How to Cite:
Preetha, S., & Anitha, D. (2022). Prediction of academic performance of students using machine learning. *International Journal of Health Sciences*, *6*(S1), 11772–11779.
https://doi.org/10.53730/ijhs.v6nS1.7868

Prediction of academic performance of students using machine learning

Dr. S. Preetha
Associate professor, Department of Computer Science, Sri Ramakrishna College of Arts & Science for Women
Email: preethacs@srcw.ac.in

Dr. D. Anitha
Associate professor, Department of Computer Science, Sri Ramakrishna College of Arts & Science for Women
Email: anithacs@srcw.ac.in

***Abstract*---**Data Mining is the process of analyzing data from different perspectives and summarizing it into useful information. It is the method of extracting patterns and drawing inferences from large and complex datasets. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Educational Data Mining (EDM) refers to techniques, tools and research designed for automatically extracting meaning from large repositories of data generated by or related to people learning activities in educational system settings. One of the most important functions of EDM is to predict student performance based on past activity. This could include looking at past CGPA scores or internal assessments, student demography (income levels, school type etc.) or extracurricular activities.

***Keywords*---**datamining, machine learning, logistic regression.

Introduction

The student result prediction based on the previous performance is carried out using machine learning techniques. Machine learning is a trending concept that outperforms data mining. The concept behind machine learning is that a model or a machine that automatically learns from past experiences to predict the unseen data. The algorithms of machine learning analyze a large amount of data to predict new information. The application of machine learning has a huge impact in the field of education, medicine, automobile, etc, improving the student performance and enhancing individuals' outcomes.

Method

IoT is an innovative technological platform to fight with COVID-19 pandemic and can fulfil significant challenges during the lockdown situation. This technology is helpful to capture the real-time data and other necessary information of the infected patient shows the significant process.
Educational data mining is used to developing methods for discovering knowledge from data that come from educational environment. Educational Data Mininghas an important role in students academic performance prediction. In this paper, there is a evaluation between students association rule mining algorithm, K-means clustering algorithm and Decision tree.

Figure 1.1

There are many technique used to predict students performance. It includes K-means clustering, decision trees, association rule mining algorithm etc. But these all have different efficiencies.

Clustering algorithm and decision tree of data mining technique is helpful for future prediction of students performance. Data clustering is a method of extracting valid, unknown, useful and hidden patterns from large data sets. It is most widely used technique for future prediction

Existing System

There are many data mining methods for the prediction of students academic performance. It includes the following types:

- Decision trees
- Clustering
- Association rule

Decision trees

A decision tree is a predictive way used in clustering, classification and prediction tasks. It is appropriate for small data sets because it is suitable for discovering knowledge derived from the experimental data. Decision tree is faster to make and easier to know.

Decision tree technique uses tree arrangement to build classification models. In this technique dataset is divided into smaller subsets and at the same time an associated decision tree is incrementally developed.

Trees root and each internal node are labeled with a question. The arcs from each node represent each possible answer to the associated question. That result in a tree having decision and leaf nodes. A decision node is one which has two or more branches. Leaf node represents a prediction, decision or classification. The root node known as a best predictor isthe uppermost decision node in a tree.

Decision trees hold both numerical data and categorical data. Decision tree is a well-known and effective technique which builds classification models in the form of a tree. A decision tree is developed through recursive methods which break down the set of training data into separate groups with the objective to maximize space among groups.

The final result is a tree with leaf nodes and decision nodes where the leaf represents a prediction, decision or classification.

Classification and prediction also play a very interesting function in the field of education. Different decision tree algorithm can be applied to predict and identify the failure risk of student at education level. Predicting the academic result of a student desires lots of parameters to be considered. Predicted students academic performance using the CGPA grade system where the data set comprised of the students gender, his parental education details, his financial background etc. In [8] the author has explored the various variables to predict the students who are at risk to fail in the exam.

The solution strongly suggests that the earlier academic result strongly plays a main role in predicting their current result. In accordance with [9], the marks obtained by the students during the internal examination will play an essential role in predicting the outcome of the student in the main examination. The internal marks for the subjects MCA11, MCA12, MCA13, MCA14, MCA15 for a maximum of 100 marks and a result of Pass/Fail depending upon a minimum of 50 marks from each subject is fed as input and a decision tree is obtained using C4.5.The output should compared with the original marks received and result obtained by the student in the university examination

Figure 1.2

Proposed System

The proposed system for predicting the student final system has been designed in two methodologies:

- Prediction of student's 6th semester percentage
- Predicting whether the student eligible to get the degree or not

The first methodology is to predict the student's 6th semester percentage .the student data set like name,regno,class,all the five semester percentage marks are collected .out of these 15 attributes the reg.no, mark is taken for the prediction.in this method is based on the classification problem .Hence the output of the prediction will be known in the advance. The student's five semester mark percentage is Calculated and based on the pattern of every individual scoring in the sixth semester mark is predicted approximately.

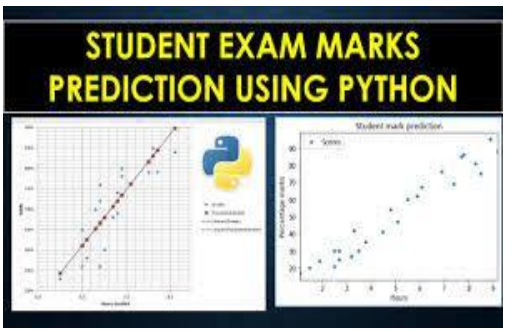

Figure 1.3

The method of the prediction:

(i) Method 1

This project is compiled using the google colab. The logistic regression is used for predicting. The logistics regression modelis the statistical model that models the probability of one event taking place by having the log adds for the event is a linearcombination of one or more independent variables.it estimates in the logistic model. The collected data set is divided into testing and training dataset.

the testeddata is used for the prediction of percentage marks. The project is designed using the python code and pandas and numpy libraries are used.

(II) Method 2

The second method is predicting whether the student will obtain the degree or not.This method is also compiled on the google colab using thelogistic regression. It is also a classification based-problem. The difference is that in this method the binary values are given as the classification variable. If it is "0 "the student is not eligible to get the degree. If the value is "1" then the student is eligible to get the degree. This criteria is fixed based on the percentage of marksobtained till the semester 5, the python coding are used for designing the project. The condition for pass percentage is above 30% .if the studentobtains the mark below the 30% then they will be not eligible to get the degree.

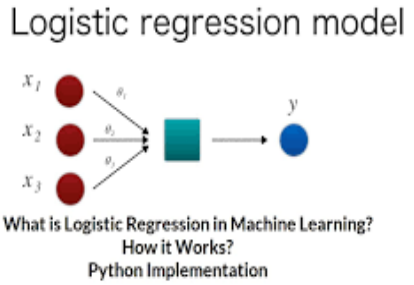

Figure 1.4

Experimential Work Of The Project

The working steps of both the projects are the google colab an open source and online tool is used for prediction the pandas and numpy libraries are imported now we need to save the collected data set in the .csv format and upload it into the google colab IDE.it is an online tool and the data gets stored in the server now print the loaded data set and also print the first and last 5 records of the data set for verification this is done using the headand tail commandassince python is rich of libraries use the python language to make the prediction.The sklearn model is imported to split up the train and test data setthe test datasize value is given as 0.25 where,out of 100 data sets the 25 % will be test data sets and75% will be the train datasets. The standard scalar in python is used for preprocessing and making the machine to learn the values to predict based on theloaded dataset. Logistic regression model is imported and since python is rich in libraries the prediction is done. The roll.no and the percentagemarks are taken for prediction.Using the "if statement" the prediction is performed and the output is printed.

The advantages of the project over the existing one

The existing systems use decision trees,clustering and association rules for predicting. while this project is based on the classification problemwere the result can be known in the advance.the KNN algorithm,naves bays are used in existing ones,but here the logistic regression is used .

Benefits of the project

This project will be useful to college institutions to track the students' performance and make them to improve

- Better results to institutions
  Each and every student can analyze their percentage patterns and focus on them to get the better results. For the institutions is useful to monitor and trach each student or the class student's performance and motivate them to get better results.

Figure 1.5

- To the faculty: The staff in charge can analyze the student's performance and make them to improve. Also the slow learners can be analyzed and separate attention can be given to them to get better results.

- To the parents: The parents can know about their child's pattern of scoring the marks

Conclusion

The paper focuses on the student academic growth and the analysis using the machine learning. Machine learning's effectiveness in predicting student output depends on the good use of algorithms for data. To achieve the best results, it is important to choose the correct machine learning approach for the right data.

Instead, it is possible to predict the result of the final year students through this trained machine since it shows more accuracy. So it can be concluded that this prediction system using machine learning techniques is one of the needs of the hour in this pandemic period.

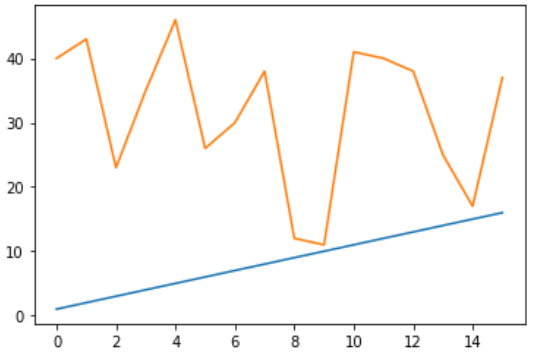

Figure 1.6 The visualized result of mark prediction

The Future Scope of the Proect

In future this project can be re-designed for predicting large amount of the dataset and other algorithms, predictive models can be used for the prediction .also in this project now it is predicted only for the integer values. Further it can also predicted for the float values.

References

1. N.V.Krishna Rao, Dr.N.Mangathayaru,Dr.M.Sreenivasa Rao," Evolution and Predictionof RadicalMultiDimensional E-Learning System with Clusterbased Data Mining Techniques",International Conference on Trends in Electronicsand Informatics,2017, PP.701-707.
2. Pushpa S.K, Manjunath T.N, "Class resultprediction using machine learning",InternationalConference On Smart TechnologyforSmart Nation,2017,pp1208-1212.Micheal Bowles, Machine Learning in Python:Essential Techniques for Predictive Analysis.John Wiley & Sons, Inc. 2015.
3. TrishulChilimbi, Yutaka Suzue, Johnson Apacible,and KarthikKalyanaraman. Project Adam:Building an efficient and scalable deep learning trainingsystem. In 11th USENIX Symposiumon Operating Systems Design and

Implementation (OSDI 14), pages 571–582, 2014.www.usenix.org/system/files/conference/osdi14/osdi14paper-chilimbi.pdf.

4. Jack Clark. Google turning its lucrative web search over to AI machines, 2015.www.bloomberg.com/news/articles/2015-1026/googleturning- its-lucrative-web-search-over-to-aimachines.
5. J. Xu, K. H. Moon, and M. Van Der Schaar, "A Machine Learning Approach for Tracking andPredicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol.11, no. 5, pp. 742–753, 2017.
6. K. P. Shaleena and S. Paul, "Data mining techniques for predicting student performance," inICETECH 2015 - 2015 IEEE International Conference on Engineering and Technology, 2015,no. March, pp. 0–2.
7. A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's PerformanceUsing Data Mining Techniques," in Procedia Computer Science, 2015.
8. Y. Meier, J. Xu, O. Atan, and M. Van Der Schaar,"Predicting grades," IEEE Trans. Signal Process.vol. 64, no. 4, pp. 959–972, 2016.
9. P. Guleria, N. Thakur, and M. Sood,"Predicting student performance using decision tree classifiersand information gain," Proc. 2014 3rd Int. Conf.Parallel, Distrib. Grid Comput. PDGC 2014,pp. 126–129, 2015.
10. P. M. Arsad, N. Buniyamin, and J. L. A. Manan,"A neural network students' performanceprediction model (NNSPPM)," 2013 IEEE Int. Conf. Smart Instrumentation, Meas. Appl.ICSIMA 2013, no. July 2006, pp. 26–27, 2013.