

Predicting Student Academic Performance Using Learning Vector Quantization And Probabilistic Neural Network

OMKAR DESHMUKH¹, YASH MISHRA², SWETA NIGAM³

^{1,2}*Master of Computer Application, University of Mumbai, Aditya Institute of Management Studies and Research*

³*Professor, University of Mumbai, Aditya Institute of Management Studies and Research*

Abstract- Early prediction of student performance is an important task in the field of education, as it helps teachers identify the performance of the students and provide them with the required academic support. This paper proposes a framework based on Artificial Neural Network (ANN) techniques, namely Learning Vector Quantization (LVQ) and Probabilistic Neural Network (PNN), for the classification of the students based on the behavioral, academic, and demographic features of the students. The performance of the students is determined by using the historical data of the students. The proposed framework is based on the Streamlit platform, which is used for the visualization of the performance of the students.

Index Terms- Student Performance Prediction, Machine Learning, Data Preprocessing, Feature Engineering, Predictive Modeling, Classification.

I. INTRODUCTION

The increasing adoption of data analytics and machine learning techniques in the education sector has provided new opportunities for improving learning outcomes. Predicting the academic performance of students is an essential task in the education sector. The datasets in the education sector have various attributes such as demographic features, academic records, and behavioral characteristics, which affect the academic performance of the students. However, analyzing the complex datasets and extracting useful insights is still an open problem.

The proposed research suggests a machine learning framework for predicting the academic performance of the students using Artificial Neural Network models. The research work is based on the application of two classification techniques, namely, Learning Vector Quantization and Probabilistic

Neural Network, for analyzing the historical datasets of the students and classifying the academic performance of the students into various grade levels.

The proposed research work has been tested using the UCI Student Performance datasets, and the results show that the neural network models can effectively analyze the datasets of the education sector and provide reliable predictions.

The proposed framework is useful in the education sector for improving academic planning and supporting early intervention strategies to enhance the academic performance of the students.

II. LITERATURE REVIEW

Many researchers have worked on the idea of using machine learning in predicting student academic performance, and data sets are available for educational data mining. The availability of data has increased, and data sets include academic performance, demographic information, and behavioral data. Data mining and machine learning are found to be significant in analyzing student performance and identifying factors related to academic performance. Researchers have used traditional statistical approaches and machine learning algorithms like Logistic Regression, Decision Trees, and Support Vector Machines, and the results are promising in classifying performance levels.

Recent studies have used advanced machine learning algorithms like Random Forest and Artificial Neural Networks to improve prediction accuracy and handle complex relationships between variables in academic performance data. The neural network model is found

to be effective in handling non-linear relationships in educational data mining problems. The neural network model analyzes multiple attributes and reveals hidden patterns in data, unlike traditional statistical approaches.

Recent research has focused on the application of classification algorithms such as Learning Vector Quantization and Probabilistic Neural Networks, which are based on neural network approaches to prediction analysis. Learning Vector Quantization, which is a supervised learning method, uses prototype vectors to represent each class and improves them for better accuracy. Probabilistic Neural Networks are based on Bayesian classification theory and use kernel density estimation for accurate classification with relatively fast training.

Studies have indicated that neural network approaches are superior to traditional approaches for prediction analysis on educational data. However, existing approaches are limited with respect to feature selection, interpretation, and scalability, particularly with large datasets. This again emphasizes the need for further research to develop effective prediction approaches that can analyze the data related to students with high accuracy and provide valuable insights to educators.

The recent literature emphasizes the need to incorporate a prediction system based on machine learning approaches to improve the academic planning and decision-making process.

III. PROBLEM STATEMENT

Despite the rapid progress being made in the fields of machine learning and data mining in the educational sector, predicting academic performance remains an uphill task. Educational institutions generate vast amounts of data, and academic performance is no exception. However, analyzing these data and predicting academic performance remains an uphill task due to the complexity and number of factors that influence academic performance. Conventional evaluation practices are based on periodic exams and analyzing the data, but they fail to identify early symptoms of poor academic performance.

The educational scenario requires intelligent systems that can analyze vast amounts of data and identify

students in need of early intervention. There are several factors that influence academic performance, and they include study habits, educational levels of parents, social activities, attendance, and lifestyle. These factors are difficult to analyze due to their non-linear nature, and conventional statistical techniques are not effective in analyzing such data. Therefore, there is a need to develop sophisticated machine learning techniques that can analyze these data and accurately predict academic performance.

Moreover, educational institutions require predictive systems that will enable teachers to monitor students' progress and take necessary interventions to improve students' performance. By applying machine learning techniques such as Learning Vector Quantization and Probabilistic Neural Networks, it is possible to analyze historical data to classify students' performance in a relevant manner. Hence, a reliable predictive system is necessary to support educational institutions.

IV. OBJECTIVE OF STUDY

This paper aims to examine how machine learning techniques are employed in predicting student academic performance using educational data. As student-related data becomes more accessible in academic environments, the need for predictive systems that consider academic and behavioral factors is significant in decision-making processes. The research aims to explore different neural network techniques and their efficiency in identifying patterns that influence student academic performance. The research aims to achieve the following objectives:

- Analyze the factors that influence student academic performance by analyzing educational data such as study habits, parental educational level, attendance, lifestyle, and academic achievement.
- Implement and evaluate classification algorithms in machine learning, particularly Learning Vector Quantization and Probabilistic Neural Networks, in predicting performance levels based on historical data.
- Evaluate the efficiency of the proposed system using standard evaluation parameters such as

accuracy, Cohen's Kappa, and Confusion Matrix in predicting academic performance.

- Identify key patterns and trends in educational data that can be used to identify underperforming students and provide academic support.
- Design a predictive system that helps educational institutions track student performance, improve academic planning, and support early interventions.

V. PROBLEM STATEMENT

A. Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) is a supervised neural network classification technique that uses prototype vectors to classify each class in the given dataset. In the context of predicting student performance, LVQ uses various aspects such as academic, behavioral, and demographic factors to classify students belonging to certain classes of performance levels. The technique compares the feature vectors with prototype vectors for each class to improve the classification accuracy.

LVQ has several benefits, including the ability to define clear decision boundaries between classes of student performance. During the training process, if the predicted class matches the actual class, the prototype vector will be moved closer to the input vector. In cases where the predicted class does not match the actual class, the prototype vector will be moved away from the input vector. The process will continue to improve the classification ability for new student feature vectors. LVQ has an easy architecture, which makes it suitable for multi-class classification problems like predicting student grade levels.

LVQ's performance is dependent upon the quality of prototype vectors used in the classification process. Poor learning parameters can affect classification accuracy, especially when dealing with complex problems involving multiple factors.

B. Probabilistic Neural Network (PNN)

Probabilistic Neural Networks are feed-forward neural networks that are developed based on Bayesian classification theory and kernel density estimation. PNN models are commonly used for pattern recognition and classification problems due to their fast-training speed and solid probabilistic foundation. In the context of predicting student performance, PNN models evaluate student characteristics to determine the probability that the student belongs to a given performance group.

A PNN has four layers: the input layer, pattern layer, summation layer, and output layer. The input layer receives feature vectors representing student characteristics, the pattern layer compares the input feature vector with the feature vectors of the training samples, the summation layer sums up the probabilities for each class, and the output layer produces the final decision based on the highest probability.

The advantages of PNN include the fact that it can classify with speed and little training compared to other neural networks. PNN can also produce reliable classification results regardless of the complex patterns and non-linear relationships between variables in the dataset. However, the performance of PNN depends on the optimal kernel parameter setting to accurately compute the probability.

C. Comparison of LVQ and PNN for Performance Prediction

Both Learning Vector Quantization and Probabilistic Neural Networks are very effective in predicting student performance. LVQ provides clear prototype-based classification, which can be very helpful in understanding the patterns in each type of performance. On the other hand, PNN has excellent probabilistic modeling capabilities and faster learning.

In this study, both techniques will be used to evaluate the effectiveness of the student performance dataset in identifying patterns in the data, which can be very helpful in predicting the performance of students. Comparing these two models can be very helpful in understanding their advantages and disadvantages, which can be very helpful to researchers in choosing

the right machine learning technique to be used in the analysis.

VI. SYSTEM ARCHITECTURE

The proposed architecture for predicting student academic performance involves processing educational data, training machine learning models, and predicting academic performance based on student attributes. The process begins with collecting historical student data, which includes student attributes, academic performance, and behavioral attributes. The data is then preprocessed, removing any incomplete data and converting categorical data into numerical data.

After preprocessing, the data is then fed into feature engineering, where relevant attributes are generated that can be used to enhance the predictive ability of the machine learning model. The data is then divided into training and testing data. The training data is used to train the machine learning model, such as the Learning Vector Quantization and Probabilistic Neural Network, while the testing data is used to evaluate the performance of the model.

After training the model, the system evaluates the relationship between different student attributes and academic performance categories. The system also uses the trained model to predict the performance of new students based on their academic and behavioral attributes. The system also includes an interface that enables users to input data and obtain real-time predictions.

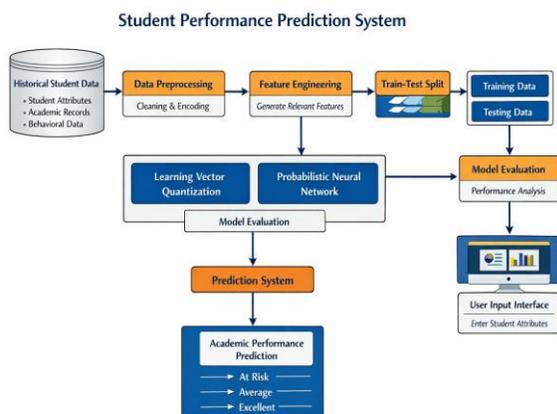


Fig. 1. Architecture of Student Performance Prediction System

In this architecture, the machine learning models act as the primary component that deals with the processing of student data. With the classification results, educators will be able to identify students who may require additional academic support. With the integration of the data preprocessing component, the training component, and the prediction component, the proposed system provides an efficient way of analyzing student performance.

VII. COMPARATIVE ANALYSIS OF ENCRYPTION TECHNIQUES

Machine learning algorithms are fundamental for the prediction of academic performance for students by analyzing the educational data and recognizing patterns that influence academic performance. This study examines two classification algorithms based on the neural network model, Learning Vector Quantization and Probabilistic Neural Networks, with emphasis on prediction accuracy, training speed, and reliability for prediction purposes.

Learning Vector Quantization enables the model to use vectors to represent each class of academic performance. This makes it easier to interpret results and applicable to multi-class classification. However, its accuracy might be limited to the use of the appropriate prototype vectors.

Probabilistic Neural Networks use Bayesian decision theory and probability density estimation for classification. PNNs are known to be faster to train and can be effective for complex nonlinear relationships. They are usually more accurate by a small percentage due to their probabilistic nature.

Table 1. Comparative Analysis of Machine Learning Techniques for Student Performance Prediction

Technique	Security Level	Performance Accuracy	Scalability	Key Limitation
Learning Vector Quantization (LVQ)	Medium	High	Medium	Sensitive to initial weights and needs to be tuned
Probabilistic Neural Network (PNN) High	High	Medium	Medium	High computational cost and memory usage
Decision Tree	Medium	High	High	Tends to overfit with complex data
Random Forest	High	Very High	Medium	Increases training time and complexity
Support Vector Machine (SVM)	High	High	Low	Not suitable for large data
Artificial Neural Network (ANN)	Very High	Very High	Medium	Requires large data, long training time

In summary, both techniques demonstrate high potential for prediction of academic performance. Although LVQ offers a comprehensible classification method with high learning efficiency, PNN offers a reliable model for probabilistic prediction and classification. The choice of method may be determined by the characteristics of the dataset, computational power, and the prediction system.

VIII. PERFORMANCE ANALYSIS

The performance of the proposed student performance prediction system is evaluated based on standard evaluation parameters of machine learning. The evaluation of the performance of the proposed system is essential in identifying the efficiency and accuracy of classification-based prediction approaches for student academic performance. In the proposed system, the performance of the proposed models is evaluated based on parameters such as accuracy, Cohen's Kappa, and confusion matrix evaluation.

The evaluation of the proposed system involves training the proposed Learning Vector Quantization and Probabilistic Neural Network-based prediction models using the training data and analyzing their predictive ability using test data. The accuracy of the proposed system evaluates the number of correct predictions, and the confusion matrix evaluates the number of correct and incorrect classifications for each performance level. Cohen's Kappa score

evaluates the agreement between the prediction and actual values, considering the chances of random prediction.

Fig. 3 below shows the comparative results of the machine learning models used in this study. From the results, the machine learning models can predict patterns in educational datasets with reliable accuracy. Moreover, the results show that the Probabilistic Neural Network model has a better prediction accuracy compared to the Learning Vector Quantization model, as it is based on probability.

The results also show that the Learning Vector Quantization model can classify the datasets by using prototypes, as it can identify patterns for various categories of student performance. Even though the results show that the machine learning models have better prediction accuracy, the results may differ based on the size of the datasets and the parameters used.

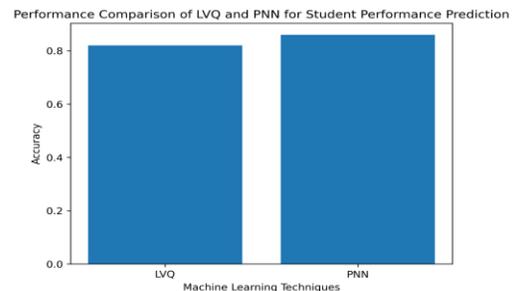


Fig. 2 Performance Comparison of LVQ and PNN for Student Performance Prediction

IX. CHALLENGES AND LIMITATIONS

Future research in predicting student performance should be geared towards improving the precision, reliability, and efficiency of machine learning algorithms in educational data analysis. The application of hybrid machine learning algorithms and feature selection methods could be beneficial in improving the performance of prediction algorithms. The use of real-time educational data from learning management systems and online learning platforms could also be beneficial in improving the efficiency of prediction algorithms in identifying underperforming students.

Despite the positive findings, there are still some challenges that need to be addressed in the application of machine learning algorithms in real-world educational data analysis. The main challenge lies in the availability and quality of data, where incomplete data could affect the performance of prediction algorithms. In educational data analysis, data could be incomplete, imbalanced, and even irrelevant, which could reduce the performance of prediction algorithms.

Another challenge lies in the ability of educators and administrators to understand the prediction model, where neural network-based prediction algorithms such as Learning Vector Quantization and Probabilistic Neural Networks could be beneficial in educational data analysis. These prediction algorithms are effective in analyzing data relationships, but they could be difficult to understand.

Furthermore, the selection of features and the proper setting of the parameters are important factors that affect the quality of the predictions. If not properly addressed, it may result in overtraining, which may affect the generalization performance of the model when it is used to predict new student data.

Some of the main issues affecting student performance prediction systems include those affecting the quality of the data, the interpretability of the models, scalability, and the fact that the models require updating to fit the changing environment.

To come up with reliable systems, these issues will have to be addressed.

X. FUTURE RESEARCH AND DIRECTIONS

Future research into the prediction of student performance should aim to improve the efficiency, precision, and application of machine learning techniques. As the use of digital platforms and management tools becomes more widespread, they will produce large amounts of student data. By fully leveraging this data with the right machine learning techniques, it is possible to create tools that produce more precise and timely predictions of student performance.

Some of the key research avenues that should be explored in the future include:

- Hybrid machine learning models. Developing hybrid machine learning models by combining the benefits of different algorithms could lead to more precise and reliable prediction outcomes. For instance, the combination of neural network architectures and other machine learning algorithms, such as Random Forest or Gradient Boosting, could lead to more reliable student performance prediction outcomes.
- Enhanced feature engineering. As a future direction, there is a need to focus more on the extraction of more informative features from educational datasets. Features such as student engagement, class participation, and ongoing assessment results could lead to more informative outcomes concerning student performance.
- Explainable artificial intelligence. Another significant future direction for machine learning and AI in the context of educational settings is the use of more interpretable machine learning models. By using more interpretable machine learning models, educators could build more trust and understanding of the factors influencing machine learning model outcomes.
- Real-time performance monitoring systems. Another significant future direction for machine learning and AI in the context of educational settings is the use of systems for

monitoring student data and generating warnings for students likely to perform poorly. Such systems could aid educators in the effective implementation of intervention programs for poorly performing students.

- Integration with educational platforms. As a future direction, there is a need to ensure the systems developed for machine learning and AI incorporate the use of learning management systems and other academic databases to aid educators in the provision of more effective analyses of student performance.

XI. CONCLUSION

The prediction of the academic performance of the students is now considered an important research domain in the field of EDM and LA. The educational institutions have access to a large amount of information regarding the academic performance of the students, as well as the behaviour and demographic details of the students. The application of machine learning algorithms on the information helps in the prediction of the academic performance of the students.

The present research paper proposes a predictive framework using Artificial Neural Network models such as Learning Vector Quantization and Probabilistic Neural Network. These models can analyse the complex relationships existing in various factors of the students and classifying them into different categories based on the performance. The results of the research show that the application of machine learning algorithms is effective in managing the educational information with reliable predictions.

The results obtained from the experiments show that both LVQ and PNN models work efficiently for multi-class classification. However, PNN works slightly better than LVQ in terms of prediction accuracy. LVQ helps the educator classify the data based on the patterns related to different levels of student performance.

The proposed system shows that the use of machine learning algorithms has a high potential to assist in the early prediction of students who need academic

support. The future possibilities for improving the accuracy of the prediction system for student performance include enhancements related to data improvement, feature engineering, and model interpretability.

XII. ACKNOWLEDGMENT

The authors would like to thank the Department of Master of Computer Applications (MCA) of Aditya Institute of Management Studies and Research (AIMSR) for their provision of essential academic resources and support, which are pertinent to the research. In addition, the authors would like to express their gratitude to Prof. Sweta Nigam, a faculty member of the MCA Department of AIMSR, Mumbai, for their valuable guidance and support during the research on the prediction of student academic performance using machine learning techniques

REFERENCES

- [1] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," UCI Machine Learning Repository, 2008.
- [2] T. Kohonen, "Learning Vector Quantization," *Neural Networks*, vol. 1, no. 3, pp. 303–305, 1990.
- [3] D. F. Specht, "Probabilistic Neural Networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.
- [4] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [5] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 40, no. 6, pp. 601–618, 2010.
- [6] R. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San

Francisco, CA, USA: Morgan Kaufmann, 2012.

- [8] S. Kotsiantis, "Use of Machine Learning Techniques for Educational Data Mining: A Survey," *Artificial Intelligence Review*, vol. 37, no. 4, pp. 331–344, 2012.
- [9] Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.
- [10] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.