Date of publication MAR-31, 2024, date of current version NOV-20, 2023.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.23.1.3441

Preventing Student's Mental Health Problems with the Help of Data Mining

MD. NAZMUL HOSSAIN, NAFIZ FAHAD, RASEL AHMED, ANIK SEN, MD. SADI AL HUDA, MD. ISMAIL HOSSEN
Faculty of Science & Technology, American International University-Bangladesh, AIUB, 408/1, Kuratoli, Khilkhet, Dhaka 1229, Bangladesh

Corresponding author: Nafiz Fahad (e-mail: fahadnafiz1@gmail.com).

ABSTRACT The increasing incidence of mental health issues among university students has become a significant concern, often referred to as a "mental health crisis" in academic settings. This study addresses the challenge of predicting mental health issues in university students using data mining techniques. The research involved the creation of a new dataset via a survey method focused on university students, covering various factors like behavioral traits, health conditions, and lifestyle choices. Data mining algorithms such as Naive Bayes, Random Forest, SVM, KNN, and Decision Tree were employed to predict mental health status. The study included dataset collection, cleaning, integration, transformation, reduction, discretization, and the use of Weka and Orange for data analysis. Therefore, exploratory analysis revealed that 53.4% of students reported depression, with a higher incidence among male students and those less involved in extracurricular activities. Predictive analysis showed Naive Bayes as the most accurate algorithm (65.91%) for this prediction task, followed by Random Forest, SVM, KNN, and Decision Tree. The performance was evaluated using accuracy, F1-Score, precision, recall, AUC, and CA. The study highlights the correlation between various aspects of university students' lives and depression. Active participation in extracurricular activities was found to lower depression risks. The effectiveness of data mining in understanding and predicting mental health issues in university students was established, with Naive Bayes being the most effective algorithm for this purpose.

KEYWORDS Mental health; university students; addiction; decision tree; psychology.

I. INTRODUCTION

WITH many universities claiming a growth in the number of students who are seeking mental health support, mental health issues are increasingly frequent among university students. In fact, some have declared a "mental health crisis" due to the rising incidence and intensity of mental health cases in academic settings [1, 2]. Therefore, mental health issues and solutions are receiving a growing amount of attention on university campuses, and it is crucial to comprehend the factors that may influence the mental health of university students.

Students' experience of multiple responsibilities and commitments at university may greatly influence their behavior, emotions, and general wellness since it is frequently during studying at university that people start to seriously confront many of life's challenges and responsibilities [3]. Both intellectually and psychologically, the university offers a very challenging journey for students. Students frequently have a number of obligations outside the classrooms that demand their focus and effort, in combination with completing a full course load and feeling pressure to earn excellent marks [4].

People develop their autonomy and make judgments about their lives and the future between the ages of 18 and 25 [5]. As a result, students are developing their identities while concurrently studying for a degree. Understanding how to live autonomously, find the job or earn, interacting with new individuals balancing extracurricular and leisure activities, and combining community and family responsibilities are just a few of the problems that students face while attending a university [6]. For students, this may be a very taxing experience. This pressure causes an increasing percentage of students to endure emotional anguish during their university studies.

No one should ignore the problems related to mental health. Nevertheless, some others view these as insignificant occurrences to the point of making fun of the individual who is suffering. This makes things harsher and could result in drastic repercussions for the victim. For students, this is particularly the case. Diagnoses of mental health problems can be challenging. The person who is suffering frequently is unaware of his condition. This study's main motivation is to examine how different personal qualities may be used to examine university students' mental health. A predictive model based on these characteristics used to identify probable mental health disorders among university students is built in this study.

II. LITERATURE REVIEW

One of the key ways to stop students from experiencing a psychological crisis is to develop a system for managing their psychological data. Many scientists are attempting to apply machine learning to anticipate complicated psychological problems that people may experience, such as anticipating stress and anxiety disorders [7]. Only the traits that are most accurate in predicting suicide risk can reach thousands, according to studies. The suicide of Weibo users is assessed in actual time using a multilayer perceptron algorithm built specifically for the social media site. The forecast accuracy percentage might be as high as 94% in terms of possibility [7]. Luo [7] created an initial data collection using the user's response outcomes. A psychological exam data set to be mined is generated after pre-processing information exchange, extraction, cleaning, and conversion. The mining process uses the decision tree method, and the results are used to classify the data and determine the user's mental health condition. A subfield of mathematics known as fuzzy mathematics examines and resolves ambiguity in mathematical situations. The introduction of fuzzy mathematics, rooted in the 1965 proposal of fuzzy set theory, presents a potential avenue for resolving ambiguity in mathematical situations, although its application in system degradation scenarios remains underexplored within the reviewed studies [8]. Furthermore, the use of a 3-layer BP neural network in research demonstrates a binary classification approach to mental health conditions. However, the review lacks an in-depth discussion on how fuzzy classifiers, particularly fuzzy decision trees or fuzzy SVM, have been employed in addressing system degradation and improving the analysis of psychological data [9]. Addressing these gaps would contribute to a more comprehensive understanding of the previous studies and their approaches to system resilience in the context of psychological data management.

The difficulties of college life were proven to make college students more susceptible to worry. The neural network exhibited the greatest accuracy, at 74%, across several machine learning techniques [10]. According to a study [11], working individuals may be predicted to be in good mental health using data mining techniques. The researches employ Decision Tree, Naive Bayes, and Random Forest; Decision Tree exhibits the best accuracy (82.2%), according to the authors. Low self-confidence and poor mental health are caused by higher BMI. Research [12] discussed a technique for spotting mental illness on social media. In order to identify a mental disorder, this study analyzed Twitter data using the Twitter API. This essay provides an explanation of the MIDAS system, which uses tweets to identify mental illness. Research [13] was undertaken with the intention of recommending a psychiatrist who predicted depression among college freshmen. The purpose of this study is to learn more about the factors that lead to depression in Bangladeshi undergraduate students. With effectiveness and F-measure of about 75% and 60%, respectively, Support Vector Machine and Random Forest were determined to be the best algorithms, however, Random Forest had a superior precision, recall, and fewer false negatives [13]. Paper [14] provides an explanation of the MIDAS system, which uses tweets to identify mental illness. Using Random Forest, predictions are achieved with an accuracy of 96%. In a similar manner, XGBoost, support vector machines, random forests, and neural networks were utilized to predict teenage mental health issues. The results show that even with a k value of 17, the accuracy is at its maximum point (0.85). Different machine learning algorithms were utilized to predict mental health difficulties in scenarios where there were [15]. In the experiment [16], implanted sensors were used to track 10 students and evaluate how stressed they were throughout examinations. Using the ECG and electrodermal activity data as input, several classification algorithms K-nearest neighbor, linear discriminant analysis, and support vector machines were applied. Results show that for the three states: relaxed, written test, and oral exam, recognition accuracy ranges from 86 to 91%. There is some knowledge available about the psychological health of Aboriginal Victorians. In a 1992 review of patient records from the Victorian Aboriginal Healthcare System (VAHS), depression was found to be the most prevalent mental condition, affecting 54% of patients. The percentages of suicidal ideations were 23% & 24%, correspondingly, among 172 Koori adolescents (12–26 years old) enrolled in VAHS, and this was associated with socioemotional well-being [17]. A research-generated [18] method involves grouping n samples into k clusters, dividing the remaining points into the closest cluster using the distance formula or another similarity calculation technique, and then determining the median of all the items within the cluster also as the new center point. The iteration process should be repeated until the goal function converges.

Surprisingly, the rate of access to public healthcare coverage care is very low when compared to the number of victims. Only 25% of victims globally receive the right course of therapy, and only half of the patients receive care through the public health services. According to the survey, many of the victims who did not use the public health services were not properly informed on how to handle their situation [19]. According to research [20], Naive Bayes provides accurate results for small datasets, but accuracy does not scale up for huge databases. Additionally, a significant number of rules are created via a priori association rule mining, which slows down searches and makes interpretation by domain experts challenging. The rules produced by decision trees are often straightforward and simple to understand, and they are more accurate than trees from logistic models. Authors of [21] employed machine learning methods to identify mental illnesses, including schizophrenia and mania. This is another fascinating example of mining mental medical data. Their argument was that more precise diagnostic categorization systems may be created by examining data or language and interactions between psychiatrists and patients. Algorithms for grouping data and classifying emotions were utilized in the study.

However, the application of data mining techniques emerges as a crucial strategy in proactively addressing mental health issues among university students. Through the analysis of factors influencing students' mental well-being, data mining provides valuable insights [1]. The creation of a specialized dataset using a survey method further refines the focus of studying mental health issues in this demographic group [2]. The integration of classification and regression techniques, coupled with student input, not only enhances the precision of the findings but also adds relevance to the study [2]. The comparative assessment of various data mining techniques, including Naive Bayes, Random Forest, SVM, KNN, and Decision Tree, aids in identifying the most effective approach for predicting mental health status [1]. This approach ensures a comprehensive evaluation of the model's performance by considering metrics such as F-1 score, precision, recall, AUC, and CA [3, 4]. Therefore, the main contributions, novelty and academic values of the studies are mentioned below:

1. This current study generates a new dataset using a survey method, adding academic value by offering a specialized resource for examining mental health issues in university students.
2. The study utilizes classification and regression techniques to analyze results, incorporating student recommendations in the survey design. This approach contributes to a deeper understanding of the factors influencing mental health in university students.
3. Research findings emphasize the importance of data mining in comprehending the factors impacting students' mental health. The study provides valuable insights into the relationships between various aspects of students' lives and depression.
4. The study compares the performance of different data mining techniques, including Naive Bayes, Random Forest, SVM, KNN, and Decision Tree. This comparative analysis contributes to the field of predictive modeling in mental health research.

III. PROPOSED METHOD

This section describes the proposed method comprising several subsections including dataset collection, data cleaning, data transformation, training model, and evaluation. The block diagram of the proposed method is shown in Figure 1.

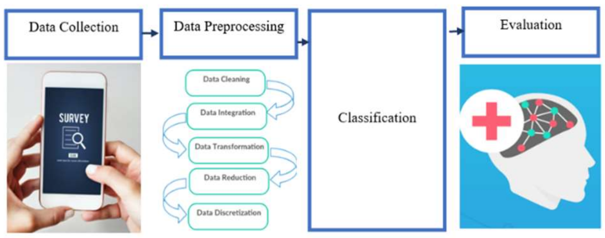


Figure 1. Block Diagram of Proposed Method.

A. DATASET

For this research, we created a questionary from which we did a survey among university students. And based on the responses we created a dataset. The dataset includes 16 variables that are associated with the student data. The variables include students' department, age, gender, year of study, involvement in projects(yes/no), influenced by anyone(yes/no), like taking initiatives(yes/no), health condition rating(1-5), self-confidence rating(1-5), have any goal in life (yes/no), concentration rating(1-5), hours of class per week, hours of sleep per day, food habit(poor/normal/over-eating), have a hobby(yes/no), involved in any extracurricular activities(yes/no), in a relationship(yes/no), any recent break-up(yes/no), anxiety level rating(1-5), have depression(yes/no). In this research, the end goal is to design a prediction model to identify mental health status.

B. DATA CLEANING

Finding data anomalies like null values, outliers, etc. is a part of this procedure. The dataset was cleaned up by removing columns that were not very useful for the study.

C. DATA INTEGRATION

Data integration is used to combine data. Whatever, the integrated dataset can then be used for analysis or further processing.

D. DATA TRANSFORMATION

The purpose of this phase was to identify words that had the same meaning but different spellings by examining separate values in each column. In this case, the cell's value must be transformed into a single form. To make use of prediction algorithms more effectively, the data is transformed into a numerical representation. All yes and no responses were replaced with 1's and 0's. The class time was broken up into 5 halves, each lasting 6 hours. A time interval from 1 to 6 hours is identical to 1, from 7 to 12 hours is comparable to 2, from 13 to 18 hours is equivalent to 3, from 19 to 24 hours is equivalent to 4, from 25 to 30 hours, and any data beyond 30 hours are denoted as 5. The sleeping time was split into 5 halves, each lasting two hours. A time interval from 1 to 2 hours is designated as 1, followed by 3 to 4 hours as 2, 5 to 6 hours as 3, from 8 to 9 hours as 4, and more than 10 hours as 5. The numbers that were ranged values have been averaged and rounded to their closest integers.

E. DATA REDUCTION

Data reduction is used for reducing the size or complexity of a dataset while retaining its meaningful information.

F. DATA DISCRETIZATION

Data discretization is used for transforming continuous data into discrete values or intervals.

G. MODELS

To make a prediction Naive Bayes, Random Forest, SVM, KNN, and Decision Tree, the five algorithms are used in this research. Table 1 provides a comparison of the prediction of models performance. The goal is to identify the model that will work best for this research challenge.

H. NAÏVE BAYES

Naive Bayes is still one of the top 10 data mining algorithms, which is frequently used in big data analysis and other domains [22, 23]. It is a straightforward yet effective algorithm with many practical uses, from recommending products and diagnosing illnesses to guiding driverless cars [24]. For building very big datasets and for additional analysis, the Naive Bayes model is appropriate. This model is a very advanced categorization system that is both easy and effective in challenging situations [25]. Finding a solid hypothesis for any circumstance or occurrence is the key characteristic of the Naive Bayes classification [26].

I. RANDOM FOREST

The Random Forest is recognized for its effectiveness and simplicity [27]. Numerous distinct decision trees work together as an ensemble to make up random forests [28]. The model forecasts which class will receive the most votes based on predictions made by each and every tree in the random forest [28]. It may alternatively be described as a Classifier Based on Decision Trees that selects most effective classification tree to use as the algorithm final classification through voting [27]. Because it has good properties like proximities, variable importance measures, out-of-bag errors, etc. Random Forest is the most often used group classification algorithm [27]. Since there are no fixed learning parameters for random forests, they are non-metric classifiers [28]. With a bigger number of variables, the feature selection problem can be handled using Random Forest, which has shown to be a very helpful approach [29].

J. SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines are the largest research area of data mining [30]. Support Vector Machines is a popular supervised learning approach and are applied to solve regression and classification problems [31]. In SVM, the dataset is often divided into a training set and a testing set. The training dataset instances all have a number of different properties or features in addition to target values. SVM's primary goal is to develop a model that forecasts the goal values of the test data using the training data based on those data features. Since SVM only utilizes a portion of the training points, it is also memory efficient. It is applied in intrusion detection systems, picture processing, and text categorization [32]. SVM has the unique ability to simultaneously maximize the geometric margin and reduce the empirical classification error [27].

K. K-NEAREST NEIGHBOR

K-Nearest Neighbor is a popular data mining algorithm, which was widely studied last decade [33]. KNN is a nonparametric technique used for regression and classification [34]. The KNN, perhaps the simplest, makes no previous assumptions regarding the probability distribution of the observed data [35].

Euclidean distance is utilized as the metric and the number of closest neighbors is set to two (k=5) in this research. In KNN classifications, Euclidean distance is the most often used distance metric. By manually changing the value of k, the neighbors were tested (from 1 to 5). According to the data, changing the value of k somewhat altered the outcome of the forecast. The optimal value of k was selected to get the best outcome based on the findings.

L. DECISION TREE

A decision tree is a basic tree-based modeling technique that categorizes or forecasts the dependent variable values through the learning of different decisions based on all relevant information [36]. A decision tree is a classification technique that appears like a tree with branches connecting nodes of two different types. The tree is made up of internal nodes that satisfy the test of logical attributes and connected branches that show the results of the test. The decision tree categorizes cases by arranging them to the leaf from the root nodes of the tree. The benefit of decision trees is that there are many effective techniques that can locate roughly optimum tree structures. Additionally, decision trees have the ability to divide the difficult challenge of decision-making into multiple more straightforward ones [37].

M. MODEL EVALUATION

To evaluate the performance of a model, a test dataset is used, which is separate from the data used to train the model. The performance of the model is assessed using various metrics, such as accuracy, F-1 score, precision, recall, AUC and CA. These metrics are calculated based on the predictions made by the model on the test dataset

1) ACCURACY

Accuracy is a metric used to assess the overall effectiveness by measuring its ability to correctly predict the class labels of a set of test data. It shows the probability of the true value of the class label, which helps to evaluate how well the algorithm is performing

2) F-1 SCORE

The F-1 score is a metric used to evaluate the performance of algorithms in classification tasks, and it is the harmonic mean of precision and recall. It works by measuring the balance between precision and recall, where precision is the proportion of true positives among all predicted positives and recall is the proportion of true positives among all actual positives. In data mining, the F-1 score is commonly used to evaluate the effectiveness of classification algorithms in identifying and correctly classifying different data patterns. It provides a more balanced evaluation of the model performance than just looking at accuracy alone, especially when the data is imbalanced.

3) PRECISION

Precision is a metric used to assess the predictive power of a particular class label (either positive or negative) by measuring the proportion of correctly predicted samples among all the samples predicted as belonging to that class. It helps to evaluate the accuracy of the algorithm in identifying true positives and minimizing false positives.

4) RECALL

Recall measures the proportion of true positive instances that are correctly identified by the model out of all actual positive instances.

5) AUC

The Area Under the Curve (AUC) is used to evaluate the performance of a binary classification model, which involves predicting whether an instance belongs to one of two classes. The AUC metric measures the overall ability of the model to distinguish between positive and negative classes by calculating the area under the Receiver Operating Characteristic (ROC) curve.

6) CA

Correspondence Analysis (CA) is a multivariate statistical technique used to analyze the relationship between categorical variables. It works by converting the data into a set of scores, which can then be visualized in a low-dimensional space, typically a two-dimensional plot.

N. TOOLS USED

Weka and Orange, the most popular data mining tools are used which provide a user-friendly interface and a wide range of algorithms for data analysis. They allow us to pre-process, visualize, and model data, making them powerful tools for exploratory data analysis and predictive modelling.

O. ML METHOD USING CAUSE

ML methods are used to find out the benchmark accuracy, f-1 score, precision, recall, AUC, CA. Based on the findings, the best result of the used model can be determined.

IV. RESULTS

There are two parts in this section; the first part discusses some exploratory examinations of university students' mental health.

The second part compares the effectiveness of the statistical models applied in this study.

- Exploratory analysis: 94 out of 176 students, that is, 53.4% of the students, have reported that they have depression. 52.6% of students under 24 said they are depressed. Among them, 73.33% are male. 69.2% of students who do not like to take initiative are depressed. 58.3% of students who are not involved in extracurricular activities are depressed. 72.5% of students who have high anxiety and sleep less have depression. 70% of people, who are in poor health and have poor eating habits, have depression. 100% of students who are in poor health and have overeating food habits have depression.
- Predictive analysis:

Table 1. Predictive analysis.

Table Head		Model				
	Accuracy	*F1-Score*	*Precision*	*Recall*	*Auc*	*CA*
Naïve Base	65.91%	0.659	0.660	0.659	0.688	0.659
Random Forest	63.07%	0.631	0.631	0.631	0.662	0.631
SVM	61.36%	0.611	0.612	0.614	0.647	0.614
KNN	57.95%	0.580	0.584	0.580	0.600	0.580
DecisionTree	55.68%	0.557	0.560	0.557	0.608	0.557

Whatever, we represent our found results below graphically in figures 2, 3, 4, 5,6,7, and 8.

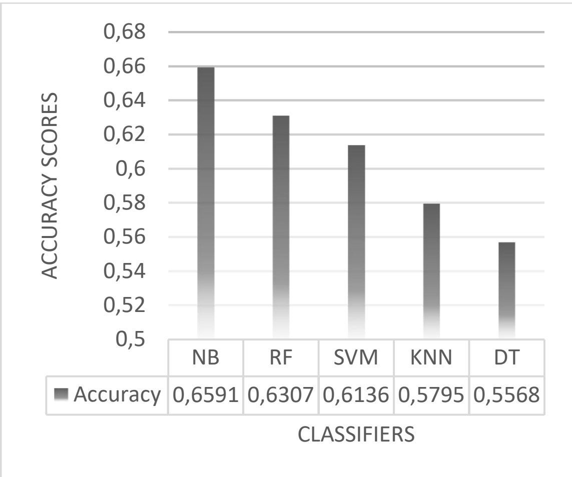


Figure 2. Accuracy vs Classifiers.

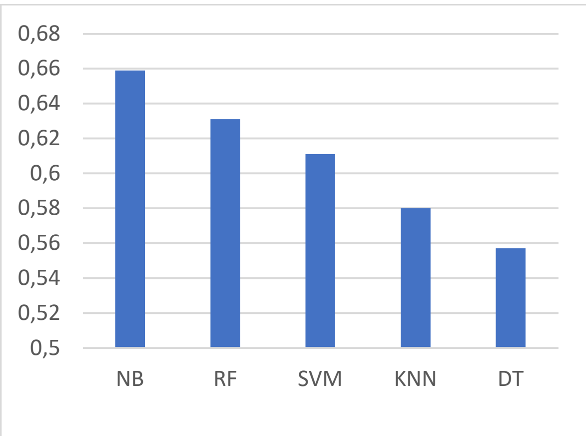


Figure 3. F-1 score vs Classifiers

Figure 4. Precision vs Classifiers

Figure 5. Recall vs Classifiers

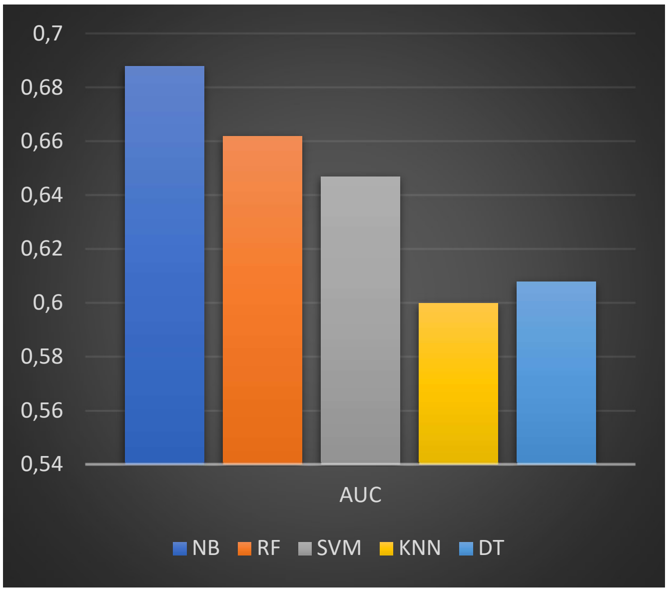


Figure 6. AUC vs Classifiers

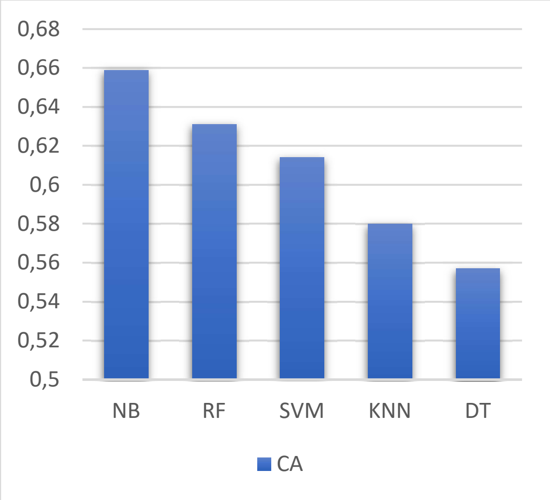
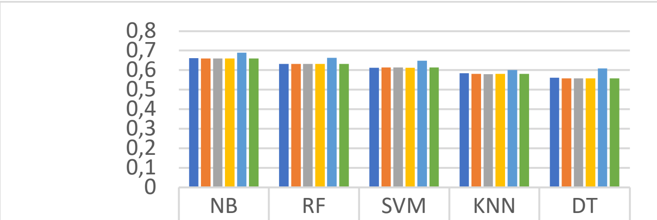


Figure 7. CA vs Classifiers.



	NB	RF	SVM	KNN	DT
Precision	0,66	0,631	0,612	0,584	0,56
Recall	0,659	0,631	0,614	0,58	0,557
Accuracy	0,6591	0,6307	0,6136	0,5795	0,5568
F-1 Score	0,659	0,631	0,611	0,58	0,557
AUC	0,688	0,662	0,647	0,6	0,608
CA	0,659	0,631	0,614	0,58	0,557

Figure 8. Summary of all Classifiers

V. DISCUSSION

Awareness of mental health is essential since it impacts every aspect of everyday living. We attempted to demonstrate in this research how data mining may be utilized to comprehend elements impacting students' mental health. We also examined the outcomes of five different data mining techniques. According to an exploratory investigation, male students experience depression at a higher rate than female students. Students who are more active have lower risks of developing depression. For a more realistic presentation, the percentage is used to compare outcomes rather than the frequency. Naive Bayes outperforms the other four algorithms, with an accuracy of 65.91%. In terms of accuracy, Naïve Bayes performs better than the other three models which is why mental health can be detected efficiently by Naïve Bayes.

VI. CONCLUSION

The paper shows the correlation between several aspects of university students' daily lives and depression. Based on gender compared to female students, male students experience more depression. Depression has a significant relationship with poor health conditions and overeating habits. According to the findings, participating in extracurricular activities lowers one's risk of developing depression. Additionally, it may be observed that most depressed students dislike taking the initiative. In order to forecast depression based on the characteristics in the dataset, the study examined five predictive modelling techniques: Naive Bayes, Random Forest, SVM, KNN, and Decision Tree. According to the performance measures, Nave Bayes is the algorithm that works best with the available dataset for this prediction task. The performance of the prediction approaches is affected by the relatively small amount of data (176 rows).

References

[1] E. Beresin, "The college mental health crisis: Focus on overall wellbeing," *Psychology Today*. [Online]. Available at: https://www.psychologytoday.com/us/blog/inside-out-outside-in/201703/the-college-mental-health-crisis-focus-overall-wellbeing.

[2] R. Kadison, *College of the Overwhelmed: The Campus Mental Health Crisis and What to do About it*, Jossey-Bass, 2004, 304 p.

[3] C. Loughlin, & J. Barling, "Young workers' work values, attitudes, and behaviours," *Journal of Occupational and Organizational Psychology*, vol. 74, issue 4, pp. 543–558, 2001. https://doi.org/10.1348/096317901167514.

[4] J. J. Arnett, "Emerging adulthood: A theory of development from the late teens through the twenties," *American Psychologist*, vol. 55, issue 5, pp. 469-480, 2000. https://doi.org/10.1037//0003-066X.55.5.469.

[5] S. E. Ross, B. C. Niebling, T. M. Heckert, "Sources of stress among college students," *College Student Journal*, vol. 33, issue 2, pp. 312-312, 1999.

[6] J. M. Sprung, A. Rogers, "Work-life balance as a predictor of college student anxiety and depression," *Journal of American College Health*, vol. 69, issue 7, pp. 775-782, 2021. https://doi.org/10.1080/07448481.2019.1706540.

[7] M. Luo, "Research on students' mental health based on data mining algorithms," *Journal of Healthcare Engineering*, vol. 2021, Article ID 1382559, 2021. https://doi.org/10.1155/2021/1382559.

[8] J. Liu, G. Shi, J. Zhou, Q. Yao, "Prediction of college students' psychological crisis based on data mining," *Mobile Information Systems*, vol. 2021, Article ID 9803212, pp. 1-7, 2021. https://doi.org/10.1155/2021/9979770.

[9] J. Qinghua, "Data mining and management system design and application for college student mental health," *Proceedings of the 2016 IEEE International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS'2016)*, December 2016, pp. 410-413. https://doi.org/10.1109/ICITBS.2016.96.

[10] H. Alharthi, "Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology," *Proceedings of the 2020 19th IEEE International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES'2020)*, October 2020, pp. 218-221. https://doi.org/10.1109/DCABES50732.2020.00064.

[11] V. Laijawala, A. Aachaliya, H. Jatta, & V. Pinjarkar, "Classification algorithms based mental health prediction using data mining," *Proceedings of the 2020 5th IEEE International Conference on Communication and Electronics Systems (ICCES'2020)*, June 2020, pp. 1174-1178. https://doi.org/10.1109/ICCES48766.2020.9137856.

[12] E. Saravia, C. H. Chang, R. J. De Lorenzo, & Y. S. Chen, "MIDAS: Mental illness detection and analysis via social media," *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'2016)*, August 2016, pp. 1418-1421. https://doi.org/10.1109/ASONAM.2016.7752434.

[13] A. A. Choudhury, M. R. H. Khan, N. Z. Nahim, S. R. Tulon, S. Islam, & A. Chakrabarty, "Predicting depression in Bangladeshi undergraduates using machine learning," *Proceedings of the 2019 IEEE Region 10 Symposium (TENSYMP'2019)*, June 2019, pp. 789-794. https://doi.org/10.1109/TENSYMP46218.2019.8971369.

[14] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, & R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PloS one*, vol. 15, issue 4, e0230389, 2020. https://doi.org/10.1371/journal.pone.0230389.

[15] U. S. Reddy, A. V. Thota, & A. Dharun, "Machine learning techniques for stress prediction in working employees," *Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC'2018)*, December 2018, pp. 1-4. https://doi.org/10.1109/ICCIC.2018.8782395

[16] A. Hasanbasic, M. Spahic, D. Bosnjic, V. Mesic, & O. Jahic, "Recognition of stress levels among students with wearable sensors," *Proceedings of the* 2019 18th IEEE International Symposium INFOTEH-JAHORINA (INFOTEH'2019), March 2019, pp. 1-4. https://doi.org/10.1109/INFOTEH.2019.8717754.

[17] K. Adams, C. Halacas, M. Cincotta, & C. Pesich, "Mental health and Victorian aboriginal people: what can data mining tell us?" *Australian Journal of Primary Health*, vol. 20, issue 4, pp. 350-355, 2014. https://doi.org/10.1071/PY14036.

[18] J. Li, "Analysis of the mental health of urban migrant children based on cloud computing and data mining algorithm models," *Scientific Programming*, vol. 2021, pp. 1-7, 2021. https://doi.org/10.1155/2021/7615227.

[19] R. Vanlalawmpuia, & M. Lalhmingliana, "Prediction of depression in social network sites using data mining," *Proceedings of the 2020 4th IEEE International Conference on Intelligent Computing and Control Systems (ICICCS'2020)*, May 2020, pp. 489-495. https://doi.org/10.1109/ICICCS48265.2020.9120899.

[20] S. D'monte, & D. Panchal, "Data mining approach for diagnose of anxiety disorder," *Proceedings of the IEEE International Conference on Computing, Communication & Automation*, May 2015, pp. 124-127. https://doi.org/10.1109/CCAA.2015.7148357.

[21] J. Diederich, A. Al-Ajmi, & P. Yellowlees, "Ex-ray: Data mining and mental health," *Applied Soft Computing*, vol. 7, issue 3, pp. 923-928, 2007. https://doi.org/10.1016/j.asoc.2006.04.007.

[22] H. Zhang, L. Jiang, & L. Yu, "Attribute and instance weighted naive Bayes," *Pattern Recognition*, vol. 111, 107674, 2021. https://doi.org/10.1016/j.patcog.2020.107674.

[23] H. Chen, S. Hu, R. Hua, & X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, issue 1, pp. 1-12, 2021. https://doi.org/10.1186/s13634-021-00742-6.

[24] I. Blanco-Montenegro, R. De Ritis, & M. Chiappini, "Imaging and modelling the subsurface structure of volcanic calderas with high-resolution aeromagnetic data at Vulcano (Aeolian Islands, Italy)," *Bulletin of Volcanology*, vol. 69, pp. 643-659, 2007. https://doi.org/10.1007/s00445-006-0100-7.

[25] V. Jackins, S. Vimal, M. Kaliappan, & M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *The Journal of Supercomputing*, vol. 77, pp. 5198-5219, 2021. https://doi.org/10.1007/s11227-020-03481-x.

[26] M. Ritonga, M. A. A. Ihsan, A. Anjar, & F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, 012045, 2021. https://doi.org/10.1088/1757-899X/1088/1/012045.

[27] N. M. Abdulkareem, A. M. Abdulazeez, D. Q. Zeebaree, & D. A. Hasan, "COVID-19 world vaccination progress using machine learning classification algorithms," *Qubahan Academic Journal*, vol. 1, issue 2, pp. 100-105, 2021. https://doi.org/10.48161/qaj.v1n2a53.

[28] P. Sadorsky, "A random forests approach to predicting clean energy stock prices," *Journal of Risk and Financial Management*, vol. 14, issue 2, p. 48, 2021. https://doi.org/10.3390/jrfm14020048.

[29] R. C. Chen, C. Dewi, S. W. Huang, & R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, issue 1, pp. 1–26, 2020. https://doi.org/10.1186/s40537-020-00327-4.

[30] Z. Xu, D. Yu, & X. Wang, "A bibliometric overview of International Journal of Machine Learning and Cybernetics between 2010 and 2017," *International Journal of Machine Learning and Cybernetics*, vol. 10, pp. 2375-2387, 2019. https://doi.org/10.1007/s13042-018-0875-9.

[31] A. Kurani, P. Doshi, A. Vakharia, & M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," *Annals of Data Science*, vol. 10, issue 1, pp. 183-208, 2023. https://doi.org/10.1007/s40745-021-00344-x.

[32] B. S. Bhati, & C. S. Rai, "Intrusion detection technique using Coarse Gaussian SVM," *International Journal of Grid and Utility Computing*, vol. 12, issue 1, pp. 27-32, 2021. https://doi.org/10.1504/IJGUC.2021.112458.

[33] D. Zhao, X. Hu, S. Xiong, J. Tian, J. Xiang, J. Zhou, & H. Li, "K-means clustering and kNN classification based on negative databases," *Applied Soft Computing*, vol. 110, 107732, 2021. https://doi.org/10.1016/j.asoc.2021.107732.

[34] X. Luo, D. Li, Y. Yang, & S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, pp. 1-10, 2019. https://doi.org/10.1155/2019/4145353.

[35] T. R. Lee, W. T. Wood, & B. J. Phrampus, "A machine learning (kNN) approach to predicting global seafloor total organic carbon," *Global Biogeochemical Cycles*, vol. 33, issue 1, pp. 37-46, 2019. https://doi.org/10.1029/2018GB005992.

[36] S. Park, S. Y. Hamm, & J. Kim, "Performance evaluation of the GIS-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling," *Sustainability*, vol. 11, issue 20, 5659, 2019. https://doi.org/10.3390/su11205659.

[37] S. Križanić, "Educational data mining using cluster analysis and decision tree technique: A case study," *International Journal of Engineering Business Management*, vol. 12, 1847979020908675, 2020. https://doi.org/10.1177/1847979020908675.

[38] F. Sahlan, F. Hamidi, M. Z. Misrat, M. H. Adli, S. Wani, & Y. Gulzar, "Prediction of mental health among university students," *International Journal on Perceptive and Cognitive Computing*, vol. 7, issue 1, pp. 85-91, 2021.

[39] N. Fahad, K. M. Goh, M. I. Hossen, K. S. Shopnil, I. J. Mitu, M. A. H. Alif, & C. Tee, "Stand up against bad intended news: An approach to detect fake news using machine learning," *Emerging Science Journal*, vol. 7, issue 4, pp. 1247-1259, 2023. https://doi.org/10.28991/ESJ-2023-07-04-015.

[40] N. Fahad, K. O. M. Goh, M. I. Hossen, C. Tee, & M. A. Ali, "Building a fortress against fake news: Harnessing the power of subfields in artificial intelligence," *Journal of Telecommunications and the Digital Economy*, vol. 11, issue 3, pp. 68-83, 2023. https://doi.org/10.18080/jtde.v11n3.765.

***MD. NAZMUL HOSSAIN**, a Lecturer at American International University-Bangladesh. His main research interests are data science, data mining, machine learning, and computer science.*

NAFIZ FAHAD, *a Student & Research Assistant, Faculty of Science and Technology, Department of Computer Science, American International University-Bangladesh. His main research interests are in the fields of machine learning, data mining, computer science, AI, sensors, image processing, and data science.*

***RASEL AHMED**, a Student, Faculty of Science and Technology, Department of Computer Science, American International University-Bangladesh. His main research interests are in the fields of machine learning, data mining, computer science, and data science.*

***ANIK SEN**, a Student, Faculty of Science and Technology, Department of Computer Science, American International University-Bangladesh. His main research interests are in the fields of machine learning, data mining, computer science, and data science.*

MD. SADI AL HUDA, *an Undergraduate Student, Faculty of Science and Technology, Department of Computer Science, American International University-Bangladesh. His main research interests are in the fields of machine learning, AI, data mining, computer science, and data science.*

Md. ISMAIL HOSSEN, *a Lecturer at American International University-Bangladesh. His main research interests are data science, data mining, machine learning, sensors, and computer science.*