

ORIGINAL RESEARCH

Understanding Psychological Dimensions of Student Life Using Machine Learning Approach

Md. Ismiel Hossen Abir¹, Taspia Salam²

^{1,2} Department of Computer Science & Engineering, International Standard University, Dhaka, Bangladesh

Abstract

Students experience feelings of depression occasionally. During such a period, they may experience a decline in their academic performance and may also exhibit antisocial behavior. This study provides a thorough examination of a dataset derived from survey response of 100 students of Computer Science, with the objective of revealing the underlying causes of their diverse human behaviors. The relationship between depression and academic performance based on sleep patterns, demanding tasks, and social engagements are examined. The study also examines the relationships among factors like age, gender, academic achievement, mental health indicators, note-taking in class and sleep cycle. The Random Forest Classifier model in a meticulously designed pipeline is employed, incorporating a preprocessing technique that encompasses the Column Transformer, Ordinal Encoder, Standard Scaler, and One Hot Encoder. The objective of this research is to identify the underlying patterns within the survey data. Every column in the dataset offers a distinct viewpoint on the different element of individuals' lives, including demographic distribution, learning behaviors, mental health, and social connections. The Random Forest Classifier, implemented using a pipeline design, achieves an accuracy rate of 85 percent. In addition, hyperparameter tuning using Grid Search CV is used to improve the performance of the model.

Keywords: Depression, Academic Performance, Socio-Behavioral Factors, Computer Science, Machine Learning

1 | Introduction

Universities are prestigious institutes where individuals cultivate and refine their critical thinking skills. The term "powerhouse" can be used to describe it as the central source of strength and influence for any nation. Nevertheless, the life of a university student can be arduous. After graduating, individuals are faced with several obligations, including family and financial commitments. Conversely, the behavior of each individual is contingent upon the surroundings, society, and economic circumstances [1-2].

In this study, a survey was administered to investigate the psychosocial challenges experienced by university students. The aim was to examine the relationship between mental

well-being and academic success, using a range of parameters. A survey of 100 students of the Computer Science department was conducted to be informed regarding depression levels, sleep patterns, academic achievements, obstacles faced in completing academic duties, and their social standing. This implementation utilizes a machine learning approach known as random forest. The preprocessing techniques utilized standard scaler, one hot encoder, ordinal encoder, and label encoder. Next, the column transformer and pipeline technique were employed. It was observed that the students who take notes during the class tend to have superior academic achievement compared to their peers.

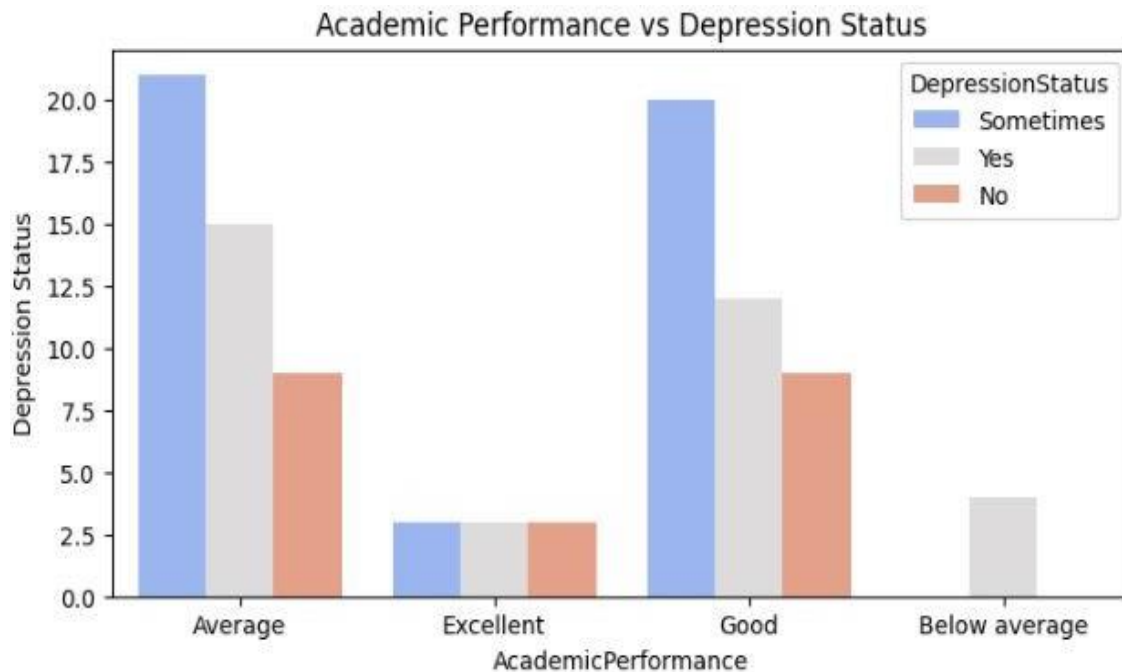


Figure 1: Academic Performance vs. Depression Status

figure-1 indicates that the pupils who experience depression have below-average academic achievement. On the other hand the students who have sufficient sleep demonstrate exceptional academic success. The objective of this study was to establish the association between academic achievement and depression, considering many associated parameters.

2 | Organization of the Paper

Many students commonly have mental health challenges. Multiple researches attempted to examine and comprehend the relationship among students' cognition, psychological well-being, and their college setting. This study seeks to investigate the influence of mental health and social behaviors of undergraduate students in the computer science department on their academic performance. We examined the impact of depression on students' academic performance. Additionally, there were some factors to consider, such as the influence of sleep patterns on academic achievement. The impact of study habits and time management skills on student achievement is evident, emphasizing the significance of employing good study tactics to attain academic objectives [3-5].

Recent research works utilized predictive modeling methodologies in the fields of machine learning and education. The objective was to identify the characteristics that have a positive impact on students'

academic achievements. For example, using documented psychological measurements enrollment in educational programs, and socio-economic information, a predictive algorithm was built to forecast pupils' academic performance. Employing machine learning algorithms to forecast student attrition rates demonstrates the efficacy of data-driven approaches in identifying students at risk of dropping out and implementing targeted interventions.

In addition, progress in the natural language processing (NLP) has allowed academics to analyze written responses from pupils. NLP provides valuable insights into students' psychological well-being and learning behaviors. For instance, employing sentiment analysis techniques can be used to analyze the emotional content of students' reflective writings, revealing connections between emotional states and academic achievements. These studies have greatly advanced our understanding of the psychological aspects of student life. Additional research is required to investigate the interaction between mental health, study habits, and academic achievement. Our research aims to address the gap by utilizing machine learning approaches to construct predictive models. This tool can access students' academic achievement by taking into account their mental health status and study habits, thereby offering important observations for tailored intervention tactics in higher education [6-12].

3 | Dataset Compilation and Preprocessing

3.1 | Dataset Compilation

During the process of compiling the dataset, a survey was conducted when the students enrolled in the Computer Science department. A total of 100 students took part in this survey. The survey consisted of 10 questions covering various aspects such as age, gender, academic performance, note-taking habits, depression status, facing obstacles in completing academic tasks (such as presentation), average hours of sleep per day, number of friends, and interest in new things. The dataset consisted of 10 columns and 100 rows.

Figure-2 shows that 56.6 percent of the participants were male students, whereas 43.4 percent were female students. Subsequently, a data cleaning technique was employed due to the fact that certain students did not accurately complete the survey questions. To address the missing values, strategy of imputing the most common value was employed as well. Within the dataset, three columns are classified as numeric data type, whereas the remaining columns are classified as object data type. Duplicate rows are eliminated to enhance the accuracy of the model as duplicate data can impede its performance.

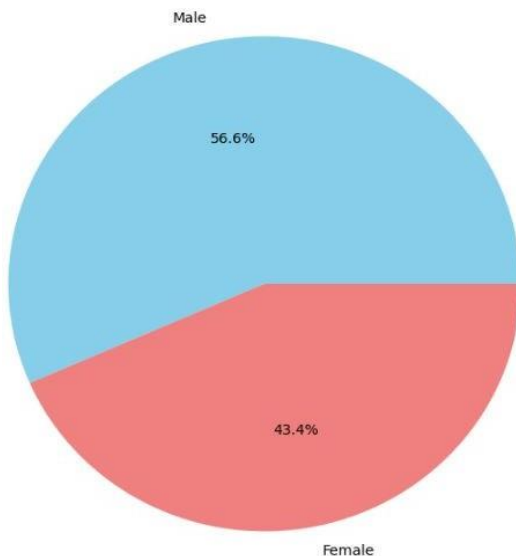


Figure 2: Pie Chart for Gender Analysis

3.2 | Data Preprocessing

Data preprocessing refers to the process of cleaning and transforming raw data into a format that was suitable for analysis. In the data preprocessing phase, the column designed to indicate depression state is the output column, while the rest of the columns were designated as input columns. One of the columns

categorizes individuals based on their gender. In order to preprocess this column, one hot encoder technique was employed due to the equal significance of both genders. The columns for sleep cycle and number of friends are numerical. Hence, it is imperative to establish a certain range in order to enhance the accuracy of the machine learning algorithm. Therefore, the conventional scalar approach was employed. For the output, the label encoder approach was used for the categorical column, while employing the ordinal encoder method for the remaining columns. Various classes were used from the scikit-learn package. Certain data visualization techniques were employed to gain insights into the data. Many data visualization libraries, including matplotlib, seaborn, and plotly express were employed as well.

figure-3 depicts a countplot illustrating the relationship between academic performance and age. Here, the academic performance of kids based on their age, is analyzed.

Based on the data presented in figure-4, it is evident that students who take notes in class exhibit superior academic achievement compared to their peers.

4 | Methodology

A supervised learning methodology was employed to construct a predictive model that comprehends the psychological aspects of student life. The scikit-learn library's Pipeline technique was employed in the model. In order to analyze the data, the column transformer approach was used to preprocess the complete dataset. Subsequently, the pipeline technique was used as well which involved a preprocessing and a random forest classifier model, as depicted in figure-5.

The preprocessor, which is a part of the pipeline, encapsulates the modifications that are done to the dataset during the preprocessing stage. Initially, the XGBoost model was used to assess the precision of the machine learning model. Next, the random forest classifier algorithm was implemented. figure-6 provides a description of the significance of random forest and the columns it encompasses.

In addition, hyperparameter tuning by grid search CV was used to enhance the accuracy of the system. Grid search CV enables the systematic search for the optimal parameters that can enhance the accuracy of the model. This leads to improved speed by employing optimization factors such as the number of trees and the maximum depth of trees. Grid search CV demonstrates string performance in cross-validation when conducting hyper parameter search.

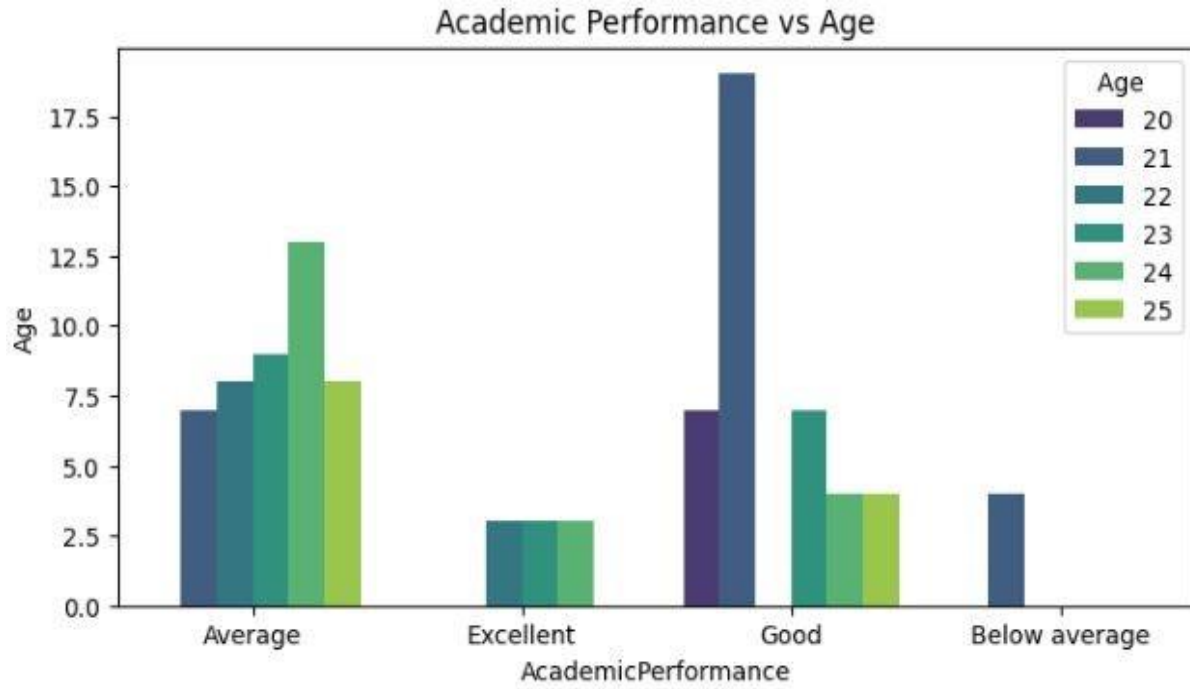


Figure 3: Academic Performance vs. Age

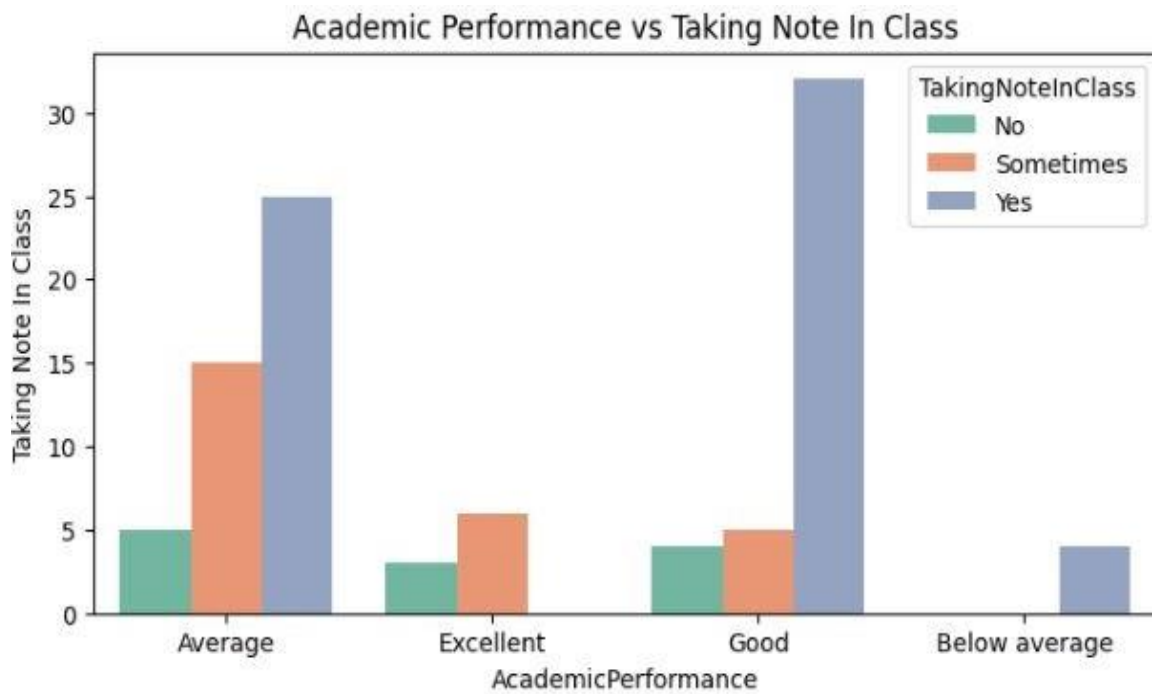


Figure 4: Academic Performance vs. Taking Note in the Class

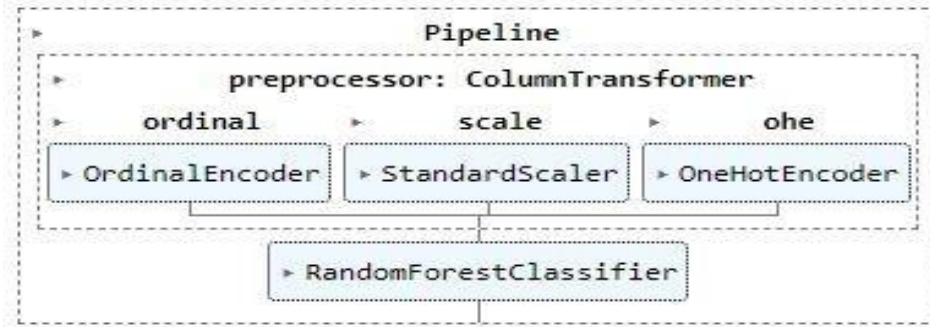


Figure 5: Pipeline and Machine Learning Model

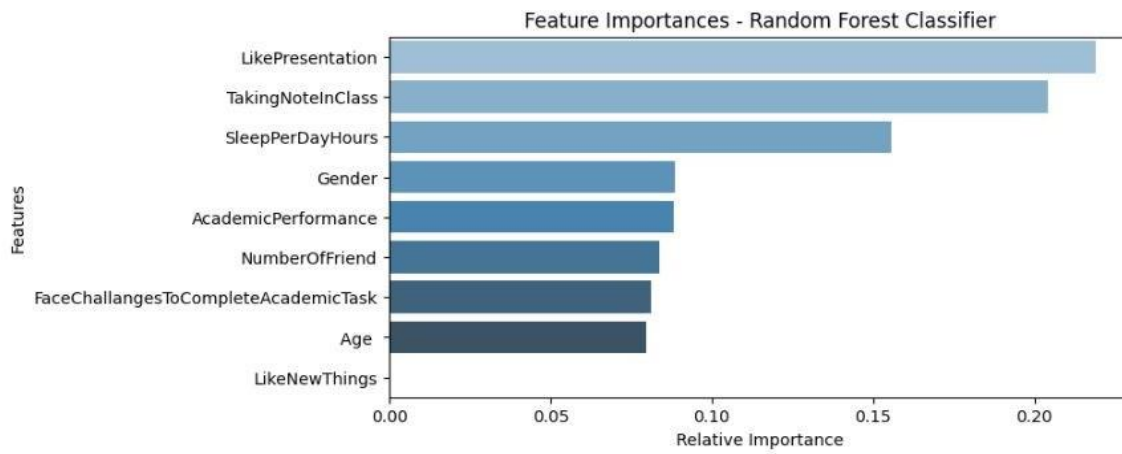


Figure 6: Features Importance Random Forest Classifier

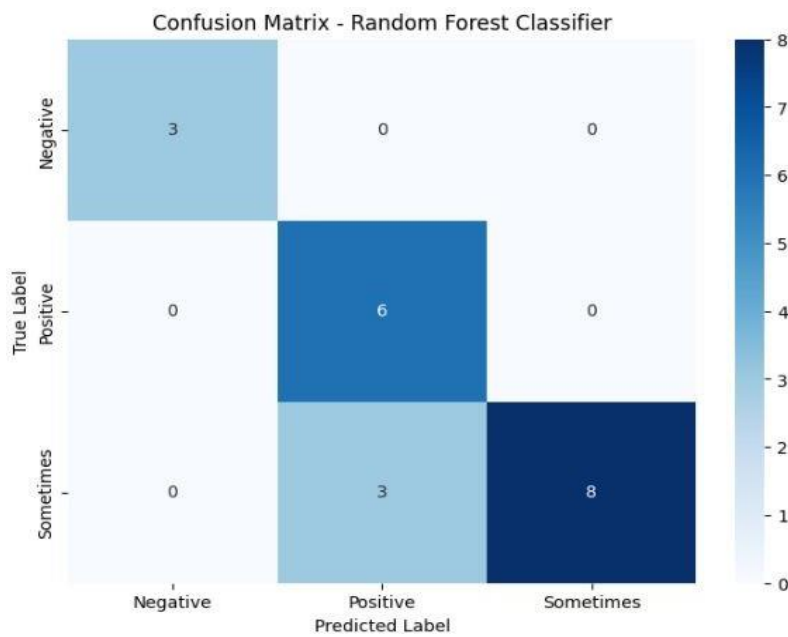


Figure 7: Confusion Matrix for Random Forest Classifier

5 | Results and Discussion

Upon executing the machine learning algorithm, it is observed that the model achieves an accuracy of 85 percent when utilizing the XGBoost algorithm. In contrast, the random forest classifier model has a slightly higher accuracy rate of 85.55 percent. After employing hyper parameter tweaking to determine the optimal parameters, the resulting accuracy is 86 percent. The accuracy rate serves as a measure of the efficiency of the method in gathering and analyzing a wide range of characteristics that impact student well-being and academic achievement.

figure 7 displays the confusion matrix for the random forest classifier algorithm. A confusion matrix provides a visual representation of the performance of the model. The matrix's rows correspond to the real classes of the data, while the columns correspond to the predicted classes. The confusing matrix indicates that the classifier accurately identified 23 instances as negative and 8 instances as positive.

Table 1: Classification Report – Random Forest Classifier

	Precision	Recall	F1-score	Support
0		1.00	1.00	3
1		0.67	1.00	6
2		1.00	0.73	11
Accuracy			0.85	20
Macro avg	0.89	0.91	0.88	20
Weighted avg	0.90	0.85	0.85	20

Table1 presents a classification report that describes the precision, recall, and F1-score of the machine learning algorithm utilized. The classifier's overall accuracy is 0.89, indicating that it accurately classified 89 percent of the cases in the dataset. The arithmetic mean of the precision, recall, and F1-score is also 0.85.

The approach effectively detects students' academic performance and its numerous relationships. The conclusive findings indicate that maintaining a consistent sleep duration of 8 hours per day, actively engaging in note-taking during class, having a smaller social circle, especially if they do not exhibit symptoms of sadness, and displaying a predisposition for embracing novel experiences, are all factors that may contribute to a high level of academic achievement. Additionally, it is noticed that even if kids adhere to all of these guidelines, their academic performance may decline if they are experiencing depression. Table1 presents a classification report that describes the

precision, recall, and F1-score of the machine learning algorithm utilized. The classifier's overall accuracy is 0.89, indicating that it accurately classified 89 percent of the cases in the dataset. The arithmetic mean of the precision, recall, and F1-score is also 0.85.

The approach effectively detects students' academic performance and its numerous relationships. The conclusive findings indicate that maintaining a consistent sleep duration of 8 hours per day, actively engaging in note-taking during class, having a smaller social circle, especially if they do not exhibit symptoms of sadness, and displaying a predisposition for embracing novel experiences, are all factors that may contribute to a high level of academic achievement. Additionally, it is noticed that even if kids adhere to all of these guidelines, their academic performance may decline if they are experiencing depression.

6 | Conclusion and Future Work

Finally, this study emphasizes the intricate connections among the mental well-being of university students (especially belonging to the Computer Science department), their academic achievements and the surrounding social milieu. This is a thorough examination of the survey feedback obtained from a sample of 100 students. It has discovered valuable insights into the factors that influence diverse human behaviors observed within this group.

It has also revealed the significance of addressing psychological and social sources of distress and implementing support systems to improve the mental and emotional well-being of students, as well as their academic achievements. The strong association between the way students take notes and their academic performance emphasizes the effectiveness of proactive learning tactics in promoting positive results. It is observed that students who consistently experience depression tend to have below-average academic performance. On the other side, if a student does not experience depression and devotes a significant amount of time to studying, his/her academic achievement will be higher than others. Thus, by giving careful consideration to university students' mental health and academic achievement, a resilient atmosphere can be created. This will guarantee them success, both in their personal lives and intellectual pursuits.

References

- [1] A. Geron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2017, ISBN: 9781491962299.
- [2] T. Chi, X. Liu, S. Takayanagi, T. Matsushita, and H. Kishimoto, "Association between mental health and academic performance among university undergraduates: The interacting role of lifestyle behaviors," *Int. J. Methods Psychiatr. Res.*, vol. 32, no. 1, p.1 e1938, 2023. DOI: 10.1002/mpr.1938.
- [3] A. Shashkov et al., "Analyzing student reflection sentiments and problem-solving procedures in MOOCs," in "Proceedings of the Eighth ACM Conference on Learning @ Sacle*", 2021, pp. 1-10. DOI: 10.1145/3430895.3460150
- [4] E. Dabbas, "Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of fully fledged fronted web framework in Python-no JavaScript required," Packt Publishing Ltd., 2021.
- [5] Plotly Express – Plotly Python Graphing Library," Plotly, [Online]. Available: <https://plotly.com/python/plotly-express/>. [Accessed: Feb. 16, 2024].
- [6] Waskom, M. (2020). seaborn: statistical data visualization. *Journal of Open Software*, 5(53), 3021. <https://doi.org/10.21105/joss.03021>
- [7] Matplotlib, "[Online]. Available: <https://matplotlib.org/>. [Accessed: Feb, 16, 2024].
- [8] Scikit-learn, "[Online]. Available: <https://scikit-learn.org/stable>. [Accessed: Feb, 16, 2024].
- [9] Breiman, L. (2001). Random forest, *Machine Learning*, 45(1), 5-32 <https://doi.org/10.1023/A:1010933404324>
- [10] Patel, J., Shah, S., Thakar, P. (2018). A review on Various Preprocessing Techniques in Data Mining. *International Journal of Computer Applications*, 181(18), 19-24 <https://doi.org/10.5120/ijca2018917592>
- [11] P. Chen, T. Chen, and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, Aug, 2016, pp.785-794.
- [12] A. Bergstra, D. Yamins, and D. D. Cox, "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures," *Proceedings of the 30th International Conference of Machine Learning (ICML)*, Atlanta, GA, USA, Jun, 2013, pp. 115-123.

Declaration of Interests

We, the authors of this research manuscript, declare that we have no financial interest. We have provided written comment to publish the paper in this journal.

To cite this article: Abir M. I. H., and Salam. T, (2024), Understanding Psychological Dimensions of Student Life Using Machine Learning Approach, *Journal of Engineering and Technology (JET)*, Vol:01, Issue: 1, page: 29-36, ISUCRDP, Dhaka.