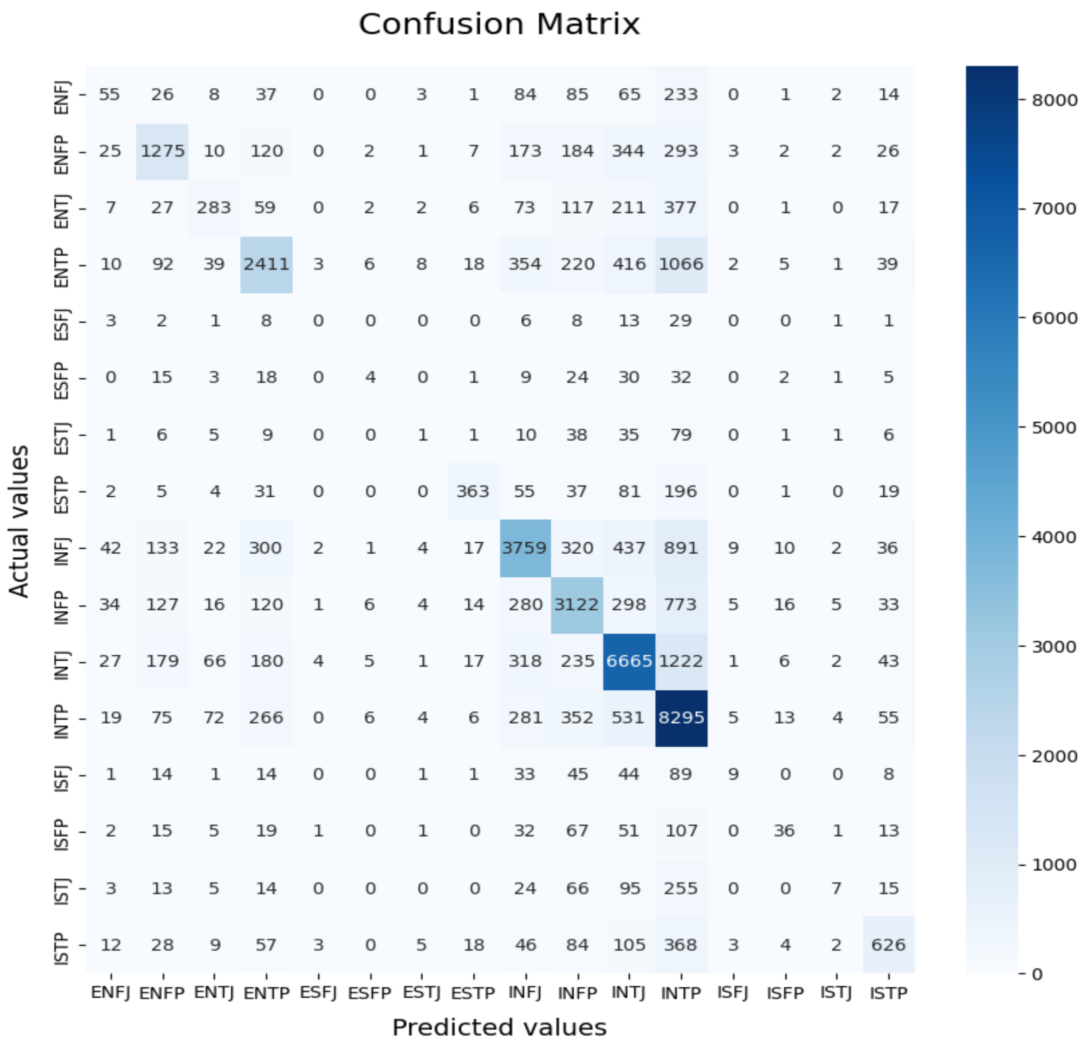


Figure 4.4: Confusion Matrix (DT)

Figure 4.4: shows the test classification report that is provided evaluates how well a model predicts sixteen different personality types. For every class, metrics such as precision, recall, and F1-score are provided; the support values represent the total number of samples. The model performs differently depending on the personality type. The INTJ class, for instance, shows good recall (0.74), precision (0.71), and F1-score (0.72), indicating successful personality type identification. However, low precision, recall, and F1-scores for some classes like ESTJ, ISTJ, and ESFJ indicate difficulties in accurately predicting these personality types. The percentage of accurate predictions for each class is represented by the overall accuracy, which is reported as 0.63. . Further information is provided by the

weighted average (0.62) and macro average (0.36), which take into account the influence of class distribution on the model's efficacy and show a moderate overall performance.

Random Forest Classifier

Achieving the Test Accuracy of RF Classifier is 64.92 % In below Figure 4:5 describing the confusion matrix of RF Classifier:

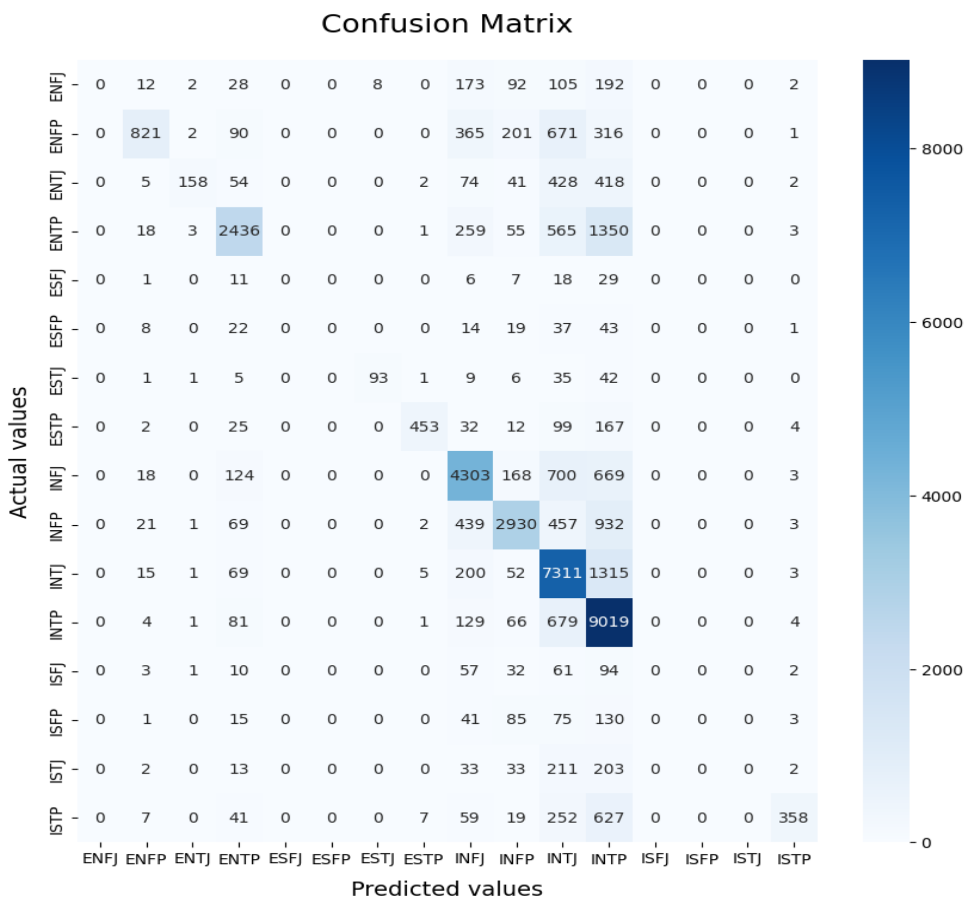


Figure 4.5: Confusion Matrix (RF)

Figure 4.5: shows the test classification report shows varying performance in various classes by evaluating a model that predicts 16 personality types. Certain personality types, such as ENFJ, ESFJ, ESFP, ISFJ, ISFP, and ISTJ, exhibit low recall, precision, and F1-scores, indicating difficulties in correctly predicting these types. However, higher precision, recall, and F1-scores have been shown by classes such as INTP, INTJ, and INFJ, indicating effective identification. The percentage of accurate predictions across all classes is 0.66, which represents the overall accuracy. Further information is provided by the

weighted average and macro average, which show that the model's efficacy is affected by class distribution and exhibit moderate overall performance.

Bi-LSTM Model:

Achieving the Test Accuracy of Bi-LSTM is 62.34 % In below Figure 4:6 describing the confusion matrix of Bi-LSTM:

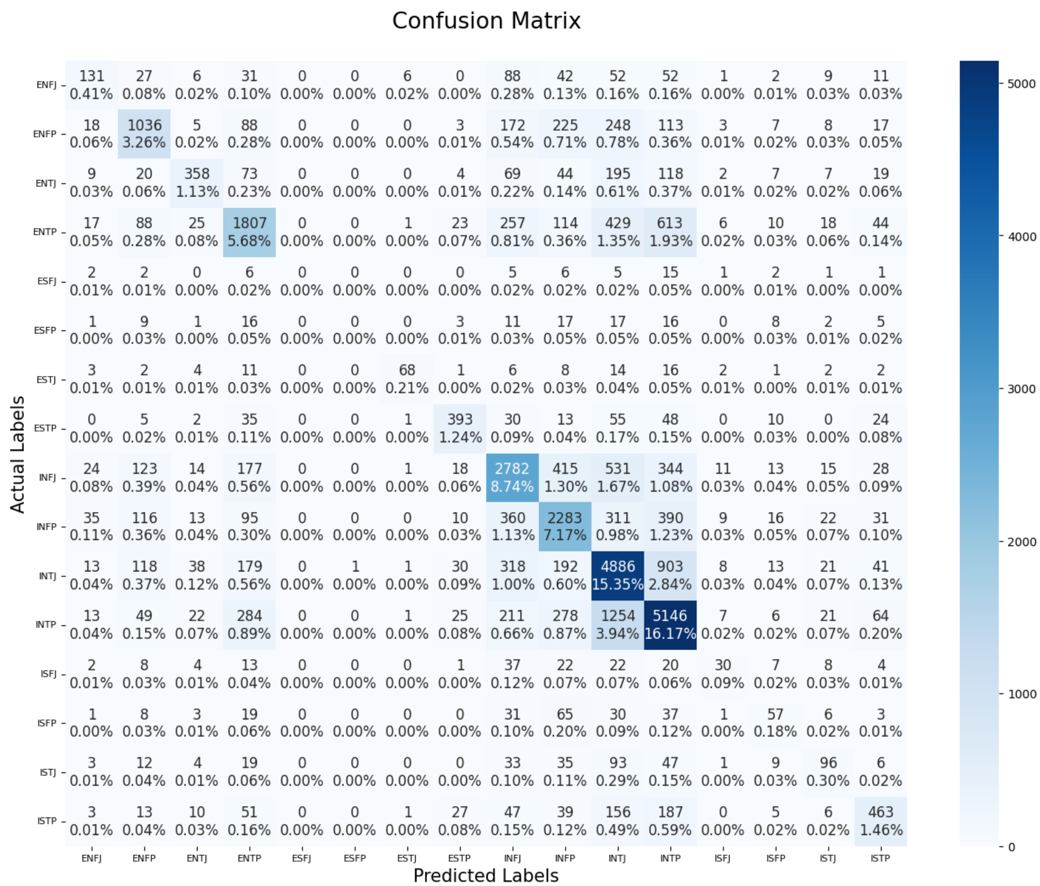


Figure 4.6: Confusion Matrix (Bi-LSTM)

Figure 4.6: shows the model's performance varies depending on the label; classes 4, 5, 12, 13, and 14 have particularly low precision, recall, and F1-scores, indicating difficulties in correctly predicting these labels. Relatively higher precision, recall, and F1-scores are shown by classes 6, 7, and 10, indicating successful identification. The overall performance across all labels is reflected in the micro average precision, recall, and F1-score, which are 0.67, 0.57, and 0.62, respectively. An unweighted average across all labels is provided by the macro average (0.44), which shows a mediocre overall performance. The weighted

average (0.61) takes into account how the distribution of classes affects the performance of the model. The samples' reported averages for precision, recall, and F1-score are 0.57, which shows that performance was consistent throughout the samples.

LSTM Model:

Achieving the Test Accuracy of LSTM is 59.62 % In below Figure 4:7 describing the confusion matrix of LSTM:

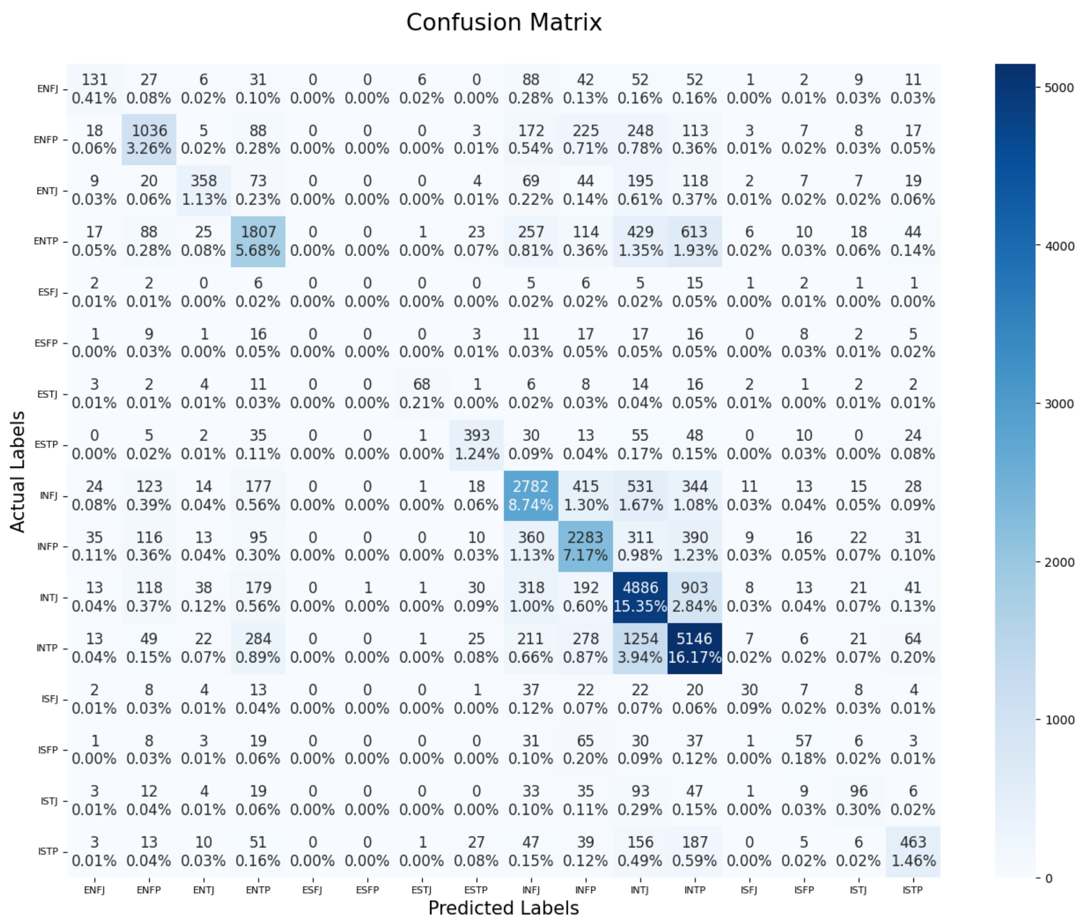


Figure 4.7: Confusion Matrix (LSTM)

Figure 4.5: shows the model's performance on a multi-label classification task is evaluated in the provided classification report. Notably, low precision, recall, and F1-scores are shown for classes 0, 4, 5, 12, 13, and 14, indicating difficulties in correctly predicting these labels. Relatively higher precision, recall, and F1-scores are shown by classes 6, 7, and 11, suggesting more successful identification. The overall performance across all labels shows

up in the micro average precision, recall, and F1-score, which are 0.67, 0.52, and 0.59, respectively. An unweighted average across all labels is provided by the macro average (0.36), which shows a mediocre overall performance. The weighted average (0.57) takes into account how the distribution of classes affects the performance of the model. The samples' average recall, precision, and F1-score are all reported as 0.52; this suggests that the samples' performance was consistent.

CNN

Achieving the Test Accuracy of CNN is 54.09 % In below Figure 4:8 describing the confusion matrix of CNN:

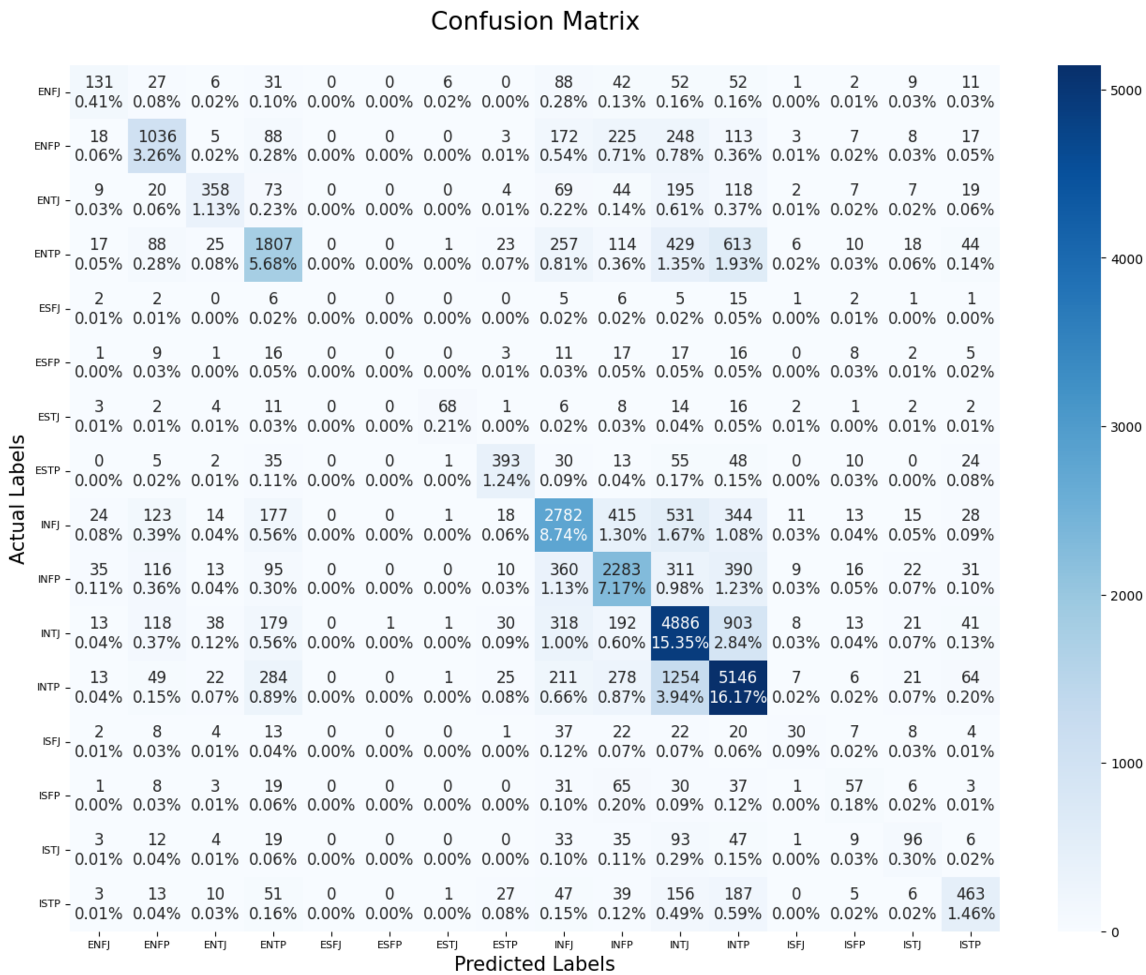


Figure 4.8: Confusion Matrix (CNN)

Figure 4.8: shows the model's performance on a multi-label classification task is evaluated in the provided classification report. Particularly low precision, recall, and F1-scores are

shown for classes 4, 5, 12, 13, and 14, indicating difficulties in correctly predicting these labels. Relatively higher precision, recall, and F1-scores are seen in classes 6, 7, 10, and 11, indicating more successful identification. The overall performance across all labels can be seen in the overall micro average precision, recall, and F1-score, which are 0.55, 0.54, and 0.54, respectively. An unweighted average across all labels is provided by the macro average (0.34), which shows a mediocre overall performance. The weighted average (0.54) takes into account how the distribution of classes affects the performance of the model. The average precision, recall, and F1-score for the samples are reported as 0.54; this suggests that the samples' performance is consistent with one another.

Accuracy: In Figure 4:9 shows the Accuracy Comparison Bar Plot Between Deep Learning And Machine Learning Models

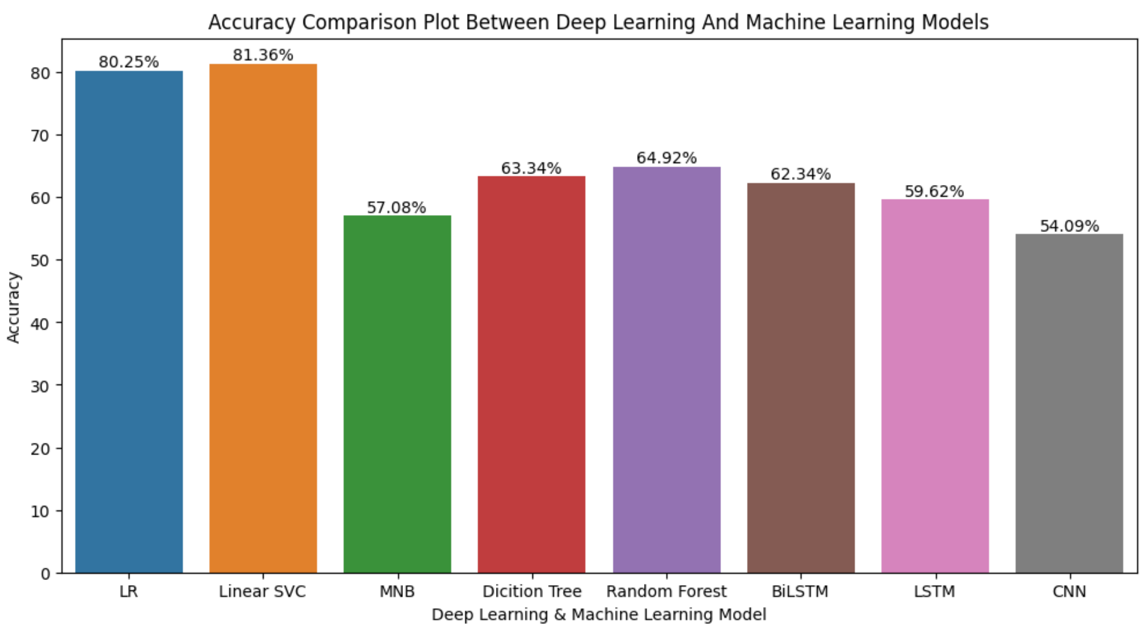


Figure 4.9: Comparative Model Accuracy Bar Plot

In figure 4.9 The accuracy comparison bar plot shows the efficacy of different algorithms in predicting personality from social media posts. Linear SVC beats Logistic Regression (80.25%) in terms of accuracy among the methods. Decision Tree (63.34%) and Random Forest (64.92%) score well in terms of accuracy, but Multinomial Naive Bayes (57.08%) performs slightly worse. Deep learning models with lesser accuracies include BiLSTM

(62.34%), LSTM (59.62%), and CNN (54.09%). These findings suggest that linear models and ensemble approaches are especially successful in capturing the complex linguistic patterns associated with various personality traits in social media data.

4.3: Discussion

Algorithms developed using deep learning are used to predict personality based on social media posts, and the conversation discusses the consequences, limitations, and future directions of this research. Analyzing the obtained data in light of the study's goals, the discussion explains the model's advantages and disadvantages.The conversation includes ethical issues related to user privacy, data security, and responsible AI use, highlighting the significance of open and moral behavior. Ensuring fairness requires tackling any biases in the model predictions and their possible effects on various user groups. Personality prediction models are examined for their potential uses in educational settings, mental health treatments, and targeted marketing. This exploration opens up new research directions and practical applications for personality prediction models. All things considered, the conversation offers a thorough analysis of the study's findings and opens the door for more developments in the field of deep learning and personality prediction using data from social media.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Deep learning algorithms that predict personality characteristics from social media posts have a wide-ranging impact on society, changing both individual and group experiences in the digital age. The fact that internet interactions are individualized is one notable effect. Social media companies may provide users with more relevant and interesting content, ads, and suggestions by using deep learning to understand personalities. By carefully targeting consumers based on expected personalities, firms can enhance their marketing techniques, resulting in more ethical and effective advertising practices. Furthermore, social media analysis can help with early diagnosis and monitoring in the field of mental health, which can lead to swift actions and assistance for individuals who may be at risk. Interestingly, issues with privacy and the appropriate application of predictive models come up. The ethical implications of using one's own data for forecasting highlight the necessity of open and responsible procedures. It is critical to strike a balance between user privacy protection and customisation in order to ensure that the positive impacts of personality prediction are consistent with ethical norms and social values. The ethical creation of personality prediction technologies into our digital interactions will be shaped by continuing conversations and legislation as society navigates this revolutionary landscape.

5.2 Impact on Environment

The application of deep learning algorithms for personality prediction based on social media posts has an influence on the environment, but mostly because of the computational resources needed for model training and deployment. Deep learning models require a significant amount of processing power, leading to higher energy use and carbon emissions, particularly when working with large datasets. Whether it is hosted locally on servers or in the cloud, the infrastructure that powers these algorithms has an opportunity to increase greenhouse gasses and harm the environment. The IT industry needs to start using sustainable methods as the requirement for processing power increases. The introduction of energy-efficient hardware, algorithm optimization for lower computing requirements, and the utilization of renewable energy sources for data centers are some of the steps taken to lessen the environmental impact. In order to bring together the goals of environmental sustainability with technology breakthroughs in personality prediction, eco-friendly techniques should be given priority by researchers and practitioners. Furthermore, it is important to take into account the long-term environmental effects of widely using deep learning models in a variety of applications. To this end, green computing principles should be incorporated into the creation and implementation of these algorithms.

5.3 Ethical Aspects

Fundamental concerns regarding user permission and data privacy are brought up by the analysis of personal data. Ensuring the anonymization of sensitive data and providing clear transparency about data usage are essential to upholding individuals' right to privacy. Another ethical concern is the transparency of the deep learning algorithms. Certain models' "black-box" characteristics can make it difficult to comprehend how decisions are made, therefore efforts to improve interpretability and explainability are required. Limiting stereotypes and unjust predictions from being maintained requires addressing biases included in training data or algorithms. The process of developing a model should incorporate fairness and equity, and it is crucial to continuously check for unforeseen effects. In addition, wise application of personality prediction models requires protections against any abuse, guaranteeing that forecasts are applied morally and do not support discriminatory actions. A proactive ethical framework is necessary to strike a balance between the advantages of individualized insights and user rights protection. Establishing and carrying out personality prediction models from social media data requires researchers and practitioners to place a high priority on openness, equity, and user consent in order to build confidence and respect moral principles.

5.4 Sustainability Plan

Applying deep learning algorithms to predict personality based on social media posts requires a sustainability plan that addresses long-term viability, ethical behavior promotion, and environmental effect minimization. First off, the carbon footprint of training and deploying models can be greatly decreased by implementing energy-efficient hardware and optimizing algorithms for lower computing demands. A long-term strategy must also take the ethical implications of data use into account. Strict privacy policies, clear consent from users, and confidentiality of personal data respect moral principles and foster user confidence. It is essential to continuously check for biases and unexpected results. The model's meeting moral standards and social conventions can be guaranteed by routine audits of its fairness and openness. In an environmental sustainability plan, increasing the use of renewable energy sources for computer infrastructure is essential. Making the switch to green computing methods is in line with more general environmental preservation objectives. In conclusion, a long-term plan for personality prediction in social media posts takes an integrated strategy that takes ethical issues, environmental responsibilities, and energy efficiency into account. By following these guidelines, deep learning model creation and implementation become more environmentally friendly and encourage the responsible use of predictive technology.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

In the end, the research on deep learning algorithms for personality prediction based on social media posts is a leading investigation into the integration of predictive analytics and natural language processing. Using an array of datasets that were obtained from sites such as Kaggle, the study used advanced algorithms like CNN, LSTM, and BiLSTM to identify patterns in textual data and uncover hidden personality attributes. The study balanced ethical issues by placing a heavy focus on model creation, fairness, privacy, and transparency. Concerns regarding the influence on the environment were addressed by optimizing the computing infrastructure, an essential element for sustainability and efficiency. The results of thorough statistical analyses and model evaluations provided insight into the model's prediction skills, highlighting its advantages, disadvantages, and potential areas for development. In summary, this study uses cutting edge methods to expand our understanding of personality inference from social media information.The results influence the conversation about tailored encounters in the digital age and add to the changing field of predictive analytics.

6.2 Conclusions

In summary, the research on deep learning algorithms for personality prediction based on social media posts shed light on the dynamic relationship between technology and human behavior. The study effectively illustrated the possibility of utilizing textual data to predict personality traits, highlighting the potential uses in mental health monitoring, education, and targeted marketing. The study was conducted with ethical issues in mind, placing a strong emphasis on user privacy, openness, and justice in model building. A common subject was the proper application of predictive algorithms, emphasizing the need for continual bias analysis and interpretability improvement. While the study presented positive results, it also recognized that there are unavoidable difficulties, such as the need for better deep learning model interpretability and a decrease of biases in training data. The results emphasized the significance of a well-rounded strategy that takes advantage of technology breakthroughs while upholding moral and ecological standards. This study contributes to the larger conversation on responsible and significant uses of deep learning for comprehending and predicting human behavior through social media information as we traverse the rapidly changing field of AI-driven personality prediction. The findings open up new avenues for investigation and support a methodical and comprehensive approach to the creation and application of predictive technologies in the age of technology.

6.3 Implication for Further Study

The effects for additional algorithms-based personality prediction research based on social media posts are significant. Although there are always opportunities for further investigation, the current study provided a framework for evaluating the predictive power of advanced algorithms in identifying personality characteristics. In the beginning, it's critical to investigate how well models translate into other languages and cultural situations. Improving the resilience and inclusiveness of personality prediction systems requires examining the extent to which models trained on a single dataset generalize to different populations. The comprehension and explanation of deep learning models can be improved with more investigation. Improving decision-making processes' transparency promotes user trust and makes responsible deployment in a variety of applications easier. Gaining knowledge on how anticipated personalities impact online interactions over time might help you better understand how the dynamics of digital communication are changing. Additionally, the study creates opportunities for multidisciplinary research in which social scientists, psychologists, and ethicists will all add different viewpoints to the conversation. Working together can result in the creation of thorough frameworks that address the psychological, social, and ethical elements of personality prediction in the digital sphere. Our understanding will be improved and safe innovation in this developing sector will be encouraged by more research in these areas.

REFERENCES

[1] Valanarasu, R. "Comparative analysis for personality prediction by digital footprints in social media." Journal of Information Technology and Digital World 3.2 (2021): 77-91.

[2] Al Marouf, Ahmed, et al. "Comparative analysis of feature selection algorithms for computational personality prediction from social media." IEEE Transactions on Computational Social Systems 7.3 (2020): 587-599.

[3] Zhu, Yu. "The prediction model of personality in social networks by using data mining deep learning algorithm and random walk model." International Journal of Electrical Engineering Education (2020).

[4] Garg, Shruti, and Ashwani Garg. "Comparison of machine learning algorithms for content based personality resolution of tweets." Social Sciences & Humanities Open 4.1 (2021): 100178.

[5] Lima, et al. "A multi-label, semi-supervised classification approach applied to personality prediction in social media." Neural Networks 58 (2014): 122-130.

[6] Al-Fallooji, et al. "Predicting Users' Personality on Social Media: A Comparative Study of Different Machine Learning Techniques." Karbala International Journal of Modern Science 8.4 (2022): 617-630.

[7] Tadesse, Michael M., et al. "Personality predictions based on user behavior on the facebook social media platform." IEEE Access 6 (2018): 61959-61969.

[8] Tandera, Tommy, et al. "Personality prediction system from facebook users." Procedia computer science 116 (2017): 604-611.

[9] Khan, Alam Sher, et al. "Personality classification from online text using machine learning approach." International journal of advanced computer science and applications 11.3 (2020): 460-476.

[10] Wei, Honghao, et al. "Beyond the words: Predicting user personality from heterogeneous information." Proceedings of the tenth ACM international conference on web search and data mining. 2017.

[11] Tay, Louis, et al. "Psychometric and validity issues in machine learning approaches to personality assessment: A focus on social media text mining." European Journal of Personality 34.5 (2020): 826-844.

[12] Imran, et al. "An enhanced approach for quantitative prediction of personality in facebook posts." International Journal of Education and Management Engineering (IJEME) 8.2 (2018): 8-19.

[13] Souri, et al. "Personality classification based on profiles of social networks' users and the five-factor model of personality." Human-centric Computing and Information Sciences 8.1 (2018): 1-15.

[14] El-Demerdash, Kamal, et al. "Deep learning based fusion strategies for personality prediction." Egyptian Informatics Journal 23.1 (2022): 47-53.

[15] Golbeck, Jennifer, et al. "Predicting personality from twitter." 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. IEEE, 2011.

[16] Sirasapalli, Joshua Johnson, and Ramakrishna Murty Malla. "A deep learning approach to text-based personality prediction using multiple data sources mapping." Neural Computing and Applications 35.28 (2023): 20619-20630.

[17] Kunte, Aditi V., and Suja Panicker. "Using textual data for personality prediction: a machine learning approach." 2019 4th international conference on information systems and computer networks (ISCON). IEEE, 2019.

[18] Salminen, Joni, et al. "Enriching social media personas with personality traits: A deep learning approach using the big five classes." Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22. Springer International Publishing, 2020.

[19] Aung, Zan Mo Mo, and Phyu Hninn Myint. "Personality prediction based on content of facebook users: a literature review." 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). IEEE, 2019.

[20] Saitov, Irek, Anatoly Surikov, and Leonid Gorokhovatsky. "Analysis of the relationship between the user's personality traits and the images they post on social media." Procedia Computer Science 193 (2021): 155-162.

[21] Maharani, Warih, and Veronikha Effendy. "Big five personality predictions based on Indonesian tweets using machine learning methods." International Journal of Electrical & Computer Engineering (2088-8708) 12.2 (2022).

[22] Kunte, Aditi V., and Suja Panicker. "Using textual data for personality prediction: a machine learning approach." 2019 4th international conference on information systems and computer networks (ISCON). IEEE, 2019.

PERSONALITY PREDICTION BASED ON SOCIAL MEDIA POSTS

ORIGINALITY REPORT

22%	18%	9%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
2	Submitted to Loomis-Chaffee High School Student Paper	3%
3	Submitted to Daffodil International University Student Paper	3%
4	Submitted to University of Bedfordshire Student Paper	1%
5	assets.researchsquare.com Internet Source	1%
6	discovery.researcher.life Internet Source	1%
7	ijsrset.com Internet Source	<1%
8	Submitted to BPP College of Professional Studies Limited Student Paper	<1%
9	Submitted to University of East London Student Paper	<1%

10	ntnuopen.ntnu.no Internet Source	<1 %
11	www.semanticscholar.org Internet Source	<1 %
12	ijisrt.com Internet Source	<1 %
13	"International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019", Springer Science and Business Media LLC, 2020 Publication	<1 %
14	link.springer.com Internet Source	<1 %
15	Michael M. Tadesse, Hongfei Lin, Bo Xu, Liang Yang. "Personality Predictions Based on User Behaviour on the Facebook Social Media Platform", IEEE Access, 2018 Publication	<1 %
16	Mohmad Azhar Teli, Manzoor Ahmad Chachoo. "Lingual markers for automating personality profiling: background and road ahead", Journal of Computational Social Science, 2022 Publication	<1 %
17	dokumen.pub Internet Source	<1 %