Exploring the Learning Analytics of Skill-based Course Using Machine Learning Classification Models

A. Vinitha[1], M. Madhusudhan[2]

[1]M.Tech. Scholar, Department of Computer, Ananthalakshmi Institute of Technology
[2]Professor, Ananthalakshmi Institute of Technology

ABSTRACT:
This study explores learning analytics of a skill-based course using various machine learning classification models, including Random Forest, Logistic Regression, CatBoost, Support Vector Classification (SVC), and Naïve Bayes. The objective is to categorize student outcomes into four classes: Pass, Distinction, Withdrawn, and Fail. The research contributes to the growing body of knowledge in learning analytics and machine learning applications in education. The findings from this investigation offer educators and academic institutions a robust framework for early identification of students at risk of underperformance or withdrawal, thereby enabling timely intervention to enhance student success in skill-based courses.

Keywords: Random Forest, Logistic Regression, CatBoost, Support Vector Classification (SVC), and Naïve Bayes,Mlp, Lda,Passive aggressive classifier.

Introduction

The increasing importance of skill-based courses in academia necessitates innovative approaches to monitor and enhance student performance effectively. This paper aims to develop and compare various ML classification models, including Random Forest, Logistic Regression, CatBoost, SVC, and Naïve Bayes, for predicting student outcomes in skill-based courses. By analyzing historical student data, the project seeks to categorize outcomes into Pass, Distinction, Withdrawn, and Fail, offering valuable insights for early identification and intervention for at-risk students.

Literature survey :-

1. **Jermine G. Valen-Dacanay., &Thelma D. Palaoag [2023]:-** Predicting student success is crucial for educational institutions to improve learning outcomes and retention rates. While big data provides a wealth of information, allowing for detailed analysis and accurate predictions, it raises concerns regarding privacy, security, and ethical use of student information..
2. **Maulud, D., & Abdulazeez, A. M. (2020). :-** Linear regression is a fundamental algorithm in machine learning, providing a straightforward approach to predicting a response variable based on one or multiple predictor variables. It establishes a linear relationship between the dependent and independent variables, often used for forecasting, trend analysis, and determining the significance of predictors. applicability to complex, non-linear relationships in data. Despite this, it remains a foundational

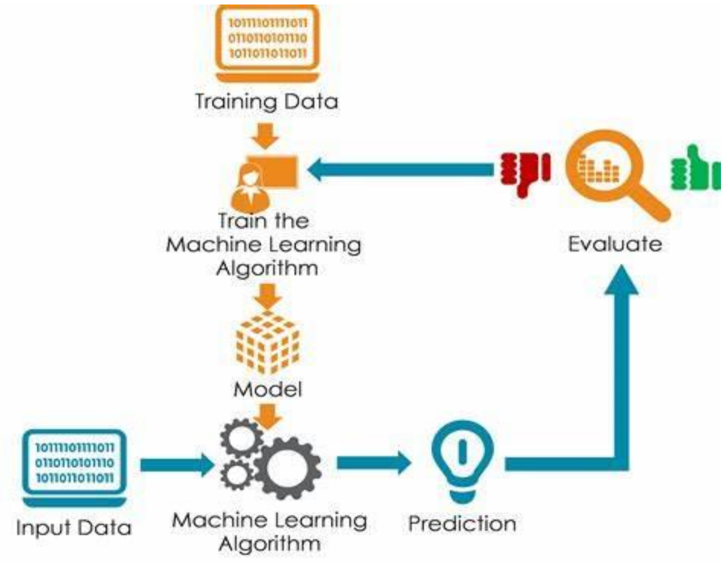
technique, often serving as a starting point for more advanced regression methods in machine learning and statistics.

3. **Peach, R. L., Yaliraki, S. N., Lefevre, D., & Barahona, M. (2019):-**Employing data-driven methods, this research utilizes unsupervised clustering to categorize online learner behavior effectively. The study leverages extensive datasets encompassing various behavioral metrics of online learners. These metrics include engagement levels, time spent on tasks, participation in discussions, and submission patterns.
4. **Waheed, H., Hassan, & Nawaz, R. (2020).:-** Predicting students' academic performance with precision is crucial for timely intervention and support. With the advent of Virtual Learning Environments (VLEs), there's an influx of data regarding student interactions and engagement patterns. This research leverages deep learning models to analyze this extensive data, offering accurate predictions of students' academic outcomes. Deep learning's ability to discern intricate patterns in large datasets makes it apt for handling the complex, multifaceted data generated by VLEs.
5. **Waheed, H., Aljohani, N. R., & Hassan, S. U. (2019):-** Cognitive Computing in Technology-Enhanced Learning (TEL) plays a pivotal role in understanding and predicting student behavior and performance. By mimicking human reasoning, this advanced technology analyzes students' learning patterns and engagement, providing a nuanced understanding of their academic journey. Cognitive computing integrates various AI technologies, such as machine learning, natural language processing, and data mining, offering personalized learning experiences and real-time feedback.

PROPOSED SYSTEM :-

The utilizes machine learning models, namely Random Forest, Logistic Regression, Support Vector Classification (SVC), and Naïve Bayes, Mlp, Lda,Passive aggressive classifier to analyze learning analytics in skill-based courses. By training and validating these models with historical student data, encompassing engagement levels and assessment scores, the system can categorize student outcomes into Pass, Distinction, Withdrawn, and Fail. The comparative analysis of each model provides insights into student performance and instructional design efficacy. This framework allows for early identification of at-risk students, facilitating timely intervention to support student success in skill-based courses.

Block diagram



The Block diagram of the Analytics of skills based of our classification models are shown in fig concretely, the systems consist of the following entities:

Training data :- is used to teach the machine learning model how to make predictions or perform a desired task. It is typically labeled, which means that the model's output is known for each data point.

Input data :- An input usually refers to an example (sometimes also known as sample ,observation or data point)x from a dataset that you pass to the model.

Machine learning Algorithms:- Machine learning algorithms are Pieces of code that help people explore analyze , and find meaning in complex data sets . Each algorithm is a finite set of unambiguous step-by-step instructions that a machine can follow to achieve a certain goal.

Prediction :- The process of using data to forecast future outcomes. The process uses data analysis, machine learning, artificial intelligence, and statistical models to find patterns that might predict future behavior.

Evaluate :- Evaluates the forecasting accuracy of each forecasted timestamp. The error metrics are aggregated over the forecasting error on each forecasted timestamp.

CONCLUSION & FUTURE WORK

This study illuminates the efficacy of various machine learning models, including Random Forest, Logistic Regression, CatBoost, SVC, and Naïve Bayes,Mlp,lda,passive aggressive classifiers in analyzing student outcomes in skill-based courses. Each model demonstrated unique strengths in predicting outcome classes, providing crucial insights into student performance and instructional design effectiveness.

Additionally, implementing a feedback loop where the models are continually refined and updated with new data will ensure that the system remains robust and relevant. These improvements aim to further support educators in identifying and assisting at-risk students, ultimately contributing to enhanced learning experiences and outcomes.

Reference

1. Jermine G. Valen-Dacanay, Thelma D. Palaoag[2023]. Analytics Of Skill-Based Course Using Machine Learning.
2. Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior, 104, 106189
3. Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.
4. Wasif, M., Waheed, H., Aljohani, N. R., & Hassan, S. U. (2019). Understanding student learning behavior and predicting their performance. In Cognitive Computing in Technology-Enhanced Learning (pp. 1-28). IGI Global.
5. Wasif, M., Waheed, H., Aljohani, N. R., & Hassan, S. U. (2019). Understanding student learning behavior and predicting their performance. In Cognitive Computing in Technology-Enhanced Learning (pp. 1-28). IGI Global.