

COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR THE PREDICTION OF ACADEMIC PERFORMANCE OF STUDENTS USING DATA ANALYTICS

Dr. Bharati Kawade

Computer Application Department, MITACSC/ SPPU Pune, India

ABSTRACT: Education is very important in shaping the future growth of the students. Educational institutions are taking efforts to provide the quality education. Faculty members play key role in the academic performance and development of the students. Educational field contains large amount of data that needs to be used for predicting the academic performance of the students. Data analytics need to be used effectively to assess the students' academic performance. It is helpful to the faculty members in improving teaching learning process. For the present research work, data is collected from UG (BCom, BBA, BBA(CA), BSc. (CS)) students. The prediction of student academic performance is done using various machine learning algorithms through classification technique. Different supervised machine learning algorithms are applied for the present study. The comparison of the algorithms is represented graphically. The interpretations of research findings of the present study are also discussed.

KEYWORDS: *Prediction, Machine Learning, Data Analytics, Classification*

INTRODUCTION

Education impacts on the cultural, social and economic growth of the society. However different educational stakeholders are facing many challenges to improve students learning in education system. It is necessary to use advanced techniques to impart quality education. Education field contains a large amount of data. It needs to be used to generate the right information for educationists to make the right decisions. There is a need to use data analytics to extract hidden knowledge and identify different patterns from the extracted data. Data analytics is used to predict the academic performance of the students. The prediction results are helpful to take certain decisions to improve the academic performance of the students.

LITERATURE REVIEW

According to Robert L. Moore [1], the rapid changes in technologies are imparting key roles for improvement in the education. Information from the educational institutions is useful for administrative and academic related decision making. He has used Khan's Learning Framework to explain the role of data analytics in education along with different opportunities limitations, and dimensions.

Nguyen, A et al discussed about information system, educational system and different challenges in terms of learning analytics. There is rapid development in the field of educational data analytics for improving student performance. Applications of data analytics are helping in education and providing new opportunities for institutional decision making. The authors have discussed educational data mining, types of learning analytics and related applications [2].

According to Rishi Raj [3], data analytics is playing key role in improving the education domain. Data analytics provides insights into students' learning to improve it. Data analytics is helpful to the faculty members to determine students' academic performance and provide analyzed details to the decision makers. Ajibola O. et al [4] predicted the student performance using linear regression algorithms using Python. They have mentioned in their research work that attendance, distance from home to school, no. of hours spent on reading, health status, education level of parents are important factors to predict the students' performance. Vairachilai S, and Vamshidharreddy [5] in their research work applied different classification algorithms such as decision tree, Support Vector Machine (SVM), and Naive Bayes to predict the student's academic performance. They evaluated the performance of the algorithms based on different measures such as confusion matrix, accuracy, precision, recall, and F1 score. From the results, they found that the Naive Bayes

classification algorithm performs better. They concluded that prediction of the student’s academic performance will help the educational system to track the student’s academic performance. The literature review shows that effective use of data analytics is needed to predict the academic performance of the students. The present research study is regarding the prediction of academic performance of the students.

MACHINE LEARNING

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. A Machine Learning system **learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it.** It is one of the most exciting technologies that one would have ever come across. It is actively being used in many more places. Machine learning algorithms use historical data as input to predict new output values [6].

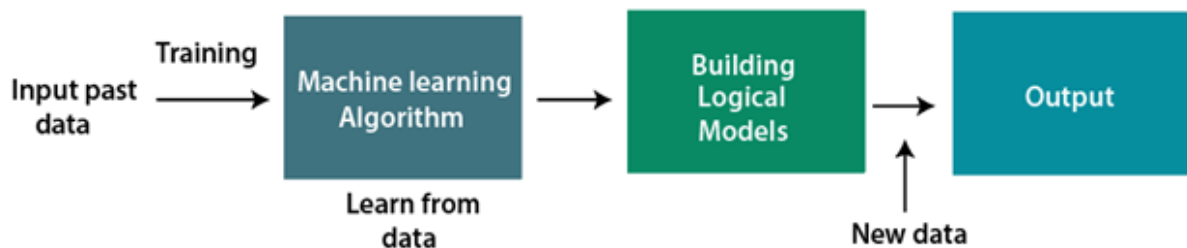


Figure 1: Machine Learning

Machine learning has given the computer systems the abilities to automatically learn without being explicitly programmed. But how does a machine learning system work? So, it can be described using the life cycle of machine learning. Machine learning life cycle is a cyclic process to build an efficient machine learning project. The main purpose of the life cycle is to find a solution to the problem or project.

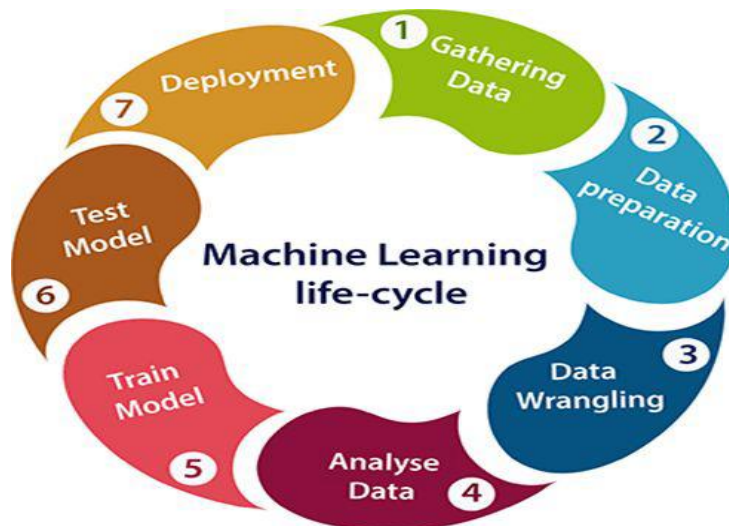


Figure 2: Machine learning life cycle

In the present study, data is collected from the students of graduation courses (BSc (CS), BCom, BBA, BBA(CA)) regarding their academic performance based on different parameters. The different machine learning algorithms are applied using Python Programming.

3.1 Different Classification Algorithms

Classification is a machine learning technique that needs the use of machine learning algorithms. It refers to a predictive modeling problem where a class label is predicted. There are different types of classification algorithms for prediction [6].

3.1.1 Naive Bayes

Naïve Bayes algorithm is a supervised learning algorithm based on Bayes theorem and used for solving classification problems. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms. It helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier and predicts on the basis of the probability of an object [7].

3.1.2 K-Nearest Neighbour

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new data and available data and put the new data into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, it does not make any assumption on underlying data [8].

3.1.3 Decision Tree

A decision tree is a flowchart-like structure in which each internal node represents a test on a feature, each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules.

A Decision Tree is a supervised learning algorithm. It is a graphical representation of all the possible solutions. All the decisions were made based on some conditions. Decision trees normally suffer from the problem of overfitting if it's allowed to grow without any control. A single decision tree is faster in computation. When a data set with features is taken as input by a decision tree it will formulate some set of rules to do prediction [9].

3.1.4 Random Forest

Random forest is a Supervised Machine Learning Algorithm. It is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems [10].

3.1.5 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, each data item is plotted as a point in n-dimensional space (where n is a number of features) with the value of each feature being the value of a particular coordinate. It is effective in high dimensional spaces. It is effective in cases where the number of dimensions is greater than the number of samples [11].

3.1.6 Ensemble Learning

Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model. Basic idea is to learn a set of classifiers (experts) and to allow them to vote. It is helpful in improving predictive accuracy but it is difficult to understand an ensemble of classifiers. The main challenge is not to obtain highly accurate base models, but rather to obtain base models which make different kinds of errors [12].

DATA ANALYSIS AND INTERPRETATION

In the present research work, various classification algorithms are used to predict the academic performance of the students. The parameters such as student learning understanding, skill and knowledge of faculty member, lecture delivery, effective use of technology, guidance and help, counselling and mentoring etc are considered to predict academic performance of the students.

4.1 Objectives

The following objectives are considered for the present study.

1. To use classification technique for the prediction of the students’ academic performance.
2. To identify different patterns through the data analytics model.
3. To compare the results obtained using classification algorithms.

The analysis of the objectives is presented as below.

4.1.1 Objective No. 1: To use classification technique for the prediction of the students’ academic performance.

The machine learning life cycle is used to predict the academic performance of the students. Classification technique is used through various classification algorithms to predict the students’ academic performance. The model is built and tested for the prediction of the results. It is helpful to the faculty members to identify the weak students who need special attention regarding the subjects. It is helping in early intervention. It is useful in decision making to improve the academic performance of the students.

4.1.2 Objective No. 2: To identify different patterns through the data analytics model.

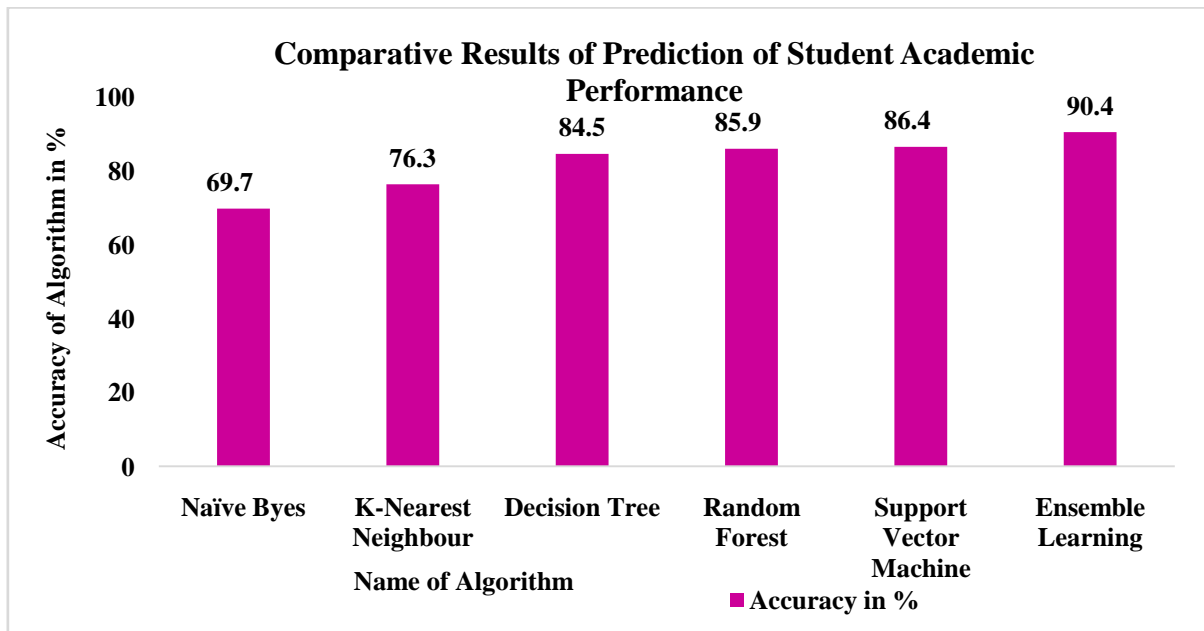
Various stages of machine learning life cycle are implemented using Python. Different classifier algorithms are applied to identify different patterns such as decision tree, byes classifier, random forest, support vector machine, ensemble learning, k-nearest neighbour. Various rules are generated based on classifiers regarding the prediction of the students’ academic performance. The identified patterns are useful to the educational authorities to take decisions to improve the academic performance of the students. The identified patterns are helpful to the faculty members to find the specific needs of the students for improving their academic performance. It helps them to identify the slow learners that are likely perform poorly and they need to improve their skills as early. The different patterns identified based on the prediction show that student academic performance in the internal examination, subject expertise and knowledge of faculty members, student understanding of lectures delivered by faculty members, effective use of current technologies by faculty members, guidance and help as well as counselling and mentoring by faculty members, practical knowledge of the students are the important parameters impacting on academic performance of the students.

4.1.3 Objective No. 3: To compare the results obtained using classification algorithms.

The results are obtained using machine learning. Different classifiers are applied using Python programming. Model accuracy is found for various classifiers to predict the academic performance of the students. The early prediction of the academic performance of the students is essential to improve their academic performance. The information generated through prediction is helpful to the faculty members and placement department to take certain decisions for academic performance improvement of the students.

Classification Result	
Name of Algorithm	Accuracy in %
Naïve Bayes	69.7
K-Nearest Neighbour	76.3
Decision Tree	84.5
Random Forest	85.9
Support Vector Machine	86.4
Ensemble Learning	90.4

Table 1: Comparative Results of Prediction of Student Academic Performance



Graph 1: Comparative Results of Prediction of Student Academic Performance

CONCLUSION

Data analytics and machine learning are playing key role in educational decision making. Prediction results are carried out using machine learning life cycle. Various classification algorithms are implemented using Python programming for the present research work. Analysis of three different objectives is done based on the data. Different patterns are identified through data analytics model. The model accuracy for different classification algorithms is found. The results are interpreted using classification techniques and are represented graphically.

REFERENCES:

Journal Papers:

- [1] Robert L. Moore, *The role of data analytics in education: Possibilities and limitations*, https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1111&context=stemp_fac_pubs, 2019
- [2] Nguyen, A., Gardner, L., & Sheridan, D. (2020), *Data Analytics in Higher Education: An Integrated View*, *Journal of Information Systems Education*, 31(1), 61-71.
- [3] Rishi Raj, *Data Analytics in Education Domain*, <https://www.magicedtech.com/wp-content/uploads/2017/12/Data-Analytics-in-Education-Domain.pdf>
- [4] Ajibola O. Oyediji, Abdulrazaq M. Salami, Olaolu Folorunsho, and Olatilewa R. Abolade, *Analysis and Prediction of Student Academic Performance Using Machine Learning*, *JITCE (Journal of Information Technology and Computer Engineering)*, VOL. 04 NO. 01 (2020) 10-15, ISSN (Online) 2599-1663.
- [5] Vairachilai S, Vamshidharreddy, *Student's Academic Performance Prediction Using Machine Learning Approach*, *International Journal of Advanced Science and Technology*, ISSN: 2005-4238, Vol. 29, No. 9s, (2020), pp. 6731-6737

Web References:

(For machine learning algorithms implementation using Python)

- [1] <https://www.javatpoint.com/machine-learning>
- [2] <https://www.javatpoint.com/machine-learning-naive-bayes-classifier>
- [3] <https://www.javatpoint.com/k-nearest-Neighbour-algorithm-for-machine-learning>
- [4] <https://www.analyticsvidhya.com/blog/2022/01/decision-tree-machine-learning-algorithm/>
- [5] <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- [6] <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>
- [7] <https://www.geeksforgeeks.org/ensemble-classifier-data-mining/>