Student Academic Performance Prediction Using Machine Learning

Talla Sai Likhitha

Dept. of Electronics and Communication Engineering Madanapalle Institute of Technology and Science,

Annareddy Samhitha

Dept. of Electronics and Communication Engineering Madanapalle Institute of Technology and Science,

Sai Sucharitha

Dept. of Electronics and Communication Engineering Madanapalle Institute of Technology and Science,

Janmoni Borah

Dept. of Electronics and Communication Engineering Madanapalle Institute of Technology and Science,

Abstract—Now-a-days people's academic overall performance prediction in huge establishments isn't always a clean process and also time taking as for large quantity of data. The potentialto expect a scholar's performance may be beneficial for movesin modern academic structures. So here machine learning algorithms play a crucial role in solving those problems. Based on those machine learning algorithms we can finish the correlation among the academic performance of a scholar and the functions of dataset. Student performance prediction model is based totallyon supervised learning decision tree classifier is used to clearup the hassle. Similarly, an ensemble technique is applied toimprove the overall performance of the classifier. Basic ensembletechniques method is used for solving of classification, prediction troubles.

Index Terms—Machine Learning, Decision Tree, Support Vector Machine, Logistic Regression.

I. INTRODUCTION

Current mastering institutions operate in a quite aggressive and complex environment. Online learning or E-learning to know is an important equipment to assist and encourage teaching and gaining knowledge. This people's academic overall performance, we are able to reduce the dropout fee and grow the overall performance of college students. Quantifying scholar educational performance of students relies upon on numerous factors. Machine learning (ML) is the examining of Machine learning algorithms that improve mechanically via experience and through use of data. Supervised Learning algorithms which include decision Tree, Support Vector Machine (SVM) and Logistic Regression schemes have been used.The goal is to examine system gaining knowledge of strategiesfor student's academic performance prediction and check out the risks of using Machine learning strategies, comparing withthe other strategies. One of the most vital use of it is to expect student's performance to improve the institutional state of affairs. Those algorithms prove their performance and ability to be used for class and prediction with appropriate accuracy. We use unique system getting to

know algorithms used for category are decision tree, SVM and Logistic regression. **LO- GISTIC REGRESSION**: Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. Logistic regression has becomean important tool in the discipline of machine learning. The approach allows an algorithm being used in a machine learningapplication to classify incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classifications within data sets. Logistic regression can also play a role in data preparation activities byallowing data sets to be put into specifically predefined buckets during the extract, transform, and load process in order to stage the information for analysis. **SUPPORT VECTOR MACHINE**: Support vector machines are a set of supervised learning methods used for classification, regression and outliers' detection. The advantages of support vector machines are effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples. The support vector method is a universal tool for solving multidimensional function estimation problems. **DECISION TREE CLASSIFIER**: A decision tree is a flowchart-like tree structure in which an internal node represents feature, thebranch represents a decision rule, and each leaf node representsthe outcome. The topmost node in a decision tree is known as the root node. The partitioning of a main node is based on the attributes of the dataset. This algorithm partitioning is basedo the recursive type and not iterative.

This paper discusses about student's performance during academic year by considering different parameters. We have used supervised machine learning algorithms. The proposed mechanism has been compared with some algorithms like support vector machine, decision tree, and logistic regression such that accuracy, precision and recall can be observed and compared. 1) The work signifies prediction of student's academic performance. 2) Different supervised machine learning algorithms has been proposed to find out which algorithm gives better accuracy.

The remaining sections of the paper could be organized as: section II consists of materials and methods which we are going to use throughout the process. Section III is composed of dataset collection, preprocessing, feature extraction, prediction and evaluation. Section IV comprises of implementation and its key functions. The simulated results and discussions of the proposed system along with the comparison of all the algorithms has been presented in section V. further the conclusion and the references are mentioned.

II.   MATERIALS AND METHODS

A. Jupyter Notebook is open-source software that enables us to create, enhance, share, and modify codes. Various pre-processing stages, like data cleaning, training, and statistical analysis, can be executed. In this application, each cell is executed separately, which is written in python language. As individual implementation is carried out, the end-user test or run the source code from scratch.

C. Feature Extraction

III.   PROPOSED SYSTEM

Recently, machine learning algorithms have played an important role in most field of science and life. These algorithmsprove their efficiency and ability to be used for classification and prediction with acceptable accuracy. Student's data istaken from student records, cooperative or interactive learning environments, or data recorded with school and universityadministrations. In our project we used three machine learning algorithms to predict student's academic performance in advance. And these three machine learning algorithms arecompared to the proposed model in our system to achieve a highest accuracy among the three algorithms used. The three algorithms used in our project are support vector machine, decision tree algorithm, and logistic regression. We obtaineda highest accuracy for decision tree algorithm which is 96.96

A. Dataset Collection

Collecting dataset from the source. Source is obtained from the student, which datset contains student personal and educational information, extra skills activity information. This datais labelled.

Name	Grade Id	Semester	Parent Survey	Label
Chinnu	4	7	0	1
Indu	7	7	0	1
Lucky	6	7	1	1
Siri	3	7	1	0
Laddu	5	7	1	1
Keerthi	4	7	1	1

B. Data Preprocessing

The main reason for testing several algorithms on the dataset was that their performance varies for the selected features. The study suggested that algorithms behave differently; Dependingon the dataset, the efficiency and performance may also vary. Preprocessing is done to remove the unwanted feature fromthe dataset, missing data imputation and noise data handling. After that, apply the unified DB module to the dataset. This module can remove repeated data. This stage includes data cleaning, transformation and filtering.
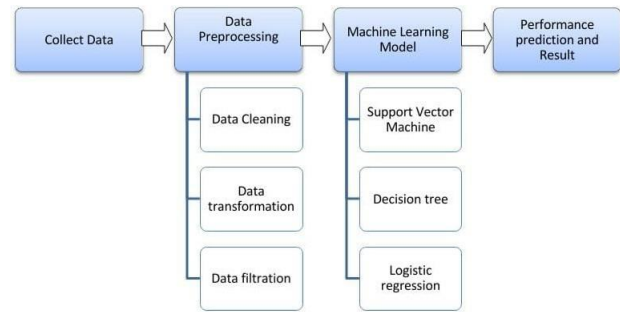

Fig. 1. **Flow Chart**

Next thing is to do feature extraction is an attribute reduction process. It enables reduced computation time, improved prediction performance while allowing a better understanding of the data.

D. Prediction and Evaluation

Model evaluation is an integral part of the model development process. It helps to find best model that representsour data. It evaluate the proposed classification model andto make comparisons. The performance of an algorithm has been evaluated based on confusion matrix, accuracy, and recall . After predicting the student performance, the system will also compare the results generated by three classification algorithms and there after determine which of them is more accurate and efficient.

E. Accuracy of the Dataset

Accuracy is given by sum of all corrected true positive and negative predictions to the total count of the given dataset. The minimum and maximum values of accuracy are given as 0.0 and 1.0, respectively. Accuracy= ((True Positive + True Negative)/ (True Positive + False Positive + False Negative + True Negative))

F. Precision of the dataset

The precision is the ratio of all the positive predictions to the sum of true and false positive predictions. Similar to accuracy, the minimum and maximum value for the precision ratio are 0.0 and 1.0 respectively. Precision = ((True Positive)/ (True Positive + False Positive)

G. Recall

The recall is the ratio of all true positive predictions to the sum of true and false negative predictions. For maximum and minimum the values are 1.0 and 0.0 respectively. Recall = ((True Positive)/ (True Positive + False Negative)

IV.   IMPLEMENTATION

Automatically generating captions to an image shows the understanding of the image by computers, which is a fundamental task of intelligence. For a caption model it not only need to find which objects are contained in the image and also need to be able to be expressing their relationships in a natural language such as English. The model should be trained more to get better efficiency. Integrating different functions is one of the crucial things in training the model because each layer should be compatible with the next corresponding layers. We have to choose right optimizer and loss function. The respective pre-trained models should be downloaded, and all the packages should be installed

B.   Key Functions

The Key functions that are to be noticed, processed, and implemented are: 1. Accessing and reading Dataset 2. Preparing the Data and processing it. 3. Developing model and extracting features. 4. Processing the data, cleaning and performing wordembeddings. 5. Developing a deep learning 6. Training the model. 7. Evaluate the trained model with test dataset. 8.Creating a dataset and providing accurate results.

V.   RESULT AND DISCUSSION

A.   Precision, Accuracy, Recall of Logistic Regression

Fig. 2. **Logistic Regression**

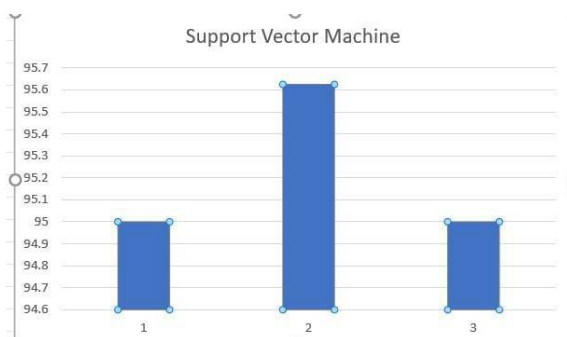### B.   Precision, Accuracy, Recall Support Vector Machine

Fig. 3. **Support Vector Machine**

A.   Introduction

C.   Precision, Accuracy, Recall of Decision Tree

Fig. 4. **Decision Tree**
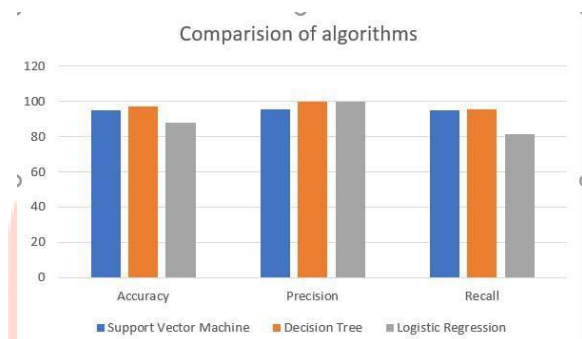
D.   Comparison of Algorithms

Fig. 5. **Comparison**

VI.   CONCLUSION

The most popular task is to predict student's performanceis classification. We analyzed the student's academic performance with various machine learning algorithms. This research study aims to gain better accuracy for the classification algorithms. This study compares the performanceof the most popular machine learning algorithms used forbinary classification. In this paper, logistic regression, SVM and decision tree algorithms are used to predict the student's academic performance. The performance of an algorithm has been evaluated based on confusion matrix, accuracy, recall andF1 score. The performance of the decision tree algorithm is better as applied to the rest of the algorithms, with accuracy of 96. Specifically in this pandemic, the students are forwarded to the next class without clearing the examinations. So in suchcases the results of the student can be predicted by our project,which is generally works on past performance and various other entities.

REFERENCES

[1]   Dindar Mikaeel Ahmed, Adnan Mohsin Abdulzzeez, Diyar Qader Zeebaree, Falah Y. H. Ahmed(2021), "predicting university's students' performance based on machine learning techniques," in international conference on automatic control and intelligent systems (I2CACIS),Shah Alam, Malaysia.
[2]   Shah Hussain Muhammad Qasim Khan (2021), "student-performulator:

predicting student's academic performance at secondary and intermedi- ate

[3]   Vairachilai S, Vamshidharreddy, Avvari Sai Saketh, GnanajeyaramanR (2020) "student's academic performance prediction using machine learning approach," in international journal of advanced science and technology, vol.29, No. 9s, pp.6731-6737.

[4]   Havan Agarwal, Harshil Maavani (2015), "student performance predic- tion using machine learning," in international journal of Engineering Research and Technology (IJERT), vol.4.

[5]   Ditika Bhanushali, Seher Khan, Mohammad Madhia, Shoumik Majum- dar (2018), "student performance prediction and analysis," in interna- tional journal of advanced research in Computer and Communication Engineering (IJARCCE), vol.7.

[6]   Osama Abdul jaleel Ali, Hussein Altabrawee, Samir Qaisar Ajmi (2019),"predicting students performance using machine learning techniques," injournal of university of Babylon for pure and applied sciences, vol.27, No.1.

[7]   Ivana Durdevic Babic (2017), "machine learning methods in predict-ing the student academic motivation," in Croatian operation research review(CRORR), vol.8, No.2.

[8]   M. Sitha Ram, V. Srija, V. Bhargav, A. Madhavi, G. Sai Kumar (2021), "machine learning based student academic performance prediction," in third international conference on inventive research on computing applications(ICIRCA), Coimbatore, India.

[9]   Preethi, S. Maheswari (2020), "prediction of student performance system using machine learning techniques," in international journal of innovative technology and exploring engineering (IJITEE), vol.9.

[10]   Roselyne Ayienda, Richard Rimiru, Wilson Cheruiyo (2021), "Predicting Students Academic Performance using a Hybrid of Machine Learning Algorithms" in Institute of Electrical and Electronics Engineers (IEEE) AFRICON.

[11]   H. M. Rafi Hasan, Mohammad Touhidul Islam, Syed Akhter Hossain (2019), "Machine Learning Algorithm for Student's Performance Pre-diction" in 10th international conference on computing, communication and networking technologies (ICCCNT).

[12]   Isma Farah Siddiqui, Qasim Ali Arain, Maleeha Anwar (2020), "Predic-itng student's Academic Performance Through Supervised Machine Learning " in international conference on information science and communication technology (ICISCT), Karachi, Pakistan.

[13]   Khalfan AI Mayahi, Mahmood AI-Bahri (2020), "Machine Learning Based Predicting Student Academic Success" in 12th international congress on ultra modern telecommunications and control systems and workshop (ICUMT).

[14]   Iti Burman, Subhrani Som (2019), "Predicting Students Academic Performance Using Support Vector Machine" in Amity International conference on Artificial Intelligence (AICAI), Dubai.

[15]   S. Tharsha, J. Dilogera, B.Mohanashiyaam, S. Kirushan (2021), "Ma-chine Learning-based Prediction Model for Academic Performance" in 3rd international conference on advancements in computing (ICAC), Colombo, Sri Lanka.

level using machine learning".