International Journal of Science and Research (IJSR) ISSN: 2319-7064 SJIF (2020): 7.803

# Educational Data Mining Using Analysis Student Learning Process

# C. Iswarya<sup>1</sup>

<sup>1</sup>MCA., M.Phil, Assistant Professor, Department of Computer Science, PSR Arts and Science College, Sevalpatti, Tamilnadu, India

Abstract: Higher education institutions are often very interested to know around the success frequency of the students throughout their study. It is vital to study and analysis educational data especially students' performance analysis. Educational Data Mining (EDM) is the field of study disturbed with mining educational data to find out interesting outlines and knowledge in educational administrations. This learning is equally concerned with this subject, specifically, the students' performance. It contains machine learning algorithms and statistical techniques to help the user for understanding of student's learning habits, their academic performance and further enhancement if required. In this study explores several factors hypothetically assumed to move student performance in higher education, and finds a qualitative model which best naïve bayes classifies and predicts the students' performance level of students.

Keywords: Data Mining, Education, Students, Performance, Patterns

### 1. Introduction

Educational Data Mining (EDM) is a new leaning in the data mining and Knowledge Discovery in Databases (KDD) field which focuses in mining useful patterns and learning useful knowledge from the educational information systems, such as, admissions systems, registration systems, course management systems blackboard, online etc..., and any additional systems dealing with students at different stages of education sector, from schools, to colleges and varies universities. The Researchers in this field emphasis on discovering useful knowledge also to help the educational institutes accomplish their students better, or to help students to manage their education and deliverables improved and enhance their performance.

Analyzing the students' data and material to classify students, or to produce decision trees or association rules, to make better decisions or to improve student's performance is an interesting field of research area, which mainly focuses on analysing and empathetic students' educational data that indicates their educational performance, and makes specific rules, groupings, and predictions to help students in their future educational presentation.

In this paper proposed system discoveries out student leanings on the basis of products of students' academic performance, strengths and weakness, hobbies, extracurricular activities. Academic data includes unit test evaluated, student's theory, practical and term effort marks. This trend will help us to track where the students outshine and where not and what are their aptitudes which can be improved. The analysis will review the outcome and will classify students based on the results. This structure will recognize notice of student in particular areas. In this work future system will also order the students those who are eligible for job placements based on working company's measures. Classification is the greatest accustomed and most actual data mining technique used to categorize and forecast standard results.

### 2. Related Work

The Data mining used higher education is a recent research field and this part of research is fast popularity because of its potentials to informative establishments.

Baradwaj and Pal [1] conducted a research on a set of 50 students enrolled in a specific course program across a period of 4 years (2007-2010), with multiple routine indicators, with "Previous Semester Marks details", " Seminar marks" "Class Test", "Attendances", "class discussion", "Assignments", "Lab test", and "End Semester Marks". They used ID3 decision tree algorithm to finally construct a decision tree, and if-then rules which will eventually help the teachers as well as the students to better understand and forecast students' routine at the final of the semester. Additionally, they defined their impartial of this study as: "in this study will also effort to categorize those students which needed distinct attention to decrease fail ration and taking proper action for the subsequent in semester examination"[1]. Baradwaj and Pal [1] selected ID3 decision tree as their data mining technique to analyze the students' concert in the selected course program; because it is a "simple" decision tree learning algorithm.

Abeer and Elaraby [2] conducted a comparable research that primarily emphases on generating classification rules and predicting students' performance in a particular course program based on earlier recorded students' concert and activities. Abeer and Elaraby [2] processed and analyzed previously enrolled students' data in a specific course program across 6 years (2005–10), with several attributes collected from the campus database. The result, this study was able to predict, to a certain level, the students' final grades in the particular sequence program, as well as, "help the student's to improve the student's routine, to identify those students which needed different attention to reduce failing ration and taking appropriate action at accurate time" [2].

DOI: 10.21275/SR21605135641

### International Journal of Science and Research (IJSR) ISSN: 2319-7064 SJIF (2020): 7.803

Pandey and Pal [3] conducted a data mining technique research using Naïve Bayes classification to analysis, categorize, and predict students as performers or below executors. Naïve Bayes classification is a modest possibility classification technique, which accepts that all specified attributes in a data set is autonomous from each other, hence the name "Naïve". Pandey and Pal [3] conducted this research on a model data of students registered in a Post Graduate Diploma in Computer Applications (PGDCA) in Dr. R. M. L. Awadh University, Faizabad, India. The research was able to classify and predict to a certain extent the students' grades in their future year, based on their grades in the previous year. Their results can be working to support students in their upcoming instruction in many ways.

Bhardwaj and Pal [4] conducted a important information mining examine with the Naïve Bayes classification technique, on a collection of BCA students (Bachelor of Computer Applications) in Dr. R. M. L. Awadh University, Faizabad, India, who appeared for the finale checkup in 2010. The questionnaire was exposed and collected after each student before the final examination, which had several personal, social, and psychological questions that was used in the study to find relations among these features and the student's performance and results. Bhardwaj and Pal [4] identified their key objects of this study as: " (a) Group of a data source of predictive variable quantity; (b) Identification of different factors, which effects a student's learning recital and performance during academic occupation; (c) Construction of a prediction seamless using classification data mining techniques on the basis of recognized analytical variables; and (d) Validation of the developed model for higher education pupils studying in Indian Universities or Institutions" [4]. They initiate that the most manipulating factor for student's performance is his grade in oldest secondary school, which connects us, that those students who performed well in their secondary school, will definitely make well in their Bachelors study. Furthermore, it was found that the living location, medium of teaching, mother's qualification, pupil other habits, family yearly income, and student family status, all of which, highly contribute in the students' educational performance, thus, it can predict a student's grade or normally his/ her routine if basic personal and common knowledge was composed around him/ her.

Yadav, Bhardwaj, and Pal [5] conducted a qualified research to exam multiple choice tree algorithms on an educational dataset to organize the educational repetitive of students. The study mainly emphases on choosing the best choice tree algorithm from among frequently hand-me-down decision tree algorithms, and provide a standard to each one of them. Yadav, Bhardwaj, and Pal [5] initiate out that the CART (Classification and Regression Tree) decision tree classification method operated recovering on the tested dataset, which was particular based on the formed correctness and accuracy using 10-fold cross validations. This study presented a good practice of categorizing the best classification algorithm technique for a particular dataset; that is by challenging multiple algorithms and techniques before deciding which one will finally work recovering for the dataset in hand. Hence, it is highly advisable to test the dataset with multiple classifiers initial, then choose the greatest accurate and accurate one in order to agree the best classification technique for slightly dataset.

# 3. Data Mining using Classification Techniques

The decision-making procedure, Data mining is a very favorable and helpful method. Classification is a same simple and frequently used data mining technique. Knowledge of training data is mandatory for empathetic of Classification.

There are two phases of classification procedure:

- 1) Growth of a model for training
- 2) Estimating the prototypical using testing data

The basis of algorithms, dissimilar approaches of classification are:

- a) *Statistical based algorithms:* Statistical procedures are generally having an accurate fundamental possibility model which provides probabilities of being in each class somewhat than just a simple classification.
- b) *Correlation Analysis:* It is a statistical method used to find the degree of association between two numerically measured, continuous variables example phase and mass is related to each other.
- c) *Regression Analysis:* This method defines that how an independent variable is mathematically associated with independent variable
- d) *Bayesian Model:* This technique uses frequent is t technique. The core of frequentist technique is to apply probability to data. Bayesian calculations drive traditional for the probability of the proposition.
- e) *Distance based algorithms:* Each item plot to a particular class can be perceive as same as other items are previously present in that class and could be differentiated from the items of other classes.
- f) *Simple Approach:* In this method, an assumption is that each class is characterized by its center. A new entry can become a associate of a class with the option of largest parallel value.
- g) *K nearest neighbors:* It is a non-parametric method which rest on the use of reserve measurement All available cases can be stored in it and whenever a new case arrived, it can be categorize created on the distance purpose.
- h) *Decision tree-based algorithms:* Giving to this method, there is a condition of construction of a tree diagram to model classification development. Two steps are important in this method of classification:
  - To Size a tree named with figure Decision Tree
  - Implementation of Decision tree to catalog
- i) *Neural Network based algorithms:* In this method, a model is formed which delivers a format for data illustration. At the while of tuple classification, all characteristics associated to that tuple are transmitted into a graph.
- j) *Rule based algorithms:* In this method, classification might be done on the source of if then else rules for data classification.

## Volume 10 Issue 6, June 2021

<u>www.ijsr.net</u>

### Licensed Under Creative Commons Attribution CC BY

# 4. Basic Terminology in Classification Algorithms

- **Classifier:** In this method algorithm that plots the input data to a thorough category.
- **Classification model:** ampere classification model efforts to draw about assumption from the input ethics given for training. It will forecast the class labels or classes used for the novel data.
- **Feature:** A feature is a exact measurable belongings of a phenomenon actuality experimental.
- **Binary Classification:** In Classification process with two possible results. **Example: Gender classification split Male or Female.**
- Multi-class classification: The Classification through more than two classes. Popular multi-class classification, separately sample is allocated to one and only single target label. Example: A animal container be a cat or dog but non both at the equivalent time.
- Multi-label classification: Classification task where each sample is mapped to a set of target labels (more than one class). Example: The news event can be about game, a individual, and position at the same period.

#### 4.1 Classification Techniques Using EDM Work

Data mining holds a collection of picturing tools and algorithms for data study and predictive modeling, together with graphical user interfaces for easy access to these functions. The organization is not so much a only program as a collection of inter-dependent programs bound together by a common user boundary. Characteristically these fundamentals fall concerned in three classes: data usual giving out, machine information systems, and creation processing.

The proposed system summaries all the students' info after the database. Then chooses the selected data, divided the dataset into test and training sets and preprocess it as filter out features. Then the data is changed into the data mining and the estimation is completed

## 5. Proposed Work

The proposed system contains a worker profile creator to deliver worker interface, worker login, pupil performance analyzer, extracurricular activities and Academic performance, exam score sheet generator, passing measures sheet and semester-wise student performance. The students will be clustered into categories example distinction, first class, higher second class, second class and pass class based on transitory conditions.

- Admin module: This module is used for login for the administrator user of the system, the administrator can accomplish staff details and create manipulate the subject in students.
- Faculty module: This module is used for login to a staff can admission the system and can add the marks to separately student for their valued subject. Faculty can add student strength and weakness finally analysis the student performance.

• Student module: This module is used for login the student and view their marks.

#### Steps for proposed system:

Procedure 1: In this purposed system the employer will primary login to the system to initiate the procedure.

Procedure 2: Then the employer will get a form which covers fields like Student Name details, Student Rollno, student department, academic year, Semester results

Procedure 3: Then from individually of the field the employer will get a data of a student individual or Semester wise or department wise or Year wise.

Procedure 4: Then it will also comprise extra co-curricular activities completed by the students and from entirely that analysis the faculty will derived to know that the student is weak or strong and the student is eligible for the placements

#### Naïve Bayes Classification

NaïveBayesclassificationisamodestprobabilityclassificationt echnique, which accepts that entirely given features in a dataset is self-determining from each other, hence the term "Naïve" Bayes classification has been proposed that is created on Bayes rule of qualified probability. Bayes rule stands a technique to evaluation the likelihood of a property specified the established of data as indication or input Bayes rule or Bayes theorem is"[4]:

$$P(h|x) = \frac{P(x_i|h_i)P(h_i)}{P(x_i|h_i) + P(x_i|h_2)P(h_2)}$$

The Resulting are the explanation for each one of the interesting Bayes Probabilities:

- a) GENDER =Male: The probability of male students to get lesser grades are knowingly higher. Affecting from higher to lower grades, the probability growths.
- b) GENDER = Female: This scenario is opposite to the previous one, where the probability of female students to get higher grades are significantly higher. The probability decreases affecting from high to low grades.
- c) HSP = Excellent: Motivating enough, students who grown excellent grades in High School had high grades in the university as fine.
- d) MOCS= Service: Interestingly, when them other occupation status is on service, it appears that students get higher grades.
- e) DISCOUNT: As illustrated earlier, students with higher grades tend to get discounts from the university more than low grades students
- f) PSM Previous Semester Marks/Grade obtained in MCA course. It is split into five class values: First – >60%, Second –>45% and <60%, Third –>36% and < 45%, Fail<40%.</li>
- g) CTG In Class exam grade obtained. In separately semester two class tests are exposed and average of two class exam are used to analyze sessional marks. CTG is divided into three classes: Poor < 40%, Average > 40% and <60%, Good->60%.
- h) SEM Seminar Performance obtained. In individual semester seminar are ordered to check the performance of students. Seminar performance is estimated into three classes: Poor Presentation and communication skill is

## Volume 10 Issue 6, June 2021

<u>www.ijsr.net</u>

Licensed Under Creative Commons Attribution CC BY

low, Average–Either presentation is fine or Communication skill is fine, Good–Both presentation and Communication skill is fine.

- i) ASSNT- In Assignment performance. In Individual semester has two assignments stand specified to students by individual teacher. Assignment performance is divided into two modules: Yes-student submitted assignment, No-Student not submitted assignment.
- j) GP In Overall Proficiency grouping performance. Example: seminar, in individual semester general proficiency exams are organized. In General Proficiency exam is divided into two classes: Yes – student participated in general proficiency, No – Student not participated in general proficiency.
- k) ATT Attendance of Student. Minimum 70% attendance is compulsory to participate in End Semester Examination. Then through in special phases low attendance students also participate in end Semester Exam on honest reason. Attendance is divided into tierce classes: Poor-<60%, Average->60% and <80%, Good ->80%.
- LW Lab Work. The Lab work is separated into two classes: Yes-student completed lab work, No-student not completed lab work.
- m) ESM The End semester and Marks obtained in MCA semester and it is declared as answer adjustable. It is divided into five categories class standards: First >60%, Second >45% and <60%, Third–>36% and <45%, Fail <40%.</li>

## 6. Conclusion

In this research paper, many data mining tasks were used toward generate qualitative analytical models which were efficiently and effectively able to predict the students' grades from a composed training dataset. First step a review was constructed that has battered college students and collected multiple personal, academic data and the second collected dataset was preprocessed and discovered to become appropriate for the data mining tasks. EDM interesting results were drawn from the classification representations, as well as, interesting patterns in the Naive Bayes model was initiate. Decision tree algorithms have been implemented with the Naïve Bayes algorithm. In the current study, it was to some extent found that the student's performance is not completely independent on their academic hard work, in spite, there are several other issues that have equal to greater influences as well. In this conclusion, this study can motivate and help colleges to perform data mining tasks on their students' data frequently to find out motivating results and patterns which can help together the university as well as the students in several ways.

## References

- [1] Baradwaj, B.K. and Pal, S., 2011. Mining Educational Data to Analyze Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.6, 2011.
- [2] Ahmed, A.B.E.D. and Elaraby, I.S., 2014. Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer

Application and Technology, 2 (2), pp.43-47.

- [3] Pandey, U.K. and Pal, S., 2011. Data Mining: A prediction of performer or underperformer using classification. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 2011, 686-690.
- [4] Bhardwaj, B.K.and Pal, S., 2012.DataMining: A prediction for performance improvement using classification. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April2011.
- [5] Yadav, S. K., Bharadwaj, B.andPal, S., 2012.Data Mining Applications: A Comparative Study for Predicting Student's Performance. International Journal of Innovative Technology & Creative Engineering (ISSN:2045-711), Vol.1, No.12, December.
- [6] Yadav, S.K. and Pal, S., 2012. Data mining: A prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT). (ISSN: 2221-0741), Vol. 2, No.2, 51-56, 2012.

Licensed Under Creative Commons Attribution CC BY