A Survey On Predicting Students' Academic Performance Earlier, Using Data Mining Techniques

Vinitha K[1], Kavitha R[2], Suruthi N3
[1, 3] Dept of Computer Science and Engineering
[2]Associate professor, Dept of Computer Science and Engineering
[1, 2, 3] Parisutham Institute of Technology and Science, Thanjavur, India

Abstract- *The methods of Data mining are used to evaluating the given data and to mine the unknown facts and knowledge which greatly supports the researchers to take effective decisions. Due to the tremendous growth in recent technology like social media, it may deflect the students from their actual track, and this is one of the reasons for the students to poor performance in academic activities and it even leads to course dropout. This paper survey the previous research works done on students' performance prediction, analysis, early alert and evaluation by using different methods of data mining.*

Keywords- Performance Prediction, analysis, Academic activities, early alert, Data Mining Methods.

I. INTRODUCTION

In the recent decays, the exploration in the educational field is increasing rapidly due to the colossal growth of data that related to the performance of students' academics. It increases the accuracy and quality of students' performance by predicting it in-advance. Research on job performance in organizations has benefitted from expanding the performance domain beyond core task performance into extra role behavior. It is now generally recognized that there are three main, broad dimensions of job performance: task performance, organizational citizenship behavior, and counterproductive work behavior [15]. The concept of combining classifier is based on the assumption that the different classifiers, which uses a different data representation, different concept and different modeling techniques are most likely to arrive at classification results with different patterns of generalization [11]. Bullying is a situation which is characterized by physical abuse, made repetitively, by one or more students against one or more peers [12]. The alarming indexes of students' academic failures, along the years, in universities' introductory programming courses have been concerning educators. Studies show that students face many difficulties during their programming activities in such a way that many of them end up failing or quitting the course at some initial stage [3]. The higher education institutions of today make extensive use of digital technology. At present, the use of database systems to manage student demographic information and academic records is widespread [1]. The educational activity involves a complex process whereby inputs are converted into outputs. By analogy with the type of production function that is typically used to analyze the technology of a firm, the labor and capital inputs used by a school are likely to influence its output [2]. Most research in this area uses undergraduate students, with graduate students being under-represented in the literature. With many occupations now requiring post-graduate degrees for entry level positions or career advancement, people are increasingly entering the workforce with graduate degrees. Because perfectionist tendencies can contribute to burnout and workplace dysfunction, it is important to understand how perfectionism manifests in graduate students to inform ways of helping these people prior to entering the workforce [7]. Baker proposes five primary categories or approaches in EDM: prediction, clustering, relationship mining, discovery within models, and distillation of data for human judgment [4]. The differential students' performance in tertiary institutions has existed and is still a source of great concern and research interest to the higher education managements, government, parents and other stakeholders because of the importance of education to the national development [9].

II. DATA MINING TECHNIQUES

The techniques of data mining are classified into two groups: they are supervised learning and unsupervised learning. In supervised learning, a model is built earlier for the analysis and then it applies the algorithm to the data in order to estimate the parameters for the model. Classification, Association Rule Mining ,Decision Tree, Bayesian Classification, Neural Networks, etc. are some examples of supervised learning. In unsupervised learning, we do not create a model or assumption earlier for the analysis; it just applied the algorithm directly to the dataset to detect its result. Later a model can be created on the basis of the obtained results. Clustering is one of the examples that falls under the category of unsupervised learning. Various data mining techniques such as Classification, Decision Tree, Bayesian Classification, Neural Networks, Clustering, Association Rule Mining, Prediction, Time Series Analysis, Sequential Pattern and Genetic Algorithm and Nearest Neighbor have been used for knowledge discovery from large data sets. Some of the

common and useful data mining techniques have been discussed. An interdisciplinary subfield of computer science, it is an essential process where in intelligent methods are applied to extract data pattern the overall goal of which is to extract information from a data set, and transform it into an understandable structure for further use.. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records, unusual records and dependencies. Data Mining usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple clusters in the data, which can then be used to obtain more accurate prediction results by a decision support system.

A. Decision Tree

Decision tree is one of the most popular technique for prediction. Most of the research have used this technique because of its simplicity and comprehensibility to uncover small or large data structure and predict the values. Decision tree in data mining is one of the simplest and easiest methods which are most frequently used by the researchers on their work. A tree has many comparisons between one thing and others in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and provide the knowledge to make the decision. The root node of the decision tree is a top node resembles simple question also called as a posture that bears multiple branches called sub nodes with answers for the root node question. In turn each answer related to a set of questions or conditions that help us to predict the data, on which the final decision is made. ID3 and C4.5 are called as induction algorithm of decision tree developed by the researcher called Ross Quinlan. Both algorithm supports greedy method, top-down recursive in divide-and-conquer manner and they does not support backtracking. C4.5 is also known as superset of ID3.The advantages of this technique are, it doesn't require detailed knowledge, it deals with complex data, these are easy to understand, and data Classification becomes simpler, makes learning easier, it produces very accurate end result.

B. Naive Bayes Algorithm

Naïve Bayes algorithm is also an option for researchers to make a prediction. Among the fifteen papers, there are four papers that have used Naïve Bayes algorithm to estimate students performance. The objective of all these four papers is to find the most effective prediction technique in predicting students performance by making comparisons. Their research showed that Naïve Bayes has used all of attributes contained in the data. Then, it analyzed each one of them to show the importance and independency of each attributes. A simple probabilistic classifier that based on applying Bayes theorem with strong independence assumptions is the Naive Bayes classifier. "Independent feature model" is a more descriptive term for the underlying probability model. In simple terms, a Naive Bayes classifier adopts a particular feature in the presence of a class that is unrelated to any other feature. For example, a vegetable may be considered to be a carrot if it is orange in color, it shaped as in cone, and about 10 to 15 centimeter long in length. If these features match on each other or match upon the existence of the other features, a Naive Bayes classifier then considers all these properties independent contribution to the probability that this vegetable is a carrot. Based on the exact nature of the probability model, the data set is then trained very effectively by a Naive Bayes classifier in a supervised learning setting. In most of the practical applications, the parameter estimation for Naive Bayes models uses the method of maximum likelihood; in other words, one can also work with the Naïve Bayes model without considering the Bayesian probability or by using any Bayesian methods. In spite of their over-simplified assumptions of a Naive design, the Bayes classifiers will work effectively in solving many complex real-world circumstances. There are several advantages of Naive Bayes classifier such as; it requires a small amount of training data set to estimate the parameters essential for classification. This is Because, the independent variables are assumed only to the variances of the variables of each class that is needed to be determined and not the entire covariance matrix, irrelevant features are removed to improves the classification performance, results High Performance and take very less computational time.

C. Support Vector Machine

Support Vector Machine is a supervised learning method used for classification. There are three papers that have used Support Vector Machine as their method to predict students performance. A powerful Support Vector Machine (SVM) which was first proposed by Vapnik and it has a great potency of interest in the machine learning research community. Several past studies have reported that the SVM generally has a proficient of delivering the high accuracy in classification when compared to other data classification

algorithms. Though, for certain datasets, the achievement of SVM is very subtle in determining the cost parameter and kernel parameters. As in the case of closure, to figure out the most encouraging condition environment the user normally needs to conduct extensive cross validation. Basically this technique is baited to as a model selection. A superior asset of this SVM technique is that, concurrent miniaturize the projected classification error and make best use of the geometric margin, So SVM is also named as paramount Margin Classifiers. It is found on the Structural Risk Minimization (SRM), SVM can be used for both classification and prediction. There are several advantages of SVM such as it uses maximum marginal hyper plane for classifying linearly separable data, Data can be separated clearly into rations, extends by itself in order to classify the linearly inseparable data. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. When data are unlabelled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The support vector clustering algorithm created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications

D. DT-J48

The J48 Decision tree classifier is an implementation of ID3 classifier used to create a decision tree based on the attribute values of the available training data set. So, whenever it encounters a set of items i.e. training set it identifies the attribute that discriminates the various instances most visibly. This feature tells us about the data instances so that we can classify them clearly and the best is said to have the highest information gain. Now, among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, then we terminate that branch and assign to it the target value that we have obtained. J48 is an extension of ID3. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. In the WEKA data mining tool, J48 is an open source Java implementation of the C4.5 algorithm.

E. Neural Networks

Neural network is another popular technique used in educational data mining. The advantage of neural network is that it has the ability to detect all possible interactions between predictors variables. Neural network could also do a complete detection without having any doubt even in complex nonlinear relationship between dependent and independent variables. Therefore, neural network technique is selected as one of the best prediction method. The area of neural networks possibly belongs to the border line between the artificial intelligence and approximation algorithm. A neural network is a pool of neurons like processing units with weighted connection between the units. It composes of many elements, called nodes which are connected in between. The connection between two nodes is weighted and by the adjustment of this weight, the training of the network is performed. There are many advantages of neural networks such as adaptive learning ability, self-organization, real time operation and insensitivity to noise. Neural networks are used to identifying patterns or trends in data and well suited for prediction or forecasting needs. There are several neural network algorithms such as Back Propagation, NN Supervised Learning, and Radial Base Function (RBF) Network etc. The term neural network was traditionally used to refer to a network or circuit of neurons.[1] The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the terms is refer to either biological neural networks, made up of real biological neurons, or artificial neural networks, for solving artificial intelligence (AI) problems. The connections of the biological neuron are modeled as weights. A positive weight reflects an excitatory connection, while negative values mean inhibitory connections. All inputs are modified by a weight and summed. This activity is referred as a linear combination. Finally, an activation function controls the output. For example, an acceptable range of output is usually between 0 and 1.Unlike von Neumann model computations, artificial neural networks do not divide memory , processing and operate via the flow of signals through the network connections, somewhat akin to biological networks.

F. Clustering

The method of grouping object that have similar characteristics is referred to us as Data Clustering. The criterion for checking the similarity is implementation dependent. Clustering is often confused with classification, but there are vast differences between them. In classification the objects are assigned to a predefined class, where as in clustering the classes are also to be defined. Clustering methods may be divided into two categories based on the structure of cluster, they are named as hierarchical cluster and partitioning cluster.so, as every other problem of this kind it deals with finding a formation in a collection of unlabeled data. A slack definition of clustering could be "the process of organize matter into groups whose members are similar in some way".

A *group* is a collection of objects which are "similar" between them and are "dissimilar" .So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. How to decide what constitutes a good clustering? It can be publicized that there is no absolute "best" measure which would be independent of the final aim of the clustering. accordingly, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs.

G. Ensemble Clustering

Ensemble learning is a machine learning model where multiple learners are trained to solve the same problem. Cluster ensembles is the assignment to pool multiple clustering's of a group of objects into a single consolidated clustering, often referred to as the consensus clustering. It is also called as aggregation of clustering. Compare to an individual clustering method it can be used to generate more stout and stable clustering results. Ensemble is to perform distributed computing under privacy or in sharing constraints or to reuse existing knowledge.

Cluster ensemble can produce either by

- Using different clustering algorithms

E.g. K-means, Hierarchical Clustering, Fuzzy C-means, Spectral Clustering, Gaussian Mixture Model…

- Running the same algorithm for many times with different parameters or initializations.

E.g., run K-means algorithm N times using randomly initialized clusters centers, or by using different dissimilarity measures, or by various number of clusters.

III. CONCLUSION

Predicting students performance is mostly useful to help the educators and learners improving their learning and teaching process. This paper has reviewed previous studies on predicting students performance with various analytical methods. Most of the researchers have used cumulative grade point average(CGPA) and internal assessment as data sets. While for prediction techniques, the classification method is frequently used in educational data mining area. Under the classification techniques, Neural Network and the Decision Tree are the two methods highly used by the researchers for predicting students performance. In this paper, we reviewed different classification method used on student database to predict the student's performance in the upcoming semester on the basis of previous student's database and the work done on this till now. As we have seen, predicting students' performance earlier is a difficult task because it is a multifaceted problem and because the available data are normally imbalanced. To resolve this problem and to improve the accuracy and quality, the Support Vector Machine algorithm can be used which is showing the greatest accuracy among other techniques. As the result of the prediction accuracy is depending on the attributes or features that were used during the prediction process. Neural Network method gave the highest prediction accuracy because of the influence from main attributes. This attributes are hybridization of two features, which are internal and external assessments. Clustering technique and ensemble cluster can also be used to fine tune the quality of resulting dataset. Information's like Attendance, Seminar and Assignment marks were collected from the student's database, to predict the performance at the mean time of the course. The other attributes are collected by students and their respective faculties who know the behavior of students. This study will help to the students and the teachers to improve the performance of the students who are at the risk of failure. This study will also work to identify those students who needed special attention to reduce fail ration and taking appropriate action for the current semester examination.

REFERENCES

[1] Augusto Sandoval, Carlos Gonzalez,Rosa Alarcon, Karim Pichara, Maximiliano Montenegro," Centralized student performance prediction in large courses based on low cost variables in an institutional context", The Internet and Higher Education(2018).

[2] Chiara Masci, Geraint Johnes, Tommaso Agasisti ," Student and School Performance Across Countries: a Machine Learning Approach, European Journal of Operational Research(2018).

[3] Evandro B. Costa, Baldoino Fonseca, Marcelo Almeida Santana, Fabrísia Ferreira de Araújo, Joilson Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses", Computers in Human Behavior(2017).

[4] Raheela Asif, Agathe Merceron, Syed Abbas Ali, Najmi Ghani Haider, "Analyzing undergraduate students' performance using educational data mining", Computers & Education(2017).

[5] Hashmia Hamsa, Simi Indiradevi, Jubilant J.Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm, Procedia Technology(2016).

[6] Carlos J. Villagra-Arnedo, Francisco J. Gallego-Dur_an, Fara_on Llorens-Largo, Patricia Compan-Rosique, Rosana Satorre-Cuerda, Rafael Molina-Carmona, "Improving the expressiveness of black-box models for predicting student performance" Computers in Human Behavior(2016).

[7] Megan E. Cowiea, Logan J. Nealisa, Simon B. Sherrya, Paul L. Hewittb, Gordon L. Flettc, "Perfectionism and academic difficulties in graduate students: Testing incremental prediction and gender moderation", Personality and Individual Differences(2018).

[8] Maria Cutumisu, Daniel L. Schwartz, "The impact of critical feedback choice on students' revision, performance, and Memory" Computers in Human Behavior(2017).

[9] Maria Goga, Shade Kuyoro, Nicolae Goga, "A recommender for improving the student academic performance" Procedia - Social and Behavioral Sciences 180 (2015) 1481 – 1488.

[10] Gabriel D. Saenz, Lisa Geraci, Tyler M. Miller, Robert Tirso, "Metacognition in the classroom: The association between students' exam predictions and their desired grades", Consciousness and Cognition(2017).

[11] Mrinal Pandey, S. Taruna, "Towards the integration of multipleclassifier pertaining to the Student's performance prediction" Perspectives in Science(2016).

[12] Felipe Resende Oliveira, Tatiane Almeida de Menezes, Guilherme Diniz Irffi, Guilherme Resende Oliveira, "Bullying effect on students' performance" economia(2017).

[13] Athanasios Kolovelonis, Marios Goudas, "The relation of physical self-perceptions of competence, goal orientation, and optimism with students' performance calibration in physical education" Learning and Individual Differences(2018).

[14] Shaina Riciputi, Kristi Erdal, "The effect of stereotype threat on student-athlete math performance" Psychology of Sport and Exercise(2017).

[15] Daniel J. Cummings, Arthur E. Poropat, Natalie J. Loxton, Nicola Sheeran, "Development and initial validation of a multidimensional student performance scale" Learning and Individual Differences(2017).