Academic Performance Prediction of Undergraduate Students using Decision Tree Algorithm

Rashmi V. Varade¹, Blessy Thankanchan²
¹,² Jaipur National University, Jaipur, India; *e-mail : rashmivvarde@gmail.com

ABSTRACT
Data mining is a technique for extracting meaningful information or patterns from large amounts of data. These techniques are frequently utilised for analysis and prediction in practically all fields around the world. It’s employed in a variety of fields, including education, business, health care, fraud detection, financial banking, and manufacturing engineering. This study explores the Decision Tree data mining methodology for predicting undergraduate students’ academic performance.

Key Words: Data Mining, Decision Tree, Classification, Educational data mining, KNN.

INTRODUCTION
A vast amount of data is generated in our educational system. This is the information about students’ academic performance, as well as information about students’ online assessments, from which we may extract useful information using data mining techniques.

We gather student data from the education system and use methodologies to analyse the data and the educational system for the betterment of students in educational data mining. We can estimate students’ academic performance, placement, and dropout rate using educational data mining, as well as which course they should take and which subjects or areas they need more attention in.

This data may be analysed by a large number of people and is crucial for the institution, teachers, and students.

We can forecast students’ academic performance using data mining methods such as Decision tree, KNN, and Nave Bayes.
techniques are used to analyse massive amounts of data in order to find hidden patterns and relationships that might aid decision-making. While data mining and database knowledge discovery are frequently used interchangeably, data mining is a distinct discipline. [1]

**DECISION TREE**

A decision tree is a flowchart-like tree structure in which each internal node represents a test on an attribute, each branch represents a test outcome, and each leaf node (or terminal node) represents a class label. The attribute values of a tuple X are evaluated against the decision tree given a tuple X.

A path is traced from the root to a leaf node that contains the tuple's class prediction. Decision trees are simple to translate into categorization rules. A decision tree is used as a predictive model in decision tree learning. It translates observations about an object to conclusions about the item's goal value. It's a type of predictive modelling used in statistics, data mining, and machine learning.

Classification trees are tree models in which the goal variable can take a finite set of values. In this tree structure, leaves indicate class labels and branches indicate feature combinations that lead to those class labels. When compared to other categorization approaches, constructing a decision tree is rather quick. [2]

**CLASSIFICATION BY DECISION TREE**

Three parameters are used to call the algorithm: D, attribute list, and attribute selection method. D is known as a data partition. It's the entire collection of training tuples and their associated class labels at first. The parameter attribute list is a collection of tuple-descriptive attributes. The attribute selection technique outlines a heuristic technique for choosing the attribute that “best” describes the situation.

An attribute selection measure, such as information gain or the Gini index, is used in this approach. The attribute selection measure determines whether the tree is strictly binary. The binary nature of the resulting tree is enforced by some attribute selection criteria, such as the Gini index. Others, such as information gain, do not, allowing for multiway splits (i.e., the growth of two or more branches).[3]

**ALGORITHM**

1) Create a node A.

2) If all the tuples in the partition are of the same class, then return A as a leaf node labelled with that class.

3) If attributes list is empty then return A as a leaf node labelled with the most common class in samples.

4) Identify the splitting attribute so that resulting partitions at each branch is as pure as possible.

5) Label node A with splitting criterion which serves as test at that node.

6) If splitting attribute is discrete valued then remove splitting attribute from attribute list.

7) Let Bi be the partitions created based on the i outcomes on splitting criterion.

8) If any Bi is empty then attach a leaf with the majority class in the partition to node A.

9) Else recursively apply the complete process on each partition.

10) Return A.

**RELATED WORK**

R. R. Kabra and R. S. Bichkar employed the decision tree method to predict engineering student success. They come to the conclusion that past academic performance can be utilised to build a model using a decision tree algorithm that can be used to predict student success in the first year of engineering test.[5]

The research of Raza Hasan1, Sellappan Palaniappan2, Abdul Rafi q Abdul Raziff 3, Salman Mahmood4, and Kamal Uddin Sarker5 looked into 8 distinct classifier algorithms that could help stakeholders enhance and improve early intervention to improve module results and increase student experience. [6]

Ms. AnkitaRaut and Dr. Anjali B Raut According to a Nichat study, identifying pupils at risk of poor performance early allows management to take proactive steps to improve their performance through extra coaching and counselling.[7]
Data mining techniques were used by Monika Goyal and Rajan Vohra to increase the efficiency of a higher education institution. When data mining techniques like clustering, decision trees, and association are used to higher education processes, it can help with student performance, life cycle management, course selection, retention rate measurement, and grant fund management. [8]

Data Mining, according to Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, could be utilised to improve business intelligence processes, including education systems, to improve efficacy and overall efficiency by maximising the use of existing resources. Students' academic achievement, as well as their entire personality development, could be enormously increased by thorough. [9]

The study by Pooja Thakar, Anil Mehta, and Manisha reveals that a significant amount of effort is put into analysing and predicting academic success, but that these efforts are separated. A coordinated strategy is clearly required. Aside from academic characteristics, there are a slew of other factors that play a part in prediction, including non-cognitive elements (a set of behaviours, skills, and attitude) [10]

**DATA SELECTION AND PREPARATION**

Since any educational institutions performance is highly dependent upon students' academic success rate. So, in this study the data is collected from the undergraduate degree colleges which are affiliated to university of Mumbai, nearly 600 student's data is collected. The student's data is collected for the year of 2019-20 and 2020-21.

The attributes like stream, age, religion, parents' education, parents' financial status, marital status and past academic performance is taken for this research.

Following attributes are taken for the research-

**Steam:** -For undergraduate there are three streams' Arts, Commerce, Science

**Age:** -Age of the student

**Gender:** -Gender of student e.g. Male or Female

**Attendance:** -Attendance of the student

**Educational Gap:** -Is there any gap during education

**Internet Connection:** -is there internet connection at home or not.

**SSC %:** -percentage obtained at SSC

**HSC %:** -percentage obtained at HSC

**Travel time:** -Travel time from home to institution

**Admission category:** -Scholarship or fees paid

**Family Income:** -Income of the family

**Family size:** -Size of the family

**Fathers’ qualification:** -Qualification of the father

**Mothers Qualification:** -Qualification of the mother

**Fathers’ occupation:** -occupation of father.

**Mothers’ occupation:** -Occupation of mother

**Frinds:** -how many friends are there?

**Maritila status:** -marital status of the student.

**Sem1, sem2, sem3, sem4:** -semester exam marks

**Remark:** -based on marks remark is given

| Name of the attribute | Values |
|-----------------------|--------|
| Age                   | 18+    |
| Gender                | Male or Female |
| Attendance            | <=75% |
| Educational Gap       | 1 or 2 or 3 years |
| Internet Connection   | Yes or No |
| SSC %                 | Above 75% From 60 to 75% From 45 to 60 Pass 35 to 45 % |
| HSC %                 | Above 75% From 60 to 75% From 45 to 60 Pass 35 to 45 % |
| Travel time           | 1 or 2 or 3 hrs |
| Admission category    | Fees paid or scholarship |
| Family Income         | Poor, good, average |
| Family size           | Small, medium, big |
| Fathers’ qualification| Literate or illiterate |
| Mothers Qualification | Literate or illiterate |
| Fathers’ occupation   | Working or business or not working |
| Mothers’ occupation   | Working or business or not working |
| Frinds                | Few or more |
| Maritila status       | Yes or No |
| Sem 1 (CGPA)          | Up to 10 |
| Sem 2 (CGPA)          | up to 10 |
| Sem 3 (CGPA)          | up to 10 |
| Sem 4(CGPA)           | up to 10 |
| Remark                | 1 to 3 Poor 4 to 7 good <7 to 10 Excellent |

**Result:** -Students created a decision tree. The academic success of pupils may be predicted with 100% accuracy using a csv file. Their academic grades
are crucial in forecasting their future success. Data from around 600 students has been obtained. The figure below displays the accuracy of a decision tree algorithm in predicting a student’s academic success.

| Actual | No. of students | Predicted | No of students |
|--------|----------------|-----------|---------------|
| Poor   | 200            | Poor      | 200           |
| Good   | 200            | Good      | 200           |
| Average| 200            | Average   | 200           |

**Figure 1.1:** Confusion matrix of three class prediction

**CONCLUSION**

This study demonstrates that the decision tree algorithm may accurately predict a student’s academic achievement. With this algorithm, the marks of first- and second-year students are taken into account and analysed. Now, students who fall into the POOR group can be given more attention to enhance their academic performance, and teachers can devise activities to improve such students’ perceptual abilities.

**REFERENCES**

[1] Brijesh Kumar Baradwaj and Saurabh "Mining Educational Data to Analyse Students Performance"International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011

[2] Himani Sharma, Sunil Kumar2"A Survey on Decision Tree Algorithms of Classification in Data Mining"International Journal of Science and Research (IJSR)ISSN (Online): 2319-7064

[3] J. Han, M. Kamber, Data Mining Concepts and Techniques, Second edition, Morgan Kaufmann, SanFrancisco, ISBN: 978-81-312

[4] Han, M. Kamber, Data Mining Concepts and Techniques, Second edition, Morgan Kaufmann, SanFrancisco, ISBN: 978-81-312

[5] R. R. Kabra , R. S. Bichkar "Performance Prediction of Engineering Students using Decision Trees" International Journal of Computer Applications (0975 – 8887) Volume 36– No.11, December 2011

[6] Raza Hasan1, Sellappan Palaniappan2, Abdul Rafi ez Abdul Raziff3, Salman Mahmood4 and Kamal Uddin Sarker5 "Student Academic Performance Prediction by using Decision Tree Algorithm" 2018 4th International Conference on Computer and Information Sciences (ICCOINS)

[7] Dr. Anjali B Raut and Ms. Ankita A Nichat" Students' Performance Prediction Using Decision Tree Technique"International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1735-1741

[8] P. Ajith, M.S.S.Sai, B. Tejaswi, Evaluation of Student Performance: An Outlier Detection Perspective, 2013.

[9] Sonali Agarwal, G. N. Pandey, and M. D. Tiwari"Data Mining in Education: Data Classification and Decision Tree Approach" International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012.

[10] Pooja Thakar, Anil Mehta, Manisha "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue"International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015