Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Evaluating and Forecasting Student Academic Performance through Educational Data Mining using Deep Learning

Mr. Andhale Rajendra[1], Prof. Dr. Monika Rokade[2], Prof. Dr. Sunil Khatal[3]

[1]Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

randhale.shivneri@gmail.com

[2]Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

monikarokade4@gmail.com

[3]HOD, Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

khatalsunil88@gmail.com

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

Frequent failures during various stages of education are a common occurrence. Numerous factors contribute to the increasing drop-out rates among students, with poor academic performance being a significant reason for school discontinuation. Many students struggle to adapt to the academic environment of their institutions, which negatively impacts their performance. Additionally, involvement in extracurricular activities and student politics often diverts focus, leading to unsatisfactory outcomes. These predictable and unforeseen factors collectively influence academic growth and development. Consequently, it is essential to analyze undergraduate performance to uncover the underlying causes of students' varying achievements. The main objective of our study is to detect the diverse factors that impact academic success at the undergraduate level. The ultimate goal of this study is to empower students by helping them understand these influencing factors, enabling them to take proactive steps to enhance their academic results. By identifying and assessing these key elements, students, educators, and institutional stakeholders can work collaboratively to create a more conducive learning environment. This paper emphasizes the critical role of utilizing student data in improving education planning. It outlines effective techniques to extract valuable insights from extensive academic databases containing student information. A deep learning-based Recurrent Neural Network classifier model is suggested to help make early predictions about how well students will do in school. The proposed approach is associated against many traditional machine learning classification models and RNN classifiers to evaluate its efficacy.

Keywords: machine learning, NLP, deep learning and Educational Data Mining (EDM)

I. INTRODUCTION

In the present era, educational sectors started focusing on improving students experience in learning. Apart from students, the other stakeholders involved in educational enterprises are parents, community and employers. The stake holders want the educational institution to give best education

to the students. The academic performance of students has long been a subject of interest for researchers and educational institutions, tracing back to the earliest efforts in education management. Academic performance is commonly measured using indicators such as grade point averages (GPAs), individual test scores, success in specific modules, or other quantifiable metrics captured at distinct intervals.

Predicting student performance has also emerged as a critical area of study, particularly with the advent of Educational Data Mining (EDM). This predictive capability is instrumental in enhancing the outcomes of higher education by identifying students' strengths, guiding them toward areas where they excel, and mitigating performance declines in challenging subjects. EDM offers a systematic approach to analyzing educational data, involving iterative processes of hypothesis generation and testing.

At its core, EDM integrates three fundamental disciplines: computer science, education, and statistics. These domains intersect to form additional subfields closely related to EDM, including computer-based education, data mining, machine learning, and learning analytics. These interdisciplinary approaches collectively drive advancements in understanding and improving educational systems through data-driven insights.

II. LITERATURE SURVEY

In [1] 2019, researchers can explore extensive educational datasets to uncover valuable insights within the field of Educational Data Mining. This domain offers powerful tools to globally enhance student engagement and participation in the learning process. By leveraging EDM and various supervised learning techniques, the research evaluates the effectiveness of multiple prediction models. The findings highlight strategic approaches that can support decision-making processes aimed at reducing student dropout rates and improving academic retention.

According to [2], developed education institutions often place significant emphasis on evaluating the overall success rates of their graduates. Accurately forecasting student achievement requires a combination of approaches, including physical assessments, statistical analyses, and advanced data mining (DM) methods. Educational Data Mining is a growing field that leverages data mining techniques alongside statistical and machine learning (ML) methodologies. These tools enable the analysis of students' study patterns, academic outcomes, and potential growth trajectories. This research explores various data mining approaches to assess and predict students' quality standards. Utilizing the Kalboard 360 dataset, decision-making methods were analyzed using the Weka platform, offering valuable insights into academic performance forecasting.

In [3] 2020, the world is generating an abundance of knowledge; however, the educational system remains unaffected by it. The compilation of data requires scrutiny due to the increasing emphasis on education. Data analysis can extract relevant information from educational datasets and reveal significant correlations among various aspects. Educational data analysis assesses student information to predict a student's behaviour or examination outcomes. Several machine learning techniques govern the assessment of educational data in this study. The primary objective is to analyse the application of various machine learning approaches to educational data for predictive purposes.

In [4] Educational Data Mining employs machine learning, data mining, and mathematical techniques to analyse many types of educational data. An application, known as a learner simulation, proposes or modifies educational system resources. This study analyses the characteristics of more than a few educational systems. The text addresses affective state modelling, predicting academic accomplishments for students, and modelling individual learning styles. The text also discusses student profiling, categorisation, and collaborative analysis.

A [5] toolkit, techniques, and methodologies for study design, educational data mining (EDM) may promptly identify patterns and correlations within extensive data sets gathered from educational environments. The vast amount of information in extensive educational databases has made the prediction of student success a significant challenge. Moreover, Learning Outcome Assessments (LOA) are crucial for evaluating and enhancing instructional quality, as well as for guiding the growth of individual students. They employ regression and several machine learning techniques to develop learning systems capable of accurately predicting a student's GPA. We also used several attribute evaluator methods to identify the factors that significantly impact a student's overall performance.

In [6] 2021, data mining and machine learning techniques have witnessed remarkable progress over the past two decades. These advancements hold significant potential to deepen academic insights into individual learning behaviors across diverse educational settings. They compare artificial neural networks and random forest algorithms for guessing how well students will do in school by looking at test scores and demographic information. The findings reveal that the ANN model outperformed the RF algorithm, achieving an accuracy of 91.08% compared to 81.35%. The results underscore the superior performance of ANN in predicting academic success and its utility in establishing early intervention systems..

In [7] 2019, the vast volume of records in educational datasets might hinder the production of high-quality data. Numerous experts in the education sector are now using the DM approach for data analysis. Consequently, completing the categorisation of the dataset requires considerable processing time, owing to its substantial computing needs. This article presents an overview of the approaches used for feature selection in the assessment of data characteristics. The proposed hybrid approach enhances the quality of students' data gathering by integrating feature extraction with wrapper-based techniques.

In [8] 2020, making appropriate decisions in university academic processes significantly enhances the quality of education and may be very beneficial for students, faculty, and the whole academic community. The study shows a decision support technology that can make learning systems more useful for people who have to make choices by giving accurate feedback, good choice help, and the ability to keep an eye on things and make plans. We achieve this by utilising a variety of machine learning approaches. The results indicate that it may predict student success, which facilitates decision-making.

In [9] Daily records of student interaction with VLEs are collected; hence, automated systems are anticipated to manage and assess this substantial volume of data. Students, educators, administrators, and other stakeholders involved in the learning activities of VLEs can use the insights obtained from

educational data. The application of machine learning methodologies can yield valuable information. Static machine learning algorithms have historically analysed educational datasets. However, because educational datasets are inherently non-stationary, managing them as data streams is the most effective approach. The vast volume of data contained within educational databases presents a significant challenge in predicting student performance, as noted in [10]. We suggest a thorough literature review to deal with this level of complexity, focussing on using data mining techniques, especially the C4.5 algorithm, to predict how students will do in school. The primary aim of this research is to explore data mining methodologies used for predicting academic achievement. This study specifically examines the application of classification techniques to organise and analyse critical aspects of student data. We can enhance student performance and support broader academic growth by leveraging educational data mining tools, particularly the C4.5 classifier. The findings indicate that the C4.5 algorithm achieved an efficiency rate of 71.9%, based on the analysed dataset. These results highlight the potential benefits for students, educators, and academic institutions in improving educational outcomes and driving meaningful advancements.

III. PROPOSED SYSTEM DESIGN

Our proposed study employs a deep learning approach, utilising an RNN-based classification algorithm to predict students' academic performance. This methodology incorporates both synthetic datasets and real-time student data to ensure comprehensive analysis. For a detailed explanation of the proposed RNN classifier design, refer to the accompanying figure. Figure 2 illustrates the architecture of the proposed system, which is centred around an RNN classifier. The data collection process involved aggregating information from various sources, including web applications, real-world student records, and synthetic datasets derived from diverse origins. This multifaceted approach ensures a robust foundation for accurate performance prediction and analysis..

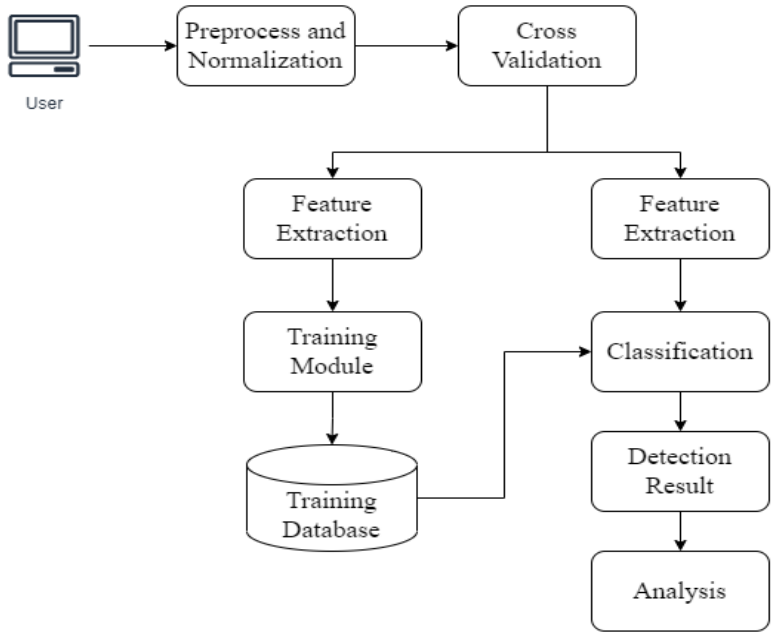

Figure 1: Proposed System Design

Implement Module

Data Collection

We collect data from multiple sources, such as the UCI Machine Learning repository, Kaggle, and various data streams. Prior to performing the classification task, it is essential to pre-process the data to optimise the outcomes. Efficient data handling is crucial to achieving the best possible results during the data mining process.

Pre-processing and normalization: There may be a lot of useless information and gaps in the data. This portion requires data preparation. We have used a variety of data pre-processing techniques at this step, including data cleansing, data transformation, and data reduction.

Data cleaning- When data is incomplete, various approaches can be employed, such as filling in the missing values or discarding incomplete entries. Null values in the data can create challenges, as they are often unrecognised by machine algorithms. This incomplete or "noisy" data may arise from issues like inaccurate data collection or incorrect data entry. To handle this, we apply techniques such as regression, clustering, and binning to clean and process the data effectively.

- **Data transformation-** This method converts the data into a suitable format for the mining process. It includes steps such as normalisation, feature selection, and discretisation to prepare the data effectively.
- **Stopword removal and stemming:** Next, we will implement several preprocessing techniques, including lexical analysis, removal of stopwords, stemming using Porter's algorithm, selection of index terms, and data cleaning, to ensure our dataset is properly prepared for analysis..

Feature extraction and Selection

This method extracts multiple attributes from the provided data. We apply a feature selection threshold to standardise the extracted features, eliminating redundant and irrelevant ones to enhance training efficiency. We use normalised data along with relational features to derive a set of hybrid attributes, followed by the selection of an optimisation strategy for training. We employ the hybrid approach to select high-quality features from the fully extracted data, thereby improving classification accuracy. During the feature extraction process, irrelevant attributes may appear, requiring their removal during selection. This approach tailors the feature selection process to each distinct feature set.

Classifier Models: Once the section has been successfully implemented, a selection of features is added to the training part, which develops complete background knowledge designed for the entire system. Once we have obtained the trained model, we can then input the test data to make classification predictions. The testing phase involves preparing the test data, vectoring it, and classifying the text. This phase assesses the system’s prediction accuracy using deep learning techniques such as RNN. The phase assesses the system's performance using various datasets.

RESULT AND ANALYSIS

The RNN model experiment gives metrics for accuracy, precision, recall, and F-scores for different cross-validation cycles. Figure 1.1 demonstrates the validation of the model using 5-fold cross-validation with the RNN classifier.

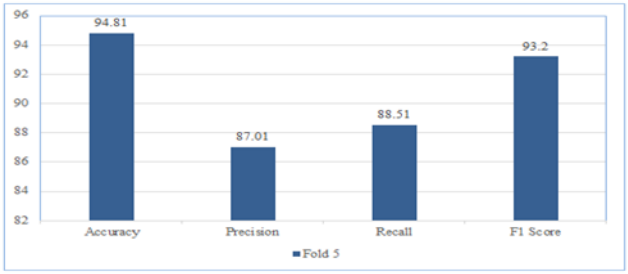

Figure 1.1: Model Validation Utilising 5-Fold Cross-Validation with an RNN Classifier

CONCLUSION

In this research proposed machine learning classification algorithms are used to predict student achievement in binary categories, such as pass or fail. On the other side, predicting success in terms of pass or fail does not provide a better insight of the accomplishment of pupils. These methodologies' failure to assess the total interrelationship of predictor elements in student data is another significant issue. Conventional ML classifiers can't accurately predict student performance based on academic information. Recently, supervised learning techniques helped academic institution to automatically retrieve high level features from the raw information. In the proposed research work, Long Short-Term Memory with Recurrent Neural Network is utilized to forecast student achievement on real-time heterogeneous data. With the developed RNN-LSTM, a comparison between Deep Learning classifiers and conventional machine learning methods is also performed.

REFERENCES

[1] Kelly J. de O. Santos, Angelo G. Menezes, Andre B. de Carvalho and Carlos A. E. Montesco. "Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout", 2019, 19th International Conference on Advanced Learning Technologies (ICALT), IEEE.

[2] Chitra Jalota and Rashmi Agrawal. "Analysis of Educational Data Mining using Classification", 2019, International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), IEEE.

[3] Dr. R. Raju, Mrs. N. Kalaiselvi, Aathiqa Sulthana M, Divya I and Selvarani A. "Educational Data Mining: A Comprehensive Study", 2020, IEEE.

[4] Nabila Khodeir. "Student Modeling Using Educational Data Mining Techniques", 2019, IEEE.

[5] Balqis Al Breiki, Nazar Zaki and Elfadil A. Mohamed. “Using Educational Data Mining Techniques to Predict Student Performance”, 2019, International Conference on Electrical and Computing Technologies and Applications (ICECTA), IEEE.

[6] Muhammad Sammy Ahmad, Ahmed H. Asad and Ammar Mohammed. “A Machine Learning Based Approach for Student Performance Evaluation in Educational Data Mining”, 2021, IEEE.

[7] Usman Ali, Khawaja Sarmad Arif and Dr. Usman Qamar. “A Hybrid Scheme for Feature Selection of High Dimensional Educational Data”, 2019, International Conference on Communication Technologies (ComTech 2019), IEEE

[8] Muhib Al-kmali, Hamzah Mugahed, Wadii Boulila, Mohammed Al-Sarem and Anmar Abuhamdah. “A Machine-Learning based Approach to Support Academic Decision-Making at Higher Educational Institutions”, 2020, IEEE

[9] Gabriella Casalino, Giovanna Castellano, Andrea Mannavola and Gennaro Vessio. “Educational Stream Data Analysis: A Case Study”, 2020, IEEE.

[10] Latifaestrelita Indi Pramesti Aji and Andi Sunyoto. “An Implementation of C4.5 Classification Algorithm to Analyze Student’s Performance”, 2020, 3rd International Conference onInformation and Communication Technology (ICOIACT), IEEE.