PREDICTIVE ANALYTICS FOR E-LEARNING SYSTEM USING MACHINE LEARNING APPROACH

S.V.N. Sreenivasu[1], M. Aparna[2]

[1,2]Department of Computer Science and Engineering, Narasaraopeta Engineering College (Autonomous), Narasaraopeta, Andhra Pradesh, India.

Corresponding Author: Dr. S.V.N. Srinivasu

E-mail: dr.svnsrinivasu@gmail.com

Abstract

Soft-learning courses are sought-after as well as late. The need to examine understudy's presentation and anticipating their exhibition is expanding alongside it. With the developing notoriety of instructive innovation, different information digging calculations appropriate for anticipating understudy execution have been surveyed. The best calculation is based on the idea of the forecast that the staff needs to make. As the measurement of understudy information broadens the need to address and manage the complexities of the information connection, it is a test for the discovery of the understudy at risk of being short-lived. In this paper covers the ID3 and C4.5 algorithms used for Predictive Analytics on understudy's presentation and Big Data with cloud.

Keywords: Soft-Learning Techniques, Machine Learning Approach, Basics of Predictive Analytics, Decision Tree Techniques (C4.5 and ID3), Big Data

I.   Introduction

Soft-Learning Systems is associating a number of educational organizations to under-studies that defeat minor area confinement and study time. Such frameworks are intended to serve people in learning subjects from distant fields separated from ordinary courses. It turned out to be workable for an educator to be linked with post-school under-study through such a framework[I, II]. The emergence of Student Data Mining (SDM) technology, distinct calculations has been updated in the examination of soft-learning Systems. The aim of SDM is to prepare predictive models with high accuracy, reliability and simplicity of elucidation. Data relapse and Data classification are the two common data techniques for distinguishing under-studies are the risk of being short-lived[IV]. The document examines the distinct calculations envisaged by analysts in anticipation of an understudy exhibition. In addition to the need to address the development of understudy information and its complexities[III, V]. Big Data is included in the soft-learning framework using the data mining system [VI].

Predictive Mining Activity:

Anticipating understudy's presentation when utilizing Soft-learning frameworks bears the capability of recognizing the understudies in danger of fizzling. In Predictive data Mining there are 3 fundamental Classifications – i. Data Classification method. ii. Data Regression method. iii. Data Density Estimation. Regarding foreseeing understudy's Exhibition the use of data Classification system and data Regression strategy are progressively prevalent. Understudies in danger of horrible showing can be recognized by applying order[V] .Then their presentation is given a selected incentive once relapse is connected. Among the soft-Learning framework with the Neural Networks (NN) and decision Tree algorithms are the foremost acceptable calculations with the discerning investigation on understudies. A little of the current properties utilized in data mining are CGPA, Internal Analysis, External Appraisal, Demographics and psychological science components[VI].
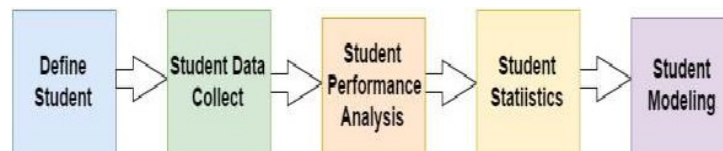
Fig. 1: PredictiveAnalyticsProcess on Data Collections

Extract The Data From Large Student Data:

To alter making ready of huge datasets we have a tendency to utilize huge information system. Here the customary calculations of knowledge mining are reborn into Map scale back calculations to run them on hadoop teams are decoding their information examination principle to the map reduce scale back work. It's to be unbroken running over hadoop data bunches[VII, VIII]. Hadoop teams are structured with Associate in Nursing extent that data store and procedure stupendous data in a very distributed process with condition. MapReduce is one among the middle hadoop half for procedure confiscate data. Map Reduce programming model contains of two separate modules. The primary is of guide work to train a dataset and it converts into another dataset. Here the individual elements are splitted down into tuples[VII, IX]. The yield from a guide is given as a contribution to minimize occupation and it consolidates those data tuples are significantly littler arrangement of tuples[V].
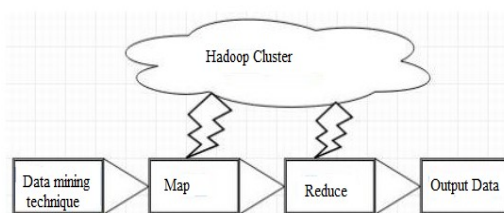
Fig. 2: Data mining technique using Map Reduce Procedure with Hadoop cluster

II.   Related Work

A wide range of information mining methodologies have been utilized to break down understudy's dropout rate. They weight on how prescient investigation can indicate in distinguishing significant properties in an understudy record set. We have attempted distinctive blend of qualities, for example, CGPA points, internal assessment marks, external assessment marks and demographics graph are used check the precision of these data mining calculations[II, IV]. Then it presumed that Decision Tree algorithms and Neural Networks are performed with high precision accuracy.   Data mining calculations techniques, for example, Decision Trees, Random Forest, Neural Networks and Support Vector Machine. It is utilized wekatool for test informational indexes on understudy records and connected different order calculations on it. It detailed that Decision Tree had the most noteworthy precision pursued by NeuralNetworks (NN). We checked on different information mining procedures connected to Soft-learning issues[IV, V].

For the understudy execution expectation class it has recorded NeuralNetworks (NN), Bayesian Network, Linear relapse Decision Tree algorithms (like ID3 and C4.5) are the most favoredcalculations[V].The data mining techniques are superior to measurable methods for getting ready monumental calculable informationacceptive this announcement as a base to differentiate an expectation model passionate about the key ascribes that are pertinent to foreseeing understudy's exhibition. Creator utilized characterization calculation in 2 stages 1) grouping each single distinguished property and apply calculations on them. 2) Applied quality determination calculation and positioned all characteristics supported their event. Here understudy dropout rate examined utilizing Rule primarily based order utilizing Trip in rail, then alternative tree calculation accustomed speak to information of characterization utilizing J48[IX, XII].

With the developing computerized understudy data, it winds up essential to avoid wasting each understudy's data. at the same time once such an oversized quantity of knowledge is to be ready for choosing up data, we'd like a mix of best distinguished calculation on massive organization. We displayed the combination of massive information with Soft-learning System. It featured the impact of utilizing three stage styles for a syndicate of schools[X]. They expressed that such a connection are going to be useful examine, mapped out and find to monumental informational indexes within the cloud condition. They need targeted on managing unstructured data utilizing the graphical device[XI, XII].In academic data processing grouping, bunching and relapse are the 3 most well-known methods utilized by various creators to survey preciseness of forecast. Utilized mixture of grouping and relapse system to look at the understudy dropout conceivable outcomes. They used discerning displaying for separating this hide knowledge from soft-learning system[V, VI]. Data is gathered from a model is readied, expectations are created and also the model is approved for lot of data is gotten too. Here a bunching technique is employed to acknowledge scholastically in peril understudies and organize the understudies in line with execution. They need used K-Means calculation for this purpose for

classification. There after multiple correlations is employed for creating predictions relating to student's dropout probabilities from the course[V, VI].

Comparative Analysis:

Aim in Title	Work	Problem Statement	Year
Predicting Analytics on Student'sPerformancevictimizationdata processing Techniques.	Concluded that C4.5 and ID3 algorithms concepts each performed with highest accuracy	That Neural Networks don't seem to be smart with psychology factors. That Neural Networks	2013
The Data Mining Techniques and its Concepts	Data Classification Technique applied.	C4.5 and ID3 algorithms are used, for to address Classification and Regression.	2011
Predictive Analytics by applying Machine Learning Techniques	Application of Data Classification using Decision Tree algorithm's	Decision Trees algorithms are working with Big Data and Classifying it.	2013
Using the Student academic Environment for Big Data	To classify the Unstructured and structured data using the graphical Tool	Extracting Data in the hadoop cluster Student academic data was not included	2013
Machine learning Approach for Data Classifying & Data Prediction on Student Result Using Map-Reduce	Data Classifying & Multiple Regression for to predict poor student failure	No comparison to made before classifying algorithm to be used	2013

Data Analysis:

In the wake of assessing each single connected paper we tend to came to admit the assorted info mining calculations that are utilized in discerning examination. It feature there is a high non-linearity and complicated association between the ward and autonomous factors, a selection tree model can beat associate degree old style font relapse strategy in foreseeing understudy execution[IV, VI]. To boot on the off probability that we've got to manufacture a model that is something however troublesome to open up to people, a selection tree model can systematically show improvement over a right away model. Selection tree models are significantly a lot of easy to translate than direct relapse. A selection tree will method each numerical and every one out info. To boot giant measures of knowledge are often poor down utilizing commonplace process assets in lesser time[VI, VII]. The usage of massive information in several areas thanks to increasing information has to boot projected a

requirement to coordinate Soft-learning Systems with massive information[III]. The massive info stage can build the productivity of the discerning mining calculation. The equipment support from massive information can facilitate in skillful factual demonstrating of the understudy information[V].
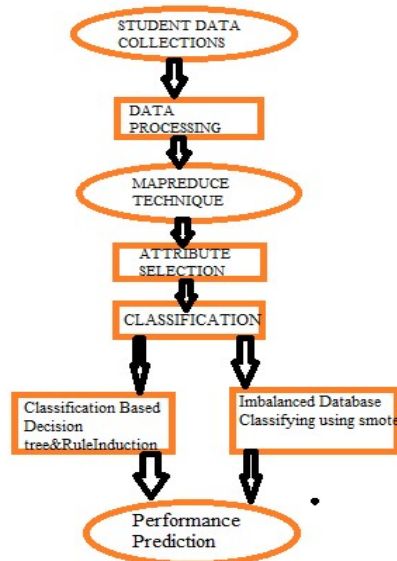
III.   Proposed System

Decision Tree based soft-learning calculations are that appear to be truly outstanding and most helpful prescient of data mining calculations. Decision Tree based strategies enable understudy prescient data models with high precision, dependability and simplicity of elucidation. In contrast to straight models, will probably delineate direct connections among understudies' properties great. They are versatile at taking care of any sort of issue close by (characterization or relapse)[I]. ID3 algorithm has been utilized so far under choice decision tree procedure to make the forecasts. It is a famous rule utilized in a wide range of information science issues[II, VI]. It works for both clear cut and consistent info and yield factors. Choice tree is one of the quickest methods to recognize most huge factors and connection between at least two factors. Clearly, with the developing cooperation of enormous information on cloud condition, essential choice tree calculations show numerous limitations[VII, VIII]. Executing a decision tree can take a great deal of time when the size of understudy dataset is very large. Subsequently new prescient system ought to be utilized for such datasets gathering. I n this paper we propose to utilize Decision Tree method's C4.5 calculation (successor of ID3) in a hadoop system to anticipate an understudy's presentation. Our purpose behind proposing C4.5 calculation is on the grounds that it does: i. treatment of both discrete qualities, and persistent characteristics; ii. It can process somewhat finish preparing informational collections with qualities not present; iii. Pruning should be possible while building the trees to forestall over-fitting issue [VI, VII].

C4.5 Algorithm Procedure:

The huge number of choice tree strategies are exist however the strategy that functions admirably by and large on a wide range of understudy's informational collections is as yet the C4.5 choice tree technique and a portion of its variations[VI]. Expect the preparation information to be a set and so forth of default characterized examples. Every example to speak the highlights of that example. The preparation test is an amount speaks to the class with which every example is related with. At each hub of choice tree C4.5 calculation chooses a solitary characteristic of the information that will part informational collection of tests SI and its littler sub-sets which can have a place with a class or the some other[VII]. It is the repetition of expected data   that gives result from choosing an element for part the information. The component reason which has the most non excess data increase is respected in basic leadership process. A while later C4.5 calculation seeks after the littler sub-records which have following most elevated non excess data gain[VII][IX].Our purpose behind proposing C4.5 calculation is on the grounds we are analyse the students soft learning system to predict the best accuracy on result analysis using Machine Learning algorithms.

Procedure for Map-Reduce:

Here to anticipate an understudy's presentation, traits like day to day performance, sample test, Seminars and Assignment imprints will be thought about [VI, I]. We propose the utilization of C4.5 and ID3 technique in Map-Reduce structure for examining the understudies exhibition. It is very well may be clarified with the accompanying stream graph.

Fig. 3: The Proposed System Data Flow Diagram

Step 1: Data extracting from the school database. Here the point is to get understudy experience and scholarly data to comprehend real component that may affect understudy outcome.

Step 2: Data gets pre-processed. While beginning all presented information can be joined as one dataset. In this action unimportant understudy's data gets expelled. Hardly any progressions are moreover done to the estimations of certain highlights. At that point, the non-stop factors are changed over the discrete variables, which give a considerably more understandable perspective of the information.

Step 3: In Map Reduce setting with Mapper isolates the understudy information with key qualities, and after that Reducer totals the outcome. From that point, the working result of MapReduce significantly is subject to how consistently it scatters a understudy's information.

Step 4: Attributes are picked in fluctuated structures like CBFS (Correlation Based Feature Selection).The properties are utilizing the Feature Selection Techniques these are called correlation based feature selection. The CFS readies a likely measure and organizes the subset of properties with the single characteristics. The arrangements of highlights are un-equivocally related with classes, notwithstanding those highlights those are in less related.

Step 5: Data acquired will be arranged for executing the decision tree and standard based enlistment. The data increase measure is taken to pick the data at each hub of the decision tree technique. The data increase determined favours to pick properties which have an enormous tally of qualities.

Step 6: The Datasets which are not adjusted to grouped with decision tree calculation. This issue with respect to imbalanced information characterization happens when include of occurrences in a single class. It is a lotof lesser contrasted with include of occasions in the rest of the classes. An answer for this issue is to work while the information is being pre-processed. It is finished by examining or adjusting of class appropriation. There exist different information adjusting or rebalancing calculations.

Step 7: At last step the presentations are estimated by checking precision related terms like True/False positive rate and True/False negative rate.

IV.   Result Analysis

**Table1: Comparison of the accuracy and execution time of ID3, C4.5 and Map reduce**

	ID3	C4.5	Map reduce
Execution time (Sec)	0.03	0.09	0.26
Recognition rate (%)	0.7509	0.8735	0.8701
Error rate (%)	0.2471	0.1359	0.1249

Fig: 4 below shows a comparison of the three algorithms of construction of decision trees that produce predictive models with the best class judicious accuracy.
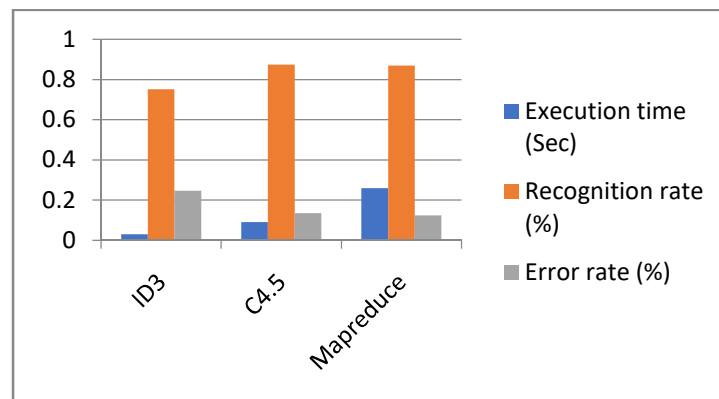
Fig. 4: Comparison of the accuracy and execution time of ID3, C4.5 and Map reduce
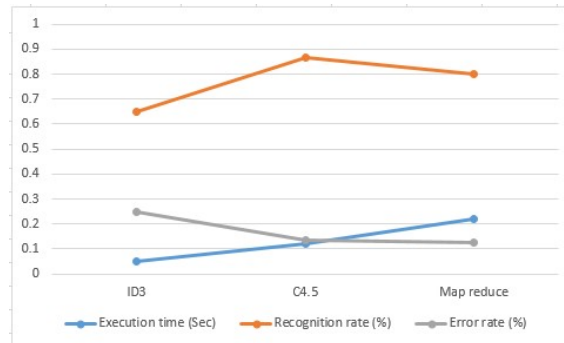
Fig. 5: Comparative Analysis of ID3,C4.5 and Map Reduce

V. Conclusion

In this we propose various information mining calculations utilized for anticipating understudies' exhibition of an soft learning system. Decision Tree Algorithms are better when anticipating understudies' exhibition was discovered better with all sorts of information extracting procedures, input esteems and exactness.C4.5 and ID3 Decision Tree Algorithm for breaking down enormous understudies' of data was presented. It holistic with map-reduce structure is proposed when contrasted with traditional decision tree execution. Hence it will prompt progressively productive prescient investigation of understudies' presentation. It will support us to recognize the understudies in danger of falling (or) dropping off from the course. As progressively number of colleges will work together to learn about understudy's presentation. Such enormous information therefore will need efficient data mining algorithms for execution. We proposed C4.5 and ID3 decision tree algorithms for   execution in Big Data's Map-Reduce structure for soft learning systems to all the more likely foresees understudies' execution

References

  I. A. M.Shahiri, W. Hussain and N. A. Rashid. "A Review on Predicting Student's Performance using Data Mining Techniques", Procedia Computer Science, vol. 72, pp. 414-422, 2015.

 II. B. Logica and R. Magdalena, "Using Big Data in the Academic Environment", Procedia Economics and Finance, vol. 33, pp. 277-286, 2015.

III. C. T. Tsai, et. al., "Exchanging course content mechanism for Moodle LMS",In: Proc. of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Huangshan, China, IEEE, pp. 464-467, 2010.

IV. D. Clanfield and J.Sivell, "Cooperative learning & social change: selected writings of CélestinFreinet. Our Schools: Canada. Firdausiah Mansur, AndiBesse, Yusof, Norazah& Othman, Mohd. Shahizan. (2011). Analysis of social learning network for Wiki in Moodle e-learning.

V. E. A. Kareem and M. G. Duaimi, "Improved Accuracy for Decision Tree Algorithm Based on Unsupervised Discretization", International Journal of Computer Science and Mobile Computing, vol. 3, no. 6, pp. 176-183, Jun. 2014.

VI. H. Chauhan and A. Chauhan, "Implementation of decision tree algorithm C4.5", International Journal of Scientific and Research Publications, vol. 3, no. 10, pp. 1-3, Oct. 2013.

VII. H. Gulati, "Predictive Analytics Using Data Mining Technique",In: Proc. of 2nd International Conference on Computing for Sustainable Global Development, New Delhi, India, IEEE, 2015.

VIII. J. Han and M.Kamber, "Data Mining Conceptsandits Techniques", Morgan Kauffmann Publishers, 2011. DOI: https://doi.org/10.1016/C2009-0-61819-5

IX. K. Kinley, "Faculty and students' awareness and challenges of e-learning in a college of education", Journal of the International Society for Teacher Education, vol. 14, no. 1, pp. 27-33, 2010.

X. M. A. Al-Barrak and M. Al-Razgan, "Prediction of Student's Final GPA implementing Decision Trees: A Case Study", International Journal of Information and Education Technology, vol. 6, no. 7, July 2016.

XI. M. G. M. Mohan, S. K. Augustin and V. S. K. Roshni,"A Big Data Approach for Classification and Prediction of Student Result Using Map Reduce",IEEE Recent Advances in Intelligent Computational Systems, Trivandrum, India, IEEE, 2015.

XII. W. Dai and W. Ji, "Implementing Map Reduce with C4.5 Decision Tree Algorithm", Journal of Database Theory and Application, vo. 7, no. 1, pp. 49-60, 2014.