(IJIEST) 2020, Vol. No. 6, Jan-Dec **e-ISSN: 2454-9584; p-ISSN: 2454-8111**

Leveraging the K-Means and K-Methods Algorithms to Analyse Student Performance

Aditya Goel

ABSTRACT

Nowadays, AI is used in every field whether it is education, defence, or enterprise. Its significance turns out to be more of a direct result of things to come of the students. Training information small scale ng is precious because the measure of information in the instruction framework is expanding step by step in advanced education is moderately new; however, its significance builds due to expanding database. There are many methodologies for estimating the student's exhibition. K-means is one of generally effective and most used technique. With the assistance of information mining, the shrouded data in the database is extracted, which enables to develop student's research. The decision tree is similarly a strategy used to foresee the student's exhibition. Latter, the most significant difficulties that educational foundations confront are the more development of information and the utilization of this information to improve quality which can help them to make better choices and judgement. Furthermore, grouping is one of the fundamental strategies used regularly in dissecting informational collections. This research applies a cluster test to portion students into groups, indicated by their qualities. Unsupervised learning, like K-means, is discussed. Training data mining is explained to examine the information accessible in the instruction field to bring the shrouded information, for example, essential and valuable data from it. With the assistance of mentioned techniques, it is anything but difficult to improve the outcome and fate of students.

I. INTRODUCTION

The machine learning clustering technique is nowadays used in various fields like- banking, credit card fraud detection, location analysis. Information mining is an information examination procedure used to distinguish concealed examples in a vast informational collection. Advanced education is significant for an understudy's life. AI gives different techniques these incorporate clustering, affiliation, k-means, DT, relapse, time arrangement, Neural Networks, and so forth.

The use of information mining in the instructive framework legitimately serves to examine members in the training framework. The students additionally suggest numerous exercises and assignments. Innumerable components could go about as a hindrance to the understudy for keeping up a high rate that reflects the general academic presentation in school. These components could be focused on by the employees in creating procedures to improve understudy learning and scholarly execution by the method of observing and dissecting the movement of their presentation. Information mining is additionally used to show how students utilize the material of a specific course. In an instructing, domain mentor can acquire criticism on students.

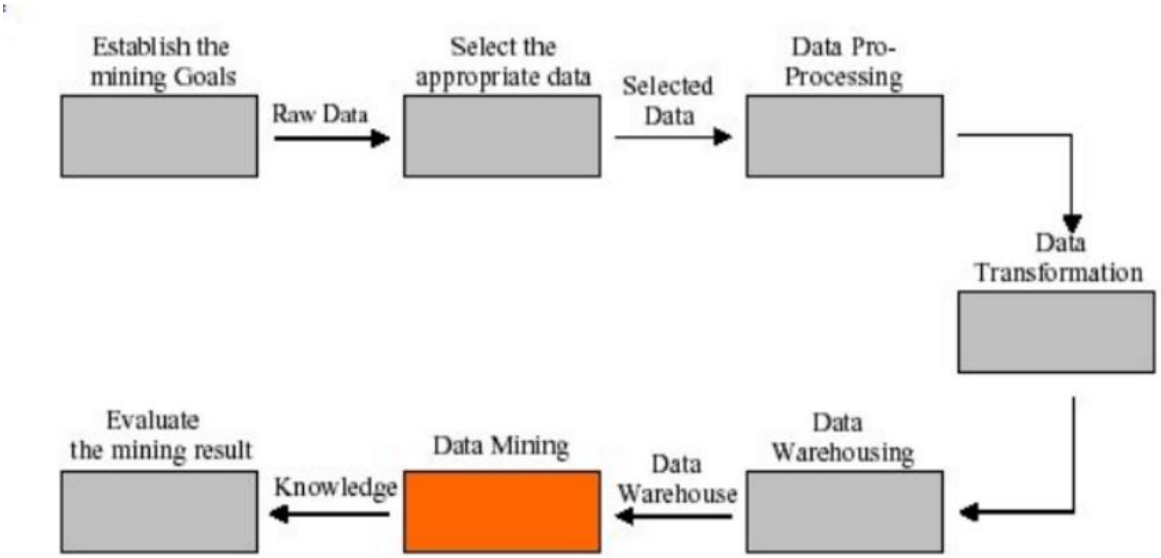


Fig: Different Stages of Data Mining Process

II. RELATED WORK

Fatma Chiheb[1]-Decision tree technique is used in this paper. A Decision tree is fabricated using the J48 calculation. Weka toolbox is used, and the CRISP-DM model is applied. It is an instance of an Algerian college. And the data is taken from the software engineering division. They tested the Decision tree and broke down the inaccuracy rates to pick the best info and yield. Various evaluations are considered as a quality, and understudy's exhibition is predicated.

V. Shanmuga rajeshwari[2]-They assess the student's exhibition utilizing characterization strategies. The info information is gathered from Ayya Nadar Janaki Ammal College, Sivakasi, from the software engineering office. For highlight determination number of techniques are discussed. Preparing report is applied to the informational index, and the classifier model has been created. Decision tree arrangement was used to anticipate the student's exhibition.

M.Durairaj[3]-Educational distinctions and execution depend on different elements like individual subtleties, social, and so forth. Weka tool is being used to gather the information of students. They Gather the data and find the relationship between learning and academic performance. The dataset contains a detailed score of each semester with subjects. To do this, they have used k-means clustering. In all this, they have taken data of 300 students out of which they have shortlisted 38 records for analysis. The Confusion Matrix shows the pass, fail, and absence for the exam. They compare the weighted mean average with DT and Naive Bayes techniques.

Mr.Shashikant Pradip borgavakar[4]-Here the information grouping is utilized as k-means bunching to assess student's presentation. Their presentation is evaluated based on a class test, mid-test, and last test. In their model, they are estimated by an inward and outside appraisal, in which they story class test marks, lab execution, analysis, and so on, and the final grade of students is anticipated. They produce the chart which shows the level of student's getting high, medium, low GPA

Edin Osman begovic[5]-In this paper, administered information mining calculations were applied. An alternate technique for information mining was thought. The information was gathered from the study led throughout the late spring semester at the University of Tuzla. Numerous factors like Gender, GPA, Scholarships, High school, entrance exams, grades, and so on are taken for the exhibition. Gullible Bayes calculation, Multilayer Perceptron, J48issued. The outcome shows that the naive Bayes classifier outflanks in the predication DT and neural system strategy. These will help the understudy for what's to come.

E. Venkatesan et al. [6]- In this article, the bunching and grouping calculations were thought about utilizing grid research centre programming for the underlying information WEKA programming is used. The informational index of students was gotten from personal expressions and science schools from Chennai city. Close around 573 students are there in the database. In the subtleties, they take the inward test and end semester test subtleties. A calculation, for example, J48, has utilized permits the information ascribe to get the older model. Grid Laboratory is being used for estimating the operational of a few information mining calculations. There is a table for mistake measures.

III. EXISTING SYSTEM

A DT is a standardized procedure, and there are numerous strategies to fabricate the DT and to foresee the exhibition. There is an immense measure of information created in the instructive framework. These can be misused to separate helpful information. In the present context, loads of strategy is utilized to predict the student's demonstration. In the current structure, the DT is assembled using the J48 algorithm. There is an instance of Algerian college where an understudy's exhibition is anticipating utilizing the DT. DT technique is sensitive because the DT offers numerous potential responses. On changing the root hub, it turns the tree and has an alternate forecast. There is a tremendous measure of information in the instructive framework of the current system; they foresee the exhibition based on the past semester result. A DT using the J48 algorithm, which is exceptionally difficult to assemble due to its parting. Tree calculation utilizes numerous tests to decide a specific split. Even before that has been resolved, the count has attempted various mixes of factors to get the best division. Weka toolbox is utilized, and the fresh dm model is applied.

IV PROBLEM STATEMENT

It is perplexing that there is an enormous measure of information in the instructive framework. Foreseeing the student's exhibition, there ought to be a strategy that is increasingly active and created favourable outcomes. A DT is a characterization strategy that is less proficient when contrasted with grouping strategies J48 is a DT calculation that is utilized for foreseeing understudy execution. Yet, it is less effective as a contrast with k-means bunching procedures. Decision trees analyze just a solitary field at once, prompting rectangular order boxes. This may not perform well with the original records in the DT. Figuring's can get unpredictable, especially if numerous qualities are not sure, and if many results are connected. A DT isn't steady; it implies that little change in the information can prompt a considerable difference in the ideal DT structure.

V PROPOSED WORK

Forecasts of students' presentations should be possible utilizing Machine Learning calculation. Bunching is an innovation wherein there is a group with a combination of comparable information. K means algorithm is used to foresee the presentation of student's. K means there is an unaided AI calculation. K means clustering set the segment of n explanations into k groups in which every perception has a place with a group with the closest mean. The group is delayed with the mean estimation of the items in a cluster, which can be seen as the cluster centroid. The thought is to characterize K focuses and one for each bunch. These middles ought to be set legitimately because a distinctive area gives diverse outcomes. So a better decision is to put them far away from one another. The following stage is to take each direct having a place toward a given informational index and partner it to the closest focus. At the point when no point is pending, the initial step is done, and an actual gathering age is finished. Now we have to recalculate k new centroids by viewing from the past advance. After we have these k new centroids, another system must be done among similar informational index focuses and the closest new focus. A circle has been produced. Because of this circle, we may declare that the k places change their area bit by bit until no more changes are done or at the end of the day communities don't move anymore.

VI COMPARATIVE STUDY OF EXISTING AND PROPOSED WORK

Decision Tree (J48 Algorithm)- A DT is a structure that incorporates a root hub, branches, and leaf hubs. Each inward centre speaks to a test on a property, each node implies the result of a test, and each leaf hub holds a class

name. The highest centre in the tree is the root hub. J48 is an expansion of ID3. The new structures of J48 are representing missing qualities, DT pruning, consistent trait esteem ranges, inference of rules, and so forth. In the WEKA information mining apparatus, J48 is an open-source Java usage of the C4.5 calculations. The WEKA device furnishes various alternatives related to tree pruning. In the event of potential overfishing, pruning can be utilized as an instrument for précising. In different calculations, the arrangement is performed recursively until every leaf is untouched; that is, the order of the information should be as flawless as expected under the circumstances. This calculation delivers the principles from which specific personality of that information is produced. The goal is slowly speculation of a DT until it picks up the harmony of adaptability and precision.

Disadvantages of decision tree algorithm-

- For data including definite variables with different number of stages information gain in decision tree is biased in favour of those attributes with more levels.
- Tree structure prone to sampling – While Decision Trees are mostly robust to outliers, due to their tendency to over fit, they are prone to sampling errors. If sampled preparation data is somewhat different than evaluation or scoring data, then Decision Trees tend not to create great results.
- Tree splitting is locally greedy – At each level, tree looks for binary divided such that impurity of tree is reduced by maximum amount.
- They are often relatively inexact. Many other predictors perform better with similar data. This can be remedied by changing a single decision tree with random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree.

K-means Algorithm- Clustering is a technique for gathering a lot of objects of a similar kind. For example, purposes in a similar gathering are increasingly like each other as a contrast with those in different collections. It is otherwise called group investigation. It isn't explicit calculation; however, the general errand to be understood. It tends to be accomplished by numerous counts that contrast in their comprehension of what makes a group and how to discover them effectively. The principle bit of leeway of bunching strategies is that it is versatile to changes that are less in arrangement procedures. It also assists single with valuable excursion highlights for various gatherings. Bunching strategy or group examination is, for the most part, utilized in applications, for example, expectation, statistical surveying, design acknowledgement, information investigation, and picture handling.

Advantages of K-means algorithm-

It is easy to implement.

When there is a large number of variables, K-Means may be computationally faster than other clustering techniques.

K-Means may produce higher clusters.

An instance can change clusters (move to another cluster) when the centroids are recomputed.

VII IMPLEMENTATION DETAILS

K means algorithm is used to anticipate the exhibition of students. k-means bunching expects to segment on perceptions into k groups in which every perception has a place with the bunch with the closest mean, filling in as a model of the bunch. The calculation is as follow-

Step 1-Select the centers cluster 'K' by the Elbow method

Step 2-Take the centroid i.e. mean value based on centers cluster by elbow method.

Step 3-Calculate the distance among each data point and the centroid (mean value).

Step 4- Assign each data item to a cluster whose distance is minimum.

Step 5- Recalculate the new mean.

Step 6- Recalculate the distance among each data point and new mean value.

Step 7- If no data point was moved then STOP, otherwise repeat the step until the convergence is met.

VIII RESULT ANALYSIS

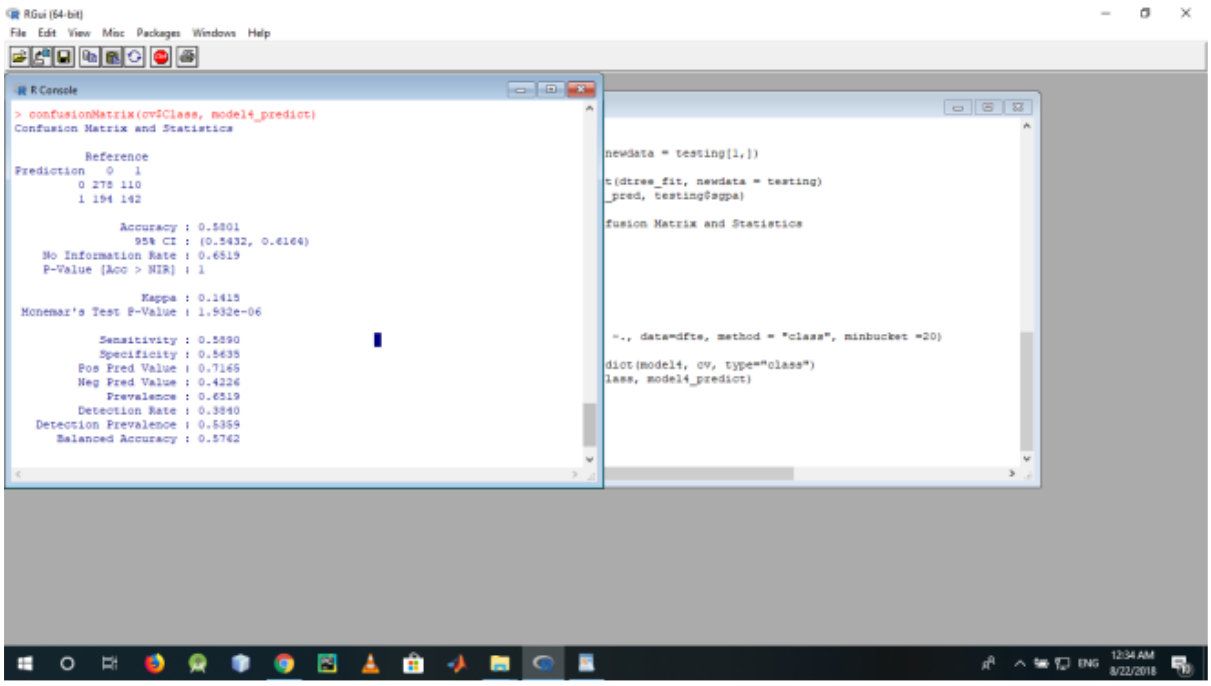
K-Means algorithm is used to predict the students' performance. It is stable and efficient as compared to decision tree. In the dataset, we take the attribute.

Student_id, -Unique id correspond to every student.

Semester (sem1-sem2)-Semester id correspond to semester i.e. (sem 1 or sem 2).

Subject-marks (sub1-sub5)-Each subject marks correspond to every student in both the semester.

Sem Result (Sgpa)–The percentage of those students in that particular semester.



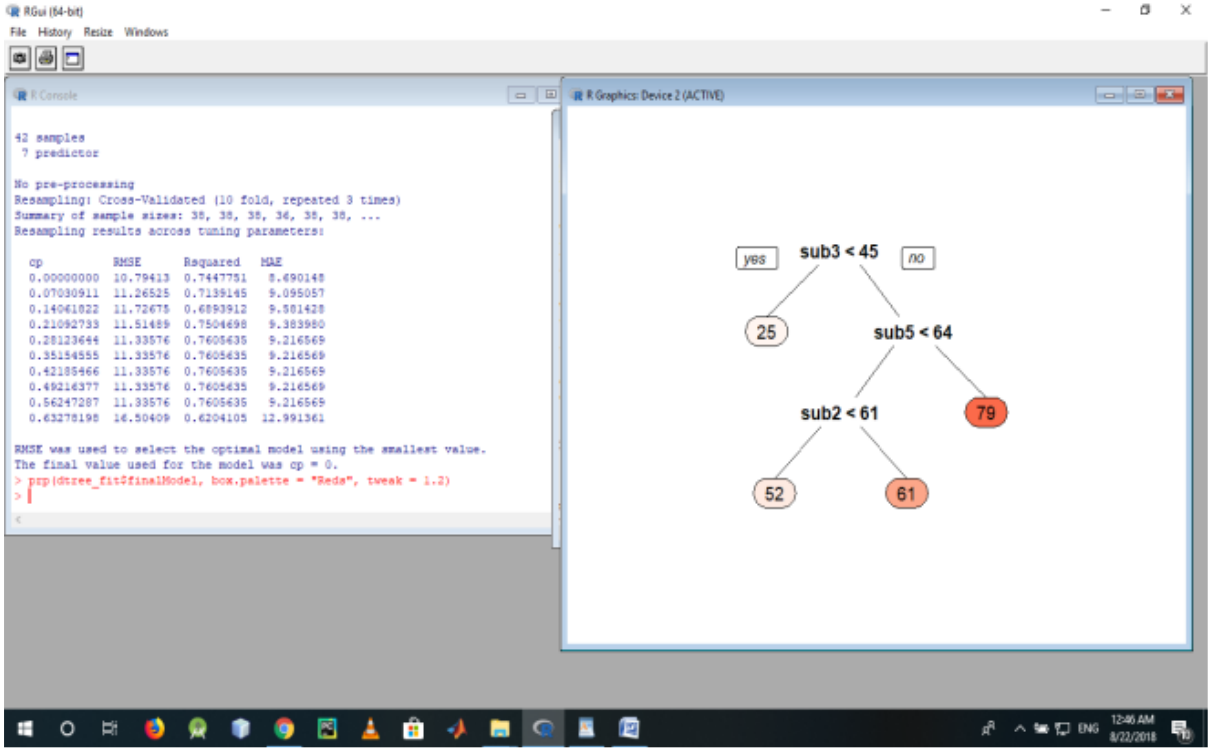


Table: Compression Study Existing Work Versus Proposed Work

Parameters	Existing Work	Proposed Work
	Decision Trees	K-Mean
Correct classified rate	Aprox. 5%	Above 71 %
Source Code Execution		10.32 Seconds

IX CONCLUSION

AI is a rising innovation as it is used widely. In the contemporary world, bank, labs, telecommunication, and mechanical uses AI to a great extent. Information mining is a piece of it that helps in expectation; the future forecast is significant in many spots, which allows to such an extent. Numerous calculation is fabricated, and increasingly more exploration is utilizing the idea of each innovation. We overview multiple papers for the expectation of students, and presentations. In addition, DT strategy is being also used in many spots; however, on contrasting with bunching procedures, K means is increasingly productive and stable. Student's presentation is so significant for their future that it helps them to understudy and assist instructors with initiating guardians. Numerous enormous organizations utilizes the idea of AI for the forecast.