Assessment Analysis and Performance Prediction using M5 Rules

Kolli Sai Poorna Chand, Prabakaran N, Ramani S, Damaraju Venkateswara Rao, Sriram Vemparala

Abstract*: With time, technology increased in a pretty decent speed. It helps us to save our lot of time and our effort in different aspects of work thus making our life much easier than it used to be. In past, Education was not given that much importance as it has now been crowned. It resulted in a lot of changes in education system. All the education institutions like colleges, schools, research colleges have seen a gradual increase in students on quality perspective. It had made to lot of works to be done in a stipulated time. Such as when it comes to evaluation of marks after the students have given their examination. For that it takes an incredible effort and time to manually calculate the marks of each student and making it more difficult with increasing numbers. To solve this, a web-based system is what came to the mind of ours to make such that they calculate the marks. A future prediction of marks has also been introduced in the system. It means that marks prediction of final exams will be done to know how the student will perform. It is much needed to have a system like this which can help to develop a warning for both student and faculty and also the parents to bring the student back on track. It will be a very helpful system in the field of education

Keywords: Educational Data Mining (EDM), M5 rules, Linear Regression, Random Forest Regression, Student performance analysis, Prediction

I. INTRODUCTION

Education has become a core part for development and betterment of nations. Education paves the way for individuals to formulate new ideas and bring innovation in the society [1]. Lack of deep knowledge and understanding of concepts can cause hindered development and may prevent system management from achieving quality objectives. In this education empowered world, the core is occupied by students who are responsible for shaping the future [3]. These students receive knowledge through universities and these universities have a well-defined system to determine a student's potential strengths and weaknesses through grades, projects, practical's and various other methodologies. This project focuses mainly on determining and analyzing the academic performance of a student based on his/her marks scored in various types of tests/exams conducted by the university. The proposed system will manage information about subjects offered in various semesters along with the marks scored by each student in different subjects in each semester. It will provide faculties with a portal to input marks for their students and view and analyze their marks and reports to identify their potential weaknesses and focus on these areas. The students will have access to their personal data and can analyze their marks and observe their potential future grades to find out where they lack and work on those areas. The admin on the other hand will oversee all the actions of both students and faculties including reports and predicted outcomes and will have the responsibility of assigning semesters and students to faculties. The predictions will be generated using M5 model trees, Linear Regression and Random Forest techniques to identify the best outcome and reports will be plotted in graphical format accordingly in the web portal.

Some of the objectives identified during development of the system include:

i. A system to store and validate information about various users
ii. To store information about subjects offered various semesters.
iii. To store marks obtained by students in each and every semester.
iv. Generation of reports from the data obtained.
v. Analysis of data based on aspects.

Data Mining

It is a term used to refer to a set of different functionalities and perspectives grouped together as a whole. Data mining techniques allows users to analyze and classify data with various dimensions and summarize the relationships recognized. Data mining is the process of identifying and finding patterns and correlations with several fields in huge databases. It is the paradigm of extracting information from large datasets. In our model data mining algorithms are applied to analyze the performance of students based on their marks obtained in various subjects as well as predict their future scores in the finals to identify whether they need additional assistance. The use of data mining algorithms proposed include M5 rules, random forest and linear regression. The results of each of these algorithms are then compared to identify the best suited algorithm and obtain the most accurate result.

II. RELATED WORK

Chew Li Sa,et al.; [1] has proposed a framework to predict the performance of students of Faculty of Computer Science and Information Technology (FCSIT) in

Revised Manuscript Received on November 14, 2019.

Kolli Sai Poorna Chand, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Prabakaran N, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
S. Ramani, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Damaraju Venkateswara Rao, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Sriram Vemparala, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

the course "TMC1013 System Analysis and Design" as the management system in the University Malaysia Sarawak (UNIMAS) does not the grant lecturers' access to the system. The framework designed by the author uses a dataset of 637 students records and applies classification algorithms such as CART, Random Tree, J48 to identify the best suited data mining algorithm for the framework and predict the performance. The primary focus of this paper was to predict the grades of students in the particular course and provide faculties with a system to view the predictions.

[2] Ishwank Singh, et al.; designed a framework which categorizes students into various clusters based on their performance in different scenarios. The scenarios are classified as: Academic Performance, Co-curricular (Projects/Internships/Skills) Performance and overall performance. The algorithm used for clustering is the k-means algorithm and the k value are calculated by repeatedly calculating the centroid. The author has used 3 clusters in this system and trained the model on a sample of 300 data and the results obtained displayed three clusters with varied predictions which showed that the median value falls in the category of good performers when compared with the benchmark values. The author has provided a well-defined system for categorizing a group of students as good or bad performers but does not focus on individual student's performance.

Bo Guo, et al.; [3] proposes a system for predicting students performances taking various attributes under consideration. The system is designed to predict a student's final grade. The author has developed a classification model using deep learning techniques with several hidden layers to achieve high accuracy. The model initially concatenates all data into a flatten vector after which features are extracted from the data using sparse encoder. The data is fed into several neuron layers with designated weights assigned. After this phase the network is fine-tuned using backpropagation method. The author has trained his model on a dataset containing 120000 records and achieves a great accuracy for each of the grades predicted which is demonstrated by producing a comparative analysis of the proposed algorithm with existing algorithms.

[7] Zhiwu Liu, et al.; developed an analysis forecasting model using the decision tree algorithm along with formulating a classification rule so that the negative factors affecting student's performance can be identified and subsequently rectified. The decision tree algorithm used is C4.5 and the model is trained on dataset formed by gathering data from actual students. Each branch is treated as a grade node with specific attributes and information gain is calculated which is used to rank the attributes. Classification rules are applied after obtaining the results of each branch and the most impactful attributes are identified.

Students' academic performance [4] play a considerable role in stakeholders' decisions for the universities. Tan Chin Hui, et al.; designed a system to focus on factors affecting students' performances and merges the findings with past academic records to study academic pattern. CRISP (Cross-Industry Standard Process for data mining) model has been utilized for the purpose of organizing, analyzing and as well as implanting results from the data. The original dataset consists of 15 attributes from which 8 major attributes have been identified and used for obtaining the results. CHAID (Chi-Squared Automatic Interaction Detection) classification techniques was used on the 2228 records classify the records on the basis of CGPA as it was found as a major factor for determining performance pattern. The author used PASW 13.0 modelling tool to visualize the results.

III. METHODOLOGY

A. Data Processing

Data Gathering

Colleges and universities are having a lot of problems in making their student to work in a way that would help them to do smart work making the overall performance of student not better. It leads to degradation of the standards of developments of that respective institution. There are many causes of degradation. First of all, the enrolment in that particular university or college decreases. Next during placement time, companies do not tend to visit those colleges in which the students do not have a good academic performance.

There are data present from different sources like different examinations in which there are different subjects for each student. But there is no organized way to make this data of efficient usage. They are just known to the faculties which are later informed to the students. To have a good usage of those marks, there is a system developed by us which will store the marks for later use which will be told soon

A web portal has been designed which will store these marks. They are also stored in an organized database such that those data are visible to the user in form of tables or graphs. Also, the data that is getting as input is on live status which means the marks of all the minor examination scores are being input by the faculty. There are some rules on which focus should be given before storing the data:

- Each faculty has to be prescribed to a particular semester
- Each student has to be prescribed to a particular faculty
- There can be more than one student under one faculty.
- Only the faculty prescribed to a particular student can upload the marks of that student

There is no need of calculation of total marks of that subject in the particular term of that semester or the total marks of all the subjects as it is automatically done by the web portal and stored in the database. There is also Display of all the marks in the form of tables in three different modules that is:

- Admin Login-In this all the data is visible. The faculties registered are visible. All the students registered in the system is displayed as a table. There is a tab in which marks of all the students are shown as marks of 3 units all having internal and external marks.
- Faculty Login-In this module, the concerned faculty login has their student stored so that the faculty can enter the marks of students thus making an increase in data. These are the essential data needed for future use in modelling.
- Student login-In this module, the data stored as marks of that student can visible to be him/her.

Now the usage of storing all this data
and displaying it should have of some

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

use. So, the data stored will be taken into account for prediction of the final marks that a student can get thus helping him /her to decide whether he would have to study that subject with more dedicated time or not. Those predicted data will be also shown to the user for each subject and for all the students.

Data Analysis

Once data has been collected it needs to be given a look of how that data can be taken for a good use. Many things need to be seen as whether all the data has to be taken in account for or there are less that of which attribute can be made for modelling or any other data mining technique uses. If all data has been taken in account which is important then there are various options available as which process to be taken so that the data can be used properly as training or testing. For e.g.-in regression, linear is used if the data is small and svm regression is taken if the datasets have lot of attributes. Also, data has to be taken together to see if any new things comes out as which can be of any use.

Data Cleaning

Sometimes the data entered into the database has lot of errors. Missing values is the one when any attribute doesn't have a value to its topic. Then there has to be some default value that needs to be given to it so that it doesn't cause any problems to the prediction process. There are outliers that needs to be not taken in account for as it causes a distraction to the data that are taken account for prediction. The dirty data are smoothened and inconsistency has to be resolved. Marks needs to be manually checked and rectified by looking into the database server and changing it to match the rest of the data making the whole input properly fit for use.

Data Preprocessing

Preprocessing of data is one of the most important tasks while building a model as the quality as well as reliability of the data is necessary to build an efficient model as data is core around which the entire system revolves in a data mining model and it directly affects the results obtained. The data needs to be cleaned and discretized before being fed into the model as ambiguous data leads to inaccurate results. The erroneous or ambiguous data is corrected or sometimes removed when they are very few numbers. In case of missing data, the data must be supplied either through normalization or standardization. After removing all the outliers, errors and noise from the data, it then undergoes transformation which is the process of converting or encoding the data into a more suitable format for processing.

Some of the preprocessing tasks applied to the data here included removal of records with missing values. The attributes considered are the marks scored in each of the 5 subjects and 2 lab sessions by a student in each of the 4 internal examinations and quizzes for every semester. The data is collected through a web interface where an initial validation is performed on the data and the users providing the data in order to check the integrity of the data. The values collected for each attribute is also subjected to a data validator to apprehend and prevent use of incomprehensible data values. The data stored in the database is used entirely as the training data for the model and each of the three algorithms are training on it.

B. *Data Modelling*

The modelling of the data is done using three different types of data mining algorithms which are namely: M5 rules, Linear Regression and Random Forest.

M5 Rules Algorithm

In this paper m5 is implemented which is a new system for predicting values by a learning model. It is a tree-based model. The m5 trees are just like piecewise linear functions. This is because in the way like regression tress has the values on their leave nodes, in M5 algorithm tress are build using multivariate linear models. One positive of using m5 is its' efficiency in learning and tackling tasks up to 100s of attributes. These model trees are usually smaller than regression trees and have been researched to be more accurate in tasks like prediction. The model trees are constructed on a collection of training cases(T) with each case is having a fixed set of attributes and a target value.

To construct a tree, the target values of training cases can be related to values of some other attributes. Accuracy of prediction can be judged by the capacity to predict the target values of the cases that are not seen.

Divide-and-conquer methodology is used for the construction of these trees. The set T is split recursively into subsets matching to the test outcomes. Sometime we get over elaborate structures because of the splitting for which pruning can be done. The next step is calculating the standard deviation of the target values of cases in T. T is then divided based on the results of a test. The result of the test gives expected reduction in error and then m5 selects in which the expected error reduction is highest. Using standard regression techniques for only the attributes that are referenced by tests, multivariate linear model is made for every node of tree. Starting from the bottom, each non-leaf node of the model tree is examined to select the last model for this node. The last model can be the linear model which is simplified or the model subtree depending on which has the least estimated error. If linear model is the one having least error then pruning is done.

Linear Regression

Linear regression has been used for discovering a linear link amongst the target and at least a single predictor. Simple Linear Regression is used for the purpose of this paper. Simple linear regression is useful when one has to make a connection between two consecutive variables. The first is independent variable and the second is dependent. It searches for a statistical association and not a deterministic association.

The core thought would be to acquire a line that would ideally fit the given data. The ideal fit line would be the one which completes the predicted error as little as could be allowed. The error is the separation among the points to the regression contour.

$$Y_{pred} = b_0 + b_1 * x$$

The qualities b_0 and b_1 must be picked with the goal that they should minimize the error in the prediction. If we take the summation of the squared error as a measurement to assess the model, then the

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

objective to get a contour which would decrease the error that is produced as much as possible.
In case error is not squared, at that point positive and negative point will counterbalance one another. For model with one predictor,

$$b_0 = \bar{y} - b_1\bar{x}$$

Intercept Calculation

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Random Forest
This is a classification algorithm based on the concept of supervised learning. It comprises of a number of trees. The more are the trees in a forest, more will be its' accuracy.
Pseudocode:

1. We have "p" features in total from which we choose any random "l" features [condition: l<<p]
2. Find a node "o" using best split' point among "l" features,
3. Split node into daughter nodes using the same point
4. Recursively so step 1-3 until "l" nodes have been reached.
5. Repeat 1-4 steps to construct random forest for "m" times to make "m" trees.

The first step of building a random forest is selecting "l" features from a total of "p" feature. Next use "l" features which are chosen randomly to discover root node applying the best split method. Then find daughter nodes using same method again. Reiterate all previous steps till the result is a tree having root node and the target value as leaf node. In the end, repeat all 4 steps to generate "m" trees. This eventually builds a random forest.

Prediction pseudocode:

1. Pass some test attributes through the rules of every tree which is randomly created so as to get a predicted result. Store this value as Target.
2. Voting is done taking all the predicted outcomes as candidates.
3. The target with the highest votes (majority voting) is considered to be the final prediction of the random forest.

Advantages

- The over fitting issue will never come when the random forest algorithm is utilized in any classification issue.
- Random forest algorithm can be utilized for both classifications as well as regression task.
- It can be utilized for feature designing. Which implies recognizing the most vital features out of the available features from the training dataset.

C. *Implementation*

The system is entirely web based and all the processing are done and executed on the server and the output is displayed in the portal. The entire portal is designed using bootstrap and JSP for frontend, Java for backend, MySQL for the database and Apache Tomcat for server. The portal is divided into three units on the basis of type of access provided, namely: admin, Faculty and Student. Any new student or faculty has to first register themselves into the portal before being provided access. The registration process is validated using several underlying algorithms and the new users have to provide detailed information about themselves and after all the details have been verified by the underlying system, the user is assigned an id and password to login in their respective portal. The three units are kept separate from each other with well suited authentication system which grants the user access to only the features available to the category the user falls into.

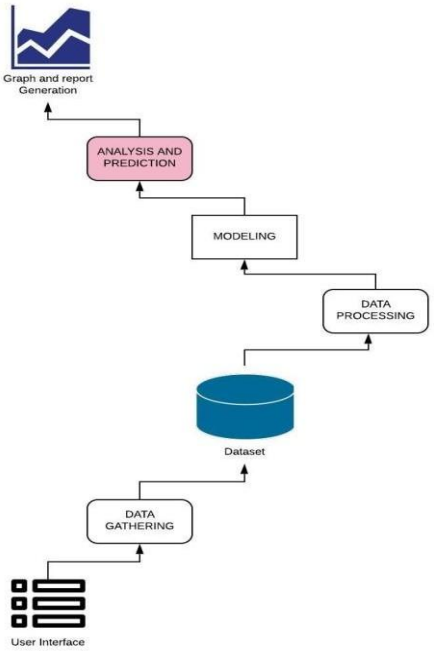

Fig 1.0 Proposed System

The admin holds the highest authority and has access to all the databases in the system. The admin's responsible for assigning semesters to each of the faculties as well as can view all of the faculty and student's data. The admin can also view the predictions generated and the graph reports of the students which can help the admin identify the potential classes with poor performance and guide the respective faculties to take appropriate steps. The different sections available in the admin portal include the home page which has a table for assigning semesters to any new faculty registered in the system, faculty details section which displays all the faculties registered in the system along with the semesters assigned to them, student details section which exhibits every students marks data to the admin and a report section to visualize every students reports and performance graphs.

Faculties are provided with a separate access to the portal and they are responsible for entering marks of the students assigned to them in the database. The system is designed to use live data for training the models and using this data to carry out the predictions. The faculties input marks scored by each of their students in each of the tests, assessments, quizzes and lab assignments. The

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

faculty portal has different sections: home for viewing the faculties personal information, Mark entry section to allow faculties to input marks of their students, Mark details section to view their student's marks and the predicted results as well as a report section to view the graphs and reports of each of their students. The feature to view each of their student's reports and predicted performances allows the faculties to identify potential weak students and take necessary measures to help them improve.

Student section of the portal provides students with access to view their own marks as well as predict their future potential scores in the final examinations and visualize their personal performance through a graphical representation. New students on registration have to select their semesters and are then assigned faculties by the admin. These faculties are responsible for grading the student and entering his scores into the database. The various sections available in the student portal include the home page with student's information and his/her scores in the respective subjects. The students are provided with an option to predict their future scores in the upcoming examinations based on their current performance and the predicted data is added into the table with student's previous scores. The predictions are carried out using data mining techniques which utilize the student's previous marks data as the training data. The students are also provided with a report section which helps them view their reports and performance graphs. Students do not have the authority to modify any data. Students can use the results of the analysis to identify their weaknesses and work on them.

The data collected from each of the users is stored securely in the database which is connected to the system via JDBC connection. The three factions of the system have different levels of access to the database. The admin can view all data whereas the faculty can only view specific data points and modify them. Students can only view their own data and have no controls over modifications. All the data is fed into the model which makes use of three data mining algorithms and after comparing these algorithms on various parameters, the best suited algorithm is proposed. The different parameters include:

Accuracy, it is defined as the percentage of instances correctly classified by a model out of a set of given instances. It is a great measure to compare algorithms as the greater the accuracy results in better predictions since the algorithm with higher accuracy classifies more instances correctly than the ones with lower accuracy and hence the former algorithm is more suitable and has a higher probability of generating correct results.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Where, TP = True Positive, TN = True Negative, FP=False Positive and FN = False Negative.

MSE, Mean Squared Error is used to measure the average square difference between the actual values and the predicted values. MSE is used to define the quality of a model and is a non-negative quantity. The lower the value of MSE, the higher reliability of the algorithm. The algorithm with the lowest MSE is classified as the best algorithm.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2$$

Where, n Total number of instances, X = Original Value and $\bar{X}$ = Predicted Value

RMSE, Root Mean Squared Error is the square root of the mean square error and also an important measure for comparing algorithms. Similarly, to MSE the RMSE is also a non-negative quantity whose value closes to zero with increasing accuracy. Hence the algorithm with least RMSE will be best suited algorithm.

$$\text{RMSE} = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2}$$

Where, n Total number of instances, X = Original Value and $\bar{X}$ Predicted Value

IV. RESULTS

The dataset used in this experiment was created by our own service. There is a front end in which the marks can be given. The same would be stored in our own database so that it can be used later. Not only marks but details of the faculty and students were also enclosed in that database through our user interface. To make it as a live project, we thought of using this idea in which as soon as the marks were given of the previous exams it would predict the results of the final examination.

The prediction was also done by more than one algorithm keeping the one as our proposed algorithm and the rest as for comparative analysis. The proposed one is M5 Rules which is considered as one of the best algorithms with a good efficiency and a real output which can be related to the values of the parameters given. The other two is -Linear Regression -It is the basic data mining technique which has been taken as it counts as a base for comparison to other algorithms and the other is – Random Forest Regression this is considered to be a complex algorithm. So, it is taken to know how it works with this live data set.

Here is a table in which marks has been shown of how the three algorithm gives result in their own domain.

Student Name	Course	Dept	Sem	Subject 1	Subject 2	Subject 3	Exam	Algorithm Used
Piyush Jaiswal	BE	CSE	1	85	95	86	Unit 1	Unit1
Piyush Jaiswal	BE	CSE	1	80	56	96	Unit 2	Unit2
Piyush Jaiswal	BE	CSE	1	89	71	90	Unit 2	Unit3
Piyush Jaiswal	BE	CSE	1	68	100	84	Model	Model
Piyush Jaiswal	BE	CSE	1	82	78	88	Final	M5Rules
Piyush Jaiswal	BE	CSE	1	83	77	87	Final	Linear Regression
Piyush Jaiswal	BE	CSE	1	81	80	85	Final	Random Forest

Fig 2.0 Marks predicted by the three algorithms of a Student

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

It is visible how the marks are been predicted by the three algorithms for each subject as a final exam. There are three rows given one to each algorithm at the end to show the predicted value. To give the marks to the user it will take the average of all the three marks given by each algorithm and the value will be the total marks taken in account of all the three algorithms. And for a good visual display, graphs have also been introduced to look at the variation in the marks and also between the predicted value by the three algorithms. The graphs have also been shown for the particular student's marks.

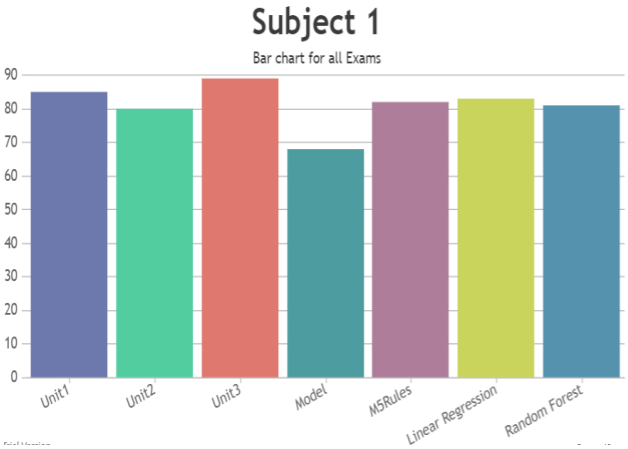

Fig 3.1 Graphical representation of marks of subject 1 of the Student

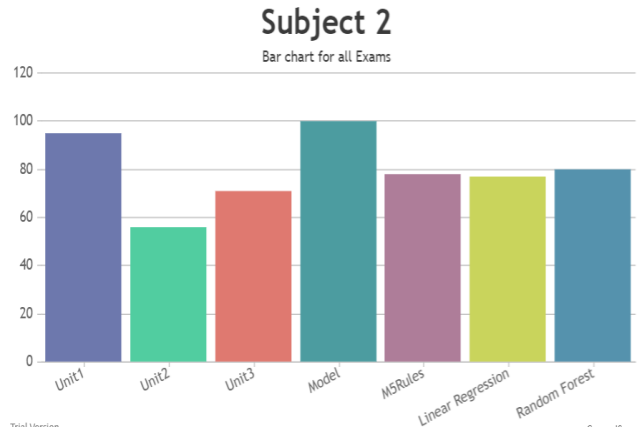

Fig 3.2 Graphical representation of marks of subject 1 of the Student

The second purpose of taking the three algorithms was to compare the success of these algorithms. So, taking the average values as real marks we calculate the accuracy of the results with the basic formula of it. Now to know a detailed performance of the algorithms we also find the RMSE and MSE values with the help of accuracy.

Algorithms	Comparative Methods		
	Accuracy (%)	MSE	RMSE
Linear Regression	84.0	0.1921	0.4382
Random Forest	87.0	0.1563	0.3953
M5 Rules	89.2	0.1296	0.3600

Table 1.0 Comparison of algorithms used

A particular algorithm is better if the accuracy is high as it tells that how correct the prediction is off that algorithm. But it is opposite in case of MSE and RMSE the lower the value of these two the better the performance is. Here it is found the accuracy of M5 is the best i.e. 89.2% and also the MSE and RMSE values are also least for this one, so it is the best algorithm for this system. After this the accuracy for Random Forest is 87 % so it is the second best. And it was known that linear would have been least as it is the basic algorithm.

V. CONCLUSION

Data mining techniques helps to get us results which may help to change things before it gets worst. Just in this case we predict the mark of the student so that it helps him to look into his studies and he gets to know where he should focus to get a desired and satisfied score. Even if the student ignores these things, we have the facility that will bring these matters to the faculties who are responsible for these students and they will help their students to get them out of the past record which are a big threat to the student education performance and help to improve the overall education system.

The work done by us helped us to generate a few inferences which can be useful in this educational field. There were different divisions made which were equally beneficial for its own purpose. A web portal was designed was made which helped to act as a user interface for the process. It also had been used as live status of the marks as we can input the marks of the student directly in the web portal which act as a parameter for modelling the data. There is a tab in which prediction is directly shown on the web portal such that it can be displayed directly to the users. We have also displayed graphs for better view of increased and decreased performance of the student in the particular subject.

The other important thing done was that we took more than one algorithm i.e. three to be precise and did the modelling by all the three algorithms. It helped in many ways as only one algorithm may predict the wrong output but by taking three algorithms, we get to know more accurate results for prediction. Also, we did comparative analysis of algorithms to know which one is best suited for the particular data. We used different formulas for getting to know how much accuracy are we getting for various algorithms such that it gives us an idea which alone can be used later on if the system does not need a heavy load and eliminate the algorithms which are less needed.

Retrieval Number: D9583118419/2019©BEIESP
DOI:10.35940/ijrte.D9583.118419

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VI. FUTURE WORK

As it is said that there is no end to what can be achieved. Thus, there are possibilities which can enhance the success of this system. So, as a thought given how this can be achieved, there were different ways we can do it

1. A better algorithm can be developed which would predict the results with a better accuracy.
2. There can be more attributes which can be taken as input parameter such that it will make the system more real scenario based.
3. To include text analysis as it would help to explore the areas which are not easily to interpret so that it would make the modelling more complex.
4. Time plays an important role. So, we need to work on things that would help to predict the results much quicker than we can now.
5. As now we have a web portal designed for user interface, we can add app development in the basket of user interface as nowadays there are apps for everything.

REFERENCES

1. Chew Li Sa, Dayang Hanani bt. Abang Ibrahim, Emmy Dahliana Hossain, Mohammad bin Hossin, "Student Performance Analysis System (SPAS)", The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M),
2. Ishwank Singh, A Sai Sabitha, Abhay Bansal, "STUDENT PERFORMANCE ANALYSIS USING CLUSTERING ALGORITHM", 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)
3. Bo Guo, Rui Zhang, Guang Xu, Chuangming Shi, Li Yang, "Predicting Students Performance in Educational Data Mining", 2015 International Symposium on Educational Technology (ISET),
4. Alan Cheah Kah Hoe, Mohd Sharifuddin Ahmad, Tan Chin Hooi, Mohana Shanmugam, "Analysing Students Records to Identify Patterns of Students' Performance", 2013 International Conference on Research and Innovation in Information Systems (ICRIIS),
5. M. Nasiri, B. Manasi, Fereydoon Vafaei, "Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining", 6th National and 3rd International Conference of E-Learning and E-Teachin.
6. Anjana Pradeep, Smija Das, Jubilant J Kizhekkethottam, "Students Dropout Factor Prediction Using EDM Techniques", 2015 International Conference on Soft-Computing and Networks Security (ICSNS)
7. Zhiwu Liu, Xiuzhi Zhang, "Prediction and Analysis for Students' Marks Based on Decision Tree Algorithm", 2010 Third International Conference on Intelligent Networks and Intelligent Systems
8. Chen Zhibo, Han Hui, Wang Jianxin. Data Warehouse and Data Mining. Qinghua University press. Beijing.2009.
9. Liao Kaiji, Liu Fengying, Hu Jianjun. Data Warehouse and Data Mining. Beijing University press. Beijing.2008.
10. ZhaoHui Tang, Jamie MacLennan. Data Mining with SQL Server 2005. Qinghua University press. Beijing.2007.
11. J.R. Quinlan, C4.5 Programs for Machine Learning [EB] 1993
12. Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", Second Edition, University of Illinois at Urbana-Champaign, 2006
13. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", Addison Wesley Publishers,
14. Boston, USA, 2006
15. Romero, C. and Ventura, S. (Eds.), "Data Mining in e-Learning", 2006, pp. 261-278
16. Hanna, M., "Data mining in the e-learning domain",
17. Computers & Education Journal, 42(3), 267–287, 2004
18. C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Syst. Appl., vol. 33, no. 1, pp. 135–146, 2007.
19. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegel Meyer, "Synthetic minority over-sampling technique,"
20. J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002.
21. J. R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufman, 1993.
22. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. New York, USA: Chapman & Hall, 1984.
23. Y. Freund and L. Mason, "The alternating decision tree algorithm," in Proc. 16th Int. Conf. Mach. Learn., 1999, pp. 124–133.

AUTHORS PROFILE

Kolli sai Poorna Chand

He is a student of Vellore institute of tech. have developed deep interest in area of data science from my schooling. This interest in data mining made me choose computer science and engineering in my graduation. Post completion of two years in graduation, I have explored lot of areas in computers science and engineering and I have completely engrossed myself in deep diving into data science. With the help of my professors from my college I am working towards publishing a paper in data science.

Prabakaran N

He is an Assistant Professor (Senior) at School of Computer science and Engg, VIT University, Vellore, India. He received his B.E and M.E degree in Computer science and Engg from anna university, Chennai 2009 and 2011 respectively. He received his PhD from VIT, Vellore in Computer science by 2017. His research activities are carried out in pervasive computing and context aware computing using sensors and networks

Ramani S

He is an Assistant Professor (Senior) at School of Computer science and Engg, VIT University, Vellore, India. His research activities are carried out in machine learning, data mining and prediction algorithms.

Damaraju Venkateswara Rao,

He is a student of Vellore institute of tech. have developed deep interest in area of data science from my schooling. This interest in data mining made me choose computer science and engineering in my graduation. Post completion of two years in graduation, I have explored lot of areas in computers science and engineering and I have completely engrossed myself in deep diving into data science. With the help of my professors from my college I am working towards publishing a paper in data science.

Sriram Vemparala,

He is a student of Vellore institute of tech. have developed deep interest in area of data science from my schooling. This interest in data mining made me choose computer science and engineering in my graduation. Post completion of two years in graduation, I have explored lot of areas in computers science and engineering and I have completely engrossed myself in deep diving into data science. With the help of my professors from my college I am working towards publishing a paper in data science.

Retrieval Number: D9583118419//2019©BEIESP

DOI: 10.35940/ijrte.D9583.118419

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication