Application of Decision Tree Algorithm for prediction of Student's Academic Performance

Jeromie Reggie Ebenezer, R Venkatesan, K Ramalakshmi, Jewel Johnson, Glen P I, Vibin Vinod

Abstract: With the increasing need of imparting quality education and evaluating students on all round development and performance, researchers and academicians have started applying data mining in the educational field, which now is also called as Educational Data Mining. The technique of educational data mining is used to evaluate and to extract facts, hidden knowledge and pat-terns from the available data of students which are used by researchers, academicians, educational panels to improve academic policies and student's academic performance. Student details such as gender, marks, attendance, attitude towards study, family income etc. have been used by various re-searchers to analyze and extract facts to improve the education system. In this paper we use student's attendance and various marks scored to evaluate and predict their performance. The model being created will be very useful for all the stakeholders of education system in improving student performance, teaching methodologies and Institution ranking.

Index Terms: Academic Performance, Data Mining

I. INTRODUCTION

A well-educated population is necessary for the growth of a country. With large number of investments in educational field, the quality of education being imparted has taken a hit. With increase in rote learning and demand of well-formed minds, researchers, academicians and governments have started investing in re-searches to find solutions and methods to improve the process of imparting education and to improve student's academic performance without affecting the quality of the education.

The use of Educational data mining has been in the field for quite long time. Each student every year generates a lot of data which once analyzed can give a great input about their academic growth and performance. This paper attempts to show the capability of educational data mining by producing a data mining model which can predict the academic performance of a student.

This model then can be used by academicians, educational institutes to improve the process of imparting knowledge and improve their students' academic performance. it helps in early prediction of student's performance which can then be

Jeromie Reggie Ebenezer, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

R Venkatesan, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

K Ramalakshmi, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

Jewel Johnson, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

Glen P I, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

Vibin Vinod, Department of Computer Science and Technology, Karunya Institute of Technology and Sciences, Coimbatore, India.

used by mentors or teachers concerned to alert and help the student improve. Early detection which give enough time for students and teachers to work together to improve the performance. The prediction of performance can also be used to analyze the effectiveness of new methods of imparting knowledge or evaluation methods hence helping researches and academicians to formulate different models and to help in improving educational policies to suit student's needs and improve the methods of imparting knowledge and evaluation of students.

In this work, we propose to use data mining approach to predict student's academic performance. The prediction is done using the academic record of students with the help of Decision Tree.

II. LITERATURE REVIEW

The essential need of filtering out students who might be weak in certain subjects by comparing their performance from their graduate degree evaluation and using the current performance review obtained from the faculties in real-time. This is a strategy that is used to maintain the educational standards and uplift the quality of education, prominently in private institutions. This method makes use of the Educational Data Mining (EDM) Technique.

[1]2014, Arora, Rakesh and Dharmendra Badal used Association analysis algorithm to evaluate the scores of the graduate degree and current performance in post graduate courses which will help in uncovering information that queries and reports won't be able to provide.

[2]2014, A. Acharya, D. Sinha use the concepts of Decision Tree in obtaining a refined data set and further there are many factors that weigh on the final result. Them being, previous academic records, economic status, family background, mid semester performance and many other available or relevant data.

The academic staff and other relevant roles play an important part in it too by putting in their best effort to fulfil their responsibilities. These activities will give out an enormous amount data which can be used in a recurring cycle which progressively improves the educational system by suggesting factors and areas of improvement. [3]2012, Osman Begovic E, Mirza S used Naïve Bayes algorithm to classify data which is based on the principles of probability and assumption. The Decision Tree in the end is used to give out a clear projection of the data sets acquired.

[4]2005, Rasmani, Khairul A., and Qiang Shen used fuzzy approaches to classify student's academic performances. Instead of using methods which solely depended on expert opinions, they used fuzzy rule-based models

which combined the expert opinions as well as the knowledge extracted from the student's data.

[5]2009, Zekic-Susac M, Frajman-Jaksic A, Drvenkar N dealt with creation of model for prediction using neural networks and decision tree and analyzed fac-tors which affected a student's performance. The model they created was based on available demographic data of students, their behavior and attitude towards studying like the importance of grades for them, attendance during exercise and tests to classify students into two categories of success.

[6]2010, Ayesha, Shaeela, Tasleem, Ahsan, Inaya used K-means clustering method to extract knowledge from educational data of students. They used class test, mid and final exams, assignments to analyse and evaluate hence helping teachers to improve students' performance.

[7]2006, S. T. Hijazi, and R. S. M. M. Naqvi used linear regression with factors like student's attitude towards attendance, marks, time allocation studies, parent's income, parents' education level to analyze student's academic performance based on socio-economic and psychological factors.

[8]H Agarwal and Harshil M used Neural Network to predict student performance. They used marks of 6 subjects of 80 students for this purpose. They concluded that performance of students is primarily dependent on their past performance.

III. DECISION TREE DATA SET SELECTION AND ANALYSIS

In this study we used data from the 2013-2017 and 2014-2018 undergraduate batch of Computer Science and Engineering Department. We used their Subject wise attendance, internal and external examination marks from first year till the final year. We also included the marks scored in class 12th board exams. In the attendance data, we also included medical leaves and On Duty leaves and were appropriately compensated during the prediction.

The marks were divided into 5 classes – Class A (95-100%), Class B (85-94%), Class C (75-84%), Class D (65-74%), Class E (60-64%), Class F (45-59%) and Class G (0-44%). This was done to classify all marks using a single method as classification differs both at School Board level and College level.

Table III contains the Grade which were given to the newly classified marks.

Similar class intervals were also used for subject wise attendance of the students. Special considerations were given for days with Medical leaves and On Duty leaves.

A sample dataset is given in the following Table I, II, and III.

Table 1. 12th Board Marks

Subject	Marks	Grade
English	95	O
Physics	88	S
Mathematics	75	A
Chemistry	80	A
Computer Science	87	S

Table II. Semester Marks

Name of Subject	Internal 1	Internal 2	Internal 3	Grade
Chemistry	38	35	37	S
Basic Electronics	34	30	31	A
Basic Mathematics	27	30	29	B
Basic Mechanical	31	37	31	A
Physics	35	37	35	S

Table III. Grade Point

Grade	Grade Point	Marks
O	10	≥ 95
S	9	≥ 85 & < 95
A	8	≥ 75 & < 85
B	7	≥ 65 & < 75
C	6	≥ 60 & < 65
P	5	≥ 45 & < 60
R	0	< 45

IV. PROPOSED METHODOLOGY

For this study the data has been collected from the University Database. The original names and registration number have been replaced with dummy codes.

The dataset collected can be messy and inconsistent. Hence, they are not fit for use for data analysis. Data is first processed where incomplete, inaccurate or corrupt records are found. Then we perform Data Cleaning where we replace or correct the records. Data which aren't needed for our analysis are also removed in this process. The cleaned data set is used for data analysis.

Various algorithms like Linear Regression, Decision Tree, Neural Network, Naive Bayes etc. are used to extract knowledge from datasets.
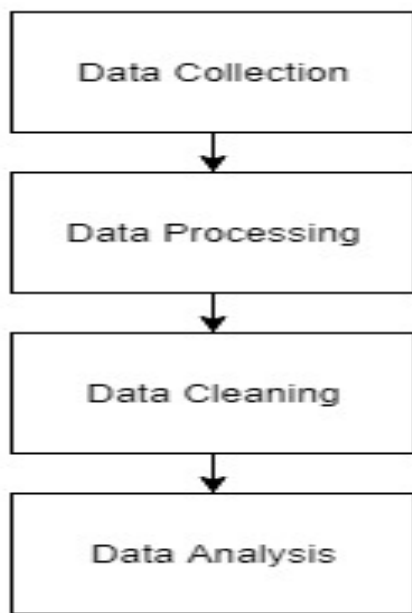
Fig 1: Data mining flowchart

A. Classification

[9]Classification is a method of categorizing objects of a dataset. It uses a pre-classified set of data to create a model which can classify a large amount of given data. This approach consists of two processes- Learning and Classification. In Learning, a training data set is used. The data set is then analyzed by the classification algorithm. In Classification, a test dataset is used, using which an accuracy of classification rules is estimated. If the accuracy of the classification rules is acceptable, then these rules can be used for new datasets. There are many classification algorithms available such as Naïve Bayes Classifier, Neural Networks, Decision tree, Nearest Neighbors etc. In this paper we will be using ID3 algorithm of Decision Tree.

We divided all the subjects the students appeared for into different group. In each group the subjects present are related to each other. Some subjects require prerequisite knowledge in certain subjects. We have ordered and grouped the subjects in a certain order. For example, in Table IV the mathematics subject group has been ordered according to its teaching order.

Table IV. Subject Group

Order	Subject Name
1	Basic Mathematics
2	Calculus and Statistics
3	Mathematical Transforms
4	Discrete Mathematics
5	Probability, Random Process and Numerical Methods

Hence from the given table we can derive that each of these subjects require the knowledge of subjects above them as pre-requisite.

Following this pattern all subjects can be divided into different groups and the marks scored can also be associated with it. Using the marks scored in earlier subjects we can predict the outcome of the newer subjects.

V. CONCLUSION

In this paper we analyzed the methodologies and basic concepts of the algorithms applied by other researchers. Using the newly gained knowledge we used a dataset which included marks of related subject and used decision tree algorithm to predict student's academic performance. We also found that marks of related subjects can affect the performance of the students in a given subject.

This study will further help different stakeholders of educational sector to improve student performance.

REFERENCES

1. R. K. Arora and D. Badal, "Mining Association Rules to Improve Academic Performance," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 1, 2014.
2. A. Acharya and D. Sinha, "Early Prediction of Students Performance using Machine Learning Techniques," *Int. J. Comput. Appl.*, vol. 107, no. 1, pp. 975–8887, 2014.
3. E. Osmanbegović, M. Suljic, and M. Suljić, "Data Mining Approach for Predicting Student Performance See Profile Data Mining Approach for Predicting Student Performance," *Econ. Rev. Econ. Bus.*, no. 1, 2012.
4. K. A. Rasmani and Q. Shen, "Subsethood-based Fuzzy Rule Models and their Application to Student Performance Classification," *14th IEEE Int. Conf. Fuzzy Syst. 2005. FUZZ '05.*, pp. 755–760, 2005.
5. M. Zekić-Sušac, A. Frajman-Jakšić, and N. Drvenkar, "Neuronske mreže i stabla odlučivanja za predviđanje uspješnosti studiranja," *Ekon. Vjesn.*, vol. 2, pp. 314–327, 2009.
6. S. Ayesha, T. Mustafa, A. Sattar, and M. Inayat Khan, "Data mining model for higher education system," *Eur. J. Sci. Res.*, 2010.
7. S. T. H. and S. M. M. R. Naqvi, "Factors Affecting Students' Performance," *Bangladesh e-Journal Sociol.*, vol. 3, no. 13621, pp. 1–17, 2006.
8. Havan Agrawal and Harshil Mavani, "Student Performance Prediction using Machine Learning," *Int. J. Eng. Res.*, vol. V4, no. 03, pp. 111–113, 2015.
9. A. B. E. D. Ahmed and I. S. Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method," *World J. Comput. Appl. Technol.*, vol. 2, no. 2, pp. 43–47, 2014.

AUTHORS PROFILE

Jeromie Reggie Ebenezer Currently pursuing his B. tech in Computer Science and Engineering at Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India.

R Venkatesan has completed his B.E. degree in Madurai Kamaraj University, M.E. with distinction from Anna University Chennai and Doctorate in Information and Communication Engineering. Completed his Ph.D. in Task Scheduling in computational grid environments. His area of interest is Task scheduling in computational environments, big data, data analytics, and internet of things.

K Ramalakshmi has completed her B.E. degree in Madurai Kamaraj University, M.E. with distinction from Anna University Chennai and Doctorate in Information and Communication Engineering. Completed her Ph.D. in wireless sensor network and datamining. Her are of interest is database technology, datamining, big data analytics, wireless sensor networks, IOT.

Jewel Johnson Currently pursuing his B. tech in Computer Science and Engineering at Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India

Glen PI Currently pursuing his B. tech in Computer Science and Engineering at Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India

Vibin Vinod Currently pursuing his B. tech in Computer Science and Engineering at Karunya Institute of Technology and Science, Coimbatore, Tamil Nadu, India