

STUDENT SEMESTER MARKS PREDICTION USING MACHINE LEARNING

Dr.M.CHANDRA KUMAR PETER

ASSISTANT PROFESSOR (SG)

DEPARTMENT OF SOFTWARE ENGINEERING

PERIYAR MANIAMMAI INSTITUTE OF SCIENCE & TECHNOLOGY -

VALLAM, THANJAVUR, TAMILNADU

J.SANJAIKUMAR

FINAL YEAR

DEPARTMENT OF SOFTWARE ENGINEERING

PERIYAR MANIAMMAI INSTITUTE OF SCIENCE & TECHNOLOGY -

VALLAM, THANJAVUR, TAMILNADU

Abstract:

Student marks Prediction is a way of predicting a student marks based on his/her previous marks. This also makes the teachers to know whether students are in a position to reach his/her expected marks or not. If this model shows that student needs to improve then that student can prepare more for that semester so that he/she can reach their expected marks or grade. Main objective for this project is to help teachers to analyze students performance easily and if needed they can help her/him to improve their student's performance by taking some actions like increasing their reading hours, giving some assignments etc..

Keywords — Student marks Prediction, analyze students performance, teachers to analyze students.

INTRODUCTION

This research looks at how machine learning applications effect higher education teaching and learning, as well as how to improve the learning environment. Data can be anything related to population, academic data of students, and interests of people. New data appears from time to time, as we can see.. Analysing the data is the difficult task for humans. So here comes the computer, which can analyse data faster than people because it is stored digitally and in a well-formatted manner. This is where the machine learning emerged. Machine learning is the branch of Artificial Intelligence that provides ability to automatically learn from past experiences. Here the machines do get programmed explicitly. It gives the computer the power to make humans and machines look alike in terms of learning, as the name implies.

Machine learning is divided into two categories based on the nature of the learning signal: supervised learning and unsupervised learning. This study focuses on supervised learning, more specifically on predictive analysis. When it comes to making predictions about future outcomes, predictive analysis is crucial. Predictive analysis has a wide range of applications. Predicting a student's academic success is critical since it can alert professors to students who may drop out of the course, and it can provide valuable information. Additional help to the scholars who want to enhance their educational performance. This have a look at is on implementation of system mastering in education. The outcome of this study is to predict student's academic performance. Students' data is utilised to create a model that can predict a student's academic achievement based on some background information. The dataset created by the students should be used as the study's input data.

I. LITERATURE SURVEY

"Implementation of Student SGPA Prediction System (SSPS) Using Optimal Selection of Classification Algorithm" [1] In today's world, there is competition in education institution every student plays a major role in the growth of the institution. An algorithm such as Logistic Model Tree, Random tree, and REP tree is used, the data

set collected from the university may contain errors and noises which make the model less effective so data cleaning is done and the data set will reduce to 236 instances from 260 records. The REP tree algorithm has given more accuracy with 61.70%.

"Machine Learning Algorithm for Student's Performance Prediction" [2]. The performance can be improved by predicting their marks by using the previous year's marks and can groom the students to improve themselves. By using machine learning techniques, we can improve the performance of every student the dataset of 1170 data was collected from three subjects. Algorithm such as KNearest Neighbors, SVC, Decision Tree Classifier, and Linear Discriminant Analysis. The decision tree classifier model has given the highest accuracy of 94.44%.

"Prediction of Student's Performance by Modelling Small Dataset Size" [3] An educational institution's major objective is to give its students a high-quality education. Early performance forecasting for students can help them earn better grades and get into prestigious schools. The machine learning classification algorithm such as Naïve Bayes, Support vector machines, K-nearest neighbor, and Linear discriminant analysis. The Linear discriminant analysis has given accuracy of 79%.

"Prediction of Student Academic Performance Using Neural Network, Linear Regression, and Support Vector Regression: A Case Study" [4] Institutions have a significant impact on academic and pupil success. In the final year, pupils' academic standing has a big impact on their future jobs. The algorithm used is Neural Network (NN), Support Vector Regression (SVR), and Linear Regression (LR). The dataset of 134 data was collected, the linear regression has shown more accuracy compared to other algorithms.

II. PROBLEM DESCRIPTION

Existing System

The existing system is created using algorithms like Random Forest, Decision Tree, Linear Regression. They use only one algorithm for the result and not compare with all three algorithms result, this result is only based on the single algorithm because of that the existing system gives less accurate result. They only used grade system and their prediction

factors are fixed because of this the accuracy of predicted result differs. In the existing the performance is very less.

Disadvantages of Existing System

- The faculty cannot find out students abilities and their interest easily so that they can enhance them in it.
- Thus it is able to have an effect on with poor college results, placement and career of individual.
- The effect is- it assists us from satisfying mission and vision of the institute.
- If the mission gets a success then it'll be wonderful assist for school to enhance training system.

Proposed System

The studies are targeted on predicting student's overall performance the usage of personalised analytics. This paper presents two different approaches to work on the thesis. The author's initial technique is the Regression Algorithm, which is a data mining function. The root mean square method is also used to calculate the regression algorithm's error rate. In this paper the author worked on how to improve the prediction algorithms which are used to analyze and predict the student's performance. The decision trees algorithm is used in this paper's work.

This paper proposed the student Academic performance prediction using Support Vector Machine. The author compared SVM to various machine learning approaches such as linear regression, Decision Trees, and KNN and determined that SVM outperformed them.

Advantages:

- To design a user-friendly web interface/ website on which the system can be implemented.
- To be able to make the performance prediction mechanism more efficient and accurate.
- To be able to predict the student performance using the MI algorithms
- The objective of this project is to use MI algorithm to study students' performance in the year. MI provides many tasks that could be used to analyze the student performance.

- In this research, the classification algorithm is used to evaluate student's performance and as there are many processes that are used for data classification, the decision tree method is used here.

III. IMPLEMENTATION

Data Collection

The data used in this research is collected from K S School of Engineering and Management. Real historic data of the students who studied undergraduate course in computer science and engineering is collected across 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017 batches. The dataset constitutes a total of 662 instances. The information of each student include percentage procured in all the 7 semesters (Sem_1=60, Sem_5=80) as attributes and grade, this particular student is likely to secure in the final semester, as target.

Data Cleaning

Irrelevant perceptions are any sort of information that is of no usage to us and can be eliminated instantly. Structural errors that emerge during estimation, transfer of data, or other comparable circumstances are removed.

Data Pre-processing

This phase in this study deals with the missing instances and other problems associated with the dataset. Initially the missing values are filled by calculating the mean with respect to the particular row. All the rows containing less than six non-NA values are discarded. If a particular row has missing values of more than two columns, then this row is deleted.

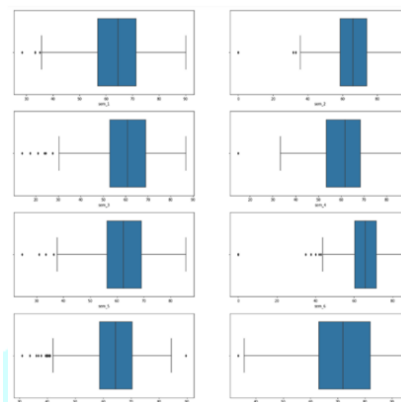


Fig: Box Plot

Outliers are the values which vary significantly from all the other values. These outliers are caused by measurement or execution error. Most of the data mining strategies eliminate these outliers, however, there are also some outliers which are found to be good type and these can be ignored. This study uses correlation and scatter plot to detect these outliers. The outliers in the dataset used for this study is displayed in above fig. IQR is utilized to measure variability by partitioning a data index into quartiles.

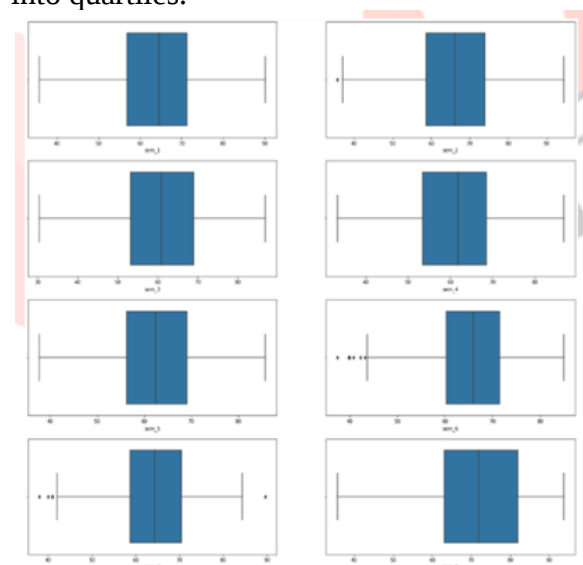


Fig : Box Plot without Outliers

The above fig. represents the box plot after removing the outliers using IQR method. The below fig. depicts the heat map that is created for this study. It shows that wherever there is a greater correlation between the points a lighter color is observed that is, 1.0 to 0.8. the darker color described that the points differ at a high level from each other. As observed in the above heat map eighth semester has a darker color because this column has a very less correlation with all the other columns in the dataset

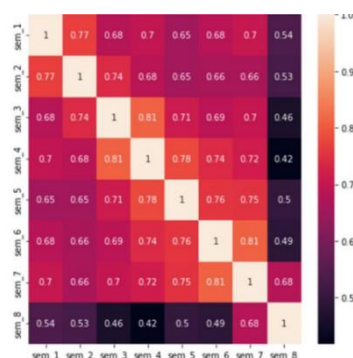


Fig: Heat Map

Data Scaling

There are numerous methods for doing this data scaling, this study uses MinMax scaler method. In this method, the lowest of feature is made equivalent to zero and the highest of feature is equivalent to one. MinMax Scaler compresses the data inside the given limit, typically of 0 to 1. It changes data by scaling features to a given limit. It scales the values to a particular value limit without changing the state of the primary distribution of the data.

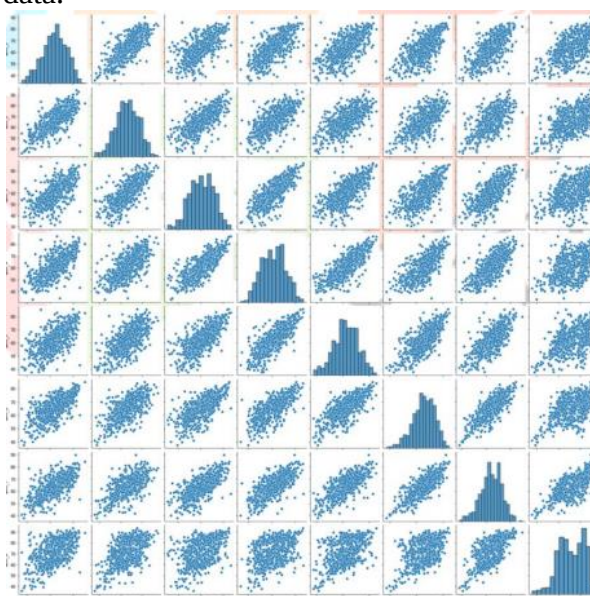


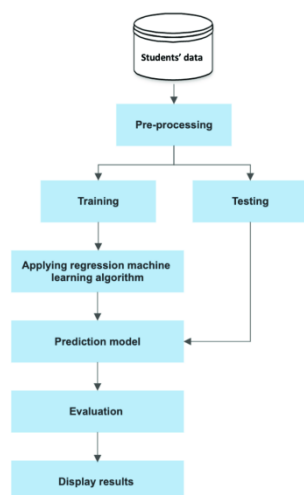
Fig: Pair Plot

In above Fig the correlation ranges between [50 - 80], positive co-relation is observed.

Cross Validation

In machine learning, we were unable to fit the model on the training data and can't say that the model will turn out precisely for the real data. For this, we should guarantee that our model got the right patterns from the data, and it isn't getting up a lot of noise. For this reason, the cross-validation

procedure is used. Cross-validation is a method wherein the model is trained using the subset of the data collection and afterward assess using the correlative subset of the dataset. This procedure primarily saves some piece of sample dataset, utilizing the rest of the dataset to train the model. The model is then testing using the piece of the dataset. This study uses the K-fold cross-validation technique for this procedure.



DATA FLOW DIAGRAM

IV. CONCLUSION & FUTURE WORK

The prediction of student performance is getting difficult day by day. In this research we have developed a linear regression based model which will help students in knowing final grade in particular subject. To accomplish this research, internal exam marks out of 30 are taken into consideration. Then the marks are converted into 100 (percentage) to have uniformity benchmark. These data is used to train the linear regression model to calculate the appropriate value of 0T and 1T. This model is a univariate i.e. it takes only one variable but it can be extended as multivariate model by adding more parameters to get more accurate results.

In future work, we can include more factors, and using more datasets the performance of the model can be improved and students' performance can be improved.

REFERENCES

- [1] L. Benko, J. Reichel, and M. Munk. Analysis of student behavior in virtual learning environment depending on student assessments. In 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA), page 16, Nov 2015.
- [2] U. bin Mat, N. Buniyamin, P. M. Arsad, and R. Kassim. An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. In 2013 IEEE 5th Conference on Engineering Education (ICEED), page 126130, Dec 2013.
- [3] Y. Y. Chen, S. Mohd Taib, and C. S. Che Nordin. Determinants of student performance in advanced programming course. In 2012 International Conference for Internet Technology and Secured Transactions, page 304307, Dec 2012.
- [4] B. E. V. Comendador, L. W. Rabago, and B. T. Tanguilig. An educational model based on knowledge discovery in databases (kdd) to predict learner's behavior using classification techniques. In 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), page 16, Aug 2016.
- [5] T. Devasia, Vinushree T P, and V. Hegde. Prediction of students performance using educational data mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), page 9195, March 2016.
- [6] S. Elnaffar. Using software metrics to predict the difficulty of code writing questions. In 2016 IEEE Global Engineering Education Conference (EDUCON), pages 513–518, April 2016.
- [7] J. Gamulin, O. Gamulin, and D. Kermek. Comparing classification models in the final exam performance prediction. In 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pages 663– 668, May 2014.
- [8] H. Goker and H. I. Bulbul. Improving an early warning system to prediction of student examination achievement. In 2014 13th International Conference on Machine Learning and Applications, pages 568–573, Dec 2014.
- [9] S. Huang and N. Fang. Work in progress: Early prediction of students' academic performance in an introductory engineering course through different

mathematical modeling techniques. In 2012 Frontiers in Education Conference Proceedings, pages 1–2, Oct 2012.

[10] K. Kaur and K. Kaur. Analyzing the effect of difficulty level of a course on students performance prediction using data mining. In 2015 1st International Conference on Next Generation Computing Technologies (NGCT), page 756761, Sept 2015.