RSP Science Hub

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

Student's Performance and Difficulties Prediction

Kumuda Kanatti[1], Kehkashaan Khanum A[1], Janani Jayashankar[1], Gayathri P [1], Ambeshwar Kumar [2]
[1]Department of Computer Science Engineering, Dayananda Sagar University, Bangalore, India
[2]Professor, Department of Computer Science Engineering, Dayananda Sagar University, Bangalore, India

Email: kumuda.kanatti02@gmail.com

Abstract

Every government and nation in the world strives to make advancements in the area of education because it's a pivotal element of society. According to studies, pupil performance has declined after the nimbus- contagion epidemic that disintegrated life in 2020, which emphasizes the need to treat this problem more seriously and seek to pinpoint both the causes and the effective treatments. The educational system has been impacted in numerous ways. By assaying and assessing scholars' academic performance while taking a variety of aspects into the account, the design aims to ameliorate academic, professional, and university guidance. Scholars will be surveyed using a questionnaire for this design, and data from the UCI Machine Learning Repository will also be used. Understanding the colorful aspects that affect scholars' performance and prognosticating scholars' success through the use of colorful machine learning algorithms to dissect pupil data and one's issues The three different orders of pupil characteristics include particular traits, academic attributes, and behavioral traits. The pupils' literacy achievements are told by their behavioral traits. It has been noted that regression and Classification models are constantly used in prediction.

1. Introduction

Education is a crucial aspect of society and governments worldwide strive to improve this area. However, since the outbreak of the COVID-19 pandemic in 2020, student performance has dropped, highlighting the urgent need to identify effective solutions and factors that affect performance. The educational system has been impacted in various ways. The goal of this project is to improve academic, professional, and university guidance by assessing student performance while taking into account various factors. The data will be collected through a questionnaire-based survey of students and from the UCI Machine Learning Repository. Three types of

student characteristics will be considered, including educational features, behavioral features, and specific aspects. It has been observed that prediction is generally made using bracket ways and regression.

2. Literature Survey

Rajalaxmi R R et al. (R) investigated the use of regression models for predicting the academic performance of engineering scholars. The study compared direct multivariate regression and support vector machine algorithms with other methods. The findings showed that while the support vector machine algorithm had lower sensitivity than other algorithms, multiple measures were used to validate the models and predict the likelihood of accurate

predictions.

Harikumar Pallathadka et al (Pallathadka) conducted research on the prediction and classification of student performance data using machine learning algorithms. The study utilized a decision tree and fuzzy genetic algorithm to anticipate academic performance based on prior academic results. The analysis of student's talents and interests in relation to their performance was also explored using the UCI machine learning dataset. (Al-Shehri et al.) This analysis could help teachers identify and provide targeted support to students who require it, ultimately leading to improved instruction quality.

The accuracy of the student's predictive model using behavioral features achieved up to 22.1% improvement compared to the results when removing such features, and it achieved up to 25.8% accuracy improvement using ensemble methods. For building a better student performance predictive model we have to make use of the behavioral features of the students. The use of ensemble methods can help us achieve a better predictive model. (Tjandra)

Pranav Dabhade et al (Dabhade et al.) presented a paper on data mining in education to predict student academic performance using machine learning algorithms where they used multiple linear regression and support vector regression. To evaluate the influence of features and attributes on the desired result, built-in models of machine learning algorithms were selected from sci-kit-learn. Past performance is most important for predicting future performance. The results obtained show that there is a relationship between student behavior and academic performance. The research can be extended using neural networks and other regression algorithms. The bigger the data set, the better the prediction. The use of behavioral elements is also important for predicting a student's future academic performance. (Van Der Schaar)

3. Design

A training model is a model used to train a machine learning algorithm. It consists of sample output data and corresponding sets of input data that affect the output. (Juan and Gómez-Pulido)The training model is used to process the input data using an algorithm to correlate the processed output with the sample output.
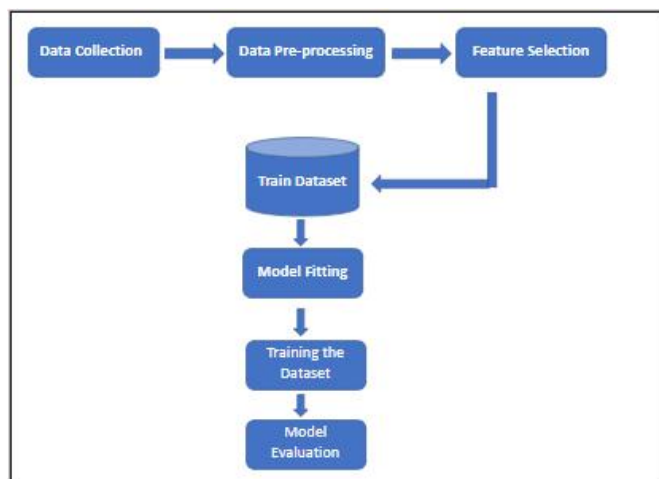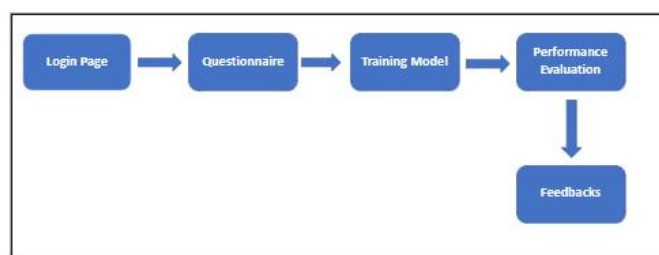

FIGURE 1. Training Model

FIGURE 2. Flow of a website

Based on the built training model, we can create a webpage that predicts student performance, allowing us to pay closer attention to weak students, the above FIGURE 2 shows the flow chart of the website.

3.1. Data Collection:

Data collection gathers and measures information on variables of interest. In our design, we substantially collect data for training our model from the UCI machine literacy depository and Google form-based questionnaire. (Nabil, Seyam, and Abou-Elfetouh) FIGURE 1

3.2. DataPre-processing:

Data preprocessing is a data mining fashion used to transfigure the raw data in a useful and effective format. Then we clean the data by removing all the null values and homogenizing them to convert it into a useful and effective format that will be useful to train our model. (Umar) FIGURE 1

3.3. Feature selection:

Feature selection is the process of reducing the number of input variables when developing a prophetic model. FIGURE 1

3.4. Training the dataset:

After a thorough literature check, we've concluded using SVM and KNN bracket models(We might change/ include any fresh models if needed). (Casillas et al.)FIGURE 1

3.5. Model Evaluation:

Model evaluation is the process of using different evaluationcriteria to understand a machine literacy model's performance, as well as its strengths andsins. (Schmidt-Thieme) FIGURE 1.

4. Methodology

Dataset obtained from UCI Machine Learning Repository. The original data set consisted of some string values so during the preprocessing we mapped the string values to numerical values and this task was done by a function named numerical_data. After the string values were mapped to the numerical values we had to normalize the variable and this task was done by a function called feature_scaling. The dataset consists of 395 rows × 31 columns. After plotting the correlation graph between the student's status(pass/fail) and the other features in the dataset. (Sekeroglu, Dimililer, and Tuncal) (Xu)

5. Result

Observations made from the dataset are as follows:

1)Students' whose parents are educated have a positive impact on the student's status.

2) Students who wanted to take up higher education also had a positive impact on the student's status.

3)Going out with friends for a longer duration of time had a negative impact.

4)Age and failures also had a negative impact on the student's status.

Features that had low or no impact on the student's status are

1)The geographical location of the students and 2)Alcohol consumption

Features that had a higher/positive impact on the student's status are

1)Spending limited time on going out with friends.

2)Is not in a romantic relationship

3)Parents are educated

4)Student has a strong desire for higher studies

5) Has access to the internet

6)Is Healthy

7)Study for at least 10 hours a week

	precision	recall	f1-score	support
0.0	0.57	0.57	0.57	30
1.0	0.85	0.85	0.85	89
accuracy			0.78	119
macro avg	0.71	0.71	0.71	119
weighted avg	0.78	0.78	0.78	119

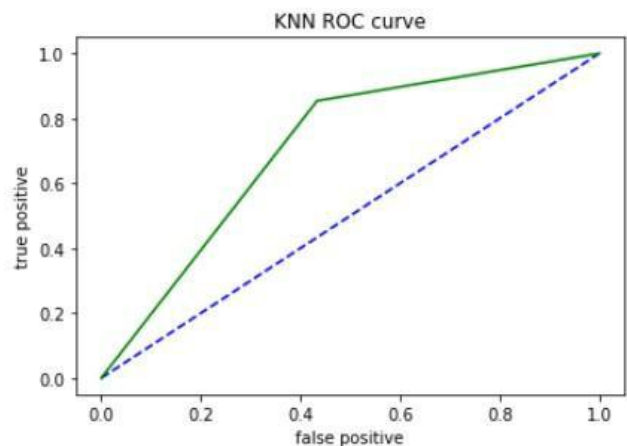**FIGURE 3.** **Classification Report of KNN Model**

FIGURE 4. **ROC Curve of KNN Model**

KNN is a machine learning algorithm that is used for classification and regression tasks. The model's accuracy has increased to 78% as a result of various factors such as data quality and quantity, feature selection, hyperparameter tuning, and model complexity. FIGURE 3, FIGURE 4
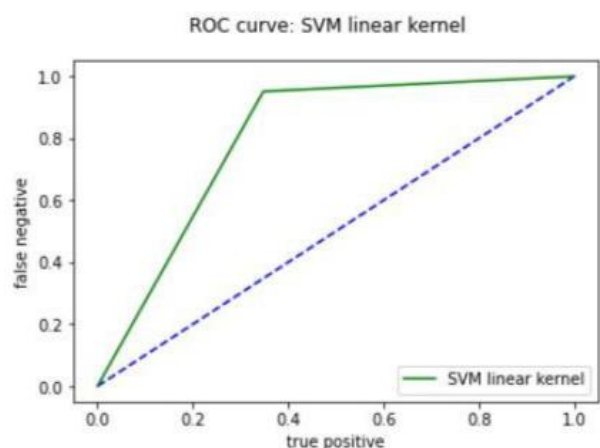

FIGURE 5. **ROC Curve of SVM Model**

Support vector machines, or SVMs, are powerful machine learning algorithms that are commonly

used for classification and regression tasks. SVM, like KNN, is affected by the same variety of factors. With the addition of the Optimal C value, the model's accuracy increased to 84%. Our model performs better with SVM linear kernel FIGURE 5 model than with other kernels.

6. Comparison of Results

The previous models only included personal and academic characteristics, whereas our model involves personal, academic, and behavioral characteristics or traits.

The addition of behavioral features improved the model's accuracy. This is a common approach in machine learning in which you try to improve the model's performance by adding new features that capture additional data information.

It is important to note that the improvement in accuracy is not always significant, and it is also dependent on the quality of the new features and their ability to capture the underlying patterns in the data. It is also critical to choose and

pre-process the new features carefully to avoid overfitting or adding noise to the model.

7. Conclusion

In conclusion, the student performance and difficulties prediction machine learning project has shown promising results in predicting academic performance and identifying potential challenges for students. By analyzing various features such as attendance, grades, and demographic information, the model was able to make accurate predictions and provide insights to support student success.

However, it's important to note that machine learning models are not perfect and there are limitations to their predictions. The accuracy of the model depends on the quality of the data used for training and testing, as well as the features selected for analysis. Therefore, it's important to continuously evaluate and refine the model to improve its accuracy and usefulness.

Overall, the student performance and difficulties prediction machine learning project has the potential to assist educators and administrators in identifying at-risk students and providing targeted interventions to support their academic success.

References

Casillas, Alex, et al. "Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior." *Journal of Educational Psychology* 104.2 (2012): 407–420.

Dabhade, Pranav, et al. "Educational data mining for predicting students' academic performance using machine learning algorithms". *Materials Today: Proceedings* 47 (2021): 5260–5267. 10.1016/j.matpr.2021.05.646.

Juan, A and Gómez-Pulido. "Optimizing Latent Factors and Collaborative Filtering for Students'PerformancePrediction". *Optimizing Latent Factors and Collaborative Filtering for* (2020). 10.3390/app10165601.

Nabil, Aya, Mohammed Seyam, and Ahmed Abou-Elfetouh. "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks". *IEEE Access* 9 (2021): 140731–140746. 10.1109/ACCESS.2021.3119596.

Pallathadka, Harikumar. "Classification and prediction of student performance data using various machine learning algorithms". *Classification and prediction of student performance data using* (2021). 10.1016/j.matpr.2021.07.382.

R, Rajalaxmi R. "Regression Model for Predicting Engineering Students Academic Performance"". *International Journal of Recent Technology and Engineering (IJRTE)* (2019): 2277–3878.

Schmidt-Thieme, Lars. "Improving Academic Performance Prediction by Dealing with Class Imbalance". *International Conference on Intelligent Systems Design and Applications* (2009). 10.1109/ISDA.2009.15.

Sekeroglu, Boran, Kamil Dimililer, and Kubra Tuncal. "Student Performance Prediction and Classification Using Machine Learning Algorithms". *Proceedings of the 2019 8th International Conference on Educational and Information Technology* (2019). 10.1145/3318396.3318 419.

Al-Shehri, Huda, et al. "Student performance prediction using Support Vector Machine and K-Nearest Neighbor". *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (2017). 10.1109/CCECE.2017.7946847.

Tjandra, Ellysa. "Student performance prediction in higher education: A comprehensive review". Ed. and others. 2022.

Van Der Schaar, Mihaela. "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs". *IEEE Journal of Selected Topics in Signal* (2017). 10.1109/JSTSP.2017.2692560.

Xu, Liancheng. ""* Student Performance Prediction with Short-Term Sequential Campus Behaviors* (2020). 10.3390/info11040201;%20https://doi.org/10.3390/info11040201.

Embargo period: The article has no embargo period.