

ANALYSIS AND PREDICTION OF STUDENTS PERFORMANCE USING MACHINE LEARNING

Velagala Sowmya Reddy¹, Barri Sandhya², K. Yashwanth Reddy³, Dr. P. Udaya Kumar⁴

^{1,2,3}Cse, B. Tech Aditya Engineering College, Surampalem

⁴Guide, M. Tech, Ph.D, Aditya Engineering College, Surampalem

ABSTRACT

Any educational institution's primary goal is providing the best knowledge and education to the students. However, student to student academic performance may vary. A single teacher can't guide each and every student at once, due to this every year many students may lag behind due to lack of proper monitoring and advice. If universities can analyze the behavior of students and can predict the low performance factors earlier, then that information can help them to take effective steps in order to increase student performance. So, Student performance analysis and prediction is very important to understand about a student progress rate. So that this helps the teacher to know about the details of the student like which student needs which kind of help. Not only students and teachers, parents also can get the knowledge on student performance in academics. Due to the abundance of data in educational databases, it is now increasingly challenging to predict students' success.. A data mining research and prediction can be used in the education system for analyzing and predicting the student performance. Using educational machine learning techniques Support Vector Machine, K-Nearest Neighbour, Decision Tree, and Random Forest The proposed system provides the efficient algorithm for analysis. Based on the previous marks of the student in each subject and student background details we will predict the marks and depending on the predicted marks the student can know in which subject he/she is lagging and teachers can easily monitor the student performance and can give the student a proper advice to the student.

Keywords—Student Performance Prediction, Machine Learning, Decision Tree, Random Forest, K-Nearest Neighbour, Support Vector Machine.

1. INTRODUCTION

A subset of artificial intelligence called "machine learning" is used to build and create algorithms that let computers predict behaviour based on incoming data. The majority of machine learning algorithms emphasise decision-making and pattern identification. Machine learning is wherever data is being used and analyzed. It is a very useful tool for exposing secret patterns and important information that can otherwise be sought and comprehended with statistical methods. In the field of education, the use of machine learning algorithms has increased significantly in recent years.. Student performance analysis and prediction is very important to understand about a student progress rate. So that it helps the teacher to know about the student like which student needs which kind of help. If universities can analyze the behavior of students and can predict low performance factors earlier, this information can help them to take effective steps to increase student performance. So, Student performance analysis and prediction is very important to understand a student progress rate. So that it helps the teacher to know about the student like which student needs which kind of help. Not only students and teachers, parents also can get the knowledge on student performance in academics. A data mining research and prediction can be used in the education system for analyzing and predicting the student performance. We can forecast student academic achievement using the educational machine learning algorithms Decision Tree, Random Forest, Support Vector Machine, and KNN classifier. The proposed system provides the efficient algorithm for analysis.

2. LITERATURE REVIEW

[1] PREDICTION OF STUDENT PERFORMANCE USING LINEAR REGRESSION-They used the linear regression approach in this paper. In this case study, supervised learning and more specifically, predictive analysis are the main topics. The ability to predict student academic success is crucial since it informs both students and teachers of those who may struggle in a course and those who need to boost their grade point average. The dataset used in this paper comprised of 100 students, which is a small dataset. Student datasets are the study's primary source of data. The dataset used for this study is tabular in nature and includes data on students' age, gender, academic performance, and health. In this case study, the implementation of linear regression to forecast the student's academic performance while taking into account the student's dataset was the main focus. [1].

[2] STUDENT ACADEMIC PERFORMANCE PREDICTION USING EDUCATIONAL DATA MINING-This study examines educational data mining, which focuses more on making predictions than on producing precise results

for future predictions. A regular analysis must be carried out in the educational field in order to keep track of the changes that are being in class patterns. In this case study exploratory design, which was made possible by gathering data sets from BMS College of Engineering, the initial findings of an educational data mining forecast are provided. Numerous classification and regression methods were examined and tried for improved accuracy in subject-by-subject analysis and GPA prediction of a student's performance. Voting on the classification algorithms was carried out for the prior ensemble system. These were also divided into other groups and contrasted according to specific criteria; the outcomes are discussed in this paper. [2].

[3] PREDICTING ACADEMIC SUCCESS IN HIGHER EDUCATION: LITERATURE REVIEW AND BEST PRACTICES-This paper primarily focuses on student achievement, which is crucial for educational institutions and usually used as a performance indicator. Prior to the identification of students who may be in danger, preventative measures can increase student achievement. Machine learning techniques have recently been applied to prediction. Although there are many success stories in the literature, "computer science" or more typically "artificial intelligence" educated educators tend to use these strategies the most. Many judgements are made possible by the use of data mining that is successful and efficient techniques, including how to define student achievement, which student characteristics to emphasise, and even which machine learning technique is best for the particular issue at hand. This study made data mining techniques more accessible to educators, allowing the full potential of their application to the given challenge in the field of education. [3].

3. PROPOSED METHODOLOGY

Students are essential for producing high quality graduates who excel in academics, and creative thought for any school, college and other educational institution. In order to get the quality students it is necessary to evaluate the output of students in order to forecast student achievement. Data abundance in educational databases has made it increasingly difficult to predict students' success. We can more effectively and efficiently increase student performance and growth by using educational machine learning algorithms. The proposed system uses machine learning methods Decision Tree, Random Forest, Support Vector Machine and KNN classifier are used to predict the student performance that is used by students, teachers and parents to get the knowledge on student performance in academics. The proposed system provides the efficient algorithm for analysis. Based on the previous marks of the student in each subject we will predict the marks and depending on the predicted marks of the student in each subject the teachers can easily monitor the student performance. This project provides a system that can analyze and predict the student performance accurately by using machine learning algorithms. The flowchart of the proposed system describes about the flow of the project. First the dataset is collected from Kaggle and the set is preprocessed which includes data cleaning and data is visualized to understand the dataset and then, the data set is split into a training and a testing dataset, and finally, the algorithms are put into practise to make predictions. The flow of the project is clearly shown in the below figure.

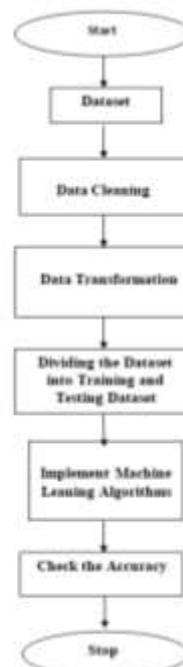


Fig 3.1 Flow Chart of the Project

Hardware Requirements:

- System : Pentium i3 Processor
- Hard Disk : 500 GB
- Ram : 2 GB

Software Requirements:

- Operating system : Windows 10
- Coding Language : Python
- IDE : Google colab or Jupyter Notebook
- Streamlit
- Students Dataset

4. ALGORITHMS USED

Decision Tree:

A supervised learning approach called a decision tree can be used to solve classification and regression problems, however it is typically chosen for classification problems. Decision trees are always used to visualise all potential outcomes and decision points that take place chronologically and are produced in both situations in the same manner. In the world of finance, decision trees are most frequently utilised in areas like budgeting, portfolio management, and loan approval

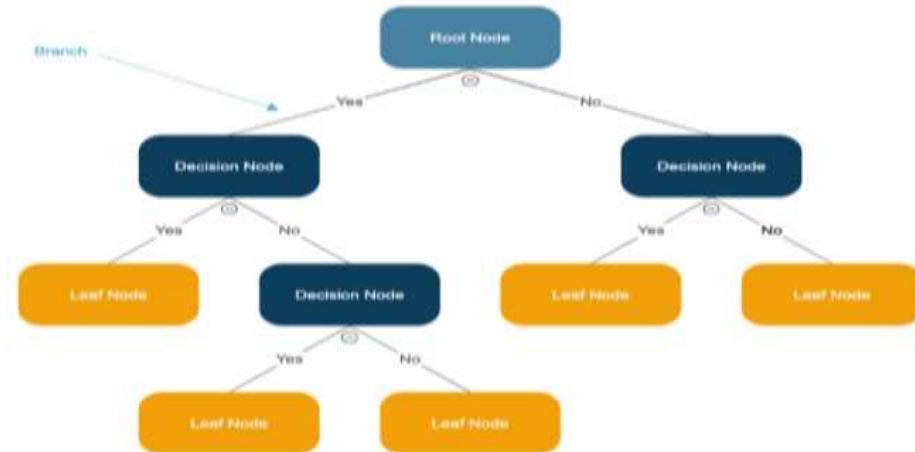


Fig 4.1 Decision Tree

Random Forest:

A very well known supervised machine learning approach called the Random Forest approach is utilised to solve classification and regression issues. We are aware that a forest has several trees, and the more trees there are, the more effective it will be. Similar to this, the accuracy and problem-solving capacity of a Random Forest Algorithm increase with the number of trees in the algorithm.

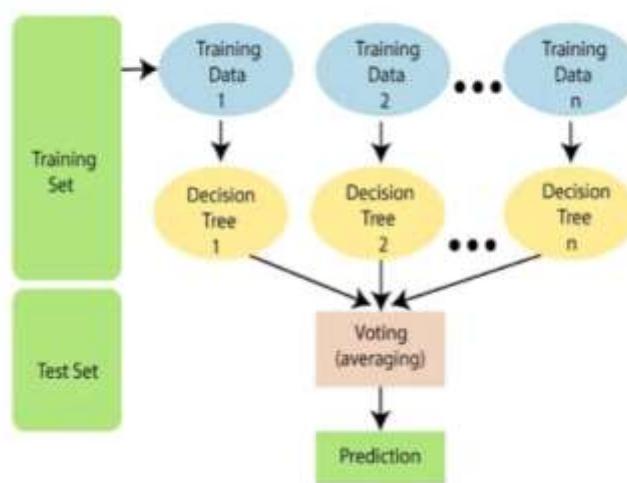


Fig 4.2 Random Forest

Support Vector Machine:

The SVM algorithm's objective is to determine the optimum line or decision boundary that can divide n-dimensional space into classes so that subsequent data points can be quickly assigned to the appropriate category. A hyperplane is the name given to this optimal decision boundary. The dataset's features determine the hyperplane's dimensions, therefore if there are only two characteristics, the hyperplane will be a straight line. Additionally, if there are three features, the hyperplane will only have two dimensions

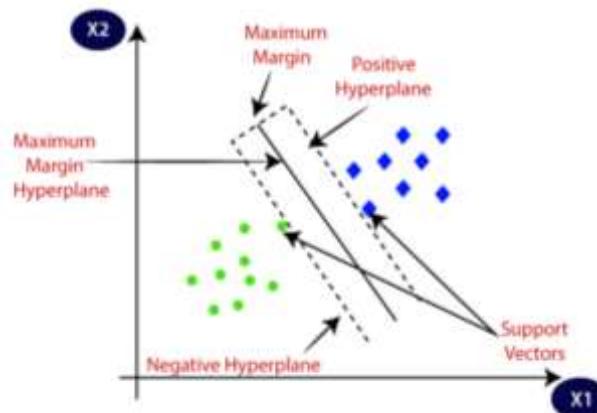


Fig 4.3 Support Vector Machine

K-Nearest Neighbor:

One of the simplest machine learning algorithms used in supervised learning techniques is K-Nearest Neighbour. The new case is placed in the category that is most similar to the existing categories after the K-NN algorithm analyses the similarity between the new data and existing cases. To put it simply, the K-NN algorithm collects all the data that is available and categorises new data based on similarities

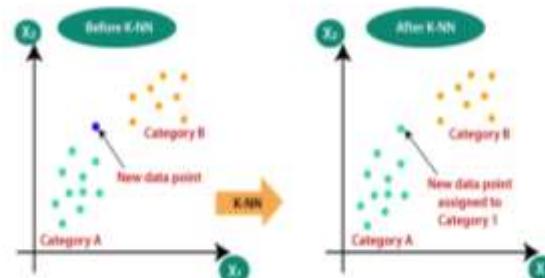


Fig 4.4 K-Nearest Neighbor

5. EXPERIMENTAL RESULTS

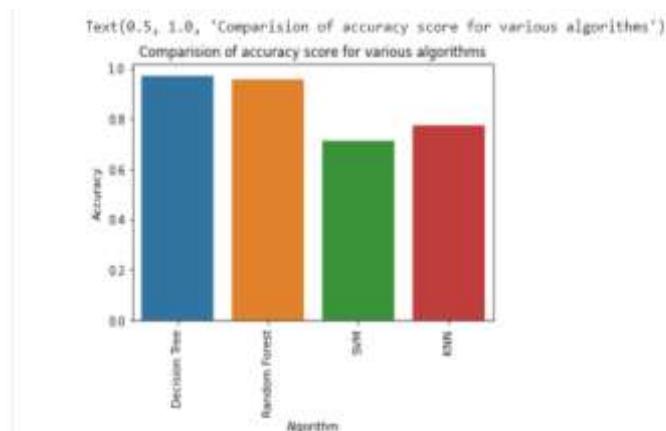


Fig 5.1 Comparison of Accuracies

Text (0.5, 1.0, 'Comparision of execution time for various algorithms')

Comparision of execution time for various algorithms

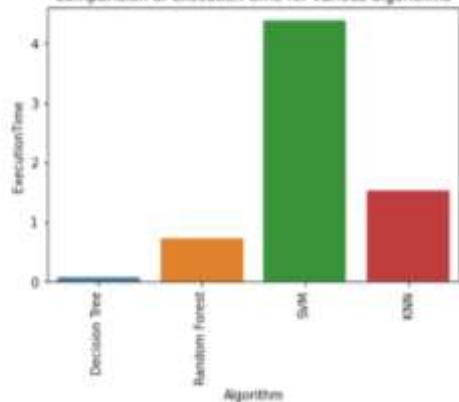


Fig 5.2 Comparison of Execution Times

Analysis and Prediction of Student Performance

Analysis and Prediction of Student Performance

Student

Student ID: 12345678901234567890

Selected Resources for University

Sex: Female

Age: 18

Address: House No. 123, Street Name: Main Street, City: New York, Country: USA

Parents' Education Status: Living Together

Mother's Education:

- Higher Education
- Higher Education

Father's Education:

- Higher Education
- Higher Education

Marital Stat:

- Married
- Married

Family Size:

- 2
- 2

Residency Status:

- Other

Student's Gender:

- Female

Travel Distance (in km): 10.0000

Present Study Year: 2020

No of past class failures: 0

Extra Educational Support:

- False

Fairly Educational Support:

- False

Extra Paid Classes While The Course Subject:

- False

Other-Curricular Activities:

- False

Attended Nursey School:

- False

Plans To Take Higher Education:

- True

Internet Access At Home:

- True

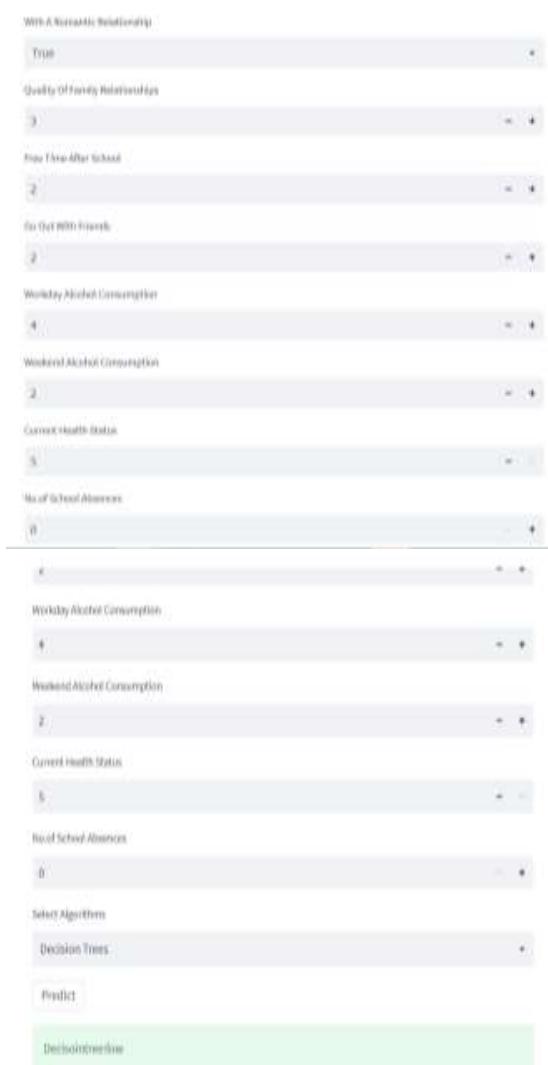
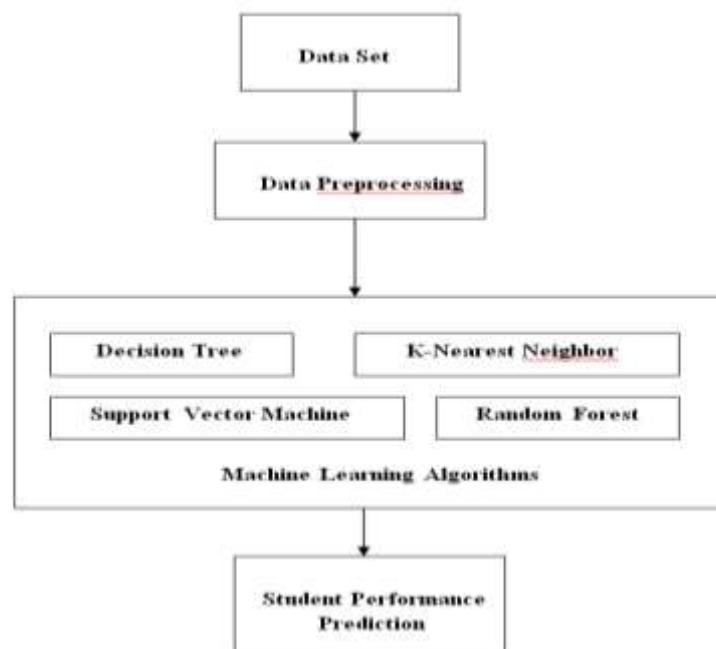


Fig 5.3 Output Screenshots

6. BLOCK DIAGRAM



7. ADVANTAGES

- The proposed system produces the results with high Accuracy
- It takes less time for training
- This system predicts better results

8. CONCLUSION AND FUTURE SCOPE

Any machine learning algorithm's efficacy in the field of education depends on the algorithm's utilisation of the dataset and the dataset itself. It's crucial to carefully select the algorithm to use when predicting student success. The machine learning method being utilised affects how accurate the forecast is. Current case studies demonstrate that a student's history and other characteristics have an impact on their academic achievement. Numerous studies support the idea that, in addition to prior academic achievements, a student's history and other characteristics may also have a substantial impact on how well they do. Data mining in education can benefit from the application of machine learning, which has been increasingly important in all industries lately. To predict student academic success in our project, we used the machine learning algorithms Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbour. The efficient algorithm for analysis and prediction is provided by the proposed system. The suggested system offers an effective analytical algorithm. Based on the previous marks of the student in each subject and some background information we will predict the marks and depending on the predicted marks the student can know in which subject he/she is lagging and teachers can easily monitor the student performance and can provide the learner with appropriate recommendations to enhance performance. Future analyses and predictions of student success should be made using an efficient and accurate algorithm, which can be used to any college website to forecast student performance there.

9. REFERENCES

- [1] Boddeti Sravani And Myneni Madhu Bala, "PREDICTION OF STUDENT PERFORMANCE USING LINEAR REGRESSION". 2020 International Conference For Emerging Technology (INCET) Belgaum, India. Jun 5-7, Pp.1-5, 2020.
- [2] D K Arun, V Namratha, B V Ramyashree, Yashita P Jain And Antara Roy Choudhary, "STUDENT ACADEMIC PERFORMANCE PREDICTION USING EDUCATIONAL DATA MINING" Int. J. Pure Appl. Math, Vol. 118, No. 18, Pp. 2703–2707, 2021.
- [3] E. Alyahyan And D. Düşteğör, "PREDICTING ACADEMIC SUCCESS IN HIGHER EDUCATION: LITERATURE REVIEW AND BEST PRACTICES" Int. J. Educ. Technol. Higher Educ., Vol. 17, No. 1, Pp. 93-97, Dec. 2020.
- [4] Manjari Chitti1 , Padmini Chitti 2 And Manoj Jayabalan: "NEED FOR INTERPRETABLE STUDENT PERFORMANCE PREDICTION." 2020 IEEE 13th International Conference On Developments In Esystems Engineering (Dese), Pp.269-272, 2020.
- [5] S. Saeed, H. Gull, And S. Z. Iqbal, "WEB 2.0 USAGE BY SAUDI FEMALE STUDENTS FOR INFORMATION SHARING IN PUBLIC SECTOR UNIVERSITY A PILOT STUDY" Int. J. Public Adm. 2017 Volume 4, Issue 3, July 2017, Pp. 79–90.
- [6] Dhanashree Mane, Pranali Namdas, Pooja Gargade, Dnyaneshwari Jagtap And S.S. Rathi "PREDICTING STUDENT PERFORMANCE USING MACHINE LEARNING APPROACH". VJER Vishwakarma Journal Of Engineering Research, Pp. 1208-1212, December 2018.
- [7] Wei Wang, Han Yu And Chunyan Miao "DEEP MODEL FOR DROPOUT PREDICTION IN MOOCS". In ICCSE 2017, Proceedings Of The 2nd International Conference On Crowd Science And Engineering, Pp, 26–32, July 2017.
- [8] A. Barredo Arrieta Et Al., "Explainable Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities And Challenges Toward Responsible AI," Inf. Fusion, Vol. 58, Pp. 82–115, 2019.