Students Exam Performance Prediction

OPEN ACCESS

Volume: 11

Special Issue: 3

Month: July

Year: 2024

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Received: 17.05.2024

Accepted: 20.06.2024

Published: 08.07.2024

Citation:
Lavanya, K., et al. "Students Exam Performance Prediction." *Shanlax International Journal of Arts Science and Humanities*, vol. 11, no. S3, 2024, pp. 14–19.

DOI:
https://doi.org/10.34293/sijash.v11iS3-July.7913

Lavanya. K
Assistant Professor, Department of AI & DS, Arjun College of Technology

G. Swathi
Department of AI & DS, Arjun College of Technology

Aparna. T
Department of AI & DS, Arjun College of Technology

Viswa Teja. P
Department of AI & DS, Arjun College of Technology

Abstract

Predicting student performance is crucial for understanding their progress and intervening effectively. This research aims to identify current student status and forecast future results to enable teachers to provide timely guidance and support. By analyzing dependencies for final examinations, we can recommend suitable courses for upcoming semesters, serving as advisor's to students. Many students struggle due to a lack of proper guidance and monitoring, as teachers can't monitor everyone simultaneously. An AI system can assist by identifying which students require specific types of support. Ultimately, the goal is to empower students to avoid predicted poor results through proactive intervention, potentially achieving accuracy rates as high as 94.88%.

Keywords: GBDT, Student Performance, Educational Data Mining, College Education, Machine Learning, Result Prediction, Kappa Statistic, F-Measure, WEKA.

Introduction

Students are a crucial component of any educational institution and the nation as a whole. In a crowded classroom, it is challenging for teachers to monitor each student individually and provide equal attention to everyone. Classrooms are composed of both introverted and extroverted students, each with unique needs. Through our experience in tutoring, we recognized the importance of closely monitoring each student's progress. It is essential for teachers to identify when a student is starting to fall behind. Therefore, we decided to conduct research on predicting a student's academic condition using Artificial Intelligence and their previous academic records.A student's academic data is crucial for this research, as it reveals a great deal about the student, such as their study habits, preferred subjects, and those they find challenging. Additionally, IQ tests and psychological assessments can provide valuable insights. Understanding how much time a student dedicates to studying versus hobbies can help determine the type of motivation they need from their teacher. Teachers act as mentors or coaches for their students.

We believe that teachers should have insight into their students' potential for future success in various fields or courses.If a teacher has access to a student's predicted results in advance, they can provide timely help and take immediate steps to improve the student's performance. Similarly, if students are aware of their predicted results, they can take proactive measures to improve. This research goes beyond merely enhancing academic results. By considering these factors, classification models using Machine Learning Algorithms (MLAs) can be developed to predict student outcomes. Researchers have utilized various MLAs for this purpose, including Decision Trees (DT), Bayesian Networks (BN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), among others..

The primary objective of this research is to determine whether students are truly learning. A student's performance in a subject depends on various factors. Typically, each subject includes marks from class tests, attendance, assignments, presentations, mid-term exams, and final exams. The total score for a subject is the sum of these components. Additionally, factors such as a student's physiological data, IQ score, and the amount of time spent studying also play a role. We have utilized the curriculum of Daffodil International University (DIU) for this research. At DIU, each course is graded out of 100 marks, which are distributed across class tests, attendance, presentations, assignments, mid-term exams, and final exams. It is evident that a student's performance in final examinations largely depends on other factors such as class tests, mid-term examinations, and attendance. Therefore, this research primarily aims to predict final examination performance based on students' past records. In the realm of Machine Learning, algorithms such as K-Nearest Neighbors, SVC, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, and Linear Discriminant Analysis can be employed to predict future results using existing student data. Academic performance is influenced by numerous factors, including previous academic records, economic status, family background, and performance in mid-semester examinations.

Problem Definition

The program combines Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), particularly Long Short-Term Memory (LSTM), to process images and generate English descriptions and captions.

CNN acts as a feature extractor, analyzing the provided image and extracting relevant features. The study focused on data mining in higher education, specifically examining university student dropout rates. Various data mining approaches were employed to analyze and predict dropout cases in university programs. The collected data included students' academic histories, transcripts of courses from the first two years of the computer science program, GPAs, and high school averages, along with a class label indicating whether the student graduated. Decision Tree (DT) and Naive Bayes (NB) techniques were used to develop the predictive model, which was tested using 10-fold cross-validation. This project conducted a comparative study of two data mining classification techniques to predict student performance in the post unified tertiary matriculation examination (PUTME). The research utilized two classification techniques: the J48 Decision Tree and the Artificial Neural Network Multilayer Perceptron, to analyze student performance trends based on their modes of entry. The study concluded that the decision tree algorithm is preferable over the neural network for large datasets, as it builds models quickly and presents results in an easily interpretable tree structure.

Methedology

There are many attempts by researchers and administrators to mine students' data because of the need to utilize the volume of data that are available in education databases for enhanced educational service delivery.

The Adaptive Neuro-Fuzzy Inference System (ANFIS) was utilized to predict student academic performance, highlighting its potential as a decision- support tool in higher education to enhance student success rates. The study also provided guidelines for identifying areas for improvement in educational processes. Additionally, two data mining classification techniques, the J48 Decision Tree and the Artificial Neural Network Multilayer Perceptron, were employed to analyze student performance trends based on their modes of entry.Concurrently, the project strives to introduce personalized learning plans, tailoring educational content to individual strengths, weaknesses, and learning styles. The incorporation of adaptive learning technologies ensures a dynamic and customized approach to each student's educational journey. Real-time progress monitoring stands as a cornerstone objective, providing educators with comprehensive data on students' academic achievements, continuous assessments, and teacher feedback. This not only allows for timely interventions but also facilitates adaptive teaching strategies to enhance overall learning outcomes. Parental engagement is a key focus, with the project aiming to establish a dedicated portal providing parents with insights into their child's academic performance, attendance records, and extracurricular activities. A real- time progress monitoring dashboard is developed, offering educators actionable insights into student achievements and assessments.

Data mining techniques were adopted to predict students' Grade Point Averages (GPAs) at graduation. These prediction techniques were implemented to help educational institutions assess the likelihood of students' GPAs at graduation. By predicting students' chances of success, those with lower predicted GPAs can be encouraged early on to invest more effort in their academic work. Neural networks and support vector machine algorithms were applied to the data of computer education and instructional technology students to forecast their potential GPAs at graduation.

Proposed Work

Our proposed methodology started with gathering data set. So we tried to collect students' class test marks, attendance mark, presentation mark, assignment marks, midterm marks and final examination marks .This is a classification problem.

Dataset

We utilized a new dataset for the proposed model, consisting of data from 1,170 students across three subjects. Processing this dataset is straightforward as it is collected from student records.

However, the primary challenge lies in obtaining real data due to the confidential nature of academic information for both students and institutions. Using dummy data proved ineffective, leading to issues with overfitting and low accuracy rates. Consequently, we prepared the dataset from scratch. Efforts to collect data from our university were unsuccessful, so we turned to our course teachers. Initially, we used real data from 30 students but found no significant patterns. It became evident that a larger dataset was necessary. After expanding the dataset to include 1,170 students.

Dataset Preparation

Data preparation was a straightforward process using data from a CSV file. Initially, we removed unnecessary columns from the dataset, including the student ID and subject columns, as they did not contribute to the desired outcome. All data was in numeric format, simplifying the process. Subsequently, we generated a correlation graph, which revealed that the final mark was predominantly influenced by mid-term and Class Test (CT) marks. Therefore, our input parameters were determined to be mid-term and CT marks.

Architecture of the Model

Our proposed model aims to predict students' final exam results, utilizing K-Nearest Neighbors and Decision Tree algorithms to achieve the highest accuracy rates. Additionally, Support Vector Classifier (SVC), Random Forest Classifier, Gradient Boosting Classifier, and Linear Discriminant Analysis are viable options for accurate implementation. K-Nearest Neighbor (KNN) stands out as one of the foremost machine learning algorithms due to its simplicity and versatility. It finds applications in various fields such as video recognition, healthcare, political science, handwriting detection, image recognition, and finance. In KNN, the model's structure is derived directly from the dataset.he parental engagement portal is designed and secured, providing parents with comprehensive insights into their child's academic journey.Concurrently, teacher support tools are developed, accompanied by training sessions to equip educators with the necessary skills to address diverse learning needs.

$$\text{nrp}(,) = nrp() - \sum \times$$
$$nrp() \ldots\ldots\ldots (2)$$

In Equation (1) and (2) nrp() = Entropy before partition, = final exam marks, value of target variable is marks of final exam,
= Marks of mid-term exam (between 0 to 25),
= Number of target variable value is 41 (0 to 40). From this equation we have created the decision tree using Class Test, Mid and Final entities.

Figure 2 Student Prediction Workflow

Results

The outlined objectives, from predictive analytics to personalized learning plans and parental engagement, collectively contribute to creating a holistic and adaptive educational ecosystem. As the system progresses, future enhancements hold the promise of elevating its capabilities to meet the evolving needs of the education landscape. Advanced predictive analytics, dynamic personalized learning plans, and seamless integration with emerging technologies will ensure that the system remains at the forefront of educational innovation. The commitment to ethical AI, transparency, and continuous improvement underscores the system's dedication to responsible technology use and addressing potential biases. Longitudinal data analysis and collaboration with career guidance systems reflect a forward-looking perspective, providing a comprehensive understanding of students' academic journeys and aligning education with future career paths.

Subsequently, in the prediction phase, the RNN effectively predicts descriptive captions based on the decoded feature vectors, ensuring the generation of captions that aptly convey the image content.

Finally, in the audio conversion step, the generated textual descriptions and captions are seamlessly converted into audio format using Natural Language Processing (NLP) techniques, thereby enhancing accessibility for visually impaired individuals through auditory perception of image content. Overall, these results demonstrate the successful execution of the proposed methodology, facilitating improved accessibility and usability of visual content across diverse user demographics.

Ultimately, the "Student Performance Enhancement System" strives to create a positive and supportive educational environment, where students thrive, educators are empowered, and parents actively participate in the learning process.Collaboration with career guidance systems will provide students with insights into the alignment of their academic pursuits with long-term career goals.

Conclusion

The proposed model demonstrates superior performance in predicting student performance over three semesters. Through optimized training and validation sets, the model achieves remarkable accuracy across various events. Specifically, the K-Nearest Neighbors and Decision Tree Classifier models yield accuracies of 89.74% and 94.44%, respectively. By excluding certain basic variables like attendance, assignment, and presentation, the model's accuracy is enhanced. However, considering all variables would likely further improve accuracy. This model represents a significant advancement compared to previous research efforts. Its implementation benefits students, teachers, parents, and institutions alike. For university department heads, having a system that can promptly identify students in need of special assistance and those falling behind is invaluable. With such a system, department heads can take immediate steps to support struggling students and easily access lists of irregular students within their department.

Engaging the expertise of a psychiatrist would greatly enhance the richness of our dataset. By incorporating psychological data based on expert opinions, we aim to enrich our dataset. Subsequently, we plan to implement additional machine learning and data mining algorithms such as Naïve Bayes and Support Vector Machine (SVM) to further improve prediction accuracy and incorporate more attributes. This expansion will enable dynamic prediction of various student events. Our intention is to develop a web-based application for implementing the model, with added features beneficial for students, teachers, and parents.

Furthermore, we aspire to expand our datasets in the future while prioritizing accuracy and outcomes. Our plan includes revising the model to incorporate additional features for academic and industrial applications. For instance, we envision implementing the model within companies to predict and analyze employee performance on an annual or monthly basis. This approach could help companies identify areas for improvement and enhance productivity. Additionally, the research findings will have broad applications across various fields, serving as a valuable alert system for students, teachers, and concerned parents regarding academic progress. Ultimately, the study's results hold significant potential for The study aimed to enhance both educational practices and related industries by enabling administrators to proactively identify students at high risk of not graduating and take corrective measures at an early stage. A comparative analysis was conducted on two algorithms, Bayes and Artificial Neural Network, using student data to predict graduation outcomes in tertiary institutions.

Reference

1. E. Parcham, N. Mandami, A. N. Washington, and H. R. Arabnia, ''Facial expression recognition based on fuzzy networks,'' in Proc. Int. Conf.Comput. Sci. Comput. Intell. (CSCI), Dec. 2016, pp. 829-835.

2. A. P. Tafti, A. Baghaie, M. Assefi, H. R. Arabnia, Z. Yu, and P. Peissig,‘‘OCR as a service: An experimental evaluation of google docs OCR, tesseract, ABBYY finereader, and transym,’’ in Proc. Int. Symp. Vis.Comput., in Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 10072.Springer, 2016, pp. 735-746.
3. S. Amirian, Z. Wang, T. R. Taha, and H. R. Arabnia, ‘‘Dissection of deep learning with applications in image recognition,’’ in Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI), Dec. 2018, pp. 1132-1138.
4. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25,F. Pereira, C. J. C. Burges, L. Bottou, and K.Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012,pp. 1097-1105.
5. M. Regneri, M. Rohrbach, D. Wetzel, S. Thater, B. Schiele, and M. Pinkal, ‘‘Grounding action descriptions in videos,’’ Trans. Assoc. Comput. Linguistics, vol. 1, pp. 25-36, Dec.2013.
6. K. Xu, J. Ba, R. Kiros, K. C. ho, A. CourvilleR.Salakhudinov, R. Zemel, and Y. Bengio, ‘‘Show, attend and tell: Neural image caption generation with visual attention,’’ in Proc. Int. Conf. Mach. Learn., 2015, pp. 2048-2057.