

Theoretical Study of Predictive Model Using Data Mining Technique on Academic Dataset

Pooja Singh, Dr. G. N. Singh, Dr. Arvind Singh

Department of Computer Science, Sudarshan College, Lalgaoon, Awadhesh Pratap Singh University Rewa, Madhya Pradesh, India

ABSTRACT

The basic purpose of an educational system's must be to prepare successful careers of students within a predetermined time period. In Present scenario economics and society is directly and indirectly linked to the effectively with educational systems by which can achieve goal. Education is a critical part of every Country and all over the world's community's overall their growth and development. To maintain and ensure the data associated with the educational institution is correctly analyzed. Many important insights which can give contribution to the continuous improvement of students in the future. Through Data mining and machine learning that are most important equipment to achieve this goal because there are various approaches that aid in finding that crucial information from that huge amount of data and databases [1]. In data mining process which employs a variety of methods that applies during the operation, such as clustering, association and classification as well as rule mining in order to fetch previously cryptic information from a database. Data mining technique is the very efficient and effective method which is used in the data analytics for measuring performance by creating models and algorithms.

Keywords : Gymnema sylvestre R. Br., diabetes, Callus induction, proliferation, biochemical analysis.

I. INTRODUCTION

The main goal of machine learning is to train computers and computer like devices by which devices can think and do behavior like human. Some of the data mining techniques and algorithm is which are often categorized based on their learning style or by similarities to each other form is purpose, such as decision trees, classification, clustering, deep learning,

regression etc. All of above mentioned techniques and methods produce a huge amount of data related to the academic development of students, regarding their grades, as well as their attendance. All information should be processed and evaluated based on techniques [2].

In modern time, the educator and teacher have access to a variety of tests and methods of evaluation to assist them in proper manner and determining the level of academic achievements

by their students. Through Data mining technique we can use for evaluating or analyzing student data. It is important and necessary to implement a comprehensive educational rejuvenate for a nation to preclude itself from the other nations. Various Data mining techniques are available which make it possible to find patterns and group of information that was hidden. We look into academic data for data mining to conclude student achievements and their qualifications. In present database, we have some features and all these features are not very important to predict by using the pattern in the dataset. Therefore some features are important and some are not important so that, we need to ignore some feature from the database for the improvement of speedy-mining process and accurate result [3]. Through, with the help of Data mining techniques we can remove some features from the database. Which feature should be selected is an important technique in the success of data mining process by which we can select the useful and relevant features in the dataset.

Through comparative study and comparing between different features selection methods is done on in this paper along with its impact on the classification technique. Every educational institution must be collected data from students by using sessional tests and exam, assignments, evaluations through laboratory, and personal interactions. All the information is used to solidify whether a student has been fail or passes test. We can analyze student data that will help academic institutions to improve students' academic performance [4]. Predicting academic performance and improvement of the students using various data mining techniques we

followed different research and review papers. Most of the researchers are trying to construct a system which can be effectively and efficiently predict the academic performance of the students. For this purpose there are various techniques available like classification algorithms which effectively predict on the class level data [5].

With data mining techniques, we can understand the learning processes by identifying and measuring its various parameters [6]. Data mining means in the educational environment is called to as Education Mining. There are many facts that affect a student's academic performance.

Institutional costs can be less through the use of data mining tools and techniques that provide more personalized learning, increased system efficiency, and a more versed education stride. Through this as guidance, we can improve student retention and academic achievement as a result of student learning [7]. Increasing student exam numbers, data mining prediction techniques can play important role in identifying the important aspects that can affect a student's score on every test exam. It can give us and the public a new perspective on the school system that was not practical before.

II. REVIEW OF LITERATURE

We have done review of literature which can affect in such a way where most of the students are affected by the academic performance. For this reason classification techniques are used by researcher to give prediction and academic performance of the student. In many research papers we found that selection feature increase the performance of prediction model.

In data mining techniques there are five types of methods: clustering, prediction, relationship, model discovery and Human judgment. F. Castro et al. identified that there are four categories of EDM tasks were such as applications that evaluate student learning performance, adaptation of course and recommendations based on learning behavior of student. It is a good practice to develop any method for the purpose of evaluating online content of courses and gathering students to take feedback from teachers in online courses and to detect plagiarism. Rahila Asif et al. have reported and stated in their paper that a study program should examine the academic achievement of students. Each year of students degree, students should be assigned to one of four groups that reflect their academic status for example x_1, x_2, x_3, x_4 . Two successful clusters were compared where X-means clustering. Both group showed that many students prefer to stay in the same clique during their studies.

Kumar, M., et al. has done research, he made use of many number of different algorithms, which includes Naive Bayes, Decision Tree, PART, Bayes Network and Random Forest. All algorithms were implemented with the help of three different and essential methods: percentage split (74%), training set and 10-fold cross-validation. There are following the examination of various metrics such as Mean Absolute Error, Time to build Classifier, Squared Error, Recall, Precision, F-measure and ROC Area founded by many data mining techniques, that was determined that which algorithm performed more precisely on the student dataset that was available [10]. About the student dataset, the Random Forest technique gives and performed better than other system, using recall value that

is very near. Invention of a various data mining techniques has shows that how these techniques increased students capacity. Pandey M. and Taruna S. Has given an idea of the RA (Rocchio Algorithm), that was implemented to do less the negative effects of the "curse of dimensionality" on the performance of PSC. In the context of the Rocchio Algorithm based PSC model of classification of problem, various different features and selection strategies are investigated. In Various research studies also used that dataset to reach the conclusion that the typical PSC is affected by the dimensionality curse [11]. Almutairi F. M. and et.al. analyzed the effectiveness of educational data mining techniques in characterizing the possibility of student failure in programming classes. Many studies have been done to examine the utility of these strategies in identifying students that who are at risk of fiasco soon enough to intervention, as well as the impact of data preparation and algorithmic tasks depends on the efficiency of the techniques [12].

There are various types of student characteristics, including personal, academic, social, family and institutional, and each of these characteristics plays a vital role in the construction of predictive models for educational achievements. Another one of the most significant technique in data mining that helps any data mining classification algorithm to achieve a good accuracy. There are methods like filter and wrapper methods, which fall under the category of feature selection. All these methods help limit the number of features in a dataset that are not particularly for making predictions about any given class. In the following review, we have selected an educational dataset from the pantry at UCI and

tried to apply the classification technique to that dataset.

In next step of our sequence and process, we need to use feature selection technique on the selected dataset. This part used as the foundation for the implementation of the proposed research project.

In this paragraph we discuss about to defining the pre-processing steps and then discuss about the final classification methods used for implementation. In this paper, we used two datasets which are publicly available that were used to predict student performance.

The dataset was contains data on a student's grades as well as their personal, regional attributes and their participation in extracurricular activities. The data was collected through the use of school and college surveys reports.

III. METHODS AND TECHNIQUES

Oracle provides good environment and capability for mining within oracle database. Predictive capabilities of oracle Data Mining to build and deploy data mining applications by which we can add various intelligent capabilities in existing applications. Oracle can also provide predictive oracle queries for data exploration and generating reports.

Oracle Data Mining technique provide us comprehensive database algorithms, by using these algorithms we can perform a various mining work, such as clustering, regression, classification etc. Oracle data mining algorithms can be used and work on all standard type of data like, star schemas, transactional data, and text forms of unstructured data. Oracle Data mining algorithm is useful on small and huge data set.

1.1. Pre-processing

There are varieties of potential grades in raw dataset, where 0 represents the worst case and 20 represent the best case. The requirement of grading, data should be converted into one category so the students final grade will be the form of integers data type and the prediction class must be in the form of categorical values. In this paper we compared two grading systems: multi-level and binary. We have parted final score into only four groups. The questions were calculated using Erasmus methodology. You can see in below Table, "Fail" grade is from 0-9, which is the lowest grade. "Good" grade is from 12-13, which is the medium grade. "Very Good" grade is from 14-15, which is the first and good grade. "Excellent" grade is from 16-20, which is the High and first grade.

I st Level	II nd Level	III rd Level	IV th Level
Excellent	Very Good	Good	Fail
A++	A	B++	C
16 to 20	14 to 15	12 to 13	0 to 9

1.2. Classification Method

In machine learning classification is method in which a algorithm of computer is learned (trained) for generating classifications and observations based on the data to be given for training. Prediction of class also based data, which is the first step of method.

1.3. Decision Tree

Decision tree is a type of supervised machine learning in which the data is continuously distinguished based on a particular parameter. There are two main properties of decision tree method: First are Nodes of tree and second are

leaves. The decision tree also called supervised learning method which could be used for regression issues and classification [20].

IV. ANALYTICS USING ORACLE DATABASE

Oracle Database used and provides various features for analytics and business sense. Oracle data mining may be united with other related analytical features of the databases for example OLAP.

- **Technology Stack:** Advantage of Oracle's technology stack is that which you can integrate data mining within a framework. This is useful for business sense and scientific inquiry.
- **Domain Environment:** Oracle data mining models are to be built, tested, validated, managed and deployed in their application domain environment.
- **Application Programming Interfaces (API):** You can use PL/SQL API with SQL language that provides access to Oracle data mining capabilities in Oracle Database.

V. DATA MINING IN THE ORACLE DATABASE

The build-in-features of oracle data mining process is that can be used to increase scalability and make efficient use of resources.

There are many advantage of data mining in oracle database

- **Movement of Data-** Movement of data in oracle data mining technique is easily can be performed.

- **Security:** Data is secured and protected using various security mechanisms.
- **Administration of Data:** Data should be filtered, cleaned, transformed, and normalized in many types. Oracle Database provides various administrative tools/techniques for managing data.

Refresh of Data: By using oracle data mining processes in Oracle Database is easily can be refreshed and can be accessed.

VI. PREDICTIVE MODEL

We present design of the predictive proposed model that works together with other important applications of data mining methods. In Figure 1, we are representing about a dataset that are interconnected with the academic achievements of the students. The preprocessing steps, we remove all discrepancies that is present in the dataset during the data were collected from the school or academic institutions. During the time of testing data sets there were two modes. The first mode was selected to testing data sets with all characteristics and the second mode was used for feature selection methods. The accuracy was found using classification algorithms by implementing the correlation with. Now we will go on the other phase, where need to select the testing mode with the classification techniques that was used in the implementation.

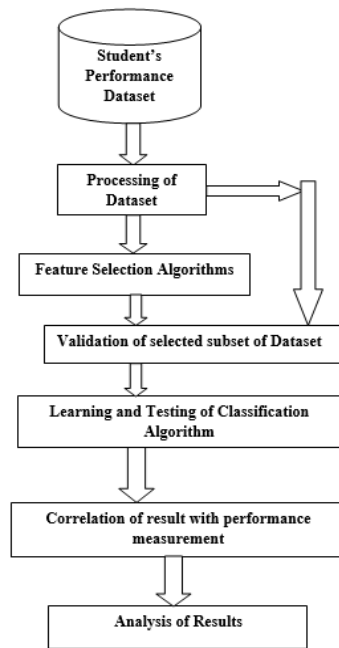


Figure 1. Predictive Design Model

VII. CONCLUSION

In this study the comparison was done by using the chunking learning techniques. After this study we have found that the classification algorithm that didn't use any feature selection technique can achieved the lowest level of accuracy compare than the other feature selection technique. Analysis of various algorithms by performance measures has taken into account. It provides us better understanding of the algorithms as well as measurement techniques. The prediction of any algorithm is laid down by comparing different types of algorithms with various metrics.

After doing review of various literatures, it is very difficult to say that a particular prediction and their performance measurements are better than other. In last my recommendation is that multiple measurements can be combined for a better algorithm result.

VIII. REFERENCES

- [1] Asif R., Merceron A., & Pathan M. K. (2015). Investigating performance of students: a longitudinal study. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 108-112). ACM.
- [2] Kumar M., & Singh A. J. (2017). Evaluation of Data Mining Techniques for Predicting Student's Performance. International Journal of Modern Education and Computer Science, 9(8), 25.
- [3] Yahya A. A. (2017). Swarm intelligence-based approach for educational data classification. Journal of King Saud University Computer and Information Sciences.
- [4] Costa E. B., Fonseca B., Santana M. A., de Araújo F. F., & Rego J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 73, 247-256.
- [5] Ghani A. A., & Mohamed R. (2017). The Effect of Entry Requirement for Civil Engineering Student Performance. Journal of Science and Technology, 9(4).
- [6] Ramanathan L., Parthasarathy G., Vijayakumar, K., Lakshmanan, L., & Ramani, S. (2018). Cluster-based distributed architecture for prediction of student's performance in higher education. Cluster Computing, 1-16.
- [7] Thomas, C. L., Cassady, J. C., & Heller, M. L. (2017). The influence of emotional intelligence, cognitive test anxiety, and coping strategies on undergraduate academic performance. Learning and Individual Differences, 55, 40-48.
- [8] F. Castro, A. Valid, `A. Neat, and F. Mugica, "Applying data mining techniques to e-

- learning problems,” in *Evolution of teaching and learning paradigms in intelligent environment*, pp. 183–221, Springer,2007
- [9] Asif R., Haider N. G., & Ali S. A. (2016). Prediction of Undergraduate Student's Performance using Data Mining Methods. *International Journal of Computer Science and Information Security*, 14(5), 374.
- [10] Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on student's performance prediction in education using data mining techniques. *International Journal of Education and Management Engineering (IJEME)*, 7(6), 40-49.
- [11] Pandey M., & Taruna S. (2016). Towards the integration of multiple classifiers pertaining to the student's performance prediction. *Perspectives in Science*, 8, 364-366.
- [12] Almutairi F. M., Sidiropoulos N. D., & Karypis G. (2017). Context-aware recommendation-based learning analytics using tensor and coupled matrix factorization. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), 729-741.
- [13] Tran T. O., Dang H. T., Dinh, V. T., & Phan, X. H. (2017). Performance Prediction for Students: A Multi-Strategy Approach. *Cybernetics and Information Technologies*, 17(2), 164-182.

Cite this article as :

Pooja Singh, Dr. G. N. Singh, Dr. Arvind Singh ,
"Theoretical Study of Predictive Model Using Data Mining Technique on Academic Dataset",
International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 2, pp. 475-481, March-April 2021.
Journal URL : <https://ijsrset.com/IJSRSET2185109>