STUDENTS PERSONALITY BASED PREDICTION OF EMOTIONAL INTELLIGENCE USING DATA MINING TECHNIQUES

B.Jayapradha[1], M.Rameshkumar[2],T.Velmurugan[3]

[1]Assistant Professor, [2]Assistant Professor, [3]Associate Professor

PG and Research Department of Computer Science [1,2,3],

[1]Dr.Ambedkar Government Arts College, Chennai, India,

[2]Government Arts College, Chennai, India,

[3]Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, India

E-Mail:[1]jayapradha.bhaskar@gmail.com, [2]proframeshkumar@gmail.com, [3]velmurugan_dgvc@yahoo.co.in

Abstract: In the current world, the behavior of students changes from time to time, based on their emotional decisions regarding the kind of activities they are involved with. Many emotions are produced by the students when they act in a particular incident-based activity. This research work, analysis such kind of emotion-based students' personalities through their activities using some of the data mining techniques, Naive Bayes and Random Forest are utilized to predict the usefulness of the students' activities, created utilizing their decisions. The data set is collected from the students for this analysis employing a questionnaire. The performance of both algorithms is evaluated based on the results of students' decision-making responses collected from the survey. The best algorithm for the analysis of emotional intelligence-based students' information is suggested based on the accuracy of the algorithms.

Keywords: Data Mining Algorithms, Emotional Intelligence, Naïve Bayes Algorithm, Random Forest Algorithm, Performance Accuracy.

1. INTRODUCTION

Today, it is the world of the digital era, where everyone's daily activities are preoccupied with digital gadgets and social media.[1] It has become very difficult for every individual to survive without the indulgence of social media in our day-to-day activities. Moreover, the current changes that have occurred due to the covid situation have also pushed everyone to be a social media addict. Students, who happen to be the future in deciding the progress of a nation, have been severely addicted to this digital threat. As a result of this, the student's mentality and behavior have changed and there is a severe wavering in their mood swings and emotions from time to time. This change in their emotions affects the efficiency of any individual in all their daily activities like making decisions, handling relationships, facing an interview, and their academic performance.[2] This study is carried out by collecting data from students through a google form. The questions were taken from a popular known technique to assess the emotions of the students, called the Maire Bridge Type Indicator (MBTI) [3] survey. The data are then analyzed, pre-processed, and evaluated based on the accuracy of the data mining algorithms.

The organization of the paper is discussed as follows. Section 2 discusses the related work via a literature survey. The Methods and Materials are given in Section 3. Section 4 has an exploratory analysis of the experiments and their deep discussions. Finally, Section 5 concludes with the research findings.

2. LITERATURE SURVEY

Several researchers have done their research work in the area of emotional intelligence to predict individual behaviors from a person's attitude and character. It is not possible to state here all the works of the different kinds of research work. However, can able to discuss some of the issues and their predictions through related work in this section. Very particularly, some of the methods of data mining algorithms utilized by different researchers are explored in this section.

The research work titled "A Novel Approach to Evaluate and Rank Candidates in A Recruitment Process by Estimating Emotional Intelligence through Social Media Data" is carried out by Vishnu M Menon et al. in [4]. This paper proposes a web application that measures a person's emotional aptitude for a job through their curriculum vitae available on Twitter. Classification algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees were applied for independent approaches and algorithms, and Random Forest and KNN algorithms were applied for dependent approaches. The research findings conclude that the former approach tends to achieve higher accuracy than the latter approach.

Another research work titled, "Deep learning-based personality recognition from text posts of online social networks" done by Di Xue1 et al. [5] has proposed a deep learning model to predict the personality of a person from their social media posts. The data set collected from a Facebook application, allows the user to test their personality via psychometric tests and donate their scores and Facebook profile data to research. A two-level hierarchical neural network based on the newly designed AttRCNN structure and a variant of the CNN-based Inception structure to learn the deep semantic representations of online social network users' text posts are suggested. Experimental evaluation shows that taking these kinds of deep semantic features as input of traditional regression algorithms contributes a lot to the performance improvement of the recognition approaches.

A research work carried out by Kim et al. [6], titled, "A deep learning model for detecting mental illness from user content on social media" collected data from Reddit posts related to specific mental disorders. The Porter Stemmer, a tool to specify the guidelines for exploring word meaning and source, was used to convert a word to its root meaning and to decrease the number of word corpus. Six classification models for each symptom such as autism, depression, anxiety, etc., were developed using the SMOTE algorithm. Among the six different subreddits, autism showed the highest accuracy (96.96%) in the CNN model, and depression showed the highest performance scores in terms of precision (89.10%). Anxiety, bipolar disorder, and borderline personality disorder also showed high accuracy with CNN models, 77.81%, 90.20%, and 90.49%, respectively and the model can accurately detect potential users who may have psychological disorders.

Rajat Yadav et al.,[7], in their research titled, "Identifying and sensing emotional quotient weightage in the outcome using Speech Dialogue", has proposed a machine learning-based approach to detect emotion from the text of a speech dialogue. This research falls under the category of multiclass or multinomial classification where there are more than two classes available for the classification task. Each sample can only be assigned one class label out of all the available class labels. The emotion dataset from CrowdFlower was used in this study and contained 2550 sentences. After training, the algorithm's scores of the Random Forest Classifier performed slightly better than the Multinomial Naïve Bayes and Support Vector Classification algorithms.

Anastasia A. Shvetsova et al.,[8], in their research work titled, "Content Analysis of Messages in Social Networks, Identification of Suicidal Types", proposes a content analysis of text to identify suicidal tendencies and types and also describes how to make a sentence classifier using neural network for machine learning. Analysis of existing information about so-called "groups of death" and its distribution on the Internet is conducted. Among all models for the content analysis of the text, this one differs in that it classifies suicidal sentences, which is extremely useful for reducing the death rate from suicide. This neural network can be used to check the text in social networks to block posts containing suicidal overtones, which will prevent the spread and promotion of suicide among young people.

Raman Goelwe et al.,[9] in their research work titled," Emotion-aware transformer encoder for empathetic dialogue generation ", has proposed an emotionally intelligent virtual assistant, a chatbot that understands the emotions in the dialogue of the users and answers them depending upon their emotions. The proposed model integrates both the semantic and affective aspects of the input utterance and the Transformer-XL architecture successfully emulates a human agent that can perceive user emotion and respond empathetically to the user input.

3. METHODS AND MATERIALS

The applications of classification algorithms are used in data mining and its applications. A Classification algorithm consists of a set of rules that are used in data mining. The following describes two machine learning algorithms, the first being Naïve Bayes and the second algorithm being Random Forest are discussed in this section. The MBTI tool, used for the conduction of the survey, which serves as a measure of the students' mentality is also discussed in this section.

3.1 Naïve Bayes Algorithm

The simple probabilistic classifiers are the Naïve Bayes classifier from Bayes' theorem in machine learning which has strong (naive) independence assumptions between the features. [10] It is very useful in text categorization with word frequencies. Naïve Bayes classifiers are highly scalable requiring many parameters in linear problems. An advantage of Naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification [11]. Naïve Bayes is a simple yet powerful probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. It is a supervised learning algorithm, trained with labeled data, and can be used to predict categorical classes. The steps listed here illustrate the working of the algorithm [12].

Step 1: The training data is used to estimate the probability distribution for each feature given in each class.

Step 2: The Bayes theorem is used to calculate the posterior probability for each class given the feature values of the dataset.

Step 3: From the above-mentioned steps, we can calculate and then choose the class with the highest posterior probability [13] as the prediction.

An important advantage of the Naïve Bayes algorithm is that it is simple to implement and can be trained very efficiently, especially when the number of features is large [14]. The algorithm also performs well when the features are independent and the number of examples is small.

3.2 Random Forest Algorithm

Random forest is an ensemble machine learning algorithm that is used for classification and regression tasks. It works by constructing a multitude of decision trees at training time and outputting the class that is the mode of the class(classification) [15] or the mean prediction (regression) [16] of the individual trees. The decision trees are constructed using a randomized subset of the data and a random subset of the feature, which makes the algorithm less prone to overfitting. When the decision trees are aggregated, they form a forest that is more robust and has higher performance compared to a single decision tree. The following steps convey the working of the algorithm which takes decision trees as its base learner.[17]

Step 1: The algorithm starts by selecting a random subset of the training data. The subset is used to train each decision tree in the forest.

Step 2: For each decision tree, the algorithm also selects a random subset of the feature. The selected feature is used as the split points at each node of the decision tree.

Step 3: The decision tree is trained using the selected subset of the data and features and is used to make predictions on the trained data.

Step 4: This process is repeated for a specified number of decision trees. The final prediction of the random forest is the mode of the predictions made by each decision tree (for classification) or the mean of the predictions made by each decision tree (for regression).

Step 5: The random forest algorithm is trained and the decision trees are aggregated to form the final model. This model can then be used to make predictions on new, unseen data.

The main advantage of the algorithm is that it reduces overfitting by training the decision trees on a random subset of the data and using a random subset of the feature at each split. This helps to reduce the variance of the model, which makes it more robust and better able to generalize to the new data.

3.3 The Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator (MBTI) is a psychological assessment tool that is based on the theory of psychological types developed by Carl Jung. It is designed to help people understand their personality traits and how to interact with others. The MBTI assessment

consists of a series of questions that measures an individual's preferences across four dimensions. This research work classifies the personalities of the students into the following types.

Type 1-Extraversion(E) versus Introversion(I): This dimension measures how an individual derives their energy and how they tend to interact with the outside world.

Type 2-Sensing(S) versus Intuition(N): This dimension measures how an individual processes information and makes decisions.

Type 3-Thinking(T) versus Feeling(F): This dimension measures how an individual makes decisions and deals with emotions.

Type 4-Judging(J) versus Perceiving(P): This dimension measures how an individual approaches their external environment and makes plans.

Based on the answers to these questions, the students are classified into one of the 16 personality types, each of which is represented by a combination of four letters (e.g., ENFJ, ISTJ, etc.) The goal of the MBTI is to provide an individual with insight into their personality and how it may impact their relationships, communication style, and career choices.

4. RESULTS AND DISCUSSIONS

This section details the dataset from which the research was conducted and the results of applying the two classification algorithms, Naive Bayes and Random Forest on the dataset for the prediction of the emotion among students. The results discussed predicts the best algorithm for the same.

4.1 Dataset Description

This research was carried out by distributing google forms to students between the age group 16 to 24. As this research work concentrates on the student community, data was collected from college and school students. The study commenced with the creation of the google form, which consisted of questions based on what actions the students would take on a certain scenario. It consisted of such related 24 questions which were used to assess the mentality of the students who attempted the survey. Based on the answers given by them, they were classified among these 8 different personality types using machine learning algorithms (Naïve Bayes and Random Forest). The classes in which the students were classified are given in table 1.

Table 1. Dataset with Respect to the types of personality

No.	Type	Class	Average count
1	Type1	Extraversion	104
2		Introversion	81
3	Type 2	Sensing	96
4		Intuition	89
5	Type 3	Thinking	73

6		Feeling	112
7	Type 4	Judging	125
8		Perceiving	60

Table 1 describes the 8 classes in which the students were categorized and also the average count of students who fall under each category in the given dataset collected. It is observed that among the students a larger part of the group is judgemental and they try to judge people on their behavior and have a pre-defined notion of the people whom they meet. The least number of students fall under the class of thinking where they try to make decisions based on their thinking capacities. This denotes that most of the students cannot access an individual. The same is depicted in the following graph for better understanding. Figure 1 gives an overview of the average count of students and also the number of students who belong to the particular class.
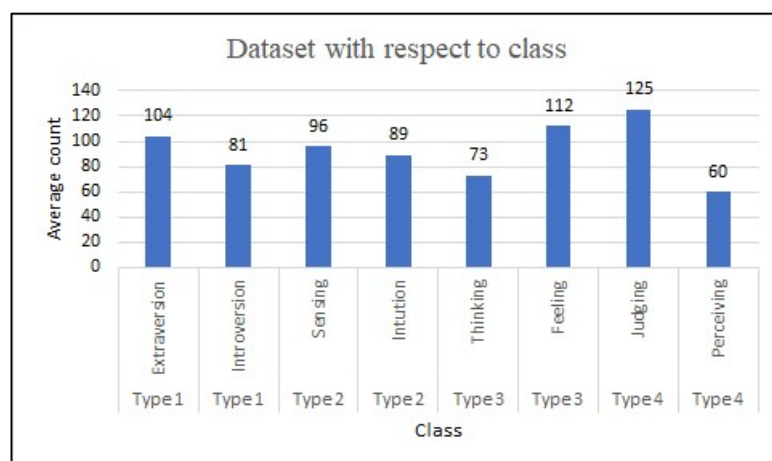

Figure 1: Dataset with respect to Class

The data set collected is then pre-processed and the above-mentioned algorithms are applied to the dataset and predictions are made on each of the four types of personalities of the students. Table 2 describes Personality Type 1 (Extrovert or Introvert). Among the 5 questions which were used to predict whether the person belongs to type 1 personality, a major part of the group fall under the extrovert type and they are more open in their decisions meaning that most students are extrovert in their behavior.

Table 2: Personality Type 1 (Extrovert-Introvert)

S. No.	Question No.	Extroversion	Introversion
1	6	110	75
2	9	75	106
3	13	126	59
4	16	114	71
5	24	89	96

Table 2 entries are pictured in figure 2 where the fact that most of the students are categorized as Extravert is analyzed.
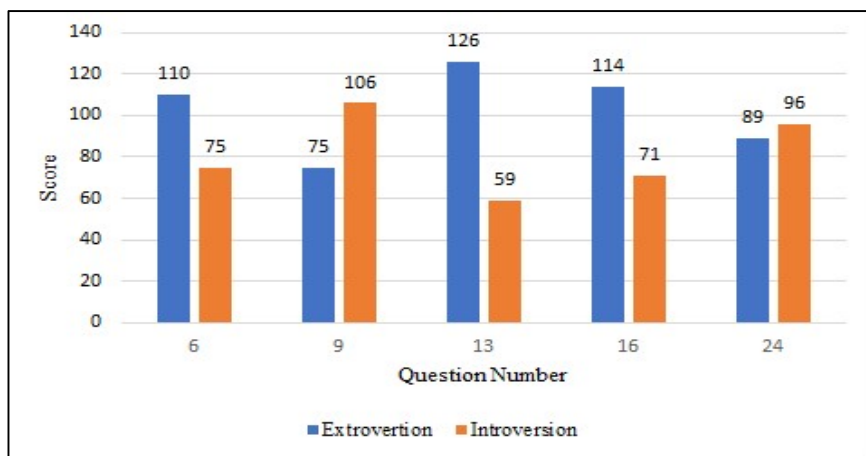

Figure 2: Personality Type 1 (Extrovert-Introvert)

Table 3 illustrates the second type of personality among the students that is whether the student is sensational or Intuitive in their nature. About half of the students tend to be sensational and tend to create a mark and want to be notable. The rest of the students tend to be more intuitive in predicting the future. They seem to be more involving themselves in activities that rely upon their future.

Table 3. Personality Type 2(Sensational-Intuitive)

S. No	Question No.	Sensational	Intuitive
1	2	119	66
2	5	54	131
3	10	91	94
4	12	151	34
5	15	79	106
6	20	131	54
7	23	48	137

The information provided in table 3 is graphically illustrated in figure 3. The current student's mentality is that they are in deep thought about their future and most of the activities they do focus on activities involving their future.
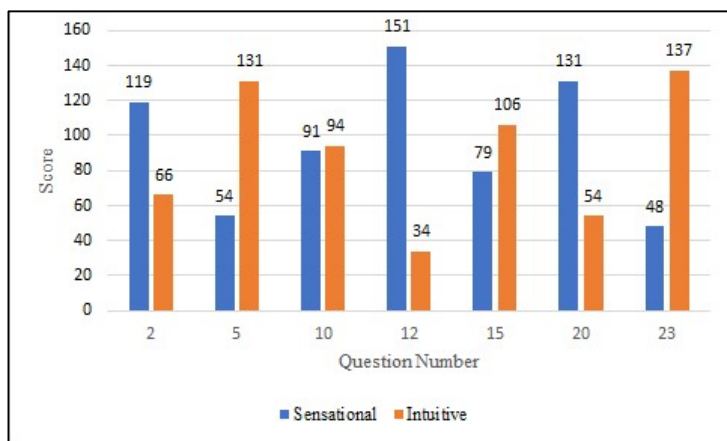
Figure 3: Personality Type 2 (Sensational-Intuitive)

Table 4 depicts Personality Type 3 where most of the students in the group overthink any decisions they take and therefore put forward a lot of importance to thinking when compared to feeling.

Table 4. Personality Type 3 (Feeling-Thinking)

S. No	Question No	Feeling	Thinking
1	4	73	112
2	14	60	125
3	22	87	98

Figure 4 visualizes the third type of personality. This conveys the fact that most students group of the current era, are overthinkers, and their intentions give a lot of importance to thinking when compared to the importance they give to feeling.
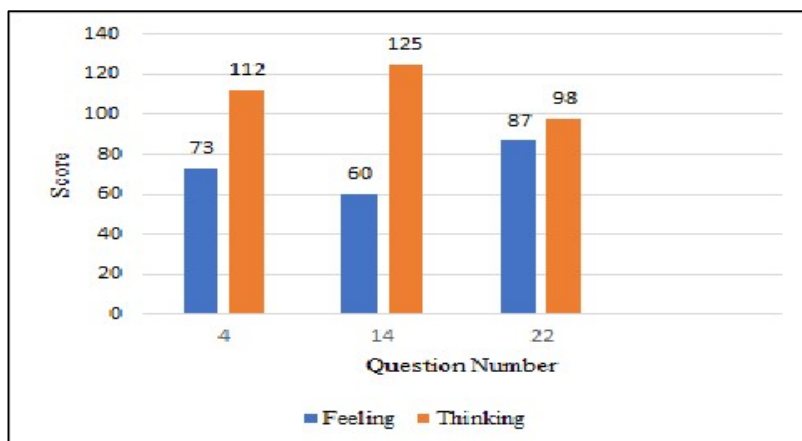

Figure 4: Personality Type 3(Feeling-Thinking)

Table 5 describes the last type of personality (Judgemental or Perspective). Most of the students are judgemental and they judge people and have their ideas of the person and have a pre-defined notion about the person with whom they are interacting.

Table 5. Personality Type 4 (Judgemental-Perspective)

S. No	Question No.	Judgemental	Perspective
1	1	138	47
2	7	82	103
3	8	126	59
4	11	83	102
5	17	145	40
6	18	153	32
7	19	147	38

The fourth type of personality is demonstrated in Figure 5. The figure demonstrates the score for each of the questions from the survey which provides the type 4 personality of the students, that is whether they are judgemental or Predictive in their nature and behavior.
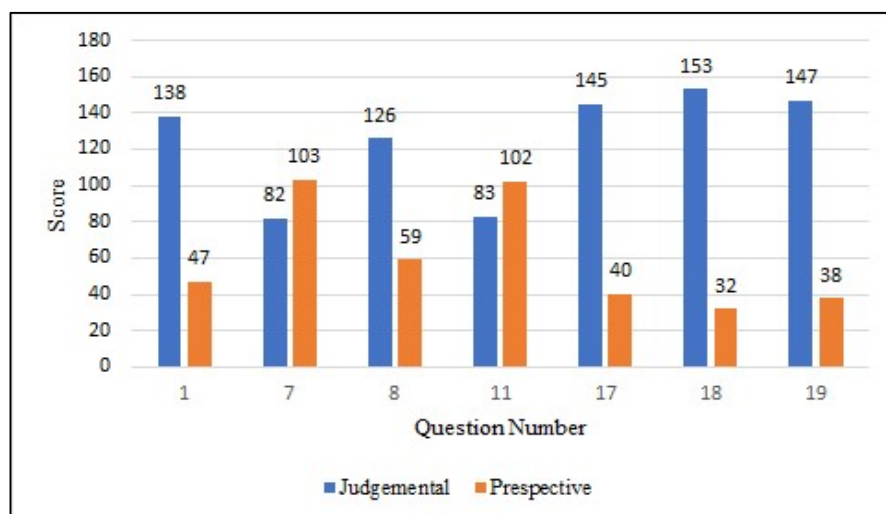

Figure 5: Personality Type 4(Judgemental-Perspective)

The data collected is implemented using Naive Bayes and Random Forest algorithms. The accuracy of each of the personality types 1,2,3, and 4 are displayed in Table 6.

Table 6. Results of the algorithms

Personality Type	Naïve Bayes	Random Forrest
Type 1	92.44	89.89
Type 2	93.68	98.52
Type 3	94.94	95.13
Type 4	89.36	95.10

From Table 6, it can be concluded that the Naïve Bayes algorithm can predict students' Personality Type 3 with an accuracy of about 94.94%, and the Random Forest algorithm can

predict the same with the highest accuracy rate of about 98.52%. The same is also pictured in the graph below in Figure 6.
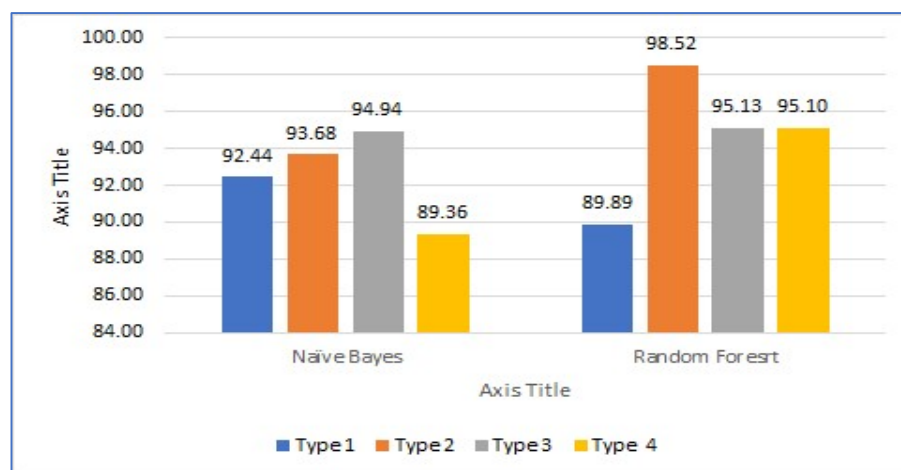

Figure 6: Performance Comparison of the types of Personality

5. Conclusion

Predicting the emotional intelligence from the emotions of different groups of students yields to identify the general behavior and attitudes so this research work is carried out by collecting information through questionnaires about the students' activities and their responsibilities. It is important to state that the performance of the classification algorithm can depend on many factors, including the characteristics of the dataset and the specific parameter of the algorithm. It's common for different algorithms to perform differently on the same set of data set and it's always a good idea to try multiple algorithms and compare their performance to determine which is the best one suited for the given date set. In this research work which classifies the students among the four types of personalities using machine learning algorithms, the Random Forest algorithm shows a better performance among three different personality types in predicting when compared to the Naïve Bayes algorithm. The latter algorithm tries to meet its performance only in one type of personality type classification.

REFERENCES

[1] Lau, Wilfred WF, "Effects of social media usage and social media multitasking on the academic performance of university students", *Computers in human behaviour,* Vol. 68, pp. 286-29, 2017.

[2] Amin, Z., Mansoor, A., Hussain, S. R., & Hashmat, F, "Impact of social media on student's academic performance", *International Journal of Business and Management Invention*, *5*(4), pp.22-29, 2016.

[3] Stein, R., & Swan, A. B, "Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology", *Social and Personality Psychology Compass*, *13*(2), e12434, 2019.

[4] Menon, Vishnu M., and H. A. Rahulnath, "A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social

media data", IEEE, *International Conference on Next Generation Intelligent Systems (ICNGIS)*, 2016.

[5] Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., & Sun, J, "Deep learning-based personality recognition from text posts of online social networks", *Applied Intelligence*, *48*(11), pp. 4232-4246, 2018.

[6] Kim, Jina, "A deep learning model for detecting mental illness from user content on social media", *Scientific reports*, pp.1-6, 2020.

[7] Yadav, Rajat, and Anurag Mishra, "Identifying and sensing emotional quotient weightage in the outcome using Speech Dialogue", *IOP Publishing Conference Series Materials Science and Engineering*, Vol. 1022, No.1, 2021.

[8] Shvetsova, Anastasia A., and Maria A. Antropova, "Content Analysis of Messages in Social Networks, Identification of Suicidal Types", *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2020.

[9] Goel, R., Susan, S., Vashisht, S., & Dhanda, A., "Emotion-aware transformer encoder for empathetic dialogue generation", *IEEE International Conference on Affective Computing and Intelligent Interaction Workshops and Demos,* pp. 1-6, 2021.

[10] Vembandasamy, K., Sasipriya, R., & Deepa, E, "Heart disease detection using the Naive Bayes algorithm", *International Journal of Innovative Science, Engineering & Technology*, *2*(9), pp. 441-444, 2015.

[11] Latha, U., and T. Velmurugan, "Analyzing agricultural text data using classification algorithms", *International Conference on Small and Medium Business*, pp. 339-344, 2016.

[12] Stern, Mia, Joseph Beck, and Beverly Park Woolf, "Naive Bayes classifiers for user modelling", *Center for Knowledge Communication, Computer Science Department, University of Massachusetts* ,1999.

[13] Yang, F. J, "An implementation of naive Bayes classifier", IEEE, *International conference on computational science and computational intelligence (CSCI)* ,pp. 301-306, 2018.

[14] Taheri, S., & Mammadov, M, "Learning the naive Bayes classifier with optimization models", *International Journal of Applied Mathematics and Computer Science*, *23*(4), pp. 787-795, 2013.

[15] Belgiu, M., & Drăguţ, L, "Random Forest in remote sensing: A review of applications and future directions", *ISPRS journal of photogrammetry and remote sensing*, *114*, pp. 24-31, 2016.

[16] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees, and support vector machines" *Ore Geology Reviews*, *71*, pp.804-818, 2015.

[17] Jaiswal, J. K., & Samikannu, R, "Application of random forest algorithm on feature subset selection and classification and regression", IEEE *World congress on computing and communication technologies (WCCCT),* pp. 65-68, 2017.