Prediction and Improving Academic Performance of Students Using Various Classification Techniques

[1]Yuwaraj Khadke, [2]Dr. Dharmedra Choukse

Dept. of Computer Science and Engineering
Swami Vivekananda College of Engineering
Indore, Madhya Pradesh, INDIA

***Abstract*- One of the ultimate goals of the learning process is the success of student learning. Using data and students' achievement with machine learning to predict the success of student learning will be a crucial contribution to everyone involved in determining appropriate strategies to help students' performance. Data mining combines machine learning, statistics and visualization techniques to discover and extract knowledge from large database. One of the biggest challenge that in technical education faces is to improve students' performance. This study found that the classification machine learning algorithm was most often used in predicting the success of students' learning. Four algorithms that were used most often to predict the success of students' learning are KNN, Naive Bayes, SVM and Decision Tree.**

Keywords: students' learning, technical education.

I. INTRODUCTION

Students' academic performance has always been a challenging task in prediction. Many systems and classification algorithms have been proposed in the past years on prediction of students' performance. Researchers had been worked on different tools and by applying different techniques, but very less work has been performed on fundamental analysis. So, this research work has been conducted on students' performance based on fundamental analysis. Recognition of students' performance is a complicated task due to the unconstrained factors such as fundamental as well as technical factors. Also, this prediction depends much on factors and because we do not always predict the same way, so this prediction system in every application is not possible.

The use of machine learning algorithms for predicting student performance has gained significant attention in recent years. This approach is driven by the need to provide more accurate and comprehensive insights into student outcomes, and to enhance the educational process. Traditional approaches to predicting student performance have relied on historical data such as grades and attendance records, which only provide a limited snapshot of a student's academic progress. Machine learning algorithms, on the other hand, can analyze a wide range of factors such as socio-economic status, demographics, and learning style to provide a more comprehensive prediction of student performance.

In the context of diploma students, performance prediction using machine learning algorithms can provide valuable insights for educators and administrators. Diploma programs are typically designed to prepare students for a specific career path and require a significant investment of time and resources. Predicting the performance of diploma students can help educators identify areas where students may be struggling and provide targeted interventions to support their progress. Additionally, performance prediction can help administrators identify students who may be at risk of dropping out or failing to complete the program, enabling them to provide additional resources and support to improve outcomes. The use of machine learning algorithms for performance prediction of diploma students is still a relatively new field, and there are many challenges and limitations to be addressed. However, with advances in data analytics and machine learning, this approach has the potential to revolutionize the way we think about student outcomes and the educational process.

The main objective of using machine learning algorithms for performance prediction of diploma students is to provide more accurate and comprehensive insights into student outcomes. The traditional approach to predicting student performance, which relies on historical data such as grades and attendance records, has limitations and does not take into account other factors that may influence student performance. Machine learning algorithms can address these limitations by incorporating a wide range of factors such as socio-economic status, demographics, and learning style to provide a more comprehensive prediction of student performance.

II. RELATED WORKS

Some of the research papers worked on performance prediction of students.

1. "A comparative analysis of machine learning algorithms for student performance prediction" by A. R. Jadhav and S. S. Patil: This paper compares the performance of various machine learning algorithms for predicting student performance, including for diploma students.
2. "Prediction of students' academic performance using machine learning algorithms" by R. Singh and S. Gupta: This paper presents a study on the prediction of academic performance using various machine learning algorithms, including for diploma students.
3. "Machine learning approaches for predicting student performance: A comparative study" by A. F. Alqahtani: This paper compares the performance of different machine learning approaches for predicting student performance, including for diploma students.

4. "Predicting academic performance of engineering students: A comparative study of different data mining techniques" by N. N. H. M. Rani and M. M. Hafiz: This paper compares the performance of different data mining techniques, including machine learning algorithms, for predicting the academic performance of engineering students, including diploma students.
5. "Predicting student performance using machine learning techniques" by A. Singh and A. B. Gupta: This paper presents a study on the use of machine learning techniques for predicting student performance, including for diploma students.
6. "A study on the performance of machine learning algorithms for student performance prediction" by S. Balaji and S. Ramasamy: This paper presents a study on the performance of various machine learning algorithms for predicting student performance, including for diploma students.
7. "Predicting academic performance of students using machine learning" by A. Subasi and N. Al-Bataineh: This paper presents a study on the prediction of academic performance using machine learning algorithms, including for diploma students.
8. "Machine learning for predicting academic performance of students" by V. A. Kumar and A. Gopalakrishnan: This paper presents a study on the use of machine learning for predicting academic performance of students, including for diploma students.
9. "Prediction of academic performance of student's uses machine learning techniques" by S. K. Dash and R. K. Swain: This paper presents a study on the prediction of academic performance of students using machine learning techniques, including for diploma students.
10. "Performance prediction of engineering students using machine learning algorithms" by S. Selvam and K. Vijayalakshmi: This paper presents a study on the performance prediction of engineering students using machine learning algorithms, including for diploma students.
11. "Performance prediction of diploma student's uses machine learning algorithms" by S. Garg and N. Arora: This paper presents a study on the performance prediction of diploma students using machine learning algorithms.
12. "Machine learning for predicting student performance in higher education" by N. N. H. M. Rani and M. M. Hafiz: This paper presents a study on the use of machine learning for predicting student performance in higher education, including for diploma students.
13. "Predicting academic performance of diploma student's uses machine learning algorithms" by S. S. Bhatia and M. Gupta: This paper presents a study on the prediction of academic performance of diploma students using machine learning algorithms.
14. "Predicting academic performance of diploma students using decision tree algorithm" by P. R. Gupta and N. K. Tiwari: This paper presents a study on the prediction of academic performance of diploma students using the decision tree algorithm.
15. "Performance prediction of diploma student's uses artificial neural networks" by R. Arora and S. Saini: This paper presents a study on the performance prediction of diploma students using artificial neural networks.

III METHODOLOGY

Data mining is the knowledge discovery process from a huge data volume. The mechanism works in large dataset where the student performance in the end semester examination is evaluated.

A. Data selection and transformation

We propose a new system for performance prediction of diploma students that addresses the limitations of the existing systems. The proposed system will use a combination of regression analysis, decision trees, and neural networks to predict the performance of diploma students accurately. The system will also allow educators and administrators to customize the model based on the unique characteristics of their students.

The proposed system will collect data on various factors that can impact the academic performance of diploma students, including their academic background, study habits, socio-economic status, and other relevant factors. The collected data will then be analysed using various statistical techniques to develop a performance prediction model.

B. Algorithms

1. Decision Tree-

Decision Tree classifier is the regression model which is represented in the form of tree structure. The purpose of Decision Tree classifier is to breakdown the dataset into smaller subset. The tree consists of decision nodes and leaf nodes.[19] In our proposed architecture the attribute which delivers maximum information will act as a decision node. The node which is present as the top most of the decision node acts as a predictor which is called as root node. The node which cannot be further divided is known as leaf node. The steps involved in the decision tree are specified below:

Process 1: Start the root.
Process 2: Perform the test.
Process 3: Follow the edges corresponding to the outcome.
Process 4: go to step 2 until reaches leaf node.
Process 5: Predict the outcome associated with the leaf.

2. K-Nearest Neighbour-

K-Nearest Neighbour is one of the basic and essential classification algorithms in machine learning. It is non-parametric and makes any underlying assumptions about the distribution of data.

The steps involved the KNN is listed below:
File the training data in a sample points array.
The Euclidean distance measure.
Make the least distance range available.

IV. RESULTS DISCUSSION AND ANALYSIS

The application was constructed ended up being a framework for binary student data to be inputted and outputted in the form of a prediction value for the likelihood of clearing a target course. The framework was constructed with the help of the sklearn framework as we as Keras API, plus the dataset that was fetched from kaggle to train the ML models. The purpose of the application was twofold, one being technical in that it provided a scientific comparison study on how these ML model performed from these certain parameters in this certain environment (datatype/data size), the other one being to provide a method of predicting the likelihood of whether certain students would clear a given course, given their previous courses. As far as those end goals/purposes are concerned, the quota has been fulfilled sufficiently and correctly. As to why Decision Tree outperformed KNN, can be theorized because the number of parameters in an Decision Tree increases linearly with the amount of the input. On the other hand, a KNN model does not. The Decision Tree method is an excellent classification algorithm. It's a supervised learning method that's mostly used to divide data into categories. A collection of label data is used to train Decision Tree. However, it is believed that large data sets are not suited for the Decision Tree method. When the data set contains additional noise, such as overlapping target classes, Decision Tree does not perform well. The Decision Tree will underperform if the number of features for each data point exceeds the number of training data samples.

The chosen method was a quantitative approach due to the nature of computer science and machine learning problems. They are inherently quantifiable and measurable. The method was broken down into small achievable and measurable milestones that could be worked on separately. The approach however was done in an ad hoc basis, as it wasn't clear at the start which software tools and hardware to use, whether to use a cloud-based solution or a native one. All these approaches were figured out on the spot when the thesis needed it. The metrics that were measured (Accuracy, Classification Error ,Sensitivity, Specificity) were chosen due to the abundance of resources surrounding them, the sheer volume of mathematical equations describing them, and due to measuring just the accuracy of a model doesn't tell the whole story.

The first research question of the thesis was to figure out if ML algorithms were an appropriate way to find and identify these student patterns. To that extent, the study deems it to be an appropriate way. As the approach was when a pattern was generated, then the research question was fulfilled, the study was successful in that regard.

The second research question was to figure out how to implement the Decision Tree, SVM, Naïve Bayes, and KNN models, and to figure out which model performed best in terms of accuracy and other parameters on an open-source ML framework using student data from higher education data.

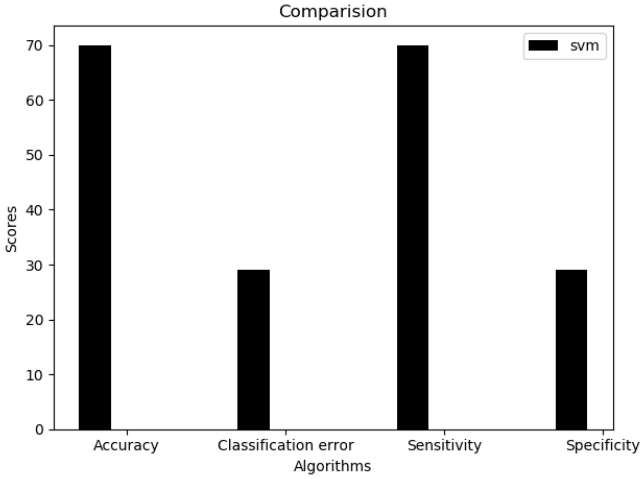

Fig.1: SVM Algorithm

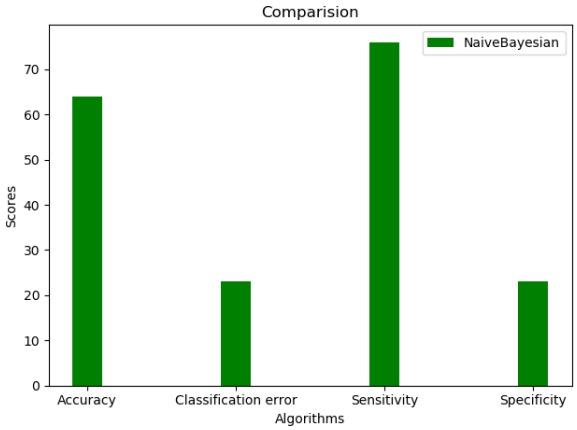

Fig.2: Naïve Bayes Algorithm

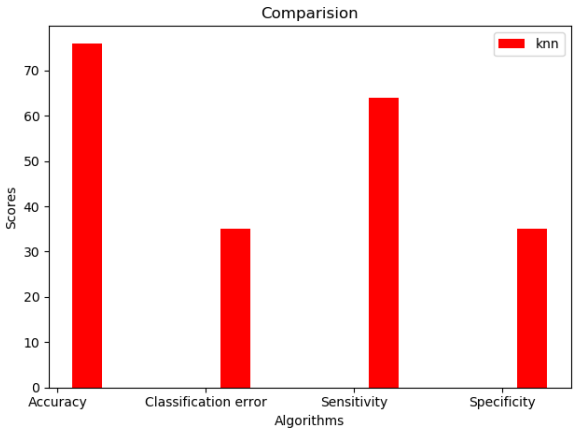

Fig.3: Knn Algorithm

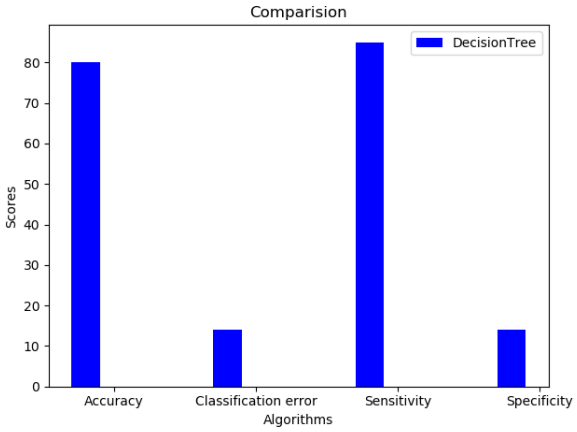

Fig.4: Decision Tree Algorithm

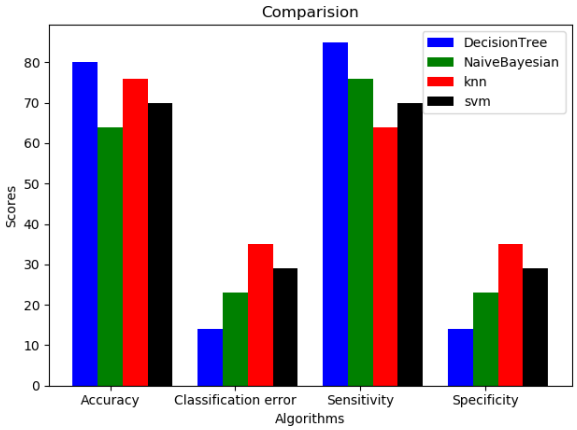

Fig.5: Comparison of Algorithms

V.CONCLUSION

This research aims to systematically examine the extent to which the implementation of machine learning algorithms and modeling gives out best results in predicting student's performance. The results showed that the most widely used type of machine learning algorithm to predict student's performance was the algorithm. The most widely used machine learning algorithms to predict student's performance are K Nearest Neighbor Support Vector Machine (SVM), Naïve Bayes, and Decision Tree algorithms. The data used in this research article uses a variety of datasets. In general, predictions of student's performance are classified into 2 to 3 classifications such as pass/fail; or file/pass/excellent.

The findings from this implementation are that Decision tree has outperformed all other algorithms at least for the given dataset. But the results may change when different size and type of dataset is used.

REFERENCES:

1. S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintela "Preventing Student Dropout in Distance Learning Using Machine Learning Techniques". Conference: Knowledge-Based Intelligent Information and Engineering Systems, 7th International Conference, KES 2003, Oxford, UK, September 3-5, 2003,
2. B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," Int. J. Adv. Computer. Sci. Appl., vol. 2, no. 6, pp. 63-69, 2011.
3. Erkan Er "Identifying At-Risk Students Using Machine Learning Techniques: A Case Study with IS 100". International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012
4. L. Coburn., (1984), "Student Evaluation of Teacher Performance, ERIC/TME Update Series," [Online].
5. S. A. Radmacher and D. J. Martin, "Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis," J. Psychol., vol. 135, no. 3, pp. 259-269, 2001.
6. Agaoglu, Mustafa. "Predicting Instructor Performance Using Data Mining Techniques In Higher Education". IEEE Access 4 (2016): 2379-2387. Web. 16 Nov. 2016.
7. N. Delavari, S. Phon-Amnuaisuk, and M. R. Beikzadeh, "Data mining application in higher learning institutions," Inform. Edu.-Int. J., vol. 7, no. 1, pp. 3154, 2007.
8. Chin Chia Hsu and Tao Huang (2006): "The use of Data Mining Technology to Evaluate Student's Academic Achievement via multiple Channels of Enrolment." An empirical analysis of St. John's University of Technology.
9. Osofisan A.O. and Olamiti A.O. (2009): "Academic Background of Students and Performance in Computer Science Programme in a Nigerian University." European Journal of Social Science. Vol. 33 Issues 4. 2009.
10. MardikyanS., and Badur B. (2011). "Analyzing teaching Performance of Instructors Using Data Mining techniques. Informatics in Education," 2011, Vol. 10, No. 2, pp 245 – 257.
11. Surjeet K.Y and Saurabh P (2012): "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification," World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012.
12. JIANG, Yuheng Helen, Sohail Syed JAVAAD, and Lukasz GOLAB. "Data Mining Of Undergraduate Course Evaluations". Informatics Educations 15.1 (2016): 85-102. 11 Nov. 2016.
13. C Marquez-Veera, C Romero and S Ventura."Predicting School Failure Using Data Mining" -2011
14. Din Ahmed, Ibrahim Sayed. "Data Mining: A Prediction for student's performance using classification method", World journal of Computer Application and Technology-2014
15. Sajadin Sembering, M.Zarlis, "Prediction of student academic performance by an application of data mining techniques", International conference on 2011.