Analysis of Factors Influencing Students' Academic Challenges and their Impact on Outcomes

Swati Joshi, Dr. Sanjay Kumar Sharma

Department of Computer Science,BanasthaliVidyapith(Rajasthan) India
Email: cse.swatijoshi@gmail.com

Abstract: The primary goal of educational institutions is to offer a favorable learning environment and impart valuable knowledge to their students. The student's achievement relies on academic performance and it is affected by the psychological problems encountered during the studies. In this paper, we conducted an analysis of the psychological issues that characterize students' experiences, including heavy workloads, insufficient or excessive sleep, mental stress, depression, feelings of pressure, diversity-related issues, negative emotions, and other study-related problems. We then performed a classification to determine the level of stress experienced by the students and examined its impact on academic performance. The dataset used in current work is collected using research methodology, and performed preprocessing. A model is developed to classify/predict the stress level. Classification and prediction techniques employed are Random Forest, K-nearest neighbors, Naïve Bayes, SVM, and ANN. We compared the performance using the metrics accuracy, precision, and recall. According to our extensive experiments, the accuracy of the Random Forest model is 98.16%, which demonstrates superior performance compared to Naïve Bayes (96.78%) and k-NN (95.41%). The ANN model accuracy is 98.16%.The Random Forest model performance is best as compared to other. Statistical method used to find the impact of students' stress level on academic achievement.

Keywords: Machine Learning, supervised learning, academic performance, study behaviors.

1. Introduction

Educational institutions are playing critical role in youth grooming and nation building. The learning model for achieving quality in education is constantly changing by many crucial factors like students' retention, absenteeism, students' academic performance and students' success rates. So in achieving quality in education and learning, identification of high risk students and identification of issues that students met with while enrolled in universities are very important. After identification there arises the necessity of providing extra support for high risk students. The provision for assistance for students is possible by a comprehensive classification of students' problems, classification of their learning behaviors and by prediction of their academic performances. The academic performance may be influenced by learning and psychological behavior of the students and therefore it is important to recognize the problems /apprehensions occur during getting the education and analyze its impact on student performance.

Numerous studies described the various aspects of stress and the ways it affects students' academic achievement. Multiple studies have been conducted to investigate the influence of various stress-related factors on students' academic performance. When someone hears the word stress, they may think of disorders like anxiety, depression, and other potentially fatal illnesses. Stress affects everyone and students are more affected by it during study life. This stress could create diverse effect on their physical and mental health. The student experience stress due to several reasons. These includes the characteristic of connections with family and friends, deadlines of exams, ineffective time management, social media, unstable mind, future job, depression, anxiety, and other issues. Because study environments are highly

competitive in an institution, students need to use their coping mechanisms to manage these issues during getting education.
This paper's main objective is to acquire information about psychological issues that students face while pursuing academic success in an institutional setting. The psychological issue arises in the students due to any reason affect their academic performance. The other factors comprising of unfavorable study environments, heavy workload, learning difficulties, lack of learning resources and social activities also have great impact on academic achievement. The students' psychological issues may be resolved by assisting students in gaining social skill and personality development skills. For students, being able to quickly find solutions to issues they encountered while studying is essential. As such, it is imperative that educational institutions establish a unit of psychological faculty members who can engage with students, comprehend their problems, and offer appropriate advice or solutions to address the difficulties at hand, guiding them toward academic achievement.
The process of developing a model for a computer that can execute computation more accurately and effectively than a human is known as machine learning [1]. Supervised, Unsupervised, Semi-supervised, and Reinforcement learning are the four categories into which machine learning techniques fall. The supervised approach uses known input-output variables to train the data, which is then used to predict or classify fresh, unknown data. k-Nearest Neighbors, Random Forest, Naïve Bayes, Support Vector Machine, and Decision Tree are a few examples of supervised learning approaches. Decision tree decides the output by tracking output at each node in the binary tree [1]. One of the well-liked supervised learning techniques that may be applied to regression and classification is Random Forest (RF). By using predictions from each decision tree and forecasting the final output based on each prediction vote, the RF classifier optimized the outcome.
K-Nearest Neighbors (k-NN) method of machine learning is also used for classification and regression both. The input in both scenarios is made up of a data set's k closest points. The data item that is chosen based on the votes of its neighbors is given the class membership.
Artificial Neural Network (ANN) is a deep learning method which is motivated by biological neural network, consisting interconnection groups of artificial neurons. The ANN learns the connection weights from training pattern available, and performance is improved by iteratively updating the weights over time.
This study focuses on concern of students getting education in the institutions, students' learning experiences and academic performance. The data related to the psychological issues occur during studies and the academic performance is collected through the research methodology. The data preprocessing performed and the model has been designed to classify and predict students' stress level. Many characteristics were employed in the research, including a high workload, insufficient sleep [1], excessive sleep, mental stress, depression, pressure, lack of group activities[1], diversity issues, and negative feelings associated with the study.
The developed model is able to classify/predict the students' stress level and will help in identifying the students' issues. The model will also provide support to overcome their issues as a result the student would get better percentage in exams. Evaluation metrics have been used to assess the classification model's performance. We also employed statistical methods such as correlation analysis to determine the relationship between students' academic performance and stress level.

2. Literature Review

Dorina Kabakchieva[3] suggested utilizing data mining classification algorithms to forecast student achievement. This data mining study is done at an university in Bulgaria to show the great potential of data mining applications to manage educational institutions in order to attract the most qualified applicants and enhancing campaign effectiveness.
C. Anuradha and T. Velmurugan [4] performed a comparative analysis of different classification methods. The purpose of this study is to analyze undergrad students' final year results using data mining techniques. Three private universities in the Indian state of Tamil Nadu hosted the research. Researchers' main goal was to forecast students' final academic achievement at the end of the academic year by using classification techniques. The decision tree method C4.5 (J48), the k Nearest Neighbor algorithm, Bayesian classifiers, two rule learner algorithms: OneR and JRip, and the decision tree are specifically used to categorize student performance and build a model of student performance predictors.

Dr. R. Ravanan, R. Karthiya Banu [5] suggested a novel approach to student empowerment using e-learning and online social media. The students can simply access these learning resources at any moment. This study assists students and educators in selecting e-learning solutions according to their needs and also encouraging them to use data mining techniques to automate the learning process.

Joane Jonathan, Shaleeza Sohail, et al. [6] introduced learning analytics as a tool for measuring and predicting student success. According to the author's recommendation, learning analytics data cannot be utilized to assess student progress in addition to ensuring that the institutional strategy target and teacher performance in the classroom are aligned.

A. Farouk and B. Prasanalakshmi developed a model for King Khalid University's academic performance classification and prediction [7]. The top ten machine learning algorithms for classification and prediction are investigated in this study. The accuracy and other evaluation metrics of the categorical prediction of student performance are assessed using the WEKA tool. The WEKA tool was used to compare the classification accuracy of twelve classifiers: Stacking, AdaBoost, Logistic, SMO, Rep Tree, Naive Bayes, J48, Bagging, lBK, Multilayer Perceptron, Random Forest, Random Tree, and AdaBoost. The datasets underwent different numbers of cases for analysis. Ultimately, the top 5 classification techniques were selected, and the complete dataset's prediction outcomes were compared.

In order to forecast and categorize student performance in two different datasets, Boran Sekeroglu and Kamil Dimililer[8] developed a prediction model that makes use of five machine learning approaches. Preliminary results from eighteen trials suggest that student performances could be better classified by pre-processing the raw data before applying machine learning, and that student achievements might be predictable.

Dede Kurniadi and Asri Mulyani[9] described a prediction technique that used K-nearest neighbor and a strength and problems questionnaire. Using a dataset of one hundred people, they used the K-nearest neighbor technique to detect problematic students. Simulation, accuracy assessment, data cleaning, and collection are the four phases of predictive computation. Utilizing the fast application development technique, the system is being constructed with the student's condition mapped using the Strengths and Difficulties Questionnaire. By teaching children about various learning styles, assisting parents in better understanding their child's personality, and helping counseling teachers put in place an early warning system, these strategies hope to support classroom teachers.

Siti Dianah Abdul Bujang, Ali Selamat[10] offers a offers a comprehensive evaluation of machine learning methods to estimate students final grades in first semester courses. Two components were the main focus of their effort. Using a dataset of 1282 students' course scores, researchers first assessed how well six well-known machine learning approaches performed: decision trees, random forests, k-nearest neighbor, support vector machines, naïve bayes, and logistic regression. The next phase involved creating a prediction model for multiple classes. To prevent misclassification and overfitting from imbalanced multi-classification, the model was based on the Synthetic Minority Oversampling Technique. A comprehensive analysis of the literature was conducted by Amita Dhankhar, Kamna Solanki, and Sandeep Dalal [11] in order to investigate or evaluate metrics, approaches and features in learning analytics and data mining also to predict students' success during their learning and studies in colleges.

A study conducted by Meghji AF, Mahoto NA, Asiri Y, and Alshahrani H[12] analyzed data from 291 university students to forecast their performance at the conclusion of a four-year degree program. They used the educational data mining method of categorization for analysis and prediction. To identify students at different level of academic success a prototype for students' segmentation was also given. This proposed segmentation prototype when tied up with prediction model, provides a working way to build instructional plans that may decrease academic failure and increase performance. The experimental outcomes depicted that the proposed paradigm is beneficial and realistic to classify students into multiple performance groups using limited set of courses or subjects that are taught in first two years of four years degree program.

Md. Abu Marjan, Md. Palash Uddin, and Masud Ibn Afjal[13] developed an approach based on educational data mining to evaluate and improve programming talents of university students.The proposed educational data mining method comprised of two mandatory modules for classification and learning process. The classification module predicts the current status of a student and learning module assists in generating relevant suggestion and feedback to improve learning quality. Six well-known machine learning algorithms were utilized by the researcher for classification: Random Forest, SVM, k-NN, Naïve Bayes, Decision Tree, and ANN. A genuine dataset was constructed by the researcher, who then used goodness of fit and assessment criteria to evaluate each algorithm's performance on the real dataset.

3. METHODOLOGY

This section depicts the machine learning method proposed to classify students' stress and prediction of their stress level based on their psychological and learning behavior. This section also shows the statistical method i.e. correlation analysis to research how stress affects students' academic performance.
Student dataset collection
The dataset is collected using the research methodology technique from the students studying in different engineering institution/colleges. After data collection comprising of 1089 records, created 19 preliminary categories, which includes issues with the syllabus, an overwhelming amount of homework, difficulties faced in the classroom study, an unbalanced lifestyle, issues with sleeping patterns, lack of motivation, concerns about one's career, future uncertainty, health issues, and stress. The data containing the 19 attributes related to psychological issues faced by the students during the studies. The dimensionality reduction is done by removing extra variables or grouping variables that belongs to same domain into one variable. We did feature selection as part of dimensionality reduction, by grouping variables HSL1, HSL2, HSL3, and HSL4 into one variable HSL (Heavy Study Load). Using same method we did grouping of other variables and finally formed the 5 variables: HSL, IS (Increased Stress), FP (Feeling Pressure), SP (Sleeping Problem) and DP (Depression) as shown in Table I. Using statistical software, we assigned 1,2,3,4 and 5 numeric values to the feature values: Strongly disagree (SDA), disagree (DA), neutral (N), agree (A), and strongly disagree (SA) respectively. After that we calculated mean of HSL1…HSL4 of every entry in the dataset and put that mean in HSL. We apply same process for the IS, FP, SP and DP. The output value is Stress level which is estimated using formula shown in equation (1) below.

$$Stress\_Level = w_1 * HSL + w_2 * IS + w_3 * FP + w_4 * SP + w_5 * DP \quad (1)$$

where, $0 \leq w_i \leq 0.5$ are the weights and $\sum_{i=1}^{5} w_i \leq 1$

TABLE I :ATTRIBUTES LIST IN STUDENT PROBLEM'S QUESTIONNAIRE

S.no.	Variable	Description	Possible Values
1.	HS	Heavy Study Load	
	HS1	Some subjects/ topics require more concentration to understand.	SDA, DA,N,A,SA
	HS2	Too many internal assessments per semesters, no sufficient time for study.	SDA, DA,N,A,SA
	HS3	No time for physical exercise (i.e. due to heavy workload)	SDA, DA,N,A,SA
	HS4	You no longer attend cultural activities and social outings due to heavy load.	SDA, DA,N,A,SA
2.	IS	Increased Stress	
	IS1	Are you not able to focus on studies for a long period of time in a single sitting?	SDA, DA,N,A,SA
	IS2	Disturbed due to busy on social sites search/view.	SDA, DA,N,A,SA
	IS3	Are you more reactive to stress?	SDA, DA,N,A,SA
	IS4	High level of stress leads to deprived fitness.	SDA, DA,N,A,SA
	IS5	Having health Issue (i.e. physically not well most of time).	SDA, DA,N,A,SA

	IS6	Appetite changes, including either a loss of appetite or overeating (depression).	SDA, DA,N,A,SA
3.	FP	Feeling Pressure	
	FP1	Having Peer pressure?	SDA, DA,N,A,SA
	FP2	Feeling of pressure due to parent's expectation (i.e. to obtain good mark in exams).	A,DA,N
	FP3	Feeling pressure to get the right career.	SDA, DA,N,A,SA
4.	SP	Sleeping Problem	
	SP1	Are you facing problem in getting sufficient sleep due to extra use of phone ,social media and video games?	SDA, DA,N,A,SA
5.	DP	Depression	
	DP1	Thoughts of helplessness and hopelessness. (Symptoms of emotions)	SDA, DA,N,A,SA
	DP2	Having trouble in concentrating or paying attention in every task.	SDA, DA,N,A,SA
	DP3	Have you resorted to binge drinking or using drugs to get over depressing feelings?	SDA, DA,N,A,SA
	DP4	Are you not enjoying activities you once loved?	SDA, DA,N,A,SA
	DP5	Have you ever had intrusive suicidal thoughts as a result of a problem?	SDA, DA,N,A,SA
6.	Stress_Level	Stress Level	{High, Moderate, Low}

Data Preprocessing

Data preprocessing is the initial step of processing raw data. It transforms the data into the format which is most appropriate for Machine Learning algorithms. Perform preprocessing, data cleaning. For the implementation, we used the librariesnumpy, pandas, train_test_split, random forest classifier,classification_report,LabelEncoder,KMeans, KNeighborsClassifier. Every class's label is stored in the label encoder and label encoded the particular column. Using an ordinal or one-hot encoding approach, it encodes categorical attributes. Replaced missing values with the mean and replaced the false values and dropped the missing values in target variable Stress_Level. Input feature is taken into X and output feature in Y. X contains the 5 features HSL,IS,FP,SP and DP and Y contains target feature Stress_ Level. The dataset is split into two sections, a 70% training set and a 30% testing set, in order to train and test the model.

Classification

The aim of classification is to construct a model that can anticipate how new data will be categorized, drawing insights from past data. This research delves into categorizing and forecasting the stress levels of college students by employing various machine learning techniques such as Naïve Bayes, Random Forest, k-NN, Support Vector Machine, and a deep learning approach known as Artificial Neural Network (ANN), depicted in Figure 1.

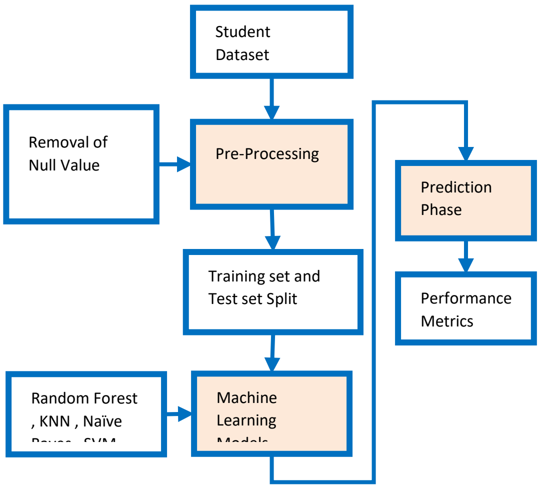

Fig 1: Flow diagram for Classification and Prediction

Random Forest:

Random Forest is aa ensemble learning technique that can be used for both regression and classification. It creates number of decision tress during training and gives the real values as a result for classification.the decision tree constructed during training have low bias and high variance.After averaging them the variance of the built RF model is reduced.

In this study, the stress level of students is classified using RF classifier, which is shown in Fig 2, when psychological and learning problems related input data variables are fed to the classification model.the classifier classified the stress level of students into three classes: Low ,Moderate and High.

```
Random Forest Performance:
Accuracy: 0.981651376146789
Recall: 0.981651376146789
Precision: 0.9836350111579469
F1-score: 0.982074805928017

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       0.89      1.00      0.94        33
           2       1.00      0.98      0.99       184

    accuracy                           0.98       218
   macro avg       0.96      0.99      0.98       218
weighted avg       0.98      0.98      0.98       218
```

Fig 2: Random Forest Classification report

The classification accuracy is 98.16% using RF classifier on the dataset.

Naïve Bayes: The Naive Bayes approach relies on the mathematical application of Bayes' theorem.

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \qquad (2)$$

Equation (2) demonstrates that to find the probability of A given B, we use Bayes theorem's formula. This formula involves multiplying the probability of B given A by the probability of A and dividing the result by the probability of B. By assuming that variables are independent of one another, the Naïve Bayes algorithm altered the learning process and provided a probabilistic analysis of the classification approach. The main benefit of NBC learning algorithms is fast and compared to numerical input variables, it is more adept at handling categorical input variables. After applying Naïve Bayes classification model on students' dataset, accuracy of 96.78 is achieved, which is shown in Fig 3.

```
Naive Bayes Performance:
Accuracy: 0.9678899082568807
Recall: 0.9678899082568807
Precision: 0.9710799186114711
F1-score: 0.9689272240252139

Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         1
           1       0.89      0.94      0.91        33
           2       0.99      0.97      0.98       184

    accuracy                           0.97       218
   macro avg       0.79      0.97      0.85       218
weighted avg       0.97      0.97      0.97       218
```

Fig 3: Naïve Bayes Classification

Support Vector Machine:

Support Sector Machine (SVM) is a non-probabilistic method which can distinguish linear and non linear data[19]. The SVM classifies the data by classifies data by determining the hyperplane that elongates the border line between the classes in training datset. the hyperplane in the training set of data that elongates the class boundaries. Here's how one could create a hyperplane:

$$f(x) = a^T x + C \qquad (2)$$

In equation (2), a is dimensional coefficient and c is the offset. The benefit of SVM, is that it can choose different kernels. It is practical to use a significantly more complicated structured data collection with many kernels. Additionally, it has fewer overfitting issues. Even though the support vector machine's kernel is its strongest point and choosing a particular kernel might be challenging. The SVM is proved to be examining the students' interaction and class participation in group activities. The classification accuracy is achieved of 96.78% on the students' dataset, which is shown in Fig 4.

```
SVM Performance:
Accuracy: 0.9678899082568807
Recall: 0.9678899082568807
Precision: 0.9630788148884349
F1-score: 0.9652961095724424

Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.94      0.88      0.91        33
           2       0.97      0.99      0.98       184

    accuracy                           0.97       218
   macro avg       0.64      0.62      0.63       218
weighted avg       0.96      0.97      0.97       218
```

Fig 4: Support Vector Machine

k-Nearest Neighbours Classifier

The k-Nearest Neighbor (k-NN) algorithm evaluates k-dataset instances that are in close proximity to the observations. Since the k-NN learns from the dataset during prediction and does not require initial training, it uses a significant amount less computing power. The distance function and the k-value are the only two values needed to implement this algorithm. It doesn't work well with data that has many dimensions. After using the k-NN classification algorithm on the student dataset, as illustrated in Fig. 5, the accuracy is 95.41%.

```
KNN Performance:
Accuracy: 0.9541284403669725
Recall: 0.9541284403669725
Precision: 0.9488483310560217
F1-score: 0.9510985216815343

Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.90      0.82      0.86        33
           2       0.96      0.98      0.97       184

    accuracy                           0.95       218
   macro avg       0.62      0.60      0.61       218
weighted avg       0.95      0.95      0.95       218
```

Fig 5: k-NN Classification

Selecting k value in k-NN :
To select value of k, we train and evaluate k-NN for each k value and draw plot between k value and accuracyAfter setting the k in the range of (2, 21) (we did not use k=1 because of the overfitting issue), we provided the k-NN model the parameters X_train, X_test, Y_train, and Y_test [20]. Figure 6 illustrates the plot, which shows that the model's accuracy peaked at k=3.

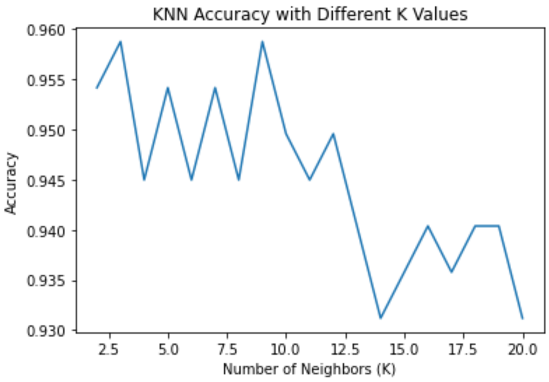

Fig 6: Plot k-NN accuracy vs. k-value

ii) Selecting value of k using Grid Search CV method:
The number of neighbors we have taken in the range of (2, 21) in order to conduct Grid search CV and obtain the optimal k value for the model. As a result, we get accuracy of 95.90% with k=3 on testing set. Hence, we determined that k=3 was the best value for the k-NN classifier by applying the two aforementioned approaches to choose k.
Artificial Neural Network(ANN)
Artificial Neural Networks are computer programs inspired by biology, that simulate how human brain process information. The individual units, also known as artificial neurons or processing elements (PE) that make up an ANN are linked by coefficients (weights), which comprise the neural structure are arranged in layers[21]. In our dataset we applied ANN model to get the evaluation of model. The ANN model is compiled and one hot encodes the target labels. The accuracy and loss of the data are then determined by evaluating this model. Fig. 7 illustrates the accuracy of the model, which uses an artificial neural network (ANN) on the dataset, which is 98.16%.

```
ANN Performance
ANN Accuracy: 0.981651376146789
ANN Recall: 0.981651376146789
ANN Precision: 0.97733405288721
ANN F1-score: 0.979460495686704

ANN Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.94      0.97      0.96        33
           2       0.99      0.99      0.99       184

    accuracy                           0.98       218
   macro avg       0.64      0.65      0.65       218
weighted avg       0.98      0.98      0.98       218
```

Fig 7: ANN Classification

Classification Result and Evaluation

Machine learning models were created to classify and predict the stress levels of students after preprocessing their dataset. Training, testing, and a comparative analysis of the models' performances using various machine learning approaches led to the identification of the best classification model. Random Forest, Naïve Bayes, SVM, k-NN and ANN classification methods are applied and among all Random Forest gave the highest accuracy of 98.16%. While SVM and Naïve Bayes both reached 96.78% accuracy, k-NN displays 95.41% accuracy when k=3. We selected k=3 for the k-NN classification model since, as Fig. 6 illustrates, the model is most accurate at this k value. The ANN classification model gave accuracy of 98.16% which is same as RF model. The evaluation matrices are also examined to compare F1 score, Recall and precision of each classification model, which is illustrated in Table III.

TABLE III: COMPARATIVE ANALYSIS OF CLASSIFICATION RESULT

Classification Method	Accuracy(%)	Precision	Recall	F1 Score
Random Forest(RF)	98.16	98.36	98.16	98.20
Naïve Bayes(NB)	96.78	97.10	96.78	96.89
SVM	96.78	96.30	96.78	96.52
k-NN	95.41	94.88	95.41	95.10
ANN	98.16	97.73	98.16	97.94

Correlation Matrix:

A correlation analysis was conducted on the dataset through the creation of a correlation matrix to delve into the impact of stress levels on academic achievement. The correlation between the variables HSL (heavy study load), IS (increased stress), FP (feeling pressure), SP (sleeping problem), DP (depression problem), Stress Value, and Percentage Marks is displayed in correlation matrix shown in Fig 8. The primary emphasis of our research revolved around exploring how the stress levels of students influenced their academic achievements. The correlation matrix result indicates a strong negative correlation (-0.65) between the percentage marks and stress value, indicating that students with higher stress values receive lower percentage marks.

```
                      HSL        IS        FP        SP        DP  \
HSL              1.000000  0.260262  0.050884  0.001209  0.156478
IS               0.260262  1.000000  0.140248  0.212264  0.143198
FP               0.050884  0.140248  1.000000  0.099968  0.012803
SP               0.001209  0.212264  0.099968  1.000000  0.032433
DP               0.156478  0.143198  0.012803  0.032433  1.000000
STRESS VALUE     0.315437  0.596112  0.270668  0.738187  0.540978
Percentage marks -0.170891 -0.387794 -0.199191 -0.472724 -0.367138

                 STRESS VALUE  Percentage marks
HSL                  0.315437         -0.170891
IS                   0.596112         -0.387794
FP                   0.270668         -0.199191
SP                   0.738187         -0.472724
DP                   0.540978         -0.367138
STRESS VALUE         1.000000         -0.650809
Percentage marks    -0.650809          1.000000
Stress Levels: ['LOW' 'MODERATE' 'HIGH']
```

Fig 8: Correlation Matrix

Students with low stress levels received better percentage scores (more than 80%), those with moderate stress levels received lower percentage marks (50 to 80%), and those with high stress levels received the lowest percentage marks (less than 50%), as indicated by the Box plot in Figure 9.

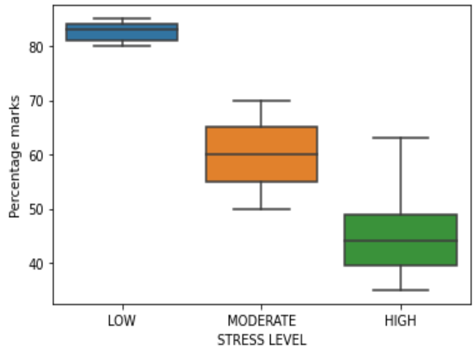

Fig 9: Box Plot(Percentage marks vs. Stress Level)

4. Conclusion

The study's main goal is to examine the psychological behavior of engineering school students and group their diverse problems into five main categories, including depression, heavy study loads, increased stress, feeling pressure, and sleep troubles. Using machine learning techniques, a model that classifies and predicts students' stress levels and academic achievement is built. Various cutting-edge machine learning methodologies such as Random Forest, Support Vector Machine, Artificial Neural Network, k-Nearest Neighbors, and Naïve Bayes are harnessed for the assessment of the model's efficiency. The accuracy of the Random Forest was 98.16 %, greater than that of k-NN (95.41%) and Naïve Bayes (96.78%). The ANN classification method gave the accuracy of 98.16%. Additionally, the influence of stress on academic achievement has been analyzed through statistical method i.e. correlation analysis and Box plots. These techniques clearly demonstrated a negative association between students' percentage marks and stress levels, meaning that when students' stress levels rises, their percentage marks decreases. This model will be used to forecast the academic achievement of the students and determine their level of stress. In order to help students succeed academically and professionally, this study will also offer additional support in overcoming psychological issues and study-related issues.

References

1. C. R. Durga devi,"A Survey on Forecasting Students Performance using EDM", IJSTE - International Journal of Science Technology & Engineering, Volume 2, Issue 01, July 2015,ISSN (online): 2349-784X.
2. Dong Yuan, Jian Huang," Improved random forest classification approach based on hybrid clustering selection", 2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109, pages 1559-1563.
3. Dorina Kabakchieva, "Student Performance Prediction by Using Data Mining Classification Algorithms", International Journal of Computer Science and Management Research Vol 1 Issue 4 November 2012,pgno. 686-690.
4. C. Anuradha and T. Velmurugan," A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance", Indian Journal of Science and Technology, Vol 8(15), IPL057, July 2015,pgno.1-12.
5. R. KarthiyaBanu, Dr.R.Ravanan, A Study on Data Mining in E-learning – Empowering Students Education through Social Networks – A Novel Approach, 2012 International Conference on Education and e-Learning Innovations, INSPEC Accession Number: 13149383.
6. Joane Jonathan, ShaleezaSohail,FadiKotob,Graeme Salter ,The Role of Learning Analytics in Performance Measurement in a Higher Education Institution,2018 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) ,Pages: 1201 – 1203.
7. B. Prasanalakshmi and A. Farouk," Classification and Prediction of Student Academic Performance in King Khalid University-A Machine Learning Approach", Indian Journal of Science and Technology, Vol 12(14), April 2019,pg.no. 1-6.
8. Boran Sekeroglu, Kamil Dimililer, Kubra Tuncal," Student Performance Prediction and Classification Using Machine Learning Algorithms", ICEIT 2019, March 2–4, 2019, Cambridge, United Kingdom ACM ,pg.no. 7-11.

9. Dede Kurniadi , Asri Mulyani , “Prediction System for Problem Students using k-Nearest Neighbor and Strength and Difficulties Questionnaire “, JOIN (Jurnal Online Informatika) Volume 6 No.1 | June 2021: 53-62.
10. Siti Dianah Abdul Bujang , Ali Selamat ,”Multiclass Prediction Model for Student Grade Prediction Using Machine Learning”,IEEE Access,VOLUME 9, 2021, Digital Object Identifier 10.1109/ACCESS.2021.3093563.
11. Jennifer S. Raj Abdullah M. Iliyasu Robert Bestak,” Innovative Data Communication Technologies and Application” Lecture Notes on Data Engineering and Communications Technologies 59, https://doi.org/10.1007/978-981-15-9651-3_11 Proceedings of ICIDCA 2020.
12. Meghji AF, Mahoto NA, Asiri Y, Alshahrani H, Sulaiman A, Shaikh A. 2023. Early detection of student degreelevel academic performance using educational data mining, PeerJ Comput. Sci. 9:e1294.
13. Md Abu Marjan, Md Palash Uddin, Masud Ibn Afjal,” An Educational Data Mining System For Predicting And Enhancing Tertiary Students’ Programming Skill “,The Computer Journal, Volume 66, Issue 5, May 2023, Pages 1083–1101.
14. Hamido Fujita, Ali Selamat, Sigeru Omatu,” Knowledge Innovation Through Intelligent Software Methodologies, Tools and Techniques, Proceedings of the 19th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SoMeT_20), Volume 327,Published 2020, ISBN 978-1-64368-114-6 | 978-1-64368-115-3 (online).
15. J.W. Han, M. Kamber, Data mining concepts and techniques, Beijing: China Machine Press, 2001.
16. X.G. Zhang, Introduction to statistical learning theory and support vector machines, Acta Automatica Sinica, vol. A26, pp, 33-42, 2000.
17. S. Ruggieri, Efficient C4.5, IEEE Transactions on Knowledge and Data Engineering, vol. A14, pp, 438-444, 2002.
18. Mohammad Rezwanul Huq, Ahmad Ali, Anika Rahman, “Sentiment Analysis on Twitter Data using KNN and SVM”, 2017, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, pp- 19-25.
19. C. Anuradha and T. Velmurugan,” A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance”, Indian Journal of Science and Technology, Vol 8(15), IPL057, July 2015,pg 1-12. www.github.com
20. Sushreeta Paul, Vijay Kumar, Priyanka Jha. "Artificial neural network and its applications: Unraveling the efficiency for hydrogen production" ,Applications of Artificial Intelligence in Process Systems Engineering, Elsevier 2021, Pages 187-206.