

STUDENT PERFORMANCE PREDICTION USING DATA MINING ALGORITHM

L. Poovarasi¹, N. Ramya²

Student, B.E - Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India¹

Student, B.E - Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India²

Abstract: In training machine evaluation and prediction of pupil overall performance is a hard task. on this paper, a model is proposed to are expecting the overall performance of students in an academic agency. The set of rules hired is a system gaining knowledge of technique referred to as Naïve Bayes and KNN. similarly, the importance of numerous extraordinary attributes is considered, in an effort to decide which of those are correlated with student overall performance. finally, the results of an experiment observe, showcasing the electricity of system learning in such an software. In attitude of this mission we are going to expect the student development and look at the greater end result thru device learning set of rules. We foresee the scholar overall performance by scanning their preceding instructional details. To execute this prediction, we've created a dataset, via the usage of this we can predict pupil information. in many corporations, records mining techniques are used for studying huge amount of to be had facts, records for choice making method. In instructional quarter, statistics mining is used for wide form of utility such as performance of the scholars like mark, attendance, personnel opinion, extracurricular activities, Ragging and strain. The facts mining strategies used for figuring out the performance of the scholar the use of Naïve Bayes and KNN algorithms. those algorithms pick out and analyses the overall performance of the student.

Keywords: KNN, Naïve Bayes, Statistics mining, Prediction, Performance.

I. INTRODUCTION

There are many tremendous improvement research interests in using data mining in educational sector. This modern emerging sector, called educational data mining, concerned with improved methods that extract knowledge from data come from the educational sector. Data mining is a technique of sorting which is actually used to extract hidden patterns from huge databases. This concept and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. Educational data mining is a new emerging or advanced technique of data mining that can be applied on the data related to the field of education. The data can be collected from past used data and operational data reside in the databases of educational institutes. The data of students can be personal information or academic performance. Further it can be achieved from e-learning database systems which have a huge amount of data and information used by most institutes. It uses many techniques for proper implementation of data mining concepts such as K-means Clustering and K-Nearest neighbour. Using these techniques different kinds of knowledge can be discovered using association rules, classification and clustering. By using this we extract knowledge that describes student performance in examination and all their detail information. From This huge amount of data, the first task is to sort them out, cluster analysis is to classify the raw data in a reasonable way. Clustering is a bunch of physical or abstract objects, as per the degree of similarity between them, divided into several groups, and makes the same data objects within a group of high similarity and different groups of data objects which are not similar.

1.1 OBJECTIVE

The main aim of this project is to improvise the student performance in studies based on some important factors. Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Now-a-days developing new methods to discover knowledge from educational database in order to analyze student's trends and behavior towards education. To analyze the data from different dimensions categorize it and to summarize the relationships. It motivated us to work on student dataset analyzation. The data collection, categorization and classification is being performed manually.

DOI: 10.17148/IJARCCE.2022.11651

1.2 SCOPE

the primary scope of this task is to apply statistics mining methodologies to study student's overall performance within the publications. data mining gives many duties that would be used to look at the pupil overall performance. in this research, the class project is used to assess scholar's overall performance and as there are numerous strategies which are used for facts classification, the choice tree method is used right here. information like Attendance, magnificence check, Seminar and project makes have been accrued from the scholar's control gadget, to predict the overall performance on the end of the semester. This paper investigates the accuracy of decision tree and Naïve Bayes strategies for predicting scholar performance.

II. ANALYSIS

2.1 SYSTEM ANALYSIS

System Analysis is a combined process dissection the system responsibilities that are based on problem domain characteristics and user requirement.

2.1.1 Problem Definition

- 1. provide a study of the prevailing ways in predicting students' success in a very course.
- 2. touch incorporate additionally non-academic options by making a model that predicts tutorial performance.
- 3. Compare and analyze with existing work.
- 4. Identify, visualize, and value key social factors that have an effect on a student's performance.

Our intensive review of previous surveys discovered that, to the most effective of our information, no systematic literature survey was allotted that specialize in the prediction of student tutorial performance from the training outcomes perspective. summarizes the distinguished surveys allotted on the prediction of student performance and emphasizes their focus and weaknesses. Indeed, our search came varied surveys on the utilization of information mining techniques in education to unravel student modelling activities and predict tutorial performance. These reviews suffered from many limitations, for they were usually broad, failed to target mistreatment student outcomes as associate indicator of student performance, suffered from quality problems (e.g., methodologies not totally defined), and weren't revealed in extremely indexed venues. These weaknesses square measure highlighted.

2.1.2 Existing System

Data mining techniques are being applied in the field of education to understand and predict student performance and this helps in focusing on the weak areas of a student as early as possible. In the current literature, the factors considered for predicting student performance are student performance in academics, teaching quality and learning methodologies. The existing system does not consider the other important elements that impact distorting performance such as the background of the student, student's involvement in social activities, the study habits of the student the learning patterns of the student, Attendance in school etc. • As of now, Existing system take only performance into consideration which is not sufficient for having system, which can help us to evaluate performance of a student. We are not having a system which would help us to integrate the performance and undesirable into consideration.

ALGORITHM USED - KNN, K-Means and Decision tree

DISADVANTAGES

- Existing system miss the undesirable data for the students.
- And it may not check the social data for the student.

•The accuracy of the model using few attributes like teaching quality, student performance and learning methodology is not satisfactory.

• Desired results are not achieved.

2.1.3 Proposed System

The work aims to develop a trust model using data mining techniques, which mines required information, so that the present education system may adopt this as a strategic management tool. The proposed system uses educational data mining techniques to evaluate performance and identify undesirable behaviour. In educational sector, Data mining is



ISO 3297:2007 Certified ∺ Impact Factor 7.39 ∺ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11651

used for wide variety of applications such as performance of the students like mark, attendance, staff opinion, extracurricular activities, Ragging and stress. The data mining techniques used for identifying the performance of the student using K-means and KNN algorithms. Educational database contains the useful information for Evaluating Students. The data mining techniques are more helpful in classifying educational database and help us in evaluating the performance and undesirable behaviour of a student.

ALGORITHM USED- KNN, K-Means clustering, Naïve Bayes

KNN- KNN (K — Nearest Neighbours) is one in each of many (supervised learning) algorithms used in processing and machine learning, it classifier formula where the coaching depends "how similar" could be a data (a vector) from different. K-nearest neighbours (KNN) formula uses 'feature similarity' to predict the values of recent datapoints that additional means that the new data are reaching to be assigned a price supported but closely it matches the points at intervals the work set.

STEP 1: BEGIN

STEP 2: Input: $D = \{(x1, c1), ..., (xN, cN)\}$ STEP 3: x = (x1...xn) new instance to be classified STEP 4: FOR each labelled instance (xi, ci) calculate d (xi, x) STEP 5: Order d (xi, x) from lowest to highest, (i = 1...N) STEP 6: Select the K nearest instances to x: Dkx STEP 7: Assign to x the most frequent class in Dkx

K-Means - k-means may be a technique for knowledge clump that will be used for unsupervised machine learning. it's capable of classifying untagged knowledge into a planned range of clusters supported similarities (k). The running of the K-Means set of rules is defined withinside the beneath steps:

Step-1: Select the quantity K to determine the quantity of clusters.

Step-2: Select random K factors or centroids. (It may be different from the enter dataset).

Step-3: Assign every statistics factor to their closest centroid, so one can shape the predefined K clusters.

Step-four: Calculate the variance and region a brand new centroid of every cluster.

Step-5: Repeat the 1/3 steps, because of this that reassign every datapoint to the brand new closest centroid of every cluster.

Step-6: If any reassignment occurs, then visit step-four else visit FINISH.

Step-7: The version is ready.

Naïve Bayes – Naïve Bayes algorithm is a supervised getting to know set of rules, that's primarily based on Bayes theorem and used for fixing category issues. It is specifically utilized in textual content class that includes a high-dimensional education dataset. It's far a probabilistic classifier, which means that it predicts on the idea of the probability of an item.

Bayes theorem affords a manner of calculating posterior opportunity P(c|x) from P(c), P(x) and P(x|c). Look on the equation below:



$$P(c \mid \mathbf{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \dots \times P(x_n \mid c) \times P(c)$$

To allow observe the below steps to perform it.

Step 1: Convert the data set into a frequency desk

Step 2: Create chance table through locating the probabilities like Overcast opportunity = 0.29 and chance of playing is 0. Sixty four.

Step three: Now, use Naive Bayesian equation to calculate the posterior opportunity for each elegance. The class with the highest posterior chance is the outcome of prediction.

© <u>IJARCCE</u> This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ∺ Impact Factor 7.39 ∺ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11651

ADVANTAGES

ΝM

- Educational database contain the useful information for evaluating students.
- The data mining techniques are more helpful in classifying educational database and help us in evaluating the performance and undesirable behaviour of a student.



III. SYSTEM DESIGN



IV. MODULES

- 4.1 Admin
- 4.2 Data Collection
- 4.3 Preprocessing
- 4.4 Data Extraction
- 4.5 Classification module
- 4.6 Prediction

4.1 ADMIN MODULE

In this section, admin can add the student data, Create a new marksheet, Remove the student data, View student data, Search student data, and can predict the accuracy.

Add student data: In this section, we can add the new student data.

Create Mark Sheet: In this section, we can create a marksheet by giving the marks for each subjects.

Remove Student Data: In this section, we can remove the particular record.

View Student Data: In this section, we can view all the student data.

Search Student data: In this section, we can search the particular student record.

prediction: In this section, we can predict the accuracy level of the student performance.

4.2 DATA COLLECTION

In this module, student data's will be collected from the college. Student's data like mark, attendance, staff opinion extracurricular activities, ragging and stress.

271

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ∺ Impact Factor 7.39 ∺ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11651

4.3 PREPROCESSING

Data pre-processing is finished to get rid of the unfinished clamorous and inconsistent knowledge. knowledge should be pre-processed before exploitation in feature choice task.

4.4 DATA EXTRACTION

Data Extraction upon applying the ultimate pool of elite studies was completely analyzed to extract the info that assist in responsive the analysis queries.

The extracted information included:

• General data regarding the publication, as an example, publication year, venue type, country of publication, and variety of authors;

• instructional dataset and context of prediction (e.g., students, courses, school, university, etc.);

• Input variables used for student outcomes prediction and also the kind within which they were predicted;

• Intelligent models and approaches used for the prediction of educational performance;

• important predictors of learning outcomes. the info were classified and categorized in line with the themes reportable within the results section. However, it had been not possible to hold out the meta-analysis of the chosen studies, in the main as a result of most instructional datasets were either personal or impractical to get. Below we have a tendency to detail the results of our synthesis analysis.

4.5 CLASSIFICATION MODULE

The data mining techniques used for identifying the performance of the student using K-means and KNN algorithms. These algorithms identify and analyses the performance of the student.

4.6 PREDICTION

In this module, to predict the student performance based upon student mark, attendance, staff opinion, extracurricular activities, Ragging and stress. This section reports general information about the surveyed articles, the forms in which the student outcomes were forecasted, the intelligent models developed for performance prediction, and the predictors of student attainment of learning outcomes.

V. RESULT AND DISSCUSSION

The students' performance prediction provides wonderful advantages for increasing student retention rates, effective enrolment management, alumni management, improved targeted promoting, and overall academic institute effectiveness. The intervention programs in colleges facilitate those students United Nations agency square measure in danger of failing to graduate. The success of such programs relies on correct and timely identification and prioritization of the student's requiring help. This section presents a written record review of revealed literature from 2009 to 2021, documenting at-risk student performance victimisation mil techniques. analysis studies associated with dataset kind, feature choice ways, criteria applied for classification, experimentation tools, and outcome of the planned approaches are summarized.

VI. CONCLUSION

In this paper, the classification task is employed on student info to predict the scholar division on the premise of previous info. As there square measure several approaches that square measure used for information classification, the Naïve Bayesian Classifier and Weighted Naïve Bayesian Classifier square measure used here. Information's like attending, Class test, Seminar and Assignment marks were collected from the student's previous info, to predict the performance at the tip of the semester. This study can facilitate to the scholars and therefore the lecturers to enhance the division of the scholar. This study also will work to spot those students that required special attention to cut back fail ration and taking applicable action for consequent semester examination. this could facilitate the scholars improve in their lecturers, that eventually ends up in an honest performance in their finish examinations. By this the suicide rates of scholars also will get reduced since the strain is reduced. this might facilitate in our country development by providing smart and economical engineers to the country.

ISO 3297:2007 Certified ∺ Impact Factor 7.39 ∺ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11651

REFERENCES

[1] M.Durairaj, C.Vijitha .Educational Data mining for Prediction of Student performance Using Clustering Algorithms. The data mining techniques are more helpful in classifying educational database Which contain the useful information for predicting a student's performance.

[2] Kin Fun Li, David Rusk and Fred Song.Predicting Student Academic Performance,The performance predictors, if identified, can then be used effectively to formulate corrective action plans to improve the attrition rate.

[3] E.Osmanbegovic, Mirza Suljic.Data Mining Approach For Predicting Student Performance,Using data mining the aim was to develop a model which can derive the conclusion on student's academic success.

[4] A. Banumathi, A. Pethalakshmi. A novel approach for upgrading Indian education by using data mining techniques.

[5] Anoopkumar M ; A. M. J. Md. Zubair Rahman, A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration.

[6] U. K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.

[7] Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009..

[8] Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005.

[9] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

[10] Kumar, V. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. IJACSA - International Journal of Advanced Computer Science and Applications, 2(3), 80-84. Retrieved from http://ijacsa.thesai.org.