Detecting Self-Esteem Level and Depressive Indication Due to Different Parenting Style Using Supervised Learning Techniques

by

Abdullah Al Taawab
20341043
Mahfuzzur Rahman
18101485
Zawadul Islam
18101005
Nafisa Mustari
19101086

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
May 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Taawab

Abdullah Al Taawab
20341043

Samir

Mahfuzzur Rahman
18101485

Zawad

Zawadul Islam
18101005

NAFISA

Nafisa Mustari
19101086

Approval

The thesis titled "Detecting Self-Esteem Level and Depressive Indication Due to Different Parenting Style Using Supervised Learning Techniques" submitted by

1. Abdullah Al Taawab(20341043)
2. Mahfuzzur Rahman(18101485)
3. Zawadul Islam (18101005)
4. Nafisa Mustari (19101086)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May, 2022.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

24/5

Shaily Roy
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Uprising a child is a psychological construct of parents, which is a combination of factors that evolves over time with the growth and development of the child. Parenting style represents a set of strategies that have diverse influences on children. These approaches can create depressive symptoms in children's minds, which can last even if they become adolescents. Moreover, these indications may affect their level of self-confidence. In this research, supervised learning models are used to detect different parenting styles, depression indications of adolescents due to parenting and the level of their self-esteem. Due to the absence of publicly available data, we created our own data set of about 500 survey responses. Additionally, eleven psychological and nine linguistic attributes of Linguistic Inquiry and Word Count (LIWC) have been used to identify depression indications. Among all the supervised models, the Logistic Regression (LR), Gradient Boost Classifier (GBC) and Bi-Directional LSTM (Bi-LSTM) provide better results than other models. This research is capable of helping the parents to know their children's psychology in a better way and make them have a more profound discussion on practical life.

Keywords: Machine Learning, Deep Learning, LIWC, NLP, Depression, Parenting style, Self Esteem

Acknowledgement

First of all, all praise to Allah (SWT) who is the most merciful and helped us to complete this study thoroughly.
Secondly, special thanks to our supervisor Md. Golam Rabiul Alam and co-supervisor Shaily Roy who assisted us by providing support and necessary resources whenever we needed.
Finally, thanks to our parents who gave us hope and inspired us to do more hard work. With their prayer and support, we have reached where we are now.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$BI - LSTM$ Bidirectional Long Short-Term Memory

CNN Convolutional Neural Network

DL Deep Learning

DT Decision Tree

GBC Gradient Boosting

GRU Gated Recurrent Unit

$LIWC$ Linguistic Inquiry and Word Count

LR Logistic Regression

$LSTM$ Long Short-Term Memory

ML Machine Learning

MNB Multinomial Naive Bayes

RF Random Forest

RNN Recurrent Neural Network

SVM Support Vector Machine

$TF - IDF$ term frequency-inverse document frequency

Chapter 1

Introduction

Parenting is the act of raising a child and caring for their needs so that they can grow up healthy and happy. The attribute of parenting is to process the development of a child's growth, education, and health which is implicated in the child's life undertaken by the parental figure [8]. Parents exert direct and substantial influence on their children through socialization. Parenting style can be described as a set of methods that the parents use to raise their kids. Baumrind found three central controls of parenting: Authoritative, Authoritarian, and Permissive [5]. On the other hand [12], Maccoby and Martin figured out one more parenting style, which they defined as Uninvolved. Each parenting style has a unique approach to raising children, and these differences can be seen in a variety of ways. There are short-term and long-term consequences on children's development of a parent's style. The authoritative parenting style allows children's independence while simultaneously enforcing support rather than punishing discipline measures. Some studies have shown that non-authoritative parenting has more beneficial effects than authoritative parenting despite the strong positive effect of authoritative parenting [2]. On the other hand, Authoritarian Parenting imposes rules that are more punitive than supportive. As a result, they forbid their child from developing characteristics and impose rules that are more punitive than supportive. In addition, this type of parent sets a one-way type of communication for their child and expects him or her to follow them without question. Moreover, the Permissive parenting style encourages their children to communicate openly with each other instead of enforcing, which results in fewer instances of discipline being used. Finally, there are parents who provide for their child's fundamental requirements while remaining apart from their child's day-to-day activities. They are referred to as "uninvolved parents" because of their lack of involvement.

Furthermore, research has found a link between parenting methods and the child's mental health. Mental illness occurs from different prejudice and discrimination that creates self-stigma, having various negative consequences including social isolation [9]. The development of a good mental state in children was facilitated by positive parenting techniques. Parents who do not understand and accept their children the way they are consequently their children may suffer from mental illnesses like neurosis and depression. As a result, the impact of parenting styles on mental health as well as the mechanisms that exert their influence should be studied. Parenting style remains the predicting factor of a child's mental health, and certain parenting styles can be a risk factor for a child's mental and behavioral disorder which can last from childhood to adulthood [21]. The interpersonal theory of acceptance and rejection impacts children's mental health. Parents being an important attachment of figures their positive attachment gives a positive effect. On the contrary, their rejection can have a negative effect on a child's psychology [26].

As parental activity has effects on mental health, it also influences one's self-esteem and decision-making abilities. Self-esteem is a psychological phrase that can be described as how much a person values and likes himself no matter what [22]. Michal et al. claim that better self-esteem leads to better mental health and social behavior. At the same time, poor self-esteem is associated with a broad range of mental disorders [3]. Numerous research and reviews have shown that an individual's origin of family and experiences as a family member have an impact on the general behavior and adjustment [6]. Furthermore, an adolescent's self-esteem might be high or poor, depending on parenting styles.

When a child grows up and becomes a young adult, he develops a sense of self-confidence. High self-esteem is associated with improved mental health, academic performance, proactive stress management, and low levels of trivializing difficulties [16] where low self-esteem is connected to self-damaging behaviors, anxiety, and depression. Also, people with high self-esteem has higher success rate and better socializing skills [7] [16]. As it is mentioned before that children's self-esteem varies upon their parent's guidance, several studies have been performed specifically finding the effect of four parenting styles [10]. In 2007, Martinez and Garcia found that children who have indulgent parents have high levels of self-esteem, whereas the authoritarian parent's children had the lowest. In the following year, they found out adolescents having indulgent parents had self-esteem levels that were equivalent to or higher than those with authoritative parents [4]. Pinquart and Gerke found a moderate amount of correlation between authoritative style and higher self-esteem among adolescents [16]. But these researches ensure one thing that some specific qualities of the parents can lead to the higher or lower quality self-esteem of their children.

1.1 Research Problem

Despite the fact that the world is progressing, the parenting style and its effects still appears to be taboo in many parts of the world. People are less vocal about this issue and hardly take any steps. As a result, the adolescent becomes more depressed and mentally vulnerable. In addition, their level of self confidence also falls down drastically. According to a report of WHO, one out of every seven (14%) adolescents is associated with depressive symptoms [28]. Another study done by Harter et al, low self-esteem affects one-third to one-half of adolescents, especially those in their early adolescence. Despite this, a significant number of them went unnoticed and untreated. It is difficult to determine whether an individual is in danger or needs care in the psychological health research field as there are some complexities and different aspects linked with it. Suicidal behaviour is more likely to be linked to low self-esteem and stress in adolescents. Our work contributes substantial fresh ideas and perspectives. In AI literature, self-esteem has been only referenced once or twice as part of a debate on depression.

An exclusive interview with the BRAC university counseling unit was conducted in this study. According to the counseling unit, simply determining whether or not a young individual is depressed is insufficient to recover by medical science. They also claimed that the children are not comfortable sharing all their information with the counselor. So, it becomes tough to find out the root causes and the impact of the parents on these reasons. It is important to figure out the root causes of depression and low self esteem in the first place, and then see how much effects parents have on it.

1.2 Research Objectives

The fundamental goal of this research is to identify various parenting methods, as well as depression indications due to different parenting styles and self-esteem levels using Machine Learning, Deep Learning, and data analysis approaches by filling up some Open-ended questionnaires. It can be anticipated that this system will be helpful to the mental health counselors and the parents. The proposed model will be able to determine whether an adolescent has depression indication or not as a result of parenting, in addition to whether his self-esteem is high or low. Furthermore, there are not sufficient studies conducted on this topic. An appropriate dataset will be developed in order to accomplish the research's goal and evaluate the introduced classification models.

1.3 Contribution

The following are the major contributions of this work:

1) A survey was conducted applying a questionnaire consisting of some open-ended questions in this study. The questionnaire was created being inspired by DASS42, PAQ, and Rosenberg's self-esteem questionnaire. This questionnaire was reviewed and evaluated by the counseling unit of BRAC University. Then after several months of survey, around 500 responses of adolescent people were recorded which was utilized to build a ground truth dataset. The counseling team evaluated the dataset and provided some labeling instructions. Based on their suggestions, we labeled the dataset accordingly.

2) The LIWC package was used on the collected data. We utilized evidence from a wide range of sources, including expression of emotion, user involvement, and egocentric social network, to uncover depression-related reactions. These measures include eleven psychological attributes from the LIWC dictionary. Additionally, nine more linguistic features were picked for characterizing the user's responses. By using these features, another dataset was developed to build and test the models.

3) Finally, several supervised models were implemented on these three datasets to detect parenting style, depression indications due to parenting and levels of self esteem.

We believe that the dataset we established through the survey and the information derived from this study will be valuable in future research. In addition, those adolescents who experience emotional imbalancement due to parenting style would also be benefited.

Chapter 2

Related Work

In order to detect low self-esteem in youths, Zaman et al. have used several machine learning algorithms [18]. The data were collected from 108 individuals, where 40 people were male, and the rest were female. The medium they used for collecting the data surveyed and search history from the participants. After collecting data, the dataset was divided into two parts- low self-esteem (LS) and no low self-esteem (NLS). Google cloud NLP API classified the searches into categories. And it is found that LS groups search more for educational topics at late night, which leads them to mental disturbance. Also, there was the use of the LIWC text analysis toolkit to find different variables. It was observed that family-related variables were lower, which means family impacts a lot. For training and testing data, the researchers used LR, SVM, and HyBaR (Hybrid Bayesian Regression) algorithms. HyBaR algorithm is a hybridization of Bayesian linear and Bayesian logistic regression. Because of additional information, HyBaR has a benefit over Maximum Likelihood Estimation. This algorithm has delivered better performance than SVM and LR. When the algorithms were used on linguistic attribute features, the F1 score of HyBaR was 0.72. Where SVM and LR scored 0.61 and 0.46, respectively. In the search category attribute, the HyBaR also performed better than the other two algorithms.

Islam et al. considered social networks as a promising instrument to detect depression [13]. Islam et al. built a Machine Learning model for detecting Depression indicative posts. He has used NCapture for collecting data from Facebook. LIWC was used to analyze the raw data, and then he built the ground truth dataset. In this paper, LIWC was also used for feature extraction, and four popular supervised ML models like SVM, DT, KNN, and Ensemble Classifiers were used to detect depression indicative posts. He has discussed the temporal process, linguistic style, and emotional process for classifying depression indicative posts. Among all these ML models, DT gains the highest result.

Orabi et al. studied how depression can be detected using Twitter data [14]. In this study, two publicly available datasets were used - CLPsych2015 and Bell Let's Talk. They encoded the data using several word embedding techniques. Then they used three varieties of CNN and one variety of RNN. The AUC score was used to evaluate the model's performance. The AUC score was calculated on the validation set and averaged over the five splits with standard deviation. Among all these models, the CNNWithMAx performed with the highest AUC score, which is 0.951.

Subreddits, or smaller communities within Reddit, are popular places for people to hold online discussions on a wide range of topics. Stigmatized subjects are frequently discussed on this platform because of their complete confidentiality. Tadesse et al. focused on the detection of depression-related posts among Reddit users [17]. He has used a publicly available dataset and built a classification model to detect depression. He claimed that a list of terms was more commonly used by those who are depressed. To evaluate the performance of the applied ML models, he has utilized both single features (bigram) and combined features (LIWC+LDA+UNIGRAM). In this study, LIWC+LDA+bigram with MLP model gained the highest accuracy for depression detection. He discussed why actual feature selection and various features combination is really important to improve the performance of the classifying models.

Nowadays, emotional texts that are posted on social media gain the focus of researchers, and they analyze those posts. Liao et al. proposed to distinguish the depression tendency using text on microblogs [24]. Here, Microblogs are the source of the dataset as a group of people liked to give depressional posts regularly on this platform. Liao et al. introduced the LSTM model with the TensorFlow framework in the python tool library to detect depression tendencies. First of all, the exquisite the data, then they preprocess it. At last, they constructed the Long Short term memory model. Furthermore, this paper also worked on the word vector training tool Word2vec to train the denoised dataset. The LSTM model got the highest accuracy after 25 iterations in this paper. Moreover, they also tried SVM, but the model didn't give any significance.

GUOZHENG et al. developed two upgraded CNN models to identify depressed individuals in online forums [19]. In this work, they fetched the RSDD dataset, which contains a training, testing, and validation dataset of 3,000 depressed diagnosed users and 35,000 control user's posts. In addition, they used The eRisk 2017 dataset as well, consisting of training and testing sets of 135 depressed diagnosed users and 752 control user's posts. Moreover, they used a sparse bag of words feature and emotion lexicon feature for data extraction. For the depression detection task, they performed BoW-SVM, Feature-rich-SVM, User-model CNN, Bi-LSTM, Bi-LSTM attention, and many more models. However, they proposed two CNN models: MGL-CNN and SGL-CNN, which performed comparatively better than other models. Finally, they encouraged everyone to explore MGL-CNN and SGL-CNN models to find out about depressed users on social media.

Guangyao Shen et al. suggested a multimodal depressive dictionary learning model detect depression via harvesting social media [11]. They established two benchmarks, a well-labeled dataset, D1, which has 1402 Twitter user's posts, and D2, which contains 300 million Twitter user's posts. However, the depressed users are too few in D1, so they constructed an unlabeled dataset D3 with 36,993 depressed Twitter users candidates. For the feature extraction, they used some important features such as emotional features, which extracts features using LIWC, topic-level feature using LDA, user Profile feature, social network feature, visual feature, and domain-specific feature. Moreover, they performed some classification models such as NB, MSNL, WDL, and their proposed Multimodal Depressive Dictionary Learning (MDL). According to their claim, MDL outperformed all the classification models, and they suggested using this model for further depression detection research in computer science and psychology.

Uban, A.S. Rosso, P. conducted a study to predict self-harm and depression levels among social media users considering Machine learning classifiers and Deep Learning architectures [20]. The training datasets were collected from available Reddit posts and history. Style, content, emotion, sentiment, and the LIWC feature were utilized to determine self-harming inclinations among Reddit users. There was the implementation of several neural network architectures for sampling training data, including Bi-LSTM with attention, Hierarchical Attention Network, Transformer, and Ensemble. Various evaluation metrics were used to know if the prediction matches true labels like AHR, ACR, ADODL, and DCHR. Moreover, traditional models such as SVM and Logistic Regression incorporating emotion and style characteristics were tested in this study. Finally, the best score was obtained in the last task of SVM and Logistic Regression in terms of detecting depression levels.

The study performed by Wongkoblap et al. suggests two novel ML models detect Depressiondepression among Twitter users where one model uses anaphoric resolution, and the other does not [27]. The dataset was collected from Twitter using an API. The API helps to search and download tweets with regular expressions. After cleaning the dataset, GloVe was used to create a vector of words to transform tweets into word embedding. The MIL approach was used to develop two models: MIL - SocNet and MILA-SocNet. After the preprocessing of tweets, these models were trained and then evaluated. The evaluation was done using AIC, which is a widely used tool for comparison between models. It also calculates the model's complexity and information loss. The MILA-SocNet performed better than other models like deep learning, LIWC, and user two vec with the highest accuracy and precision.

Christopher Rauh and Laetitia Renee use a machine-learning algorithm to classify parenting behavior through unsupervised machine learning by the dataset of the following children whose age is between 5 to 29 months [23]. The dataset is collected from 1985 families who participated in an interview about the observation of family life. After sampling, they worked with 1442 mother-children pairs. They classified two behavioral types of parenting- "active" and "laissez-faire" towards their children by using an ML model based on the latent Dirichlet allocation. The final result depicts the action distribution for each type and the parent type distribution. Although it is known that there are four types of parenting styles, their research shows only two different types of parenting styles which are improved in our research. We are trying to classify four parenting styles.

Ms. Sumathi M.R. and Dr. B. Poorna proposed some machine learning techniques to predict mental health problems among children by analyzing medical data [30]. The data has been collected by an interview which is taken from a clinical psychologist to know about mental illness that often occurs among children. Ten cases were sampled to complete this study. Moreover, eight techniques were selected as they intended to produce correct results for their dataset. The selected techniques are Multiclass Classifier(MCC), AODEsr, FT, Multi-Layer Perceptron(MLP), RBF IB1, Network, KStar, FT, and LADTree. The ROC Area of 4 classifiers, Multilayer Perceptron, namely, LAD Tree, AODEsr, and K* Multiclass Classifier are between 0.8 and 0.9, so the accuracy level of these classifiers was better. In this research, the quantity of the sampled dataset was so small, and the children's appearance was missing in the process of collecting the dataset. We are trying to improve our research by questioning almost 500 to 600 students about their mental health.

Machine Learning algorithms were used in the research of Choudhury et al. for predicting Depression in Bangladeshi undergraduates [15]. The datasets were collected through a survey questionnaire consisting of 4 sets of questions. The questionnaire was inspired by the DASS21 and Beck Depression Inventory. Out of 935 responses, 577 responses were left after data cleaning. At first, there were 20 features. After the conversion of features using dummy variables, there were a total of 45 features. Random Forest and SVM show better accuracy than K-NN. The most acceptable performance was delivered by the Random Forest algorithm with an accuracy of 75%. Also, the precision, recall, and F1 score of this algorithm are higher than other algorithms. The K-NN algorithm shows less accuracy because the amount of features is a bit high. If there were a written part in the survey and NLP was used, then the dataset could be more detailed and enriched.

Nimi, Y. & Miyaji, Y. two Japanese researcher's study was to Detect Depression in Japanese sentences using a Machine learning model [25]. The data was collected from the Largest site Tobyo which has information about patient's experiences. They picked random 460 users from different Ameba blogs and retrieved data from 166,312 articles. After collecting data, the preprocessing was done by using a Japanese Tokenizer called Sudachi5. They eliminated the parts of speech, pictograms, normalized words, and converted numbers in preprocessing. Here, the elimination of parts of speech was quite easy to determine a depressing topic. Then there was a comparison between the depressed group and the control group. However, it found that eliminating Depression and depression-related topics is much easier to detect depression. So, the LDA model was used to reduce depression-related topics, TF-IDF vectorized the sentences, SVD reduced computation, and SVM did binary classification. To evaluate the accuracy and F1 value, there was a comparison between different topic models. At last, they preferred the model with limited parts of speech, depression topic withheld, and some adjustment with the accuracy of 0.956 and F1 with a value of 0.959.

Chapter 3

Methodology

The Figure 3.1 represents the Proposed Architecture Diagram of our research. Initially, a dataset was constructed based on the survey responses. During data analysis, some irrelevant data was deleted and the remaining data were labeled accordingly. Then the data were preprocessed applying some techniques such as normalization, one hot encoding and tokenization so that the computer can understand the data properly. The features were then extracted from the preprocessed data. Afterwards, the data were splitted into training and testing dataset which were used for training the ML and DL models. The performance of these models were compared and evaluated through different performance metrics.

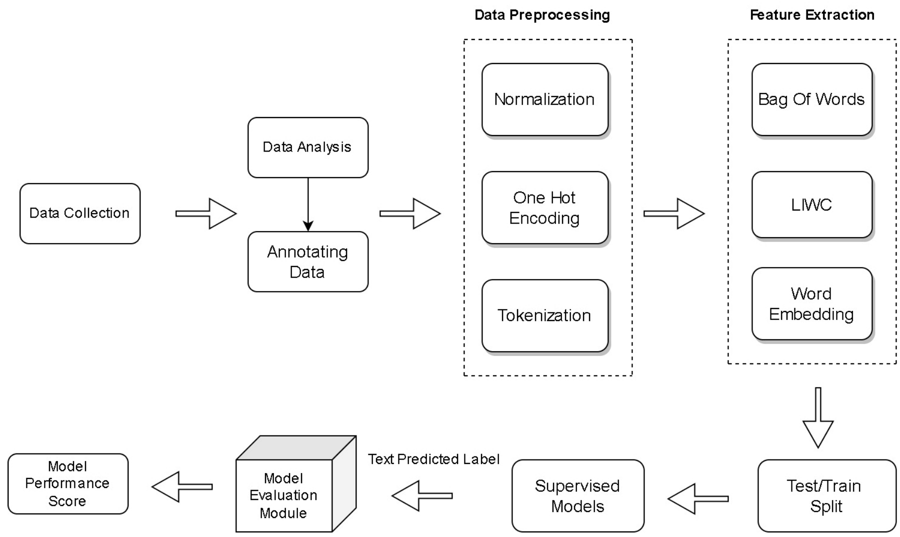

Figure 3.1: Proposed Architecture Diagram

3.1 Dataset Collection

In the AI literature, the discussion about low self-esteem and depression due to parenting style is hardly found. As a result, no publicly available datasets were found for conducting our study. Because of this unavailability, a survey through google forms was conducted. Many reputable websites and questionnaires were used in the process of producing survey questions in order to better understand parenting styles and the effects they have on children. To identify parenting styles, our survey question was influenced by the Parental Authority Questionnaire (PAQ). DASS21 and DASS42 were used as models for the mental state questions [29]. Furthermore, by getting motivated by Rosenberg's Self Esteem Scale [1], the questions of self-esteem were created. In the questionnaire, five multiple-choice questions were used to identify the different parenting styles. Then a total of eight open-ended questions were given about depression and self-esteem so that people could express their feelings freely. While providing the questionnaire, the participants were requested not to overthink and answer the questions honestly. Moreover, they were given the liberty to write as briefly as possible. About 500 data were collected after conducting the survey. Among these responses, the male response is 207, and the female response is 293.

3.2 Data Filtering

The audience used different approaches while filling out the survey. Some of the responses were written in Bangla and then transliterated into Bengali languages. These kinds of responses were eliminated. There were a few emoticons that also were taken out. The punctuation marks such as full stop, comma, and semicolons were then removed. The responses that were less than two words in length were also eliminated. Lastly, to eliminate the words which do not bear much meaning to a sentence were eliminated using stopwords. Figure 3.2 represents the workflow of the Data Filtering process.

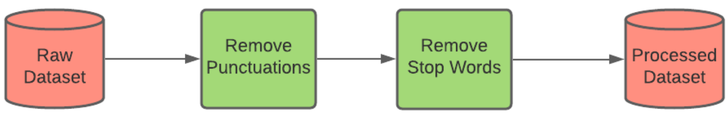

Figure 3.2: Data Filtering Workflow

3.3 Annotation Guidelines

The annotation task requires labelling the data sets. Suggestions from the counselling unit of BRAC University were taken about how the responses will be labelled. Responses related to parenting methods have been labelled authoritarian, authoritative, permissive and uninvolved. The options of each question represent the individual parenting style. The characteristics that portrayed the parenting style mostly from the audience's response were labelled accordingly. The effects of parenting on the mental state have been labelled into two categories-

- **Depression Indicative**: The responses of sadness, irritability, restlessness and anxiety are labelled as depressive. Besides, hopeless and worthless reactions are also counted as depressive.
- **Non Depression Indicative**: The responses consisting of happy, hopeful, joyful and grateful responses are labelled as non-depressive.

The responses that were collected for self-esteem are mainly used for personal evaluation. Through this, the idea of one's confidence level and self-worth can be obtained. Parenting styles have been shown in numerous studies to have a tremendous impact on developing a child's self-esteem. Moreover, self-esteem have been labeled into two categories-

- **High Self Esteem**: The responses that portray one's belief in him/herself are mainly addressed as high self-esteem. Even after knowing and admitting one's weakness, he/she does not give up and goes on. These types of responses are labelled as high self-esteem.

- **Low Self Esteem**: The responses that show a lack of confidence and belief in themselves and tend to focus only on weakness are labelled as low self-esteem. Their responses contain disappointment, frustration and gloominess.

3.4 Dataset Statistics

As it is already mentioned, about 500 responses were recorded. Among these responses, the male response is 263, and the female response is 236. Considering the parenting style, there are 132 authoritarian responses, 285 authoritative responses, 36 permissive responses and 47 uninvolved responses. Regarding the mental state section, the Figure 3.3 shows that 244 responses are non-depression indicative and 256 responses are depression indicative. Here label 0 means non depression indication and 1 means depression indication. At last, the self-esteem section consists of 293 high and 207 low esteem responses.

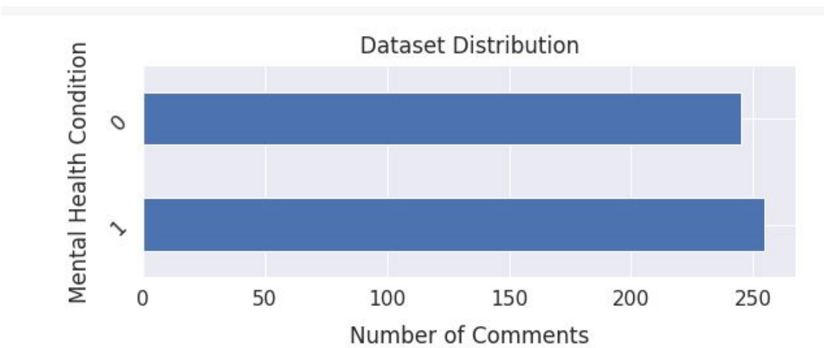

Figure 3.3: Depression indication detection Dataset Distribution

Table 3.1 and Table 3.2 illustrates the Length-Frequency distribution of depression indication and self-esteem dataset, respectively. In the depression indication detection questionnaire, the maximum length of the reply to the first question is 142 words, and the average length is 19 words. For the second question's reply, the maximum length is 109 words, and the average length is 17 words. In the next question's reply, the maximum length is 74 words, and the average length is 12 words. Finally, for the last question's answer, the maximum length is 93 words, and the average length is 12 words.

Table 3.1: **Length-Frequency Distribution of Mental Health Dataset**

Category	Question 1	Question 2	Question 3	Question 4
Maximum Length of a review	142	109	74	93
Minimum Length of a review	1	1	1	2
Average Length of a reviews	19	17	12	12

In the self-esteem questionnaire, the reply to the first question's maximum length is 173 words, and the average length is 14 words. In the second question's reply, the maximum length is 106 words and the average 12 words. Then in the third question's reply, the maximum length is found 107 words, and the average length is 17 words. In the last question, the maximum length is 126 words, and the average length is 18 words.

Table 3.2: **Length-Frequency Distribution of Self Esteem Dataset**

Category	Question1	Question2	Question3	Question4
Maximum Length of a review	173	106	107	126
Minimum Length of a review	1	1	1	2
Average Length of a reviews	14	12	17	18

3.5 Feature Extraction

To increase the accuracy of learned models and eliminate data redundancy, several strategies have been utilized to extract features. The techniques that were used for extracting the features are N-gram, LIWC and Word Embedding.

As the data set contains plain text, the n-gram model is used for exploring features. N-gram is mainly a probabilistic feature which usually works at the word level. This model is used in numerous research for examining characteristics of text-based data. At the same time, this model comes in handy to find out depression indications from the responses of the audience. This model stores n numbers of words in a token. Then these tokens are used for extracting the feature from a cell. For this study, the types of n-gram used are unigram and bigram. Unigram means a sequence of 1 word, and bigram means a two-word sequence. These n-grams are extracted from the texts, and then the TF-IDF score is calculated for each n-gram. TF-IDF is a statistical approach to highlighting the importance of a word in a particular text. These TF-IDF scores are inputted into the ML classifier. All words are transformed into lower case, and then features are extracted from the data.

The LIWC is an application that associates various lexico-syntactic features with words from the English vocabulary. This is also a widely used tool in computational studies and psychological analysis. LIWC extracts textual features by calculating the number of words that belong to LIWC lexicon categories. It delivers the percentage of words within that text that belong to one or more categories after the derivation. There are about 80 psychological, linguistic and thematic categories which represent diverse cognitive, social and affective processes. Also, LIWC categories can encompass a variety of levels of language, such as style, emotions, and topics. The style can be captured with the help of syntactic categories, where the topic can be captured from affect categories. Then the language about the topic can be known via content-oriented categories. Many research supports the fact that emotional terms reflect people's emotional feelings. For example, if someone uses more positive words in a response, it means he is feeling positive. What LIWC does is count the number of words in a specific category. Also, people leave a trace of their behavioural and social values in their responses which can later be used for figuring out connections between individuals and social activity. The features that are used for finding depression indications due to parenting and self-esteem from the texts are - tone_pos, tone_neg, emo_pos, emo_neg, emo_anx, emo_ang, emo_sad, feeling, focus past, focuses, family. These characteristics are more correlated to depression and self-esteem than others, and that's why these features are chosen over others. Additionally, nine more features are selected for linguistic features to characterize the user's responses (auxiliary verbs, prepositions, articles, adverbs, impersonal and personal pronouns, conjunctions, verbs and negations).

One of the significant successes of deep learning on complex natural language processing applications may be word embedding techniques to represent words and documents. Word embedding represents words as compact real-valued vectors with semantic information. These word embeddings, on the basis, appear to indicate implicit connections between words, which are valuable for training on datasets that can gain valuable information from contextual data. To get the intended vector representation, we first encoded all of the sentences with the vocabulary size information and afterwards padded those encoded representations. After that, any machine learning algorithm can use these embedded vectors as input. Mainly, word embedding improves performance as well as the generalization for any NLP problem and also it works well when limited training data need to be trained.

3.6 Machine Learning Models

ML classifiers are now widely utilized in various sectors, including medical prognosis. So this work makes use of a number of popular supervised machine learning classifiers, including Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Multinomial NB, Gradient Boosting Classifier, and Support Vector Machine. The ML classifiers have used the TF-IDF and Count-Vectorizer to find the best model. In terms of measurements, TF-IDF stands for term frequency-inverse document frequency. Count-Vectorizer is used to turn a given sentence into a vector-based on the count of each word in the entire sentence. Machine learning algorithms used N-Gram and LIWC to identify text features. For the parenting detection dataset, the features were encoded and then applied to these models. ML models are implemented in the sci-kit learn toolkit, and parameters are set to their default values. Dataset is split into 80% for training and 20% for testing. To mitigate overfitting in ML algorithms, the K-Fold cross validation method was implemented. Here the value of K is 10. That means the dataset was splitted into 10 folds where one fold was considered for testing and the rest folds were used for training. For machine learning algorithms,

- **Logistic Regression**: When the target is categorical, Logistic Regression(LR) is a strong supervised machine learning approach for binary classification tasks. LR is applied when the target variable is categorical. LR uses the term "logistic" to refer to a logistic function that executes classification operations within the method. To generate a binary output variable, LR uses a logistic function. The major distinction between linear and logistic regression is the range of LR is limited to 0 and 1. In addition, LR does not have any relationship between the input and the output variables.

$$SigmoidFunction, f(x) = \frac{1}{1 + \mathrm{e}^{-(x)}} \tag{3.1}$$

  Nonlinearity can be added to ML algorithms by using the sigmoid function as an activation function. Logistic regression is also a linear regression model. But the logistic regression uses a more complex cost function which can be defined as the Sigmoid Function. This function is also known as a "logistic function" instead of a linear function. Equation 3.1 shows the sigmoid function. The sigmoid function is most commonly used because it falls somewhere between (0 to 1). A threshold can be used to determine which type of data collection belongs. Classification of the estimated probability is done based on the value of this threshold point.

- **Multinomial Naive Bayes**: For Natural Language Processing (NLP), the multinomial Naive Bayes algorithm (MNB) is a probabilistic method of learning. MNB classifiers are created from multiple models, all of which have a common trait: each feature getting classified is independent of any other feature. The presence or absence of one trait has no impact mostly on the presence or absence of the other. It's easy to put into practice since all we have to do is compute the probability. To know about Naive Bayes, we must understand Baye's rule.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{3.2}$$

  Equation 3.2, the Bayes theorem estimates the probability P(c|x) in which c is the class of possible results and x is the provided instance that has to be identified, expressing some specific properties. This MNB method is appropriate for both discrete and continuous data. It is easy and can be executed fast. In real-time applications, it can be applied. It is extremely scalable and can handle enormous datasets with ease. Because of its quick learning rate and ease of design, MNB classifiers have already been widely employed in NLP challenges compared to other Machine Ml algorithms like SVM and neural networks. Despite its heavy naive assumptions, these have a higher accuracy rate in text classification.

- **Support Vector Machines**: Support Vector Machine (SVM) is one of the most widely used Supervised Learning techniques for Classification and Regression. Using the SVM technique, we represent each piece of data as an n-dimensional point (n represents the number of features), with each feature's value being the value of a specific coordinate. SVM selects the most extreme positions that help to form the hyperplane. It is possible to classify data points using hyperplanes, which are selection boundaries. Using the SVM technique to text classification issues provide extremely good results if proper vector representations can be created that encode. The SVM hypothesis is a very straight cut when weight is denoted as w:

$$\begin{array}{l} \text{if } w^T X \geq 0;\ \text{predict } 1 \\ \text{if } w^T X < 0;\ \text{predict } 0 \end{array} \tag{3.3}$$

- **Decision Tree**: The Decision Tree (DT) might be considered a useful technique when categorization and prediction are required. It is a predictive model that connects observations about an entity with estimates of its goal value. Entropy is the ML metric used by decision tree algorithm to measure impurity or unpredictability. Entropy represents the amount of details required to correctly describe a sample. Equation 3.4 represents the entropy where 'Pi' represents the maximum - likelihood probability of a class 'i'th data. In the tree structure, the internal nodes illustrate the attributes of the dataset, branches represent decision rules and the leaf nodes reflect the result in the tree structure. The decision tree begins with a single root node and then divides into possible outcomes. Later the value of the root node is compared with the attributes of the record. The main advantage of DT is it provides a clear demonstration of which elements are most significant for categorization or prediction.

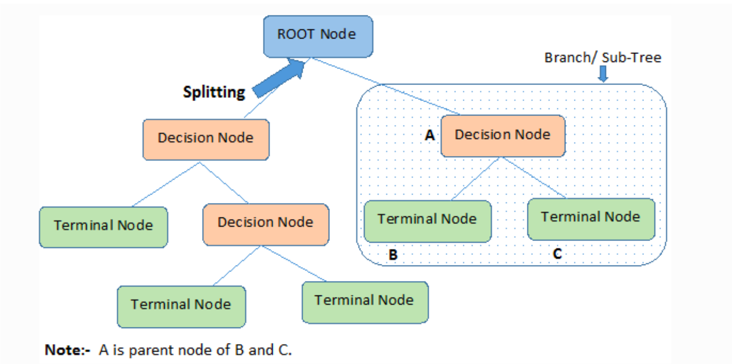

Figure 3.4: Decision Tree Algorithm

$$E = -\sum_{i=1}^{N} pi * \log(pi) \tag{3.4}$$

Figure 3.4 shows an representation of DT. It categorized samples by sorting them along the tree from the root to a leaf node. The edges which are descending from the nodes refer to the possible responses to the test case, while the nodes in the tree act as test cases for various attributes. This process continues recursively until the best result comes out.

- **Random Forest**: One of the most widely used machine-learning algorithms is random forest(RF). "Forest" is a group of decision trees, typically trained using the "bagging" technique. The bagging technique generates a new training subset with substitution from the experimental training data, and the final output is determined by majority voting from all models. A solid learner is created by combining separate trees. Regression and classification problems can be addressed with it. The regression problem is solved using the mean square error (MSE). This shows the data branches of each node. The formula of MSE is given below-

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2 \tag{3.5}$$

In equation 3.5, yi denotes the value of testing data point of a certain node and fi shows the value received from the decision tree. On the other hand, gini index is used when the RF is applied on classification based data. The formula of gini index follows as below-

$$\text{Gini } = 1 - \sum_{i=1}^{C} (p_i)^2 \tag{3.6}$$

In the given formula 3.6, pi demonstrates the relative frequency of the observing class and c is the number of classes in the given dataset. Combining learning models in a bagging method leads to a better overall outcome. However, the majority of the time, it is employed to solve classification-related issues. In a forest, the more trees there are, the stronger the forest will be. Similarly, the random forest model builds decision trees and then returns the results from each of them. Multiple decision trees are combined in a random forest in order to produce more accurate and consistent forecasts.

- **Gradient Boosting**: Gradient Boosting(GBC) can be considered a perfect approach for reducing Bias errors and Variance errors. It is commonly addressed as a greedy algorithm. This algorithm gathers weak learning models and, in an iterative process, transforms them into powerful learners. This model can perform as a regressor as well as a classifier. In regressor mode, the gradient boosting algorithm predicts continuous target variables, and Mean Square Error (MSE) is considered a cost function. But in classifier mode, this algorithm identifies the categorical target variables and the cost function for this state is Log Loss. Gradient boosting has the advantage of providing prediction accuracy. It can optimize a variety of loss functions and has a large number of hyper-parameter tuning options, which turns this algorithm into a versatile function.

3.7 Deep Learning Models

Hyperparameters are parameters that are expressly given as input to govern the learning process in Deep Learning. These hyperparameters are often used to enhance the model's learning. Thus, their values are defined before the model's learning begins. The Table 3.3 represents all the hyperparameters for the structure of the models.

3.7.1 Experimental Parameters:

Table 3.3: Hyperparameters for the structure of the models.

Parameter Name	**Parameter Value**
Learning Rate	0.001
Batch Size	64
Maximum Epoch Size	50
Dropout	0.5 or 0.6

- **Batch Size**: Batch size is the amount of training samples utilized in one iteration in deep learning networks. Simply said, the batch size is the number of samples handled before the model is modified. The batch size determines how fast the model runs and how much memory is used. As an outcome, if the batch size is excessively large, the model would use an enormous amount of memory. On the other hand, if the batch size is reduced, the model will train more quickly. In our approach, we used batch sizes 64 as experimental parameters and compared results using them while keeping the rest of the factors fixed. After analyzing the data, it was found that whenever the batch size is 64, the training period is the shortest. As a result, we've landed on 64 as the experimental batch size.

- **Epoch Size**: When the complete dataset is traversed forward and backward across the neural network once, it is termed an epoch. In the majority of deep learning models, we apply more than a single epoch to develop the model. The weights are changed on each traverse through the network, and the curve progresses from underfitting to overfitting. We examined the model's value accuracy after numerous training sessions with varied epoch sizes to see if it was overfitting or underfitting. Applying epoch size 50 with little or no overfitting and underfitting gave us the best outcome.

3.7.2 Model Parameters

- **Dropout**: The goal of the dropout layer is to ignore a specific ratio of the output value from random layers. After experimenting with various dropout ratios, it is discovered that dropout 0.5 and 0.6 is the perfect fit for these models. Here, dropout value 0.5 and 0.6 sequentially indicates that 50% and 60% of the random units will be dropped. After applying this technique, the performance of the used models will be improved since it prevents overfitting.

- **Loss Function**: We used binary cross-entropy as our loss function for each and every model. In binary cross-entropy, each of the estimated probabilities is compared with the actual class output, which can be simply 0 or 1. Moreover, both dataset has two labels, such as depressive indicative or not and high self-esteem or low self-esteem. As a result, binary cross-entropy outperforms all other loss functions in our dataset for binary prediction.

- **Optimizer**: The Adam optimizer produces better results than other optimization algorithms because it takes less time to compute and needs fewer tuning parameters. As a result, Adam is suggested as the recommended optimizer for the vast majority of applications. The default parameters for this optimizer are learning rate=0.001, epsilon = 1e-07, beta 1= 0.9, and beta 2 = 0.999 which works perfectly well with our models. Therefore, we chose "Adam" as our ultimate model optimizer after observing the performance of other optimizers.

- **Activation Function**: The activation function generates a weighted sum and then adds bias to it to determine whether a neuron should have been triggered or not. We used the sigmoid function the most out of all the activation functions because it takes input in form of any actual value and outputs values in the range of 0 to 1, which perfectly fulfills our objective. However, we used tanh and relu functions as well in some particular models as per its requirement.

3.7.3 Deep Learning Models

Traditional NLP features were mostly handcrafted, imprecise, and time consuming to develop. Neural networks can learn multilayer properties automatically and offer better outcomes. For this NLP task, there is the use of models such as RNN, CNN, GRU, LSTM, and several versions of LSTM such as Stack LSTM and bidirectional LSTM.

"Embedding Layer", "Dropout Layer", and "Dense Layer" is the common layers used to develop all models. The embedding layer utilized for our models has three parameters such as input dimensional, which is the vocabulary size (5000) for our dataset, output dimensional set as 40, which is the output vector size and input length as 200, which can be the approximately maximum encoded sentence length in input. This embedding layer was employed once in each model at the start. Although the embedding layer is not used in parenting style identification, Depression indicative datasets with LIWC features as the features were fixed and numerical, so it did not require NLP techniques. Moreover, for the self-esteem dataset, we utilized dropout layers numerous times with a rate of 0.5 and 0.6 for parenting style identification, depression indication detection, and depression indicative datasets with LIWC features. To wrap up the models, we used a dense layer imported from Keras with the sigmoid activation function.

- **RNN**: The Recurrent Neural Network is a type of neural network model that performs well with sequential data such as NLP-related data. This algorithm has access to data from the node of the preceding layer. RNN is a type of neural network that can be appropriately applied for NLP-related data because it can be trained to retain information from the past. RNN is capable to implicitly memorize information from all the previous components of the sequential input due to a well-designed connections across hidden layers, outperforming other simple neural networks in these tasks. As in the case of simple RNN, there is a recurrent hidden state,

$$h(t) = a(W_i i_t + W_h h_{t-1} + b) \tag{3.7}$$

In formula 3.7, i_t represents the input vector with m-dimensional at time t, h_t represents the hidden state with n-dimensional, and a represents the activation function point-wisely, like the logistic function, hyperbolic tangent function, or rectified Linear Unit. Also here b and W_i, W_h means the appropriately sized parameters (one bias and two weights). Moreover, W_i is a $n \times m$ matrix, W_h is a $n \times n$ matrix, and b is a $n \times 1$ matrix in this scenario.
However, when training RNNs, long-term dependability issues can arise due to gradient exploding or vanishing. Moreover, long data patterns are difficult to learn because of this gradient exploding or vanishing. For comparison with other models, we used a simple RNN neural network as well.
For our RNN model construction, we used a built-in "SimpleRNN" layer from Keras, which has 100 units attached vertically to each input in the sequence, and it passes filtered information to the next memory units.

- **CNN**: A Convolutional Neural Network is a sort of neural network which can work with one, two, or even three-dimensional data, consisting of convolutional layers that perform a convolution process. Convolution is a linear operation that involves multiplying the data matrix and a Kernel where the size of the kernel determines the kind of CNN.

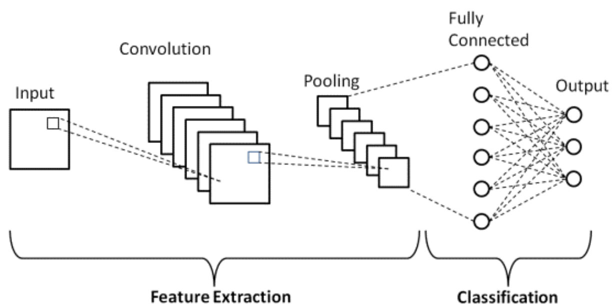

Figure 3.5: Convolutional Neural Network Algorithm

Figure 3.5 indicating, our model used a convolution layer named "Conv1D" fetched from Keras. This layer generates a tensor of outputs by combining the layer input with the convolution kernel over a single temporal dimension. We employed three parameters such as filters (128), kernel size (5) with sigmoid activation function for self-esteem, depression indicative datasets with LIWC features, and relu activation function for depression dataset. Moreover, we utilized the "GlobalMaxPooling1D" as a pooling layer to shrink the feature map's dimensions. Finally, we added it to output layer to execute our classification tasks.

- **GRU**: The upgraded version of RNN is the Gated recurrent units Algorithm(GRU). It is a recurrent neural network model that seeks to tackle the vanishing gradient issue. For solving the vanishing gradient problem, GRU uses two gates named the update gate and the reset gate. These two gates help the trained model to keep information that has been used in the past.

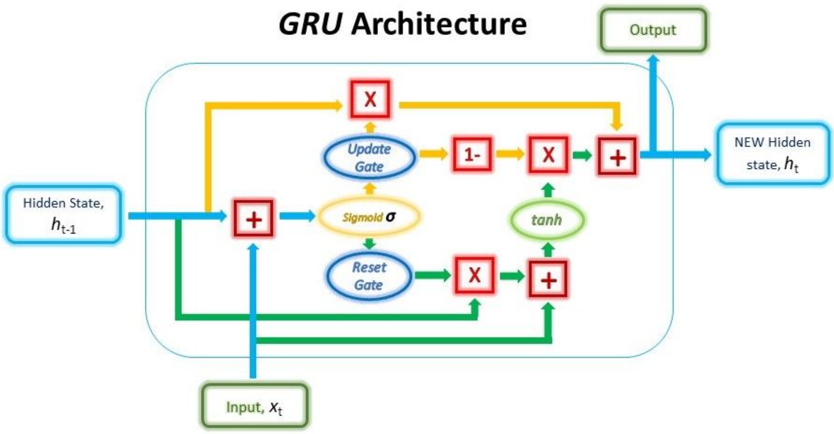

Figure 3.6: Gated Recurrent Units Algorithm

The input layer, as shown in Figure 3.6, is made up of large number of neurons, the number of which is dictated by the size of the input dimensional space. Similarly, the number of neurons in the output layer is proportional to the output space size. Moreover, the GRU network's major functions are covered by the hidden layers, which include resetting and updating cells. Those gates presented in the form :

$$n_t = \sigma(W_n x_t + U_n h_{t-1} + b_n) \tag{3.8}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{3.9}$$

The representative equations of update gate and reset gate are n_t(equation3.8) and r_t(equation3.9) respectively. W_n and W_r are the weight matrices for the update and reset gates respectively. Furthermore, the weight matrices of the update and reset gates related to the hidden state are U_n, U_r. Furthermore, the biases of the gates are represented by b_n, b_r. A sigmoid activation function is used to cover all of those gates' equations, where h_{t-1} denotes the hidden state at the previous timestamp and x_t is the input at the current timestamp t.
The reset and update gates in the cell are responsible for all cell state changes and maintenance. These changes help to prevent vanishing gradient problems. We used the GRU model to observe the performance of this model in our task. We used Keras "GRU" Layer with 100 units and the Tanh activation function.

- **LSTM**: Long short-term memory (LSTM) networks are a variant of Recurrent Neural Network model, potential of learning long-term dependencies. It is so powerful because of its feedback connection structure, and it can handle single data points to sequences of data. Long short term memory is simply an extension of RNN that uses long short term memory to overcome RNN's limitations.

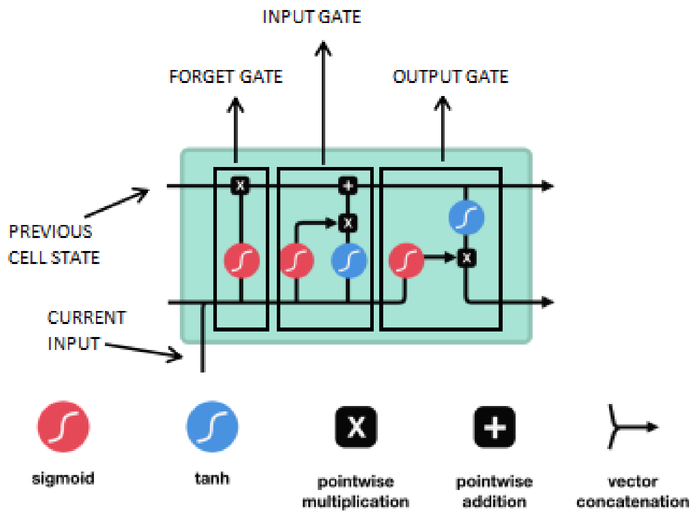

Figure 3.7: Long Short-Term Memory Algorithm

From Figure 3.7, it is visible that LSTM contains three gates: input, forget, and output gate. The vector equations are being used to express the particular mathematical structure of the gating signals:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3.10}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3.11}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{3.12}$$

Here, i_t, f_t, and o_t are sequantially the representative equations of input gate, forget gate and output gate. Here W_i, W_f and W_o represent the weight matrix for input, forget and output gate. Moreover, U_i, U_f, and U_o are the weight matrix of input, forget and output gate associated with the hidden state. In addition, b_i, b_f, and b_o indicate the biases of those gates. All those gate's equations are covered with a sigmoid activation function where x_t is the input at the current timestamp t and h_{t-1} is a hidden state at the previous timestamp. During the processing of the sequence, the cell state can carry crucial information. As a consequence, information from previous periods could reach later time periods, ignoring the impact of short-term memory. As it goes along its path, information is added to or removed from the cell state via gates. Different neural networks operate as gates, selecting which cell state information is allowed. The gates may learn what data is crucial to remember during training. These LSTM steps can avoid the vanishing gradient problem. For our NLP task, we exported the "LSTM" layer from the Keras library, which is a built-in layer with a value of 100 units as input. It performs admirably with our requirements.

- **Stacked LSTM**: A stacked LSTM architecture can be defined as an LSTM model with multiple LSTM layers. This model is an extended version of LSTM architecture. The main difference between LSTM and stacked LSTM is that stacked LSTM has multiple hidden LSTM layers with numerous memory cells in each layer, whereas the original one has only one hidden layer. These extensive layers gain higher power through their deep architecture. Sometimes Deep RNNs outperform shallower RNNs on several tasks, according to empirical evidence. Each layer gradually adds more layers of abstraction to the incoming observations. We employed three “LSTM” layers having 50 units each as parameters for our model building.

- **BI-LSTM**: Another special version of LSTM is Bidirectional LSTM. It can make the neural network have sequence information across both directions. Also, this architecture can improve model abilities in any natural language processing problems. The Figure 3.8 represents the model structure of Bidirectional LSTM. To generate our model, we used a Keras “Bidirectional” layer and applied it to a standard LSTM layer having 100 units as parameters. It is achievable to preserve knowledge from the past and future at any moment in time by combining the two hidden forward and backward states. In this method, the first LSTM studies the impact of previous words while the second understands the impact of future words for each word. Because a word’s context in a phrase contains both future and previous words, bi-LSTMs often offer slightly better performance than single LSTMs for most NLP tasks.

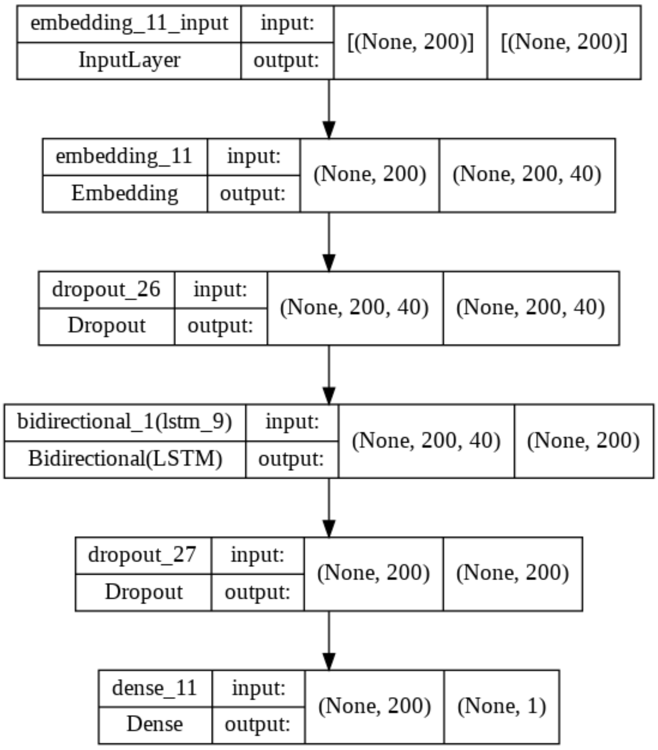

Figure 3.8: Bidirectional LSTM Model Structure

Chapter 4

Result and Analysis

The performance of DL and ML models that are used in this study are evaluated using accuracy, F1 score, recall, and precision. By using these parameters, it can be anticipated how well the models perform. Each of these parameters uses True Positive, True Negative, False Positive, and False Negative values to determine the results, except the F1 score. On the other hand, the F1 score uses the value of precision and recall.

Parenting Style Detection: From Table 4.1, it can be seen that the Gradient Boosting Classifier (GBC) algorithm has the highest accuracy and F1 score, which are respectively 95 and 83.28 among other algorithms. The precision and recall score is also higher than other models. As we know confusion matrix mainly summarizes the performance of the models. The confusion matrix depicts the various ways in which the classification model becomes perplexed while doing the predictions. This helps to overcome the drawbacks of relying just on classification accuracy. So the confusion matrix in Figure 4.1 also proves how efficient this GBC algorithm is for this dataset. In the confusion matrix, 36 samples were predicted as authoritarian by the GBC model from the prediction; 32 samples were predicted right. Furthermore, the performance of DT and RF is extremely close to the GBC algorithm. The SVM, MNB, and LR algorithm's result is also appreciable. All the ML algorithms performed well on this dataset.

Table 4.1: **Performance analysis for Parenting Style Detection For ML Models**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	87.00	68.08	65.07	66.53
DT	94.00	81.86	93.98	83.50
RF	95.00	79.77	84.02	81.44
MNB	84.00	63.57	60.44	61.51
GBC	95.00	84.72	82.40	83.28
SVM	87.00	70.27	64.75	67.30

The performance of all models with word embedding features are shown in Table 4.2. Each model has a value for accuracy, F1-score, precision, and recall which allow for a more in-depth examination of the results. Here, Stacked LSTM is the

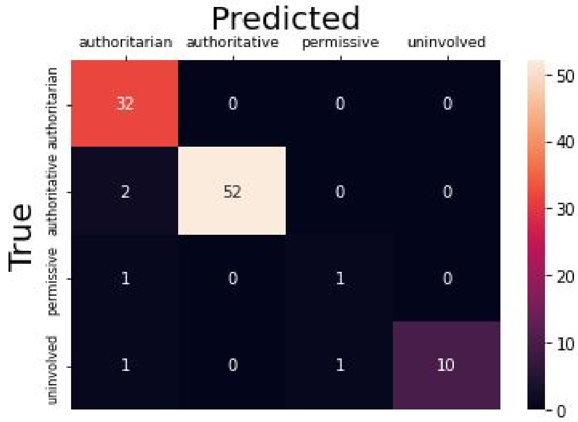

Figure 4.1: Confusion Matrix of GBC

best among all the models according to accuracy, precision, recall, and F1 score. It has sequentially 77.58 and 78.25 scores as the accuracy, precision score which outperforms the Bi-LSTM model with 71.52 accuracy as well as 72.30 as precision score consecutively. Also it outperforms every other model with better precision, recall and F1 score. Finally, we can add that Stacked LSTM is better in every aspect according to the table.

Table 4.2: **Performance analysis for parenting style detection For DL Models**

Classifiers	Accuracy	Precision	Recall	F1-score
Simple RNN	73.94	75.24	71.54	72.34
CNN	64.85	63.71	65.25	62.65
GRU	75.76	74.80	73.56	74.20
LSTM	75.15	73.65	74.75	75.05
Stacked LSTM	77.58	78.25	76.50	77.26
Bi-LSTM	71.52	72.30	71.94	72.05

Moreover, the Loss Curve is used to evaluate a model's error. If the loss is reduced, the model would perform more effectively. The training loss and validation loss curves depict the incremental evaluation of a model's performance over time.

The gap between training and testing loss curve is minimal in the Figure 4.2 which decreases to a point of stability. It means the stacked LSTM has adequately minimized loss score. So it can be assumed that this model found a global minimum of its loss function. This model has a comparatively better loss function than any other models in terms of preventing overfitting and underfitting.
Based on the results from the Table 4.1 and 4.2, the GBC shows better results than other models.

Depression Indication Detection: Based on the feature extraction techniques, the ML algorithms perform differently. Table 4.3 shows that the LR with TF-IDF(Unigram and Bigram) has gained the best score. The accuracy is 77, and the

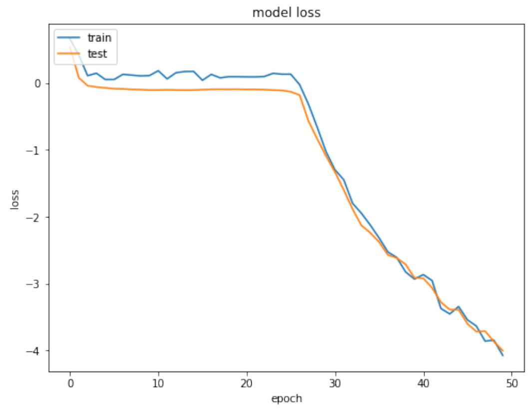

Figure 4.2: Loss Graph of Stacked LSTM for Detecting Parenting Style

precision is 80.70. The f1 score and recall of LR are also noticeable. The confusion matrix of Figure 4.3 interprets that LR with TF-IDF can detect 41 depression indicative samples among 53 samples and 39 non-depression indicative samples among 47 samples correctly. On the other hand, Table 4.4 represents the performance of ml models with Count-Vectorizer. LR with CountVec technique has the highest score with an accuracy of 83. The confusion matrix for this algorithm (Figure. 4.4) shows that it can detect 37 depression indicative samples among 46 samples and 45 non-depression indicative samples among 54 samples correctly.

Table 4.3: **Performance analysis for Depression indication detection with TF-IDF**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	77.00	80.70	78.40	76.80
DT	72.00	80.50	60.50	65.0
RF	66.00	91.70	43.60	55,50
MNB	76.30	74.50	93.70	81.80
GBC	73.00	85.30	65.4	70.90
SVM	76.00	82.10	72.00	74.60

Table 4.4: **Performance analysis for Depression indication detection with Count-Vectorizer**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	83.00	75.80	79.50	76.80
DT	75.00	71.60	69.30	66.90
RF	77.00	81.00	74.30	67.40
MNB	71.20	62.50	66.20	63.4
GBC	77.00	71.60	77.60	73.80
SVM	76.00	71.90	69.20	67.4

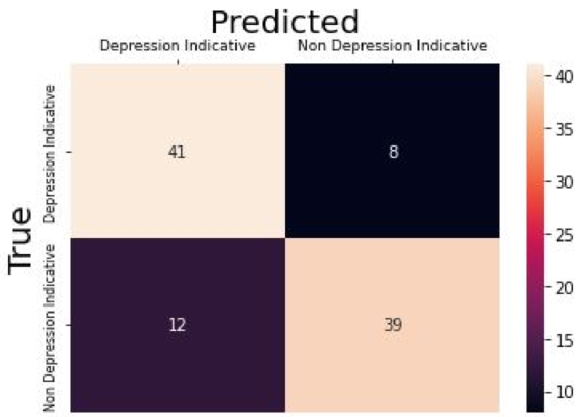

Figure 4.3: Confusion Matrix of LR with TF-IDF

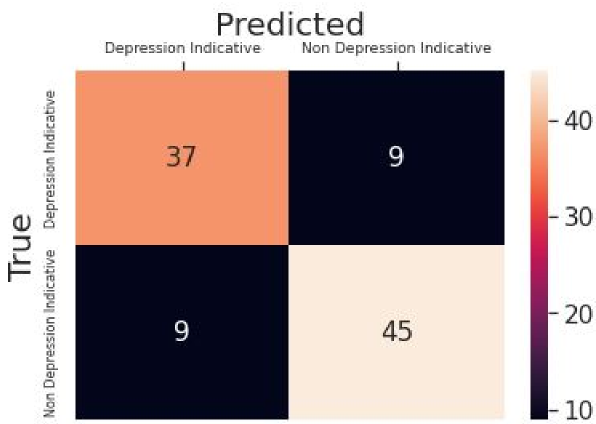

Figure 4.4: Confusion Matrix of Logistic Regression with Count Vectorizer

According to accuracy, precision, recall, and F1 score shown in Table 4.5, Bi-directional LSTM is the best of all the models. It has accuracy and precision scores of 83.21 and 83.17, respectively, outperforming the stacked LSTM and LSTM model, which have accuracy and precision values of 80.51 and 78.63, respectively. Furthermore, it surpasses all other models in terms of precision, recall, and F1 score.

Table 4.5: **Performance analysis for Depression indication detection with Word Embedding**

Classifiers	Accuracy	Precision	Recall	F1-score
Simple RNN	71.76	71.47	71.39	71.43
CNN	74.81	74.61	74.79	74.66
GRU	77.10	76.89	77.02	76.94
LSTM	78.63	78.41	78.41	78.41
Stacked LSTM	80.51	80.12	80.41	80.10
Bi-LSTM	83.21	83.17	83.50	83.16

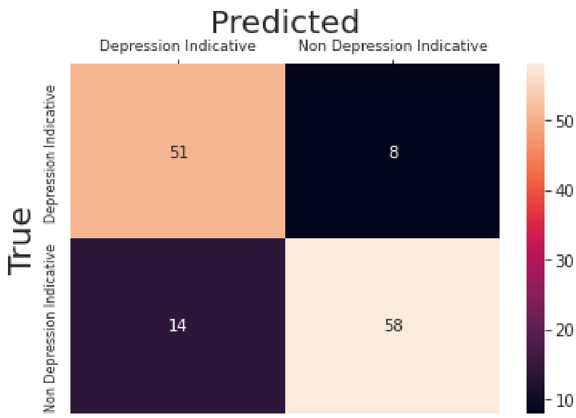

Figure 4.5: Confusion Matrix of Bi-LSTM Depression Indication Detection

The confusion matrix shown in Figure 4.5 describes the true positive (TP) and true negative (TN) rate of this model. Other elements of this figure are the False Positive (FP) and False Negative (FN) rates. Furthermore, by monitoring these properties, this matrix provides us with information about accurately identified samples. After observing this matrix we can state that this model correctly identified 51 (TP) depression indications out of 65 (TP+FP) samples and 58 (TN) non depression indications out of 66 (FN+TN) samples. Following these insights, we can conclude that this model performs well with our requirements.

In Figure 4.6, the training and testing curve gets away from each other after 15 epochs. The training loss continues to reduce with experience while the testing loss has already decreased to a minimum and then begun to rise. This indicates the model has extracted all the signals that the Bidirectional LSTM model could learn. From this information , we can confirm that this model is performing well with our requirements. This AUC-ROC curve is a performance metric for any classification

Figure 4.6: Loss Graph of Bi-LSTM for Depression Indication Detection

problem at different threshold levels. AUC represents the degree or measure of separability, whereas ROC is a probability curve. It indicates how well the model can distinguish between classes.

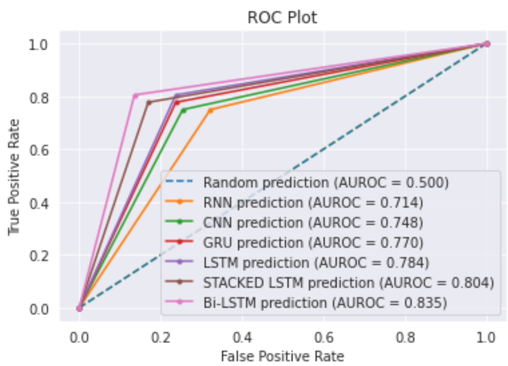

Figure 4.7: AUC-ROC Graph of BI-LSTM for Depression Indication Detection

Figure 4.7 shows that the Bidirectional LSTM graph stays more far away from the random prediction line than other algorithms. The figure also indicates that the Bidirectional LSTM has the roc auc value consisting of 0.835 score which is better than any other model such as Stacked LSTM with 0.804 and LSTM model with 0.784 score. As a result, this model's true positive rate is higher than any other model, indicating that it is the best model for this task.
In this dataset, the Bi-LSTM model shows the most improved results among all applied models.

Depression indication detection with LIWC features: In the Table 4.6, it is given that LR Classifier has the highest score in accuracy and in other parameters as well with LIWC features. The accuracy is 82.27, and recall is 86.86 for this algorithm. This algorithm can correctly figure 27 depression indications and 38 non-depression indication samples (Figure 4.8). Decision Tree gained the lowest result among these ml models.

Table 4.6: **Performance analysis for Depression indication detection with LIWC**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	82.27	82.60	86.86	84.44
DT	62.02	61.95	62.11	61.87
RF	77.21	76.91	76.91	76.91
MNB	72.15	71.94	71.20	71.37
GBC	81.01	82.07	79.74	80.20
SVM	79.74	79.61	79.18	79.34

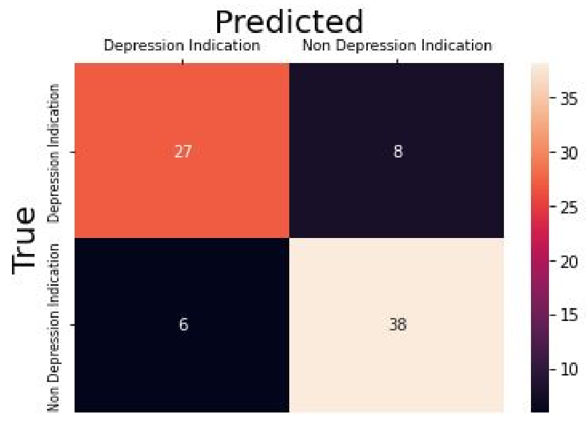

Figure 4.8: Confusion Matrix of LR with LIWC

As seen in Table 4.7, stacked LSTM beats all other models in terms of accuracy, precision, recall, and F1 score. This model's accuracy for this task is 74.81. Furthermore, with a precision score of 74.57 and recall score of 74.64 percent, this model outperforms other models such as Bi-LSTM with 72.52 accuracy , 72.30 precision and 71.94 recall score. According to all this score, it can be added that Stacked LSTM is better than other models for this dataset.

Table 4.7: **Performance analysis for Depression indication with LIWC**

Classifiers	Accuracy	Precision	Recall	F1-score
Simple RNN	67.17	66.82	66.47	66.53
CNN	69.47	75.28	66.71	65.38
GRU	67.94	67.71	67.01	67.09
LSTM	71.76	72.29	70.17	70.19
Stacked LSTM	74.81	74.57	74.64	74.60
Bi-LSTM	72.52	72.30	71.94	72.05

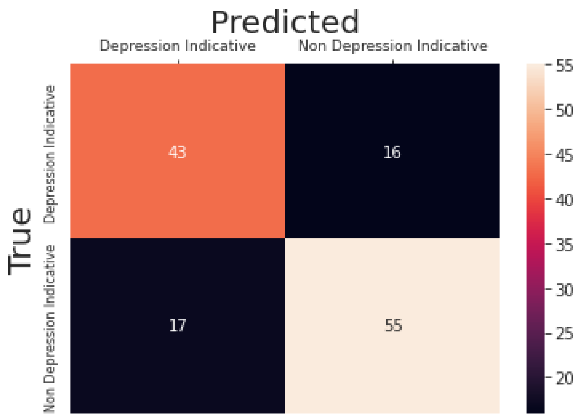

Figure 4.9: Confusion Matrix of Stacked LSTM for LIWC Features

The true positive, true negative, false Positive and false Negative rates of this model are depicted in the confusion matrix in figure 4.9. This matrix provides us with information on accurately recognized samples by monitoring these attributes. This model properly detected 43 depression indications out of 60 samples and 55 non depression indicators out of 71 samples, according to this matrix. We can conclude from these findings that this model performs better to meet our requirements effectively.

In Figure 4.10, it can be seen that the gap between the curves among 20 to 30 epochs is a bit higher. After 40 epochs, the gap keeps getting minimized. This can be anticipated that the curve is showing near optimality. Both curves reach a point of stability and have adequately minimized loss score. This minimized loss score gives us the information that this model surpasses other models in terms of loss of model.

The Figure 4.11 shows the auc roc value of all deep learning neural network models. The stacked LSTM gained the highest auc roc score which is 0.746 which surpasses

Figure 4.10: Loss Graph Of Stacked LSTM for LIWC Features

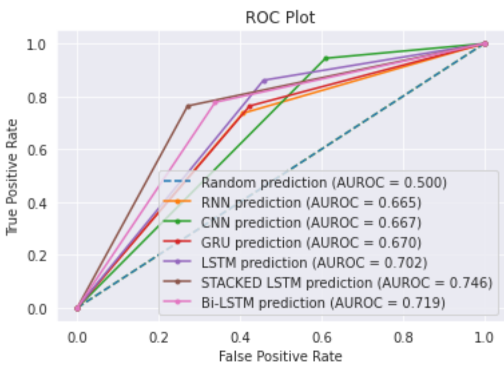

Figure 4.11: AUC-ROC graph of stacked LSTM for LIWC Features

the auc roc score of LSTM and Bi-LSTM with 0.702 and 0.719. It means this algorithm has the highest capability of detecting correct results than any other models applied to do this NLP task.
The LR provides the highest results in terms of accuracy, precision, recall and f1-score compared to other models.

Self Esteem Detection: From the self-esteem dataset, the features were extracted using the same two techniques which were used in the depression dataset. When TF-IDF was used as a feature extractor, the SVM gained the best scores. The accuracy is 81.80, and the precision is 89.50, which can be seen from Table 4.8. The Figure 4.12 contains the confusion matrix of this algorithm. The algorithm specified 47 high self-esteem samples among 61 samples and 20 low self-esteem samples from 23 samples correctly.

Table 4.8: **Performance analysis for Self Esteem detection with TF-IDF**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	74.70	93.30	51.30	58.50
DT	66.70	57.50	52.80	53.40
RF	62.80	50.00	22.00	44.80
MNB	74.70	75.80	68.30	64.60
GBC	73.60	70.00	59.70	70.00
SVM	81.80	89.50	74.80	76.20

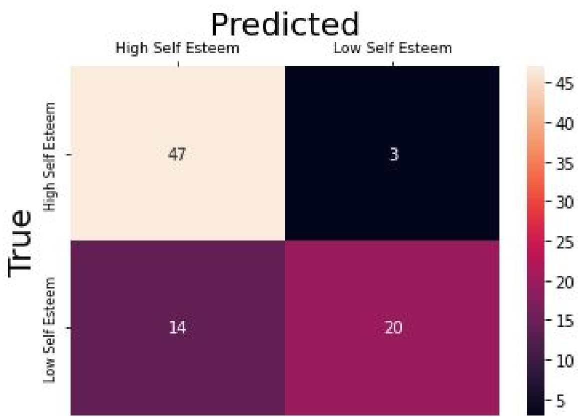

Figure 4.12: Confusion Matrix of SVM with TF-IDF

Then for the CountVec technique, Table 4.9 interprets that the LR performed the best performance. The accuracy is 83.00 and f1 score is 76.80 of this algorithm. The Figure 4.13 represents the LR algorithm with CountVec feature extractor accurately found 46 high self-esteem samples within 57 and 23 low self-esteem samples among 27 samples.

In terms of accuracy, precision, recall, and F1 score, LSTM outperforms all other models, as shown in Table 4.10. The accuracy of this model for this dataset is 74.81. Furthermore, this model surpasses other models such as Bi-LSTM which has 63.77

Table 4.9: **Performance analysis for Self Esteem detection with Count-Vectorizer**

Classifier	Accuracy	Precision	Recall	F1 Score
LR	83.00	75.80	79.50	76.80
DT	75.00	71.60	69.30	66.90
RF	77.00	81.00	74.30	67.40
MNB	71.20	62.50	66.20	63.4
GBC	77.00	71.60	77.60	73.80
SVM	76.00	71.90	69.20	67.4

Figure 4.13: Confusion Matrix of LR with Count-Vectorizer

accuracy score, 63.14 precision score where our proposed LSTM has 67.39 accuracy score and 67.09 precision score. Based on these results, it can be claimed that LSTM exceeds other techniques for this dataset.

Table 4.10: **Performance analysis for Self-Esteem detection With Word-Embedding**

Classifiers	Accuracy	Precision	Recall	F1-score
Simple RNN	61.59	61.12	59.28	58.62
CNN	55.80	54.49	52.65	53.74
GRU	63.77	63.12	62.59	62.63
LSTM	67.39	67.09	66.01	66.09
Stacked LSTM	63.04	62.43	61.42	61.33
Bi-LSTM	63.77	63.14	62.41	62.43

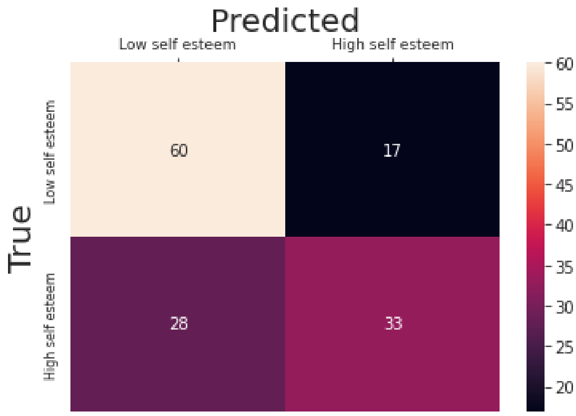

Figure 4.14: Confusion Matrix of LSTM for Self-Esteem Detection

By tracking the features of this matrix in Figure 4.14, we can extract the information about accurately recognized samples. According to this matrix, our model correctly detected 60 low self esteem out of 88 samples and 33 high self esteem out of 50 samples. We can generalize from these results that this model is more effective than any other model at meeting our needs.

The training and testing loss curves in Figure 4.15 do not have that much gap till 18 epochs. After 18 epochs the curves go in different directions. This observation indicates that the model's training loss keeps decreasing when at the same time the validation loss has already reached to minimum. From that state, the validation state starts to increase and keeps going in the upper direction. These observations represent that this model is slightly over fitting over time. But it has a better result than any other model used in this task.
In self-esteem dataset within different feature extraction techniques LR with Count-Vectorizer shows the highest result.

Figure 4.15: Loss Graph of LSTM for Self-Esteem Detection

Chapter 5

Conclusion

To conclude, this research applied supervised algorithms to assess parenting style, depressive indications related to parenting, and self-esteem among young individuals. The dataset was created through an online survey because of the scarcity of publicly available datasets. Furthermore, LIWC was applied in the dataset to detect depression indications. The features that represent the psychological conditions more accurately were used to train the models. The ML models show better results than the DL models in terms of recognizing parenting style, depressive indication, and self-esteem.

The research can be enhanced and developed in the near future by using advanced transformers and hybrid models. To enrich the dataset, more responses from the survey should be gathered. In this way, the DL models will be able to identify more precisely and provide better outcomes than before. Besides taking data from the survey, social media can also be considered a valuable source of data. The supervised models can be implemented to detect depression from social media posts. Nowadays, people express their emotions not only through texts but also through posting images. Thus various computer vision techniques can be utilized to predict depression from images.

Bibliography

[1] R. M, *Self-esteem*, 1965. [Online]. Available: https://wwnorton.com/college/psych/psychsci/media/rosenberg.htm.

[2] M. R. Gray and L. Steinberg, "Unpacking authoritative parenting: Reassessing a multidimensional construct," *Journal of Marriage and Family*, vol. 61, no. 3, pp. 574–587, 1999, ISSN: 00222445, 17413737. [Online]. Available: http://www.jstor.org/stable/353561 (visited on 04/09/2022).

[3] M. (Mann, C. M. H. Hosman, H. P. Schaalma, and N. K. de Vries, "Self-esteem in a broad-spectrum approach for mental health promotion," *Health Education Research*, vol. 19, no. 4, pp. 357–372, Aug. 2004, ISSN: 0268-1153. DOI: 10.1093/her/cyg041. eprint: https://academic.oup.com/her/article-pdf/19/4/357/1493536/cyg041.pdf. [Online]. Available: https://doi.org/10.1093/her/cyg041.

[4] I. Martínez and J. F. García, "Impact of parenting styles on adolescents' self-esteem and internalization of values in spain," *The Spanish Journal of Psychology*, vol. 10, no. 2, pp. 338–348, 2007. DOI: 10.1017/S1138741600006600.

[5] R. Hale, "Baumrind's parenting styles and their relationship to the parent developmental theory," *ETD Collection for Pace University*, Jan. 2008.

[6] P. Heaven and J. Ciarrochi, "Parental styles, gender and the development of hope and self-esteem," *European Journal of Personality*, vol. 22, no. 8, pp. 707–724, 2008. DOI: https://doi.org/10.1002/per.699. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/per.699. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/per.699.

[7] ——, "Parental styles, gender and the development of hope and self-esteem," *European Journal of Personality: Published for the European Association of Personality Psychology*, vol. 22, no. 8, pp. 707–724, 2008.

[8] S. Virasiri, J. Yunibhand, and W. Chaiyawat, "Parenting: What are the critical attributes?" *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*, vol. 94, pp. 1109–16, Sep. 2011.

[9] P. W. Corrigan and D. Rao, "On the self-stigma of mental illness: Stages, disclosure, and strategies for change," *The Canadian Journal of Psychiatry*, vol. 57, no. 8, pp. 464–469, 2012.

[10] L. C. Driscoll, "Parenting styles and self-esteem," 2013.

[11] G. Shen, J. Jia, L. Nie, *et al.*, "Depression detection via harvesting social media: A multimodal dictionary learning solution.," in *IJCAI*, 2017, pp. 3838–3844.

[12] S. Y. Arafat, "Validation of bangla parenting style and dimension questionnaire," *Global Psychiatry*, vol. 1, no. 2, pp. 95–108, 2018.

[13] M. R. Islam, A. Kabir, A. Ahmed, A. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health Information Science and Systems*, vol. 6, p. 8, Aug. 2018. DOI: 10.1007/s13755-018-0046-0.

[14] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of twitter users," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, pp. 88–97.

[15] A. Choudhury, M. R. Khan, N. Nahim, S. Tulon, S. Islam, and A. Chakrabarty, "Predicting depression in bangladeshi undergraduates using machine learning," Jun. 2019, pp. 789–794. DOI: 10.1109/TENSYMP46218.2019.8971369.

[16] M. Pinquart and D.-C. Gerke, "Associations of parenting styles with self-esteem in children and adolescents: A meta-analysis," *Journal of Child and Family Studies*, vol. 28, no. 8, pp. 2017–2035, 2019.

[17] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in reddit social media forum," *IEEE Access*, vol. 7, pp. 44 883–44 893, 2019. DOI: 10.1109/ACCESS.2019.2909180.

[18] A. Zaman, R. Acharyya, H. Kautz, and V. Silenzio, "Detecting low self-esteem in youths from web search data," in *The World Wide Web Conference*, ser. WWW '19, San Francisco, CA, USA: Association for Computing Machinery, 2019, pp. 2270–2280, ISBN: 9781450366748. DOI: 10.1145/3308558.3313557. [Online]. Available: https://doi.org/10.1145/3308558.3313557.

[19] G. Rao, Y. Zhang, L. Zhang, Q. Cong, and Z. Feng, "Mgl-cnn: A hierarchical posts representations model for identifying depressed individuals in online forums," *IEEE Access*, vol. 8, pp. 32 395–32 403, 2020.

[20] A.-S. Uban and P. Rosso, "Deep learning architectures and strategies for early detection of self-harm and depression level prediction," in *CEUR Workshop Proceedings*, Sun SITE Central Europe, vol. 2696, 2020, pp. 1–12.

[21] Ö. Azman, E. Mauz, M. Reitzle, R. Geene, H. Hölling, and P. Rattay, "Associations between parenting style and mental health in children and adolescents aged 11–17 years: Results of the kiggs cohort study (second follow-up)," *Children*, vol. 8, no. 8, p. 672, 2021.

[22] K. Cherry, *What are the signs of healthy or low self-esteem?* Apr. 2021. [Online]. Available: https://www.verywellmind.com/what-is-self-esteem-2795868.

[23] S. Hyun, "Cambridge working papers in economics cambridge-inet working papers social networks, confirmation bias and shock elections," 2021.

[24] Y. Liao, J. Shu, and S. Liu, "A text classification framework for depression tendency detection over online social platform," in *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2021, pp. 796–801. DOI: 10.1109/ICAICA52286.2021.9498200.

[25] Y. M. Y. Niimi, "Machine learning approach for depression detection in japanese," in *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, 2021, pp. 350–357.

[26] B. Peng, N. Hu, H. Yu, H. Xiao, and J. Luo, "Parenting style and adolescent mental health: The chain mediating effects of self-esteem and psychological inflexibility," *Frontiers in Psychology*, vol. 12, 2021, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2021.738170. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2021.738170.

[27] A. Wongkoblap, M. A. Vadillo, V. Curcin, *et al.*, "Deep learning with anaphora resolution for the detection of tweeters with depression: Algorithm development and validation study," *JMIR Mental Health*, vol. 8, no. 8, e19824, 2021.

[28] *Adolescent mental health.* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health.

[29] S. Lovibond and P. Lovibond, *Depression anxiety stress scales - long form (dass-42).* [Online]. Available: https://novopsych.com.au/wp-content/uploads/2021/03/dass-42_PDF_scoring_online.pdf.

[30] T. Nagar, "Prediction of mental health problems among children using machine learning techniques,"