"Predictive Analytics in Education: Applying Machine Learning to Forecast Students Academic Performance"

Submitted By:

1. Ms. Palak Patel
   Research Scholar,
   SEMCOM, CVMU.

2. Dr. Tejas H. Thakkar
   Assistant Professor,
   NVPAS, CVMU.

Abstract

In educational research, student academic achievement has been a major focus. Early detection of students who are likely to perform poorly enables prompt interventions to enhance results. This study investigates how machine learning classifiers can be used to forecast student performance based on a variety of factors, such as engagement indicators, demographic information, and past academic performance. Various classifiers, such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Neural Networks, are compared. The outcomes demonstrate how well machine learning methods can forecast student performance and provide information to the teachers for personalized and proactive interventions. *The purpose of this research is to examine various ML algorithms for predicting the student performance based on different factors and identifying the most influencing factor.*

Key Words: *Educational Data Mining, ML Algorithms, Pattern of Data, Prediction Model*

INTRODUCTION

It is essential for educational institutions to forecast student performance in order to make sure that students are getting the help they need to thrive academically. Periodic tests or end-of-term grades are frequently used in traditional assessment systems, which do not give early indicators of students who are having difficulty. Machine learning techniques can be used to forecast performance results ahead of time, enabling teachers to intervene sooner and offer more specialized help. [1]

Machine learning (ML) technologies can forecast in a variety of disciplines, including academia, sales, marketing, engineering, medical, the stock market, and weather trends. ML can forecast many events based on past data or event. ML predictions are significantly more accurate than those made by humans. For predicting the future, ML requires a large amount of data. [2]

Machine learning assists in forecasting students' academic achievements and identifying concealed patterns within the data. Due to the ability to uncover these hidden patterns, machine learning can predict the elements influencing student performance. This technique is applied to predict student outcomes based on their historical performance records. [2]

The purpose of this work is to evaluate the efficacy of several machine learning algorithms in predicting student performance using available data such as demographic information, previous academic performance, and student behavior. We will look at the following research questions:

- How accurate are machine learning algorithms in predicting student performance?
- Which features most strongly influence predictions of student performance?
- What are the implications of these findings for students and teachers?

REVIEW OF LITERATURE

In the field of EDUCATIONAL DATA MINING (EDM), Machine learning has been applied to predict student results since many years. During the studies, various types of classifiers, such as Decision Trees (C4.5), Random Forests, Support Vector Machines (SVM), and Deep

Learning models have been used, to predict performance indicators such as grades, dropout rates, and final scores. [3]

In the year 2014, Delen et al. used SVM to classify students based on early indicators of academic success. [4]

Kumar et al. (2020) used Random Forests to predict students' final grades based on demographic and behavioral features. In terms of prediction accuracy, they proposed that Random Forests perform better than other classifiers, including Logistic Regression and Naïve Bayes. [5]

Perkash et al. (2024) utilized a balanced dataset and optimized feature set with several ML models. Decision Tree classifier achieved highest accuracy of 99.06%. The study was restricted to the dataset from one higher educational institute and may not generalize to other institutions. [6]

Zhang et al. (2024) employed SVM, CNN-CBAM, XGBoost, CART. They achieved 83.64% accuracy. They integrated psychological and academic data. Study was limited to mechanical engineering students and may not apply to other disciplines. [7]

Few studies have compared a wide range of classifiers on the same dataset, and many have concentrated on small or limited datasets despite notable advancements. By contrasting several machine learning classifiers on a wide range of student-related attributes, this study seeks to close that gap.

METHODOLOGY

Data Collection

For this study, the data were collected from different university's students. The dataset consists demographic information as age, gender, socioeconomic status, academic performance like previous results, and engagement metrics - attendance, participation in online forums, etc.

- **Features**:

    - Demographic: Age, Gender, Socio Economic Status
    - Academic Performance: Previous Results
    - Study Habits: Study Methods, Study Duration
    - Engagement: Attendance, Participation in Class, Participation in curricular and co-curricular Activities
    - Behavioural and Psychological Factors: Motivation, Stress Levels, Parental Involvement and Support
    - Health: Physical and Mental Health
    - Peer Influence: Peer Pressure, Peer Support , Peer Influence

Data Preprocessing

Data preprocessing is an important process in ML. [8] The dataset was cleaned to deal with missing values, remove outliers, and to normalize continuous features. One-hot encoding is used to encode the categorical features. Feature selection techniques were applied to identify the most influence features.

Classifiers [9]

In this study, following machine learning algorithms, classifiers are utilized for predicting student performance:

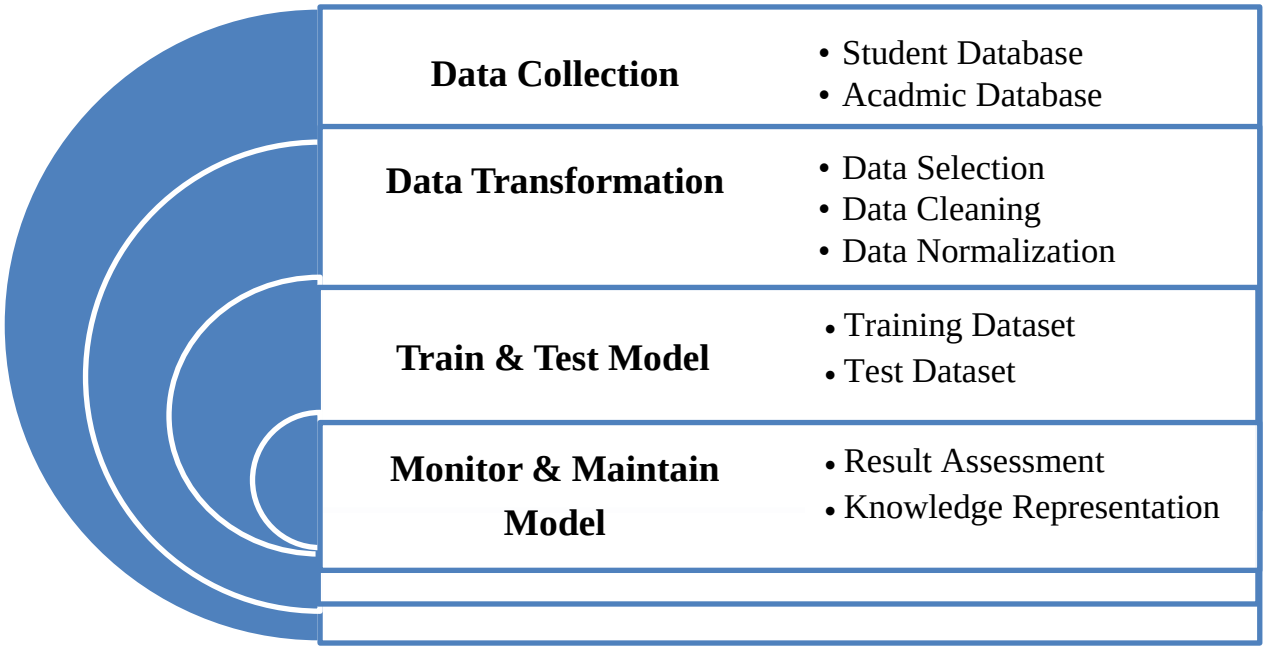
- **Decision Trees**: It splits the data based on feature thresholds.
- **Random Forests**: A method that combines multiple decision trees to improve accuracy.
- **Support Vector Machines (SVM)**: A model works well for both linear and non-linear data by transforming the feature space.
- **Neural Networks**: NN is a deep learning technique that employs back propagation to learn from the data using interconnected nodes (neurons).

Evaluation Metrics

The performance of the classifiers was evaluated using:

- **Accuracy**: The proportion of cases that were accurately predicted.
- **Precision, Recall, and F1-Score**: Metrics that work well in unbalanced datasets.
- **ROC-AUC**: An indicator of the classifier's class distinction ability is the area under the receiver operating characteristic curve.

Proposed Framework for Predicting Student's Academic Performance



RESULTS

Classifier Performance

Table 1 : Performance of the classifiers based on accuracy and other metrics.

Classifier	Accuracy (%)	Precision	Recall	F1-Score	AUC
Decision Trees	82.5	0.78	0.81	0.79	0.85
Random Forests	85.4	0.81	0.84	0.82	0.89
Support Vector Machines	80.3	0.76	0.79	0.77	0.84
Neural Networks	88.2	0.84	0.86	0.85	0.93

Influencing Factor

Using optimization technique for Random Forest and Decision Tree, the following influencing factors, affecting student academic performance are identified:

1. Previous Results
2. Engagement – Attendance and Participation in Activities
3. Peer Pressure and Influence
4. Socioeconomic status

DISCUSSION

The findings imply that machine learning classifiers are capable of accurately and successfully predicting student performance. The most accurate classifiers tested were Neural Networks (88.2%), followed by Random Forests (85.4%). Both models demonstrated the advantages of using ensemble and deep learning approaches for this kind of task by outperforming conventional Decision Trees and Support Vector Machines.

Academic history (GPA, grades) and student engagement (attendance, participation) were found to be the most significant predictors of student performance, according to the feature importance analysis. This result is consistent with earlier studies that highlight the significance of early academic performance and engagement in predicting future success.

CONCLUSION

This study shows how machine learning techniques can be used to accurately predict student performance. The most successful classifiers among those tested were Random Forests and Neural Networks. Teachers can use these findings to more successfully implement interventions and identify students who are at risk earlier. Future research might examine hybrid models, which combine several classifiers for improved prediction accuracy, and incorporate additional data sources, like social media activity or mental health indicators.

REFERENCES

1. Data Mining Tutorial, javaTpoint, https://www.javatpoint.com/data-mining
2. Shivam Arora (2023). "Data Mining Vs. Machine Learning: The Key Difference", https://www.simplilearn.com/data-mining-vs-machine-learning-article#what_is_machine_learning
3. C. Romero and S. Ventura (2010). ''Educational data mining: A review of the state of the art,'' IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601–618, Nov. 2010, https://dl.acm.org/doi/10.1109/TSMCC.2010.2053532
4. Delen, D., et al. (2014). Predicting student performance: An application of data mining methods. *Journal of Educational Data Mining*, *6*(2), 1-18.
5. Kumar, M., et al. (2020). A Random Forest approach to predict student performance in university. *Journal of Educational Data Mining*, *12*(1), 32-46.
6. Perkash, A., Shaheen, Q., et al. (2024). Feature optimization and machine learning for predicting students' academic performance in higher education institutions. Education and Information Technologies, 1-25.
7. Zhang, T., Zhong, Z., Mao, W., Zhang, Z., & Li, Z. (2024). A New Machine-Learning-Driven Grade-Point Average Prediction Approach for College Students Incorporating Psychological Evaluations in the Post-COVID-19 Era. Electronics, 13(10), 1928.
8. Data Mining Process: Models, Process Steps & Challenges Involved, Software Testing Help, https://www.softwaretestinghelp.com/data-mining-process/
9. analytics steps, Bhumika Dutta, 2022, https://www.analyticssteps.com/blogs/types-classifiers-machine-learning

10. Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215
11. Kumar, Vaibhav, et al. "A Machine Learning Approach For Predicting Onset And Progression"“Towards Early Detection Of Chronic Diseases “." Journal of Pharmaceutical Negative Results (2022): 6195-6202
12. Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.
13. Abideen, Z. U., Mazhar, T., Razzaq, A., Haq, I., Ullah, I., Alasmary, H., & Mohamed, H. G. (2023). Analysis of Enrollment Criteria in Secondary Schools Using Machine Learning and Data Mining Approach. Electronics, 12(3), 694.
14. Mehdi, R., & Nachouki, M. (2023). A neuro-fuzzy model for predicting and analyzing student graduation performance in computing programs. Education and Information Technologies, 28(3), 2455-2484.
15. Li, L., Yu, H., & Kunc, M. (2024). The impact of forum content on data science open innovation performance: A system dynamics-based causal machine learning approach. Technological Forecasting and Social Change, 198, 122936.
16. Adane, M. D., Deku, J. K., & Asare, E. K. (2023). Performance Analysis of Machine Learning Algorithms in Prediction of Student Academic Performance. Journal of Advances in Mathematics and Computer Science, 38(5), 74-86.
17. Baig, M. A., Shaikh, S. A., Khatri, K. K., Shaikh, M. A., Khan, M. Z., & Rauf, M. A. (2023). Prediction of Students Performance Level Using Integrated Approach of ML Algorithms. International Journal of Emerging Technologies in Learning, 18(1).
18. Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. Education and Information Technologies, 1-30.
19. Masabo, E., Nzabanita, J., Ngaruye, I., Ruranga, C., Nizeyimana, J. P., Uwonkunda, J., & Ndanguza, D. (2023). Early detection of students at risk of poor performance in Rwanda higher education using machine learning techniques. International Journal of Information Technology, 1-10.
20. Gilholm, P., Gibbons, K., Brüningk, S., Klatt, J., Vaithianathan, R., Long, D., ... & Schlapbach, L. J. (2023). Machine learning to predict poor school performance in

paediatric survivors of intensive care: a population-based cohort study. Intensive Care Medicine, 1-11.

21. Dai, C. P. (2024). Applying machine learning to augment the design and assessment of immersive learning experience.In Machine Learning in Educational Sciences: Approaches, Applications and Advances (pp. 245-264). Singapore: Springer Nature Singapore.