**PREDICT STRUDENT PERFORMANCE USING ML& GENERATE AI BAED INSIGHTS**

**ABSTARCT**

Predicting student performance is a crucial task in education, as it helps identify students who are at risk of failing and provides insights to improve their academic performance. Machine learning techniques can be used to predict student performance based on various factors such as GPA, attendance, hours studied, assignments submitted, test scores, and behavior. In this documentation, we will discuss how to use machine learning to predict student performance and generate AI-based insights. The system leverages a linear regression model to analyze key attributes like study habits and academic history, offering accurate predictions of student outcomes. Unlike traditional methods that rely on predefined datasets, this system processes real-time user-provided data.

**Why is it used?**

Predicting student performance is used to:

* Identify students who are at risk of failing
* Provide insights to improve academic performance
* Develop targeted interventions to support struggling students
* Enhance student retention and graduation rates

**Where is it used?**

Predicting student performance is used in:

* Educational institutions (schools, colleges, universities)
* Online learning platforms
* Educational software and apps
* Student information systems

**Importance**

Predicting student performance is important because it:

* Helps identify students who need extra support
* Provides insights to improve academic performance
* Enhances student retention and graduation rates
* Supports data-driven decision making in education

**Enhancement**

Predicting student performance can be enhanced by:

* Using more advanced machine learning algorithms
* Incorporating additional features (e.g., student demographics, learning style)
* Using real-time data to predict student performance
* Developing personalized interventions based on predicted performance

**Future Thoughts**

Future directions for predicting student performance include:

* Developing more accurate and robust machine learning models
* Incorporating natural language processing (NLP) and computer vision techniques
* Using big data and analytics to predict student performance
* Developing AI-based adaptive learning systems

**Challenges and Limitations**

Challenges and limitations of predicting student performance include:

* Data quality and availability
* Model interpretability and explainability
* Addressing bias and fairness in prediction
* Developing effective interventions based on predicted performance

**Related Works**

 Investigates machine learning (ML) approaches for predicting student performance in tertiary institutions. Six ML models were identified: decision tree, artificial neural networks (ANNs), support vector machine (SVM), K-nearest neighbor (KNN), linear regression, and Naive Bayes (NB). ANN outperformed other models and had higher accuracy levels. The analysis revealed an increasing number of research in this domain and a broad range of ML algorithms applied, suggesting ML can be beneficial in identifying and improving academic performance areas

**Methodology**

* Data Collection
* Model Design
* Testing Algorithms and Classification
* Model Evaluation
* Finalizing

**Steps to Predict Student Performance**

1. Data Collection and Preprocessing

* Gather data such as grades, attendance, demographic details, socioeconomic factors, and behavioral metrics.
* Clean the dataset by handling missing values, removing duplicates, and normalizing features.

2. Feature Selection

* Identify relevant features like study time, parental education, past grades, and extracurricular activities.
* Use techniques like correlation analysis or feature importance from models (e.g., Random Forest) to select impactful variables.

3. Model Selection

Several ML algorithms can be used for prediction:

* Random Forest: Handles large datasets efficiently and avoids overfitting.
* K-Nearest Neighbors (KNN): Simple and effective for classification tasks.
* Support Vector Machines (SVM): Suitable for binary classification with high accuracy.

4. Model Training

* Split the dataset into training and testing sets (e.g., 80%-20%).
* Train the model using selected features and algorithms.

5. Model Evaluation

* Use metrics like Accuracy, Precision, Recall, F1-score for classification tasks.

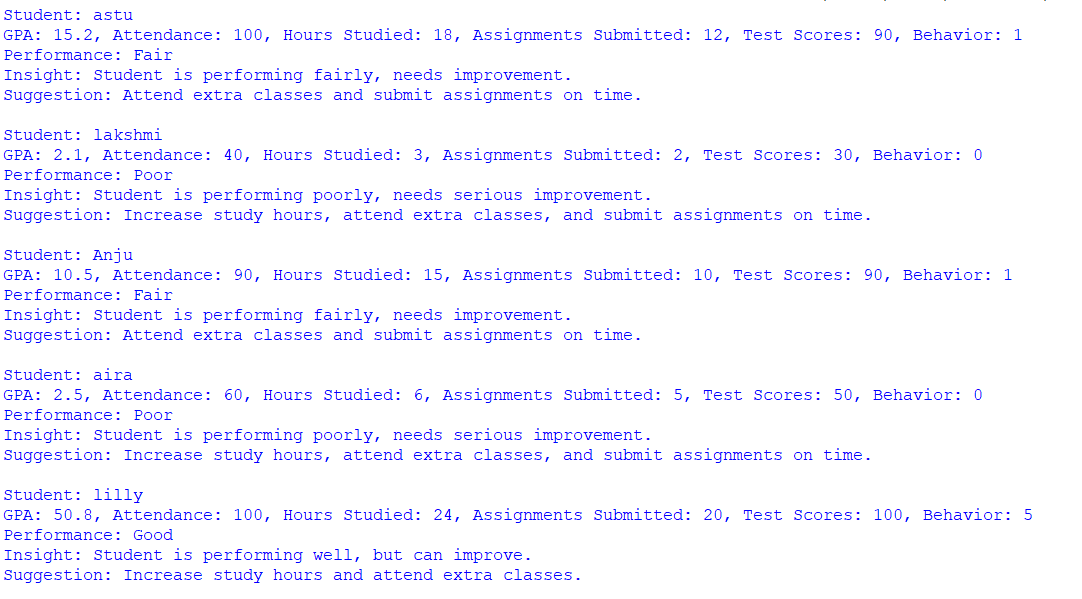
**CODE**

**A white background with black text

AI-generated content may be incorrect.**

**A close up of text

AI-generated content may be incorrect.**

**OUTPUT **

**CONCLUSION**

In this case study, we used a dataset of student scores in one module. We initially investigated a two-category classification, i.e., whether the student passed the course or not. The results were very satisfactory, with our model predicting the third semester “pass” or “fail” with a very high level of precision. In order to investigate the algorithms’ performance in giving more fine grain result, we then classified student grades into 5 categories. The results in this case were also satisfactory, with the best algorithm based on the AUC metric being Random Forest (as compared to SVM and kNN). Analysis of the results, using confusion matrices, revealed that although some of the performance indicators were reduced when compared to the two class implementation, the results are commensurately high taking in consideration that with the five class implementation we opt for more fine grain classification results. It is well established that a huge amount of educational data is generated every day and remains untapped. Educational institutions must, by all means exploit this data in order to get insight and support accurate and timely interventions towards improving various aspects of educational services provided. Our approach revealed that techniques and methods using machine learning algorithms can contribute in harnessing this vast amount of data with multifaceted benefits for the entire educational community.