iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53398

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Student General Performance Prediction Using ML Algorithm

Shikha Pachouly[1], Shubham Zope[2], Atharva Raut [3], Rupesh Rajput [4], Krishnakant Patil[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, AISSMS College of Engineering Maharashta India*

Abstract: ***We will start to examine and categorise the data once we have obtained the required information using our surveys. This plan involves locating data patterns and trends that can be leveraged to create powerful machine learning models. The accuracy, recall, and precision of these models will then be tested to ensure that they are reliable and effective. They will then be developed using a variety of techniques. This would require meticulous attention to detail when analysing datasets as well as a thorough understanding of the advantages and disadvantages of various machine learning techniques. Our ultimate goal is to develop machine learning (ML) models that accurately predict events using the data we have collected, providing insightful knowledge about our area of interest. We are confident that we can achieve this goal and significantly advance the fields of machine learning and data analysis by carefully planning and carrying out our work. For example, this technique could be used to predict how well students in a particular school district will perform academically. By collecting information on factors such as their emotional state and extracurricular activities in addition to more conventional data like grades and test results, we could develop ML models that precisely forecast which children are at risk of falling behind and how to support them. Which will enable us to identify the variables influencing students' success or contributing to their poor performance. This could make it easier for teachers to provide more individualised support for students and improve their overall academic performance.***
Keywords: Classification, LR – Logistic Regression, SVM – Support Vector Machine, ML – Machine Learning, KNN – K-Nearest Neighbours, SVM-Support Vector Machine

I. INTRODUCTION

The primary goal of this project will be the creation of features for predicting students' overall performance. We can predict the factors that will help students perform well or poorly by collecting and analysing the data. Which will enable teachers to provide more individualised support for students and improve their general academic performance. Why is there a system that determines a student's performance solely on the basis of their academic grades when every student is special, unique, and has some special talent? In the OECD countries as a whole, 56% of students agreed or strongly agreed that they worry about what other people think of them when they fail. According to a recent study, one in three college students suffers from severe depression and anxiety. Common causes of this include difficulties with college work and a decline in interest in extracurricular activities like clubs, sports, or other social commitments. The primary goals of this are to raise student academic performance and prevent dropouts. The performance of the student is dependent on a number of elements, including their mental state in addition to their grades and academic course work. In light of this, we will carry out a survey in which we will inquire about things like your home, your grades, your financial situation etc. We can determine where the learner is falling behind and where he or she needs to improve by carefully examining these answers. We can help students perform better by adding these answers as an additional feature to our machine learning model.Dropout rates and academic losses rose along with the digital divide following the pandemic and significant changes to the educational system and students' private life, such as coping with bereavement and social anxiety. Up until now, we have only evaluated a student's academic success based on the grade they received. The existing educational system results in a dreadful experience for more than 50% of students who are not good at what they are doing, according to the current academic performance evaluation method and honest feedback from a student who is average or below average in education. When more than 50% of students lack enthusiasm for studying, the educational system has failed. By carefully studying these answers, we may establish where the student is slipping behind and the areas in which the individual needs to improve. We can help students perform better by adding these answers as an additional feature to our machine learning model. This approach can lead to relevant and effective learning strategies that cater to the individual needs of each student. By incorporating this information into the machine learning model, we can create a more comprehensive and accurate assessment of the student's academic performance.

©IJRASET: All Rights are Reserved

II. RELATED WORK

With machine learning, it is possible to forecast a student's general performance and identify the factors that will decide whether they succeed or fail.. S. Bhutto, I. F. Siddiqui, Q. A. Arain, and M. Anwar, in "Predicting Students' Academic Performance Through Supervised Machine Learning," experiment with different supervised learning models and, after comparing a bunch of algorithms, conclude that the sequential minimal optimization algorithm outperforms by achieving improved accuracy as compared to logistic regression.[1]M. Crivei, G. Czibula, G. Ciubotariu, and M. Dindelegan, in "Unsupervised learning-based mining of academic data sets for students' performance analysis," have investigated the usefulness of unsupervised machine learning methods, particularly principal component analysis and relational association rule mining, in analysing students' academic performance data [2]. In "Collective Emotional Intelligence and Group Dynamics Interplay: Can It Be Tangible and Measurable?" Fotopoulou, A. Zafeiropoulos, L. Cassà, I. M. Guiu, and S. Papavassiliou conducted research on how group dynamics and human emotion can improve our comprehension of people's motivations. [3] El-Sayed Atlam, Ashraf Ewis, M.M. Abd El-Raouf, Osama Ghoneim, and Ibrahim Gad, in "A New Approach to Identifying the Psychological Impact of COVID-19 on University Students' Academic Performance," in their paper An online survey was utilised to collect facts like demographics, digital tools, sleep routine, social contact, educational achievement, psychological condition, panic, and depression scale.The major analysis is divided into two portions that are carried out using machine learning methods. .[4]Chen, P. Chen, and Z. Lin, in "Artificial Intelligence in Education: A Review," studied the effect of AI on student education of student.The study suggested that AI has been extensively adopted and used in education, particularly by educational institutions, in different forms. Machine learning has enabled personalised instruction that caters to the unique needs of each student, leading to customised courses and content designed to meet the demands of today's learners. This has resulted in a more engaged and motivated student body. [5]" Prediction of Student Academic Achievement Using the Decision Tree Algorithm," Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarkerin stated that predicting students' academic performance is important because it allows teachers to take proactive actions and develop strategies to help students learn, eventually boosting their academic performance. [6] In "Student Performance Prediction Using Classifier Data Mining Methods," Vinaya Patil, Shiwani Suryawanshi, Mayur Saner, and Viplav Patil investigated the various algorithms and their benefits and drawbacks. Because the purpose of the paper is to forecast the student's overall performance, they focus on the most effective approach to do so. [7]Alraddadi, S. Alseady, and S. Almotiri, in "Prediction of Students' Academic Performance Utilizing Hybrid Teaching-Learning-Based Feature Selection and Machine Learning Models," proposed a reliable combination of a wrapper for feature selection and a hybrid method of machine learning techniques. [8]A. Olorunmaiye, O. J. Ogunniyi, T. Yahaya, J. O. Olaoye, and A. A. Ajayi-Banji state in "Modes of Entry as Predictors of Academic Performance of Engineering Students in a Nigerian University" According to the survey, students admitted through JAMB had higher GPAs and were more likely to complete their studies within the predicted time frame of their degree. Pre-degree programmes had lower GPAs and a lower graduation rate. Universities must carefully assess their admission policies and provide enough support services to all students.[9]In Aman, A. Rauf, R. Ali, F. Iqbal, and A. M. Khattak's "A Predictive Model for Predicting Students' Academic Performance," In Aman, A. Rauf, R. Ali, F. Iqbal, and A. M. Khattak's "A Predictive Model for Predicting Students' Academic Performance," they listed the key factors that affect students' performance and then created an accurate prediction model for prediction of their performance prior to applying for admission in a desired programme, deciding to continue for higher classes and semesters in the same programme, or deciding to leave the programme. [10] Estimating student success is an important topic for learning situations such as institutions since it aids in the creation of effective mechanisms that increase overall results and reduce dropouts. Hence, carefully analysing and analysing these data can provide us with useful information about students' information and the correlation between them and academic assignments.

The proposed work's major goal is to predict and categorise student performance. The current system focuses solely on the traditional technique of grading pupils; it only examines responses supplied by a student in its program assessment; he does not have the choice to choose a subject of interest, nor is he capable of comprehending the substance interested field of study. Social media optimisation can help build a strong online presence and improve brand visibility, but it is based on parents and students' input without teachers' input. Media optimisation is a powerful tool for building a strong online presence, but it is important to consider the perspectives of all stakeholders involved. Teachers bring a unique perspective to the table, as they have first-hand experience with students and can provide valuable insights into what resonates with them. By incorporating teacher input into media optimisation strategies, schools can ensure their online presence accurately reflects their values and mission and build trust with parents and students. By taking a collaborative approach, schools can create a more comprehensive and effective online presence that resonates with all stakeholders involved. [1]

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Predicting the students' performance for courses, detecting what type of learners the students are, grouping them according to their similarities, and assisting instructors in the educational process help instructors understand and plan things and also give very similar results, so it would give more accurate results. By dividing learners, the outliner, i.e., non-group able youngsters, will be excluded from evaluation, affecting the model's accuracy. It only evaluates binary class labels for prediction, which is insufficient to obtain insight into an individual student's performance. [2] Machine learning algorithms can be used to identify the most important features and reduce the number of classes in a model without sacrificing accuracy. This can be especially useful in healthcare, where predicting patient outcomes based on emotional states can be crucial for treatment decisions. Additionally, machine learning can help identify patterns and correlations between emotions and other variables, such as demographics or environmental factors. The potential applications for emotion analysis in various industries are vast and exciting indefinitely. [3] It works similarly to a classic one; however, it requires considerable time to execute, and the parameters of the model must be carefully calibrated. Because the dataset for this study's students who completed an online COVID-19 survey was too small to generate meaningful machine learning models, all data from individual students was used. [4] Demonstrate how artificial intelligence (AI) can be used efficiently to transform the quality education, and show how we currently employ Machine learning and artificial intelligence in different forms in our framework. They used artificial intelligence to enhance the efficacy and productivity of teachers or educators in performing various administrative chores. Its traits are more traditional in nature. There is no explanation of the dataset utilised in this paper. [5] Employs a decision tree that takes less time for text processing during pre-processing than other methods. A minor alteration to the information can result in a significant modification to the framework of the decision tree, resulting in instability. [6] Predicting the type of a sample of data is simple and quick. It also excels in multiclass prediction. Predicts probability outputs shouldn't be taken too seriously. [7] Educational data mining is used. Use a hybrid model which combines an Ml algorithm with an optimisation technique. The queries in this article are just about the course and the instructor, not about their emotional situations. The use of two datasets may result in complexity. [8] Data collection is confined to the name, modalities of admission, CGPA, and degree class and does not include additional factors that influence student performance. The data is analysed using statistical analysis. [9] Educational data mining is used. Student career field forecast According to its academic achievement as well as social factors only exam results are considered academic data. There is no comprehensive examination of all extracted data. The proposed hypothesis is exclusively utilised to forecast the futures of students. [10]

III. PROPOSED SYSTEM

First, we identify the variables outside the academy that have an impact on student performance in order to forecast general student performance using a machine learning model. After considering the problem statement as we created the dataset, we used feature scaling to transform the categorical data into numerical data for the classification model, increasing the models' efficiency. It entails altering feature values to fit them within a predetermined range. This prevents any one feature from dominating the model and enhances algorithmic interconnection. We then use our model to classify data using supervised learning techniques like logistic regression, k-nearest neighbours, and SVM in order to identify the factors promoting or hindering student success. We developed three models to use in the process of predicting student performance, and by comparing the effectiveness and accuracy of the models, we chose the algorithm that worked best. SVM was more effective in our situation for categorising the factors. We utilised SVM's three kernels i.e. linear svm, polynomial svm and radial basis svm to predict student performance. Our study found that attendance and participation were the most significant factors in predicting student performance. The linear SVM kernel provided the highest accuracy rate of 81.25%, followed by the polynomial SVM kernel with 73.43% and the radial basis SVM kernel with 73.43%. Our findings suggest that SVM is a reliable algorithm for predicting student performance, and its effectiveness can be improved by selecting appropriate kernels.

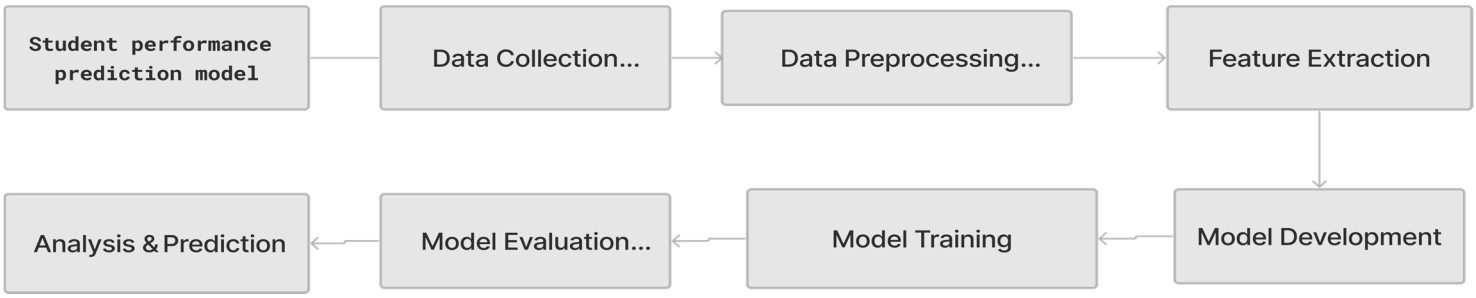


Figure 1.System Architecture

1) Logistic Regression (LR)

A mathematical model in use for binary classification situations is logistic regression. It is frequently employed when the dependent variable (the parameter we wish to forecast) is categorical and has two alternative outcomes, which are often labelled as 0 and 1. Logistic regression is a machine learning statistical modelling technique used to attempt to predict binary outcomes such as student performance. It establishes the link between independent factors and the possibility of a situation occurring. By fitting a logistic function to the data, we may evaluate the impact of various factors on student performance and make predictions based on that information.

Algorithm 1: Logistic Regression
Input: Student data
Step1: Start.
Step 2: Data Preparation Following the collection of student data from the survey,
Step 3: Read the features after importing the dataset into the data model.
Step 4: Divide the dataset into two parts: training and testing.
Step 5: If necessary, perform feature scaling or standardisation on the independent variables.
Step 6: Construct a logistic regression model.
Step 7: Using the training data, train the logistic regression model.
Step 8: Using the testing data, evaluate the model's performance.
Step 9: Apply the trained model to new data to create predictions.
Step 10: Examine the logistic regression model coefficients to better understand the factors influencing student success.
Step 11: Visualise the results by charting the link between independent variables and predicted probabilities, for example.

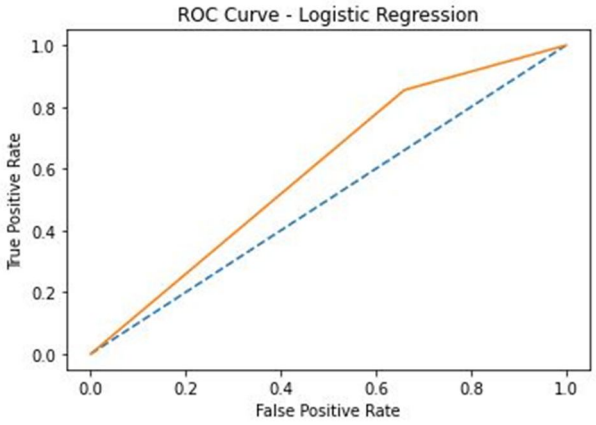


Figure 1 Logistic Regression

2) K-Nearest Neighbours (KNN)

It is a machine learning algorithm that predicts students' overall performance and identifies factors that influence their academic success. It operates by determining the k-nearest neighbours (students) to a given student based on shared qualities or features. The programme can detect trends and correlations that contribute to student success or failure by analysing the properties of these nearest neighbours. This data can assist educators and policymakers in better understanding the factors that influence student performance and making educated decisions to improve educational outcomes.

Algorithm 2: KNN
Input: Student data
Step1: Start.
Step2: Data Pre-processing: After obtaining the student data from the survey, import the dataset and preprocess it by handling missing values, encoding categorical variables, and splitting it into features and target variables.
Step3: Feature Scaling: Perform feature scaling on the data to normalize the values and bring them into a similar range.

Step4: Create KNN Classifier: Initialize the KNN algorithm with the desired number of neighbors(k) and any other relevant parameters.
Step5: Training the Model: Fit the KNN classifier to the preprocessed training data.
Step6: Making Predictions: Use the trained model to predict the outcome (pass /fail) based on the features of new, unseen student data.
Step7: Evaluating the Model: Assess the performance of the KNN model by comparing the predicted outcomes with the actual outcomes using appropriate evaluation metrics (e.g., accuracy, precision, recall).
Step8: Visualizing the Results: Visualize the results of the predictions and the performance of the model using appropriate plots or charts.

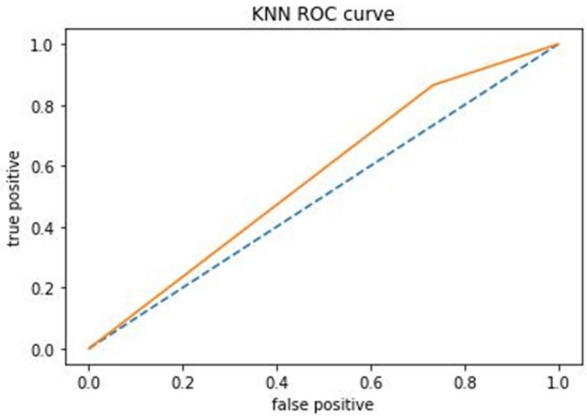


Figure 2 K-Nearest Neighbours (KNN)

3) SVM (Support Vector Machine)

It is a machine learning algorithm used for classification and regression tasks. In the context of predicting student general performance, SVM can be applied to identify the factors that affect a student's academic success. The given results show the performance of SVM with different kernel functions: linear, polynomial, and Gaussian. The linear kernel achieved the highest accuracy of 81.25%, followed by the polynomial and Gaussian kernels with an accuracy of 73.43%. Additionally, the F1 score and Roc Auc score indicate the overall performance of the SVM models, with the linear kernel achieving the highest scores. These results suggest that the linear kernel SVM is the most effective in predicting student performance based on the given metrics.

Metric	Linear Kernel	Polynomial Kernel	Gaussian Kernel
Training Time	3ms	6ms	4ms
Accuracy	81.25%	73.43%	73.43%
F1 Score	0.77	0.67	0.64
Roc Auc score	0.75	0.66	0.64

Algorithm 1: SVM
Input: Student data
Step 1: Start.
Step 2: Data Pre-processing: After obtaining the student data from the survey, import the dataset and perform necessary data cleaning and feature extraction.
Step 3: Feature scaling: Scale the features to ensure they have similar ranges and do not dominate the model.
Step 4: Splitting the dataset: Divide the dataset into training and testing sets to evaluate the model's performance.
Step 5: Create an SVM classifier: Initialize the SVM model with the desired kernel function (linear, polynomial, or Gaussian).
Step 6: Training the model: Fit the SVM model on the training data to learn the underlying patterns and relationships.
Step 7: Model evaluation: Evaluate the model's performance using appropriate evaluation metrics such as accuracy, precision, recall, or F1-score.
Step 8: Hyperparameter tuning: Fine-tune the model by adjusting hyperparameters like the regularization parameter, degree of polynomial kernel, or kernel width in the Gaussian kernel.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

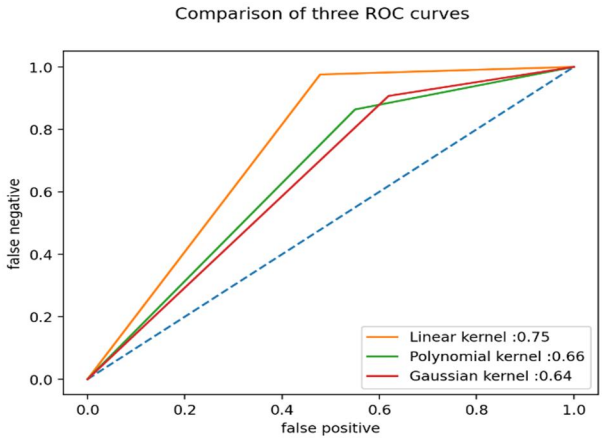


Figure 3 SVM (Support Vector Machine)

Step 9: Predictions and visualization: Use the trained SVM model to make predictions on new or unseen data and visualize the results using appropriate plotting techniques.
Step10: Repeat steps 5-9 for each kernel function (linear,polynomial,and Gaussian) to compare their performance and choose the best one based on evaluation metrics.

IV. RESULT AND DISCUSSION

Based on the students' data, the proposed model is implemented using Python to predict their general performance.
We use an algorithm that predicts the student's overall performance in our research paper. The elements that will determine whether they succeed or fail are shown in the image below from our algorithm.

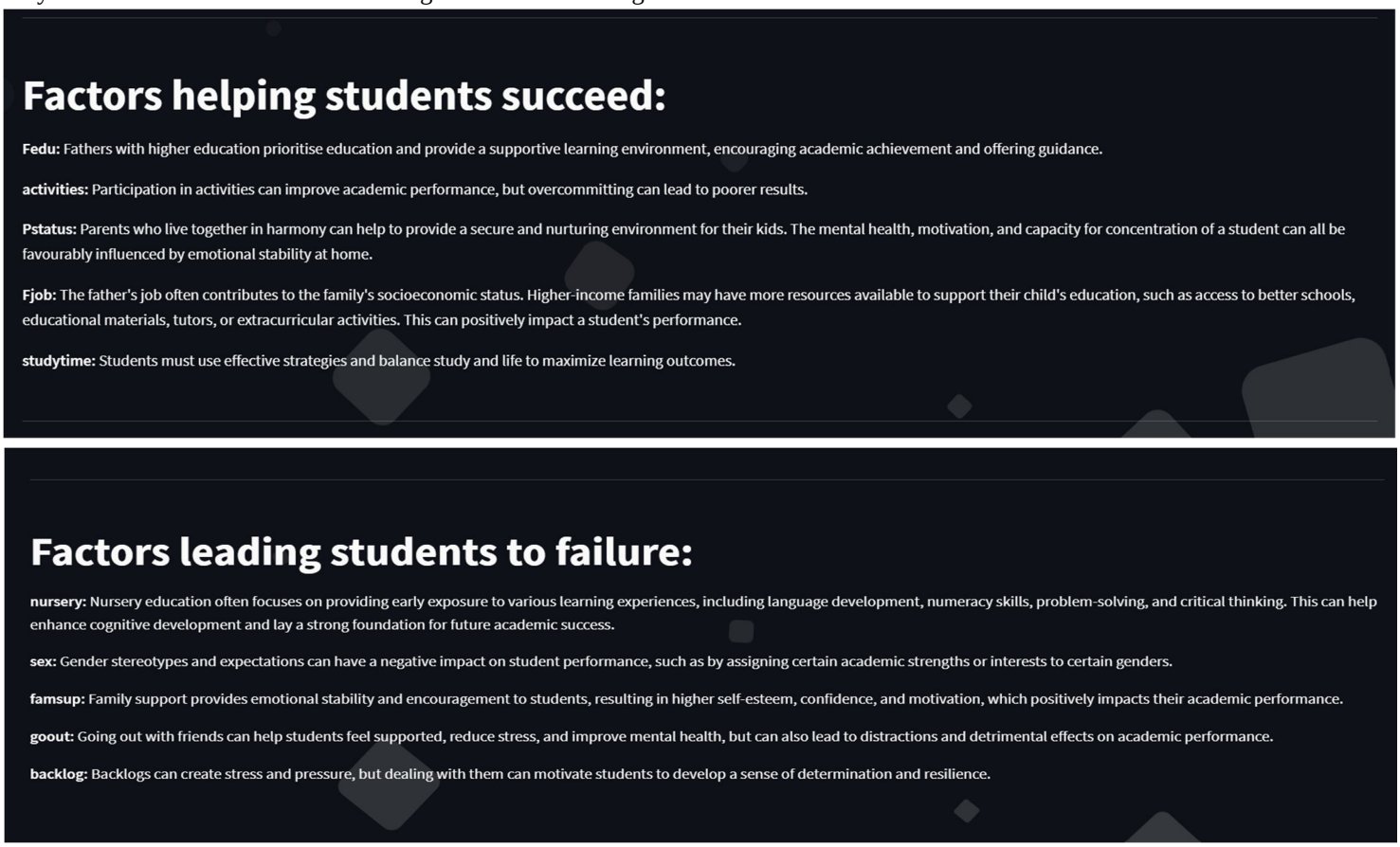


Figure 4 OUTPUT

1) *Father's Education:* Fathers with higher education prioritise education and provide a supportive learning environment, encouraging academic achievement and offering guidance.

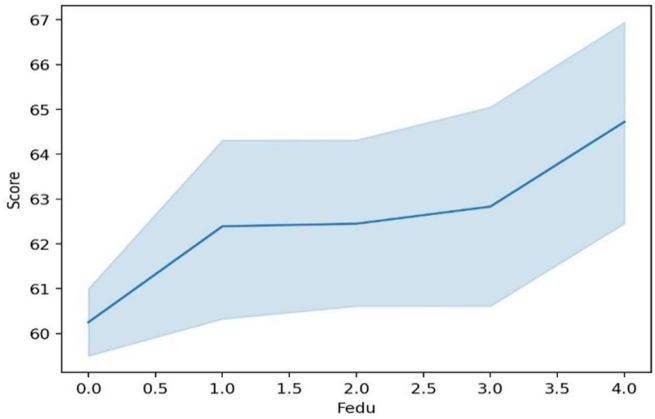


Figure 5 Father's education

2) *Parental Status:* Parents who live together in harmony can help to provide a secure and nurturing environment for their kids. The mental health, motivation, and capacity for concentration of a student can all be favourably influenced by emotional stability at home.

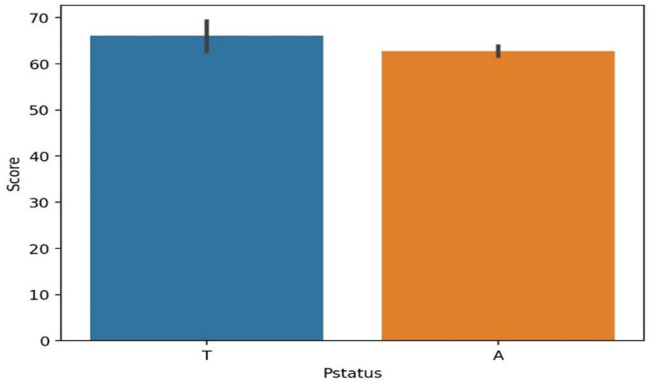


Figure 6 Parental status

3) *Father's Job:* The father's job often contributes to the family's socioeconomic status. Higher-income families may have more resources available to support their child's education, such as access to better schools, educational materials, tutors, or extracurricular activities. This can positively impact a student's performance.

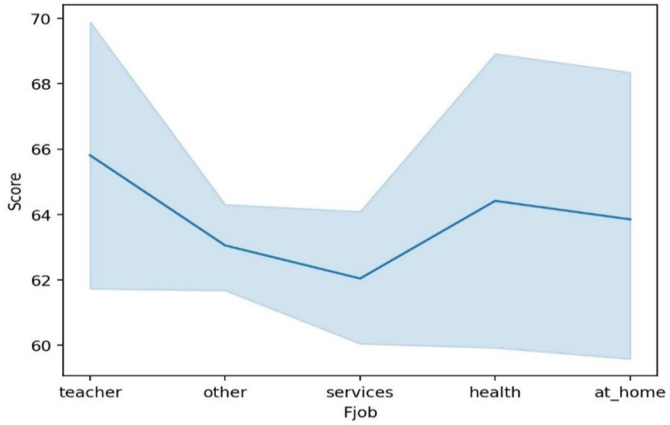


Figure 7 Father's Job

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

4) *Backlog:* Backlogs can create stress and pressure, but dealing with them can motivate students to develop a sense of determination and resilience.

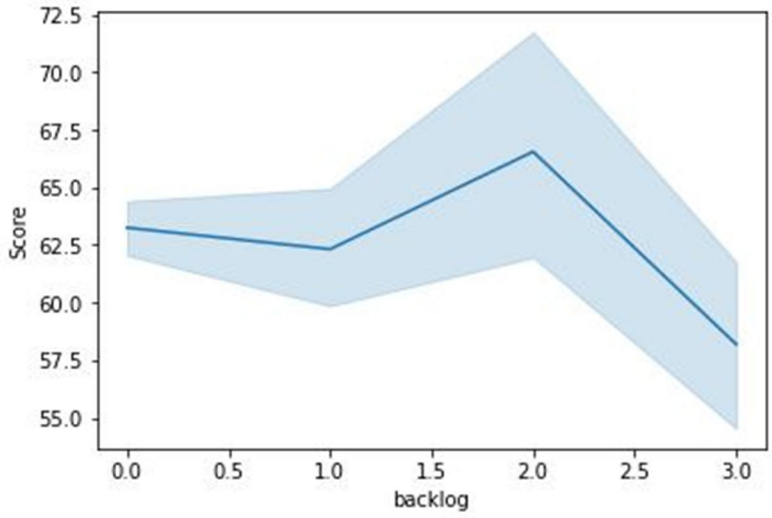


Figure 8 Backlogs

5) *Score by Gender:* Gender stereotypes and expectations can have a negative impact on student performance, such as by assigning certain academic strengths or interests to certain genders.

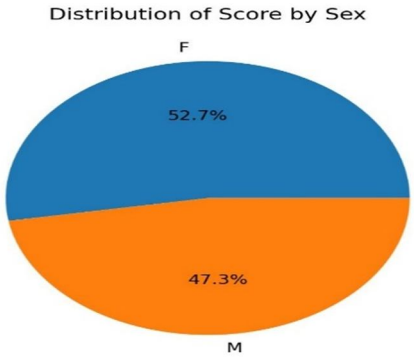


Figure 9 Score by Gender

6) *Going Out:* Going out with friends can help students feel supported, reduce stress, and improve mental health, but it can also lead to distractions and detrimental effects on academic performance.hers with higher education prioritise education and provide a supportive learning environment, encouraging academic achievement and offering guidance

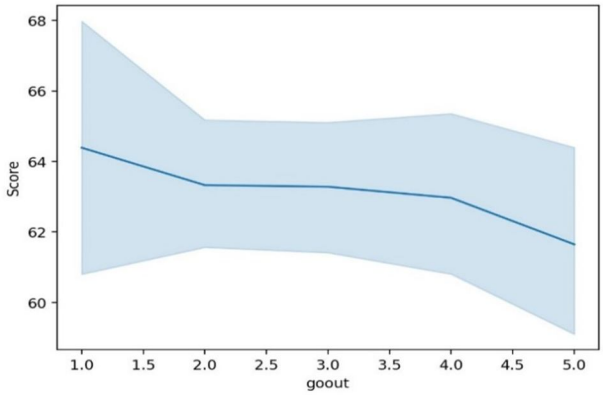


Figure 10 Going Out

V. CONCLUSION

Consequently, we can predict student general performance affecting factors using the proposed system. Using a machine learning algorithm that considers a student's academic performance as well as their geographic, social, and economic circumstances We were able to successfully collect information from students and format our own dataset. By examining the factors that help or hinder student performance, the model we have developed is best suited for an institution or college to keep track of students' performance. As part of this project, the model will be trained using a variety of algorithms to produce accurate results. It may also influence a student's choice regarding their career.

VI. ACKNOWLEDGEMENT

We appreciate Mrs. Shikha Pancholy, our project coordinator, for her support and direction in helping us finish our analysis of the subject (STUDENT GENERAL PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHM). It was a fantastic learning opportunity. Please allow me to take this opportunity to thank Rupesh Rajput, Atharva Raut, and Krishnakant Patil, who made up my group. Their teamwork and input were essential to the project review's success.

REFERENCES

[1] E. S. Bhutto, I. F. Siddiqui, Q. A. Arain and M. Anwar, "Predicting Students' Academic Performance Through Supervised Machine Learning," 2020 International Conference on Information Science and Communication Technology (ICISCT), 2020, pp. 1-6, doi: 10.1109/ICISCT49550.2020.9080033.

[2] L. M. Crivei, G. Czibula, G. Ciubotariu and M. Dindelegan, "Unsupervised learning based mining of academic data sets for students' performance analysis," 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2020, pp. 000011-000016, doi: 10.1109/SACI49304.2020.9118835.

[3] E. Fotopoulou, A. Zafeiropoulos, È. L. Cassà, I. M. Guiu and S. Papavassiliou, "Collective Emotional Intelligence and Group Dynamics Interplay: Can It Be Tangible and Measurable?," in IEEE Access, vol. 10, pp. 951-967, 2022, doi: 10.1109/ACCESS.2021.3137051.

[4] El-Sayed Atlam, Ashraf Ewis, M.M. Abd El-Raouf, Osama Ghoneim, Ibrahim Gad,

[5] A new approach in identifying the psychological impact of COVID-19 on university student's academic performance,Alexandria Engineering Journal,Volume 61, Issue 7,2022,Pages 5223-5233,ISSN 1110-0168,https://doi.org/10.1016/j.aej.2021.10.046.

[6] L. Chen, P. Chen and Z. Lin, "Artificial Intelligence in Education: A Review," in IEEE Access, vol. 8, pp. 75264-75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[7] R. Hasan, S. Palaniappan, A. R. A. Raziff, S. Mahmood and K. U. Sarker, "Student Academic Performance Prediction by using Decision Tree Algorithm," 2018 4th International Conference on Computer and Information Sciences (ICCOINS), 2018, pp. 1-5, doi: 10.1109/ICCOINS.2018.8510600.

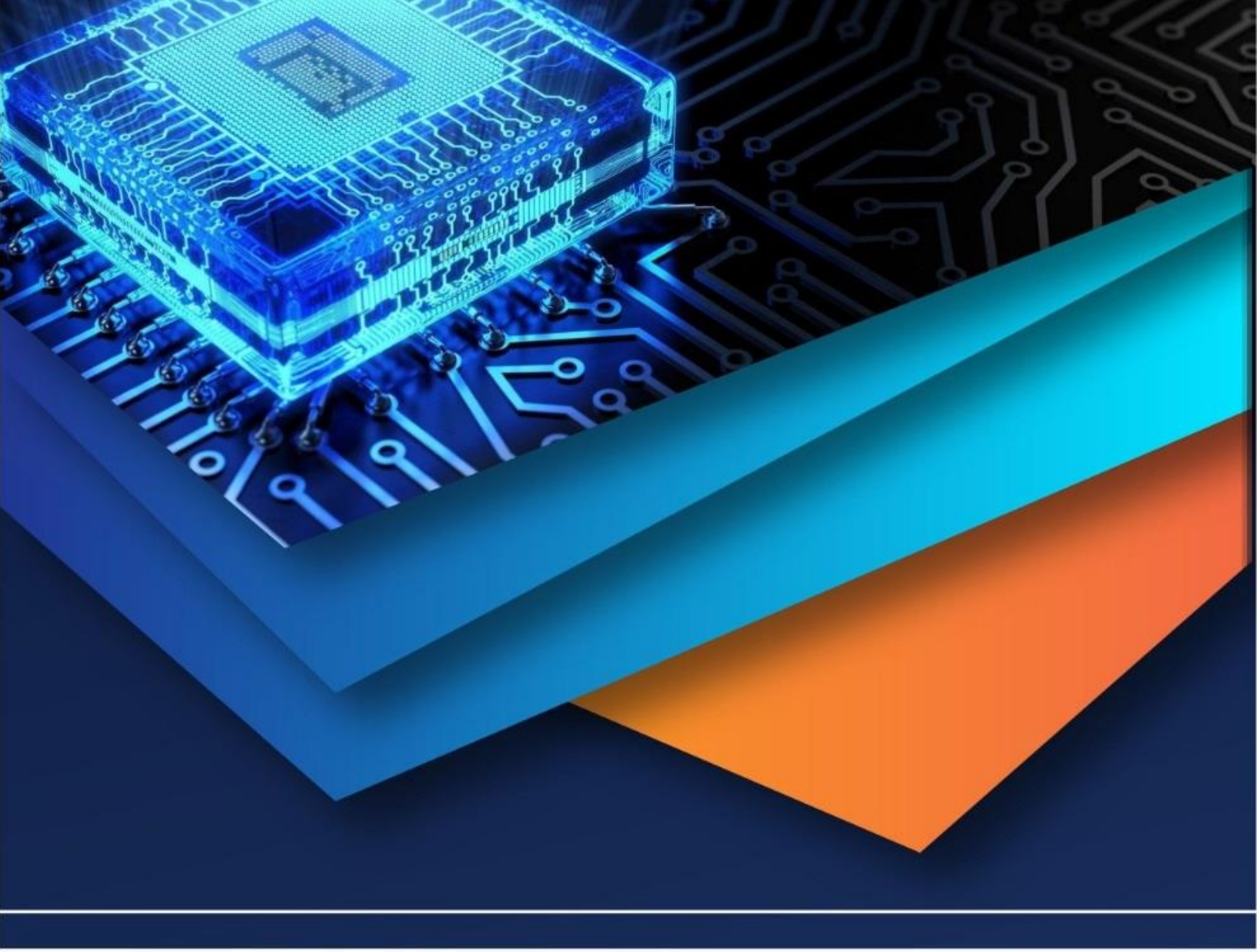
[8] Vinaya Patil,Shiwani Suryawanshi,Mayur Saner and Viplav Patil in, " Student Performance Prediction Using Classification Data Mining Techniques " International Journal of Scientific Development and Research (IJSDR) www.ijsdr.org

[9] S. Alraddadi, S. Alseady and S. Almotiri, "Prediction of Students Academic Performance Utilizing Hybrid Teaching-Learning based Feature Selection and Machine Learning Models," 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), 2021, pp. 1-6, doi: 10.1109/WiDSTaif52235.2021.9430248.

[10] J. A. Olorunmaiye, O. J. Ogunniyi, T. Yahaya, J. O. Olaoye and A. A. Ajayi-Banji, "Modes of Entry as Predictors of Academic Performance of Engineering Students in a Nigerian University," 2020 IFEES World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC), 2020, pp. 1-4, doi: 10.1109/WEEF-GEDC49885.2020.9293683.

[11] F. Aman, A. Rauf, R. Ali, F. Iqbal and A. M. Khattak, "A Predictive Model for Predicting Students Academic Performance," 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019, pp. 1-4, doi: 10.1109/IISA.2019.8900760.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)