International Journal of Innovative Research in Engineering
Volume 5, Issue 5 (September-October 2024), PP: 14-18.
https://www.doi.org/10.59256/ijire.20240505002
www.theijire.com
ISSN No: 2582-8746

Predictive Analysis of Students' Learning Performance Using Machine Learning

Sivakumar Nagarajan
Technical Architect, I & I Software Inc, 2571 Baglyos Circle, Suite B-32, Bethlehem, Pennsylvania, USA.

How to cite this paper:
Sivakumar Nagarajan, "Predictive Analysis of Students' Learning Performance Using Machine Learning", IJIRE-V5I05-14-18.

Abstract: *Several researchers and educational institutions got attracted to the domain of predicting the performance of the students to classify the educational level of student performance. Though the educational sector uses several techniques for extracting useful information about them and the measures to be undertaken in the learning process, it is necessary to develop a student performance assessment model to assist the students as well as faculty members to improve their performance to the next level. This research work majorly concentrates to examine and identify useful rules and patterns to motivate the students for handling their education as well as career in a better way and also to improve the functioning of academics to supervise the policies for students' benefit. Here, EDM is employed to identify the effective data which helps in applying the active learning in technical applications. Initially, a review of different DM, machine learning (ML), and meta-heuristic models to assess the performance of the students. the proposed models can be employed as an appropriate tool for mining educational data to achieve effective performance assessment of the students.*

Key Word : *Educational Data Mining, Predictive Analysis, Machine Learning, Classification, Feature Selection, Clustering*

I.INTRODUCTION

The educational system is the backbone of India's economy and social progress. The primary goal of an educational system is to endow students with the knowledge and good employability skills. The recent advancements in technology led most of the educational institutions to step towards Educational Data Mining (EDM). EDM is the method of applying the data mining (DM)methodologies on a large number of student details with the intention of extracting meaningful details that helps in enhancing the teaching-learning process. DM approaches gained significant attention in the educational sector. Decision making and prediction, being the important functionalities of EDM, it is rapidly gaining its importance in the educational field. Particularly, DM models offer the educational plan makers with data-based approaches required to support the motto of enhancing the effectiveness and excellence of the teaching learning process. In this view, the usage of DM models seems to bring significant changes in the global educational system. EDM assists educational institutions in offering solutions for particular problems. Therefore, the DM tools and techniques offer priceless support in predicting the student's outcomes and to take necessary decisions prior.

II. DATA MINING

In recent times, DM is a process of identifying knowledge from massive databases for uncovering the patterns and trends involved in data. Massive collection of data in the information sector is transformed into meaningful data. Apart from this, DM is an extraction of knowledge from large scale data that also computes the operations like Cleaning, Combination, Transformation, Mining, Estimation, and Projection. DM is classified into 3 phases namely, Descriptive, Classification, and Prediction. Initially, the descriptive process consults with common features of data while classification and prediction resolve the modules of data. Followed by, the descriptive function is composed of summarization and data mapping typically named data characterization as well as data differentiation. Predictive analytics is defined as a unification of Machine Learning (ML), historical information, and Artificial Intelligence (AI). Moreover, it helps examine the condition of recent details and examines the upcoming results. It becomes more prominent in applications like the economic sector, marketing, medical science, social media, and so on. The execution of predictive analytics is one of the trivial processes as it has massive challenging issues. The key objective of predictive analytics is to earn better profit with cost-effective and limited threats. There is a better solution for all predictive issues. It is also used for mysterious data in past, present, or future.

The fundamental modules involved in DM and corresponding phases are defined in the following:
Data Pre-processing: Data in the practical world is inconsistent and partial with a massive error. Preprocessing is commonly employed in converting the real data to a specific behaviour that is learnable. Some of the pre-processing steps are cleaning, integration, transformation, reduction, and discretization. The unwanted and repeated data tends to develop the values from

(- 100) and impossible data unifications (Gender:Male Pregnant: Yes). Data Cleaning: This process is applied for cleaning the partial (no attribute values), noisy (errors/outliers), and inconsistent (difference in codes/names) data. It occupies the missing values by applying the attribute mean and detects the missing values along with a learning model. Moreover, it finds the noises by binning (sorting the attributes), clustering (grouping the attributes), and regression (predicting the attributes).

Data Integration: The data from diverse sources are integrated from various databases, data cubes, or flat files. Metadata can capture the data which is applicable in resolving the issues of redundancy in data combination.

Data Transformation: This step is used for changing the data from one format to alternate for computing proper mining. Some of the steps involved are normalization, aggregation, generalization, and parameter development.

Data Reduction: This process is involved in limiting the massive data for computing the productive storage. Data Transformation: This step is used for changing the data from one format to alternate for computing proper mining. Some of the steps involved are normalization, aggregation, generalization, and parameter development.

Data Reduction: This process is involved in limiting the massive data for computing the productive storage.

Data Discretization: it is applied in transforming frequent attributes into general attributes using supervised as well as unsupervised mechanisms.it is applied in transforming frequent attributes into general attributes using supervised as well as unsupervised mechanisms.

III.EDUCATIONAL DATA MINING (EDM)

In the last decades, EDM has attained massive attention from researchers due to the existence of massive educational details which is accessible from many sources. The main aim of EDM is to make DM models more effectively in order to safeguard the numerous amount of educational information and to develop a protective atmosphere for the student's learning. In this approach, diverse models have been deployed for DM and its analytics. Moreover, prediction models were used namely, Classification, Regression, and Latent factor evaluation technologies.

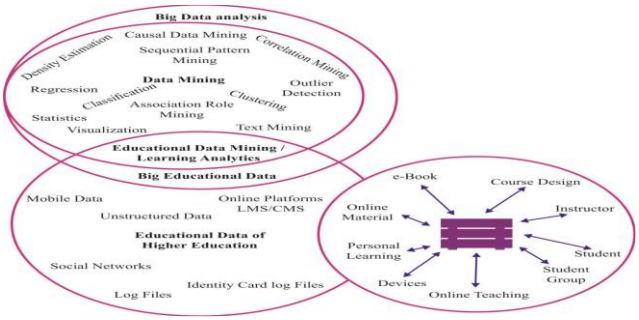

Figure 1: An illustration of DM use in HE

The DM tools are collaborated with academics in enhancing the student's learning methodologies by exploring, filtering, and estimating the parameters relevant to student's features or behaviors. The major challenges faced by any educational institutions lies in the number of placements it gets and the number of successful graduates it produces.

IV.LIFECYCLE/ PROCESS INVOLVED IN EDM

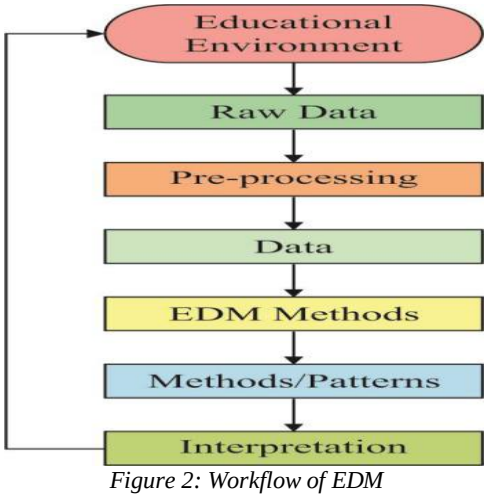

Figure 2: Workflow of EDM

The application of EDM models is composed of various phases (Figure 2). At the initial phase, the method is developed with the responsibility of identifying essential data. Then, data has been filtered from an accurate educational platform. Subsequently, data has to be pre-processed, as it is aroused from diverse sources with distinct templates and hierarchy levels. The identical patterns are attained while using EDM models which is interpreted. Finally, the simulation outcome recommends that using changes in the teaching process is not an appropriate solution and the analysis is carried out after changing the teaching process previously.

V.EDM APPLICATIONS

1. Student modelling: student details (knowledge, motivations, and so on.) and EDM methods might be employed in developing a customized learning process by labelling the variations among students.
2. Modelling the knowledge structural way of the field, which integrate the psychometric modelling approaches with space-searching technologies are developed in identifying the data-related applications.
3. Pedagogical care: Offers effective educational care.
4. Scientific researches: domains are employed for developing and sample educational scientific strategies and to make novel hypotheses.

VI. AN EDM EXAMPLE

In this method, consider that a tutor is permitted to use multidimensional information regarding the learning nature of the students in the online learning model. Especially, the student value of sign-in has been saved and the proportion of lesson accomplishments the value of student development. The input data point is represented as a black dot. This figure is drawn under the application of an online visualization system (Mohan, 2016). The main aim of a tutor is to categorize the student's behavior. But, labelling this classification is not an appropriate model. Also, it is considered to be general UL issues in ML, named as clustering. When compared with diverse clustering models, a reputed k-means clustering mechanism has been applied. The parameter k implies the count of clusters desired. Initially, k-points for cluster centres have been selected randomly. The primary centroids are illustrated as triangles.

Then, consider the points in a 2-D Cartesian coordinate plane and x coordinate point of a centroid is the mean value of x coordinate points that belongs to a cluster. The y coordinate of the cluster is estimated identically. A similar strategy uses points in dimension space. Thus, an applicable distance measure has to be defined for measuring distance among 2 points. Also, centroids are not considered as original data points. Followed by, every data point is divided into closest centroid relied on a distance measure. Diverse class labels of the points are represented in various colors. This categorization is not a novel clustering of points as cluster borders are slightly ineffective. Afterward, for all classes, a novel centroid has been re-determined. As the centroids were modified for students who signed for massive times; yet is not applicable in numerous progress, perhaps the short attention spans. According to the interpretation of 4 clusters, a tutor designs unique learning paths for various students. indicates that EDM is used for automatic identification. Then, EDM's method in classifying the problems into tiny portions as it resolves the problems using ITS. This illustration is extended to detect learning results. Finally, (Figure 1.3) depicts some of the typical approaches applied in EDM, they are Clustering and Visualization.

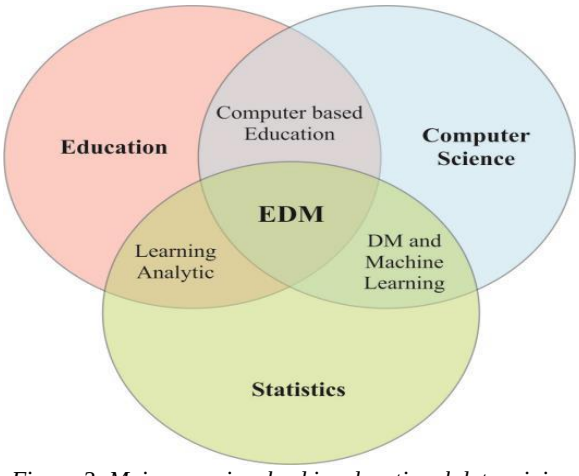

Figure 3: Main areas involved in educational data mining

VII.APPLICATIONS OF EDM IN HIGHER EDUCATION

Student modeling is defined as a way of representing the cognitive factors of student actions like examining student performance and behavior, separating the fundamental misconception, shows student aims and achievements, finding the advanced as well as acquired knowledge, retaining episodic memory, and defining personality features. In this approach, the definition of classifying EDM applications. Each domain in this class shows a technique which defines student's aims and goals. According to the study in diverse features in student modeling, such as knowledge and skill (1), error and misconception (2), learning style with priority (3), affective and cognitive aspects (4) as well as metacognitive factor (5). The taxonomy of application in EDM.

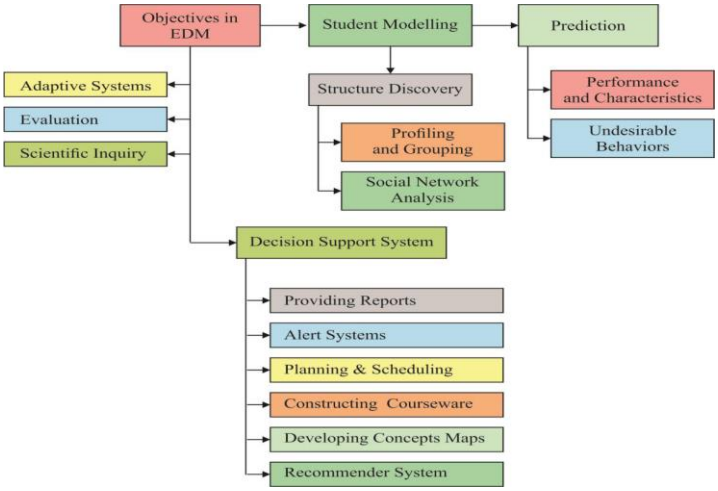

Figure 4: Classification of different EDM Applications

VIII.CLASSIFICATION PROCESS

Usually, data classification is comprised of 2-advanced operations. Initially, it is considered as a learning step that defines planned classes by examining the collection of training dataset samples. techniques, NN, rule-based methods, and DT. The process contained in data classification is illustrated

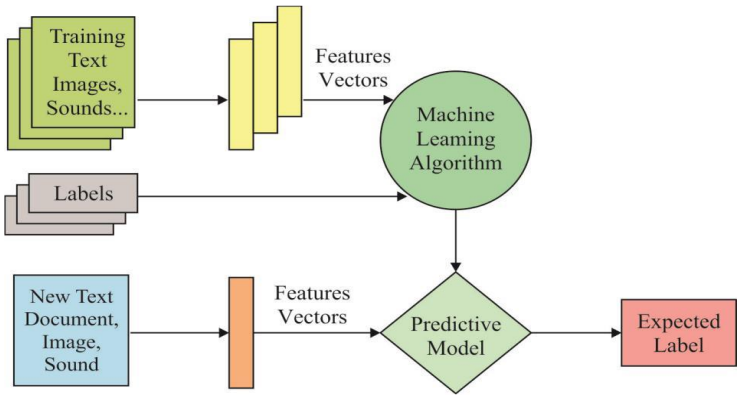

Figure 5: Process involved in data classification

Every instance is interpreted to have a location with the previous class. Using the second step, a method is sampled with the help of an additional dataset applied for assessing the classification accuracy of a method. When sufficient accuracy is attained, then this approach is employed in classifying upcoming data instances for a class label. Certainly, the classification process is performed in the decision-making process. Some of the principles applied for classification are Bayesian.

IX.MACHINE LEARNING

An ML model is defined as a computer-intensive mechanism and applies re-sampling and iterative methodologies for classification approaches. ML approaches are considered with optimal subset selection and eliminate the issues of classical classifiers like over-fitting as well as distributional demands of parameters. ML technologies that have emerged in computer science with logic and basic mathematics, statistics as ML approaches do not estimate the group features rather it is initialized with an arbitrary group separator and tunes frequently till satisfying the classification groups. ML examines the tuning variables and individual ML functions became unstable, which makes a suitable process. As the non-statistical nature is embedded, these approaches can apply the data in various formats like nominal data that generates maximum classification accuracies.

X. CONCLUSION

The assessment of student's performance in educational institutions indicates the level of efforts taken by educational institutions should do for enhancing the poor or average learner. The significance of applying EDM models, make use of the previous data of the students for predicting the unseen or upcoming performance of the students. This idea has attracted several researchers to develop classification models to predict the unknown labels of future instances. Several research people and educational institutions started to get attracted to the domain of predicting the academic performance of the students to classify the educational level of the student performances. Though the educational sector uses several techniques to extract meaningful details related to the features of the students that undertake learning procedure, it is needed

to develop student performance assessment model to assist the students as well as faculties to improve their performance to the next stage. A comparative analysis of the attained results showed superior performance of the proposed models to the existing models under different aspects. (The attained comparative results analysis stated the superior performance of the proposed models over the compared methods under different aspects.) Therefore, the proposed models can be employed as an appropriate tool for mining educational data to achieve effective performance assessment of the students.

Reference

1. *Abu Tair, M.M. and El-Halees, A.M. (2012). Mining educational data to improve students' performance: a case study. International Journal of Information and Communication Technology Research., 2(2):140-146.*
2. *Adekitan, A.I. and Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon., 5(2): e01250.*
3. *Bravo, J. and Ortigosa, A. (2009). Detecting Symptoms of Low Performance Using Production Rules. International working group on educational data mining.,4(2):31-40.*
4. *Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. Computers operations research., 13(5):533-549.*
5. *Ramos, J.L.C., e Silva, R.E.D., Silva, J.C.S., Rodrigues, R.L. and Gomes, A.S. (2016). A comparative study between clustering methods in educational data mining. IEEE Latin America Transactions., 14(8):3755-3761.*
6. *Urbina Nájera, A.B., De La Calleja, J. and Medina, M.A. (2017). Associating students and teachers for tutoring in higher education using clustering and data mining. Computer Applications in Engineering Education., 25(5):823-832.*