TEST Engineering & Management

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

Student Academic Performance Prediction Using Machine Learning

Likhitha V[1], M P Vaishnavi[2], Kamachi Gayathri[3], Shreyas V[4], Gopal Krishna Shyam[5]

[1]Student, Department of Computer Science and Information Technology, REVA University, likhithav11@gmail.com
[2]Student, Department of Computer Science and Information Technology, REVA University, vaishnavi824@gmail.com
[3]Student, Department of Computer Science and information Technology, REVA University, gayathrirathinam1997@gmail.com
[4]Student, Department of Compute Science and Information Technology, REVA University, vshreyas88@gmail.com
[5]Professor, Department of Computer Science and Information Technology, REVA University, gopalkrishnashyam@reva.edu.in

Article Info
Volume 83
Page Number: 4825-4833
Publication Issue:
May-June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Abstract

Every educational institution around the world maintain student result database which contains information about student marks, grade in different subjects in different semesters. This project is used for evaluating students' performance and predicting their scores. The programming language that will support, interface and help in building our project is Python. Data mining and machine learning are growing fastin areas of computer science with far-reaching applications. Machine learning in general means performing a specific task without using explicit instructions by humans, relying on patterns and inference to detect meaningful patterns and predicting.

Keywords: *Dataset,Data Mining, Educational Data Mining (EDM),Machine Learning, Random Forest Regressor, Support Vector Regressor, Decision Tree Regressor, Linear Regression.*

1. Introduction

Importance of Predicting Students Result

The objective of education institutions is to provide quality education to their students. A way to achieve this is by discovering knowledge by prediction depending on the count of students in each semester, isolation of traditional classroom teaching model, exposure of unfair means used in online examination, students' performance, detection of strange values in the result sheets of the students and so on. Neither faculty nor the parent cannot figure out students' abilities and their interests easily, so that they can enhance them in it. As a result, it may affect with poor university results, placements and career of an individual. At present, Educational data mining is gaining more attention because of escalation in the educational data of e-learning systems, and even progressing traditional education. Alarmed with evolving techniques for discovering the unique types of data present in academic environments, it seeks to extract meaningful information in order to advance learning processes from vast amounts of raw data. So, many researchers started exploring various data mining techniques in order to support educational institutions to evaluate and progress their respective organization. By machine learning techniques along with EDM, continuous evaluation systems are built and is practiced by numerous institutions today.

Even though there are many existing models, they still do not provide high accuracy which is a very important factor in prediction of scores and grades. So, using machine learning and EDM, we decided to build a system that would implement four different algorithms and use the best algorithm with highest accuracy and lowest error rates to go on with the prediction. The main objective of our project work is

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

to predict engineering student's current semester results base on the previous semester results with a high accuracy rate. This would help the student, teacher, parent and the educational institution to know how a student is going to perform in the upcoming exams and help them in improving their predicted score by putting in the required hard work to obtain it. Our prime motto is to aware students about their academic status and guide them for better performance. This can also decrease the ratio of self-harm in students due to low marks and study pressure up-to some extent which can be seen as a social help.

Issues In Predicting

In the existing work, analysis is statistical, it does not provide hidden information about students and does not provide much insight into a student's performance. Functionalities can be added such as gender classification and subject classifications for better predictions. Most of the present work, also does not provide accuracy, which is an important aspect of prediction.

Resolving The Issues

This paper investigates the accuracy and error values of 4 different algorithms namely, Linear Regression, Decision Tree Regressor, Random Forest Regressor and Support Vector Regressor. The algorithm with highest accuracy and minimum errors is chosen at the end.

2. Literature Survey

During our literature review study, to define the project problem as well as objectives, we went through a lot of research papers and web application statistics. Few of the paper's summarized details are as follows:

In the first research paper [1], the authors applied Standard and Hybrid approaches as well as fuzzy logic and compared outcomes. The first phase included preprocessed data which required data cleaning, transformation and data Filtration.J48 tree, LibSVM, FL and NaiveBayes Standard Classification algorithms as well as Hybrid approach LMT were chosen for better classification. Training set focused on Result attribute(nominal attribute) were chosen.J48 tree was visualized to see the rules. The rules were tested separately on another file, named test file. All the attributes in the file are same as that of training set but Result attribute was kept empty here, which they had to be predicted. After inserting test file classification was performed. Further, visualized classifier errors in the model were noted. Now either Java code was executed to obtain the predicted results or extract the predicted data by importing it in Microsoft Excel.Fuzzy Logic was used to predict students' performance in terms of result and expected Dropout rate. A fuzzy system is an information-based rule system. The core of this system was a database which was configured with if-then rules. Fuzzy inference was the procedure of expressing input/output mapping by applying the fuzzy logic. It tried to conclude answers from a knowledge base by utilizing a fuzzy inference engine. The inference engine which was examined to be the brain of the expert system gave the methodologies to reasoning around the data in the knowledge base and explains the results. The comparative results of standard and hybrid machine learning approaches concluded that the prediction results provided by Hybrid approach are much significant as compared to the standard approaches of machine leaning in terms of accuracy, correct and incorrect classification as well as other mentioned factors.

LMT exhibits higher accuracy when compared to accuracy of LibSVM, FL J48 and Naive Bayes approaches. LMT, LibSVMand FL provided significantly better accuracy as compare to J48 and Naive Bayes approaches. In terms of precision, recall, F-measure and Kappa stats also LMT outperforms as compared to all other standard algorithmic approaches.

The second paper[2] that we went through, comprises of three folds of methodology which included pre-processing phase, construction phase, and the evaluation phase. The first phase included aggregating the records using various techniques such as questionnaires, student interviewing and online student evaluation. Data was pre-processed by cleaning the dataset, removing the outliers, missing and inconsistent values. The second phase, i.e., construction phase comprised ML techniques such as:

1. Functions: This classifier was used to construct a function model. The model consisted of input features and output labels. The inputs were mapped to the outputs using neural networks which in turn used forward and back propagation for network updates which also included SVM, logistic regression, etc.
2. Meta: The prediction model's performance was increased using few ensemble techniques. This ensemble class was developed by taking a set of feeble learning classifiers from the training dataset.
3. Bayes: This is a rule-based or network model considering the probability of particular attributes. This model uses the previous probability, the odds of observing a variety of given data with the hypothesis and observed data.
4. Lazy: It is used to find the output label from training dataset by comparing the most alike labels. This family includes techniques namely K-NN, K-Star, IBk and LWL.
5. Trees: There are multiple methods under this classifier family: Decision trees, Random forest and Logistic Model Tree (algorithm that combines logistic regression and decision tree learning). It builds tree like structure where nodes are labels and edges are values of an attribute. The third phase i.e., evaluation phase they used various evaluation methods to evaluate EMT technique which provided results in

Published by: The Mattingley Publishing Co., Inc.

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

terms of accuracy, F-measure and ROC curve metrics. In the construction phase, an ensemble boosting classifier syndicated the weak learners with a coefficient weight for each learner into a strong prediction model with high accuracy performance. AdaboostM1 method was used to train more than one learner to solve the problem of student prediction performance. The best prediction model from these learners was selected as one of the EMT methods. On the other hand, the bagging methods evaluated the output of the weak learners using two techniques. EMT technique combined the best classifier of the whole families on the same training set with the best learners of AdaboostM1 method as a bagging technique, using the voting method between the two. In the last phase, evaluating phase, in order to evaluate the EMT technique to generalize it for future instances, the 10-fold cross-validation technique was used. Every 10 parts held the same number of instances divided in a random way. At each iteration, one portion is used as a validation set, while the other parts were combined as a training set. As a result, for building the classifier the training set was used, whereas the validation set was used to evaluate the classifier by predicting the class of its cases. A set of evaluation metrics were used to evaluate the EMT technique, F-measure, and ROC metrics.

The third paper [3] consisted of a number of Machine Learning methods such as Support Vector Machines (SVM), Naive Bayes, and K-Nearest Neighbor, and an array of natural language pre-processing techniques were applied to determine the optimum classification methods, and to increase the accuracy of the results. Initially, work was divided into two main parts- one is Topic Modelling-LDA and another one is prediction by classification. Data was selected and then pre-processing which includes data selecting, data cleaning and data transformation was done. The first step of evaluation of this project was Data Pre-processing. The authors have used these two categories of filters - Supervised and Unsupervised. Topic Modelling in NLP was used as it seeks to find hidden semantic structure in documents. They used classification algorithms to obtain mail goal of pass/fail which can be used for classification itself. Binary classification was the model being trained for predicting of two target classes, in this paper's view, based on student earlier result, it would predict whether the student will pass or fail. To predict students' result these algorithms were used: K-Nearest Neighbors Algorithm (k-NN), Support Vector Machines (SVM), and Naive Bayes. They said that the best accuracy for students' result prediction from these classifiers was k-NN algorithm. The correctly and incorrectly classified examples showed the percentage of test examples that were correctly and incorrectly classified. The Accuracy, Precision, Recall and F1 score were also calculated.

In the fourth paper [4], the authors started off with analyzing the data logged by a technology-enhanced learning (TEL) systems called digital electronics education and design suite (DEEDS) using ML algorithms. This technique included ML algorithms like ANNs, SVMs, Naïve Bayes classifiers and decision trees. The machine was trained with the help of ML algorithms on the data from the previous sessions and the upcoming sessions tested the algorithm on the data. By using ANNs and SVMs helped them attain higher accuracy. SVM's and ANN's can effortlessly be integrated into the technology-enhanced learning systems.
The main aim of their project were as follows:
1. To identify the most appropriate machine learning algorithms for predicting the difficulty an individual student would have in the next session of a digital design course based on prior session activities and the current session.
2. To investigate which machine learning algorithms used in the current study are appropriate for predicting student difficulty in the next session of digital design course while using the fewest features.
Their project work showed that the training and test data from the course suffered some difficulty and such methods actually do not help the teacher correctly evaluate the model accuracy in the succeeding session because student course outlines and activities change from one class to another.

In the fifth paper [5], authors used the WEKA (Waikato environment for knowledge Analysis). Data mining is done using the tool WEKA. The process of generating new information on the basis of past or previous information is called as Data mining. WEKA is a tool based on JAVA language. Classification, Clustering and many other data mining algorithms are the inbuilt tools of JAVA which is a platform independent tool. In the WEKA tool, they gave input of the past semester results on basis of which, a model was created which formed a pattern which predicted the results of the student. The behavior of the student and their marks were used as an attribute or property. The past result set was constructed based on any attribute or property of students, either their marks or behavior in class. Later, the result was predicted at any stage of course test set based on same attribute was made and tested on the model created using WEKA.

3. Implementation Method

Steps Involved

At first the data set was prepared in the form of a CSV file by collecting the data from students through a google form which was circulated through our college. The dataset included the attributes: Name, SRN(Reg. No.), semester I grade point, semester II grade point, semester III grade point, semester IV grade point, semester V grade point and cumulative grade point. Then, a machine learning model was

Published by: The Mattingley Publishing Co., Inc.

built. We analyzed the data by various factors, such as, pass or fail, count of the same score in different semesters, grade obtained, number of student that have passed or failed with respect to gender, etc by using matplotlib and seaborn libraries. We tried to implement different algorithms, such as, Decision Tree Regressor, Random Forest Regressor, Linear Regression and Support Vector Regressor. The data set that we received was fed into all the four algorithms seperately.

The first 70% of the data was used for training our model and the remaining 30% of the data was used for testing with random state. We trained the machine model based on the data set. The train model has two inputs: ML algorithms and 80% of the data split provided by the user.We tested the model using these two inputs of data: train model and 30% of the data split provided by the user.This was done on all the four algorithms. On considering the accuracy [*Table I*] and error rates [*Table II*], the best algorithm among these, Random Forest Regressor was chosen as our main algorithm. Then, the input is taken from the user for the first 4 semesters to predict their semester V grade point. The predicted score is produced with the distinction achieved by the student. The CGPA including the predicted semester grade is also shown.

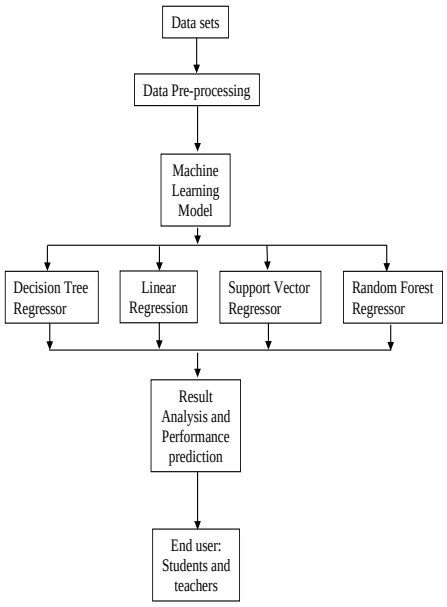

Figure 1: Flow of data while building the system.

Technologies and Libraries Used

- Jupyter Notebook,
- Python 3,
- Numpy,
- Pandas,
- Seaborn,
- Matplotlib,
- Sci-kit Learn.

Data Analysis

Analyzing of data is very important as it involves the process of cleaning data, transforming it and then modeling the data to identify hidden information which help is making decisions required. Comparison on various data helped in discovering the various factors affecting the students.

From the dataset we used, the first few data items were plotted into a line graph[*Fig. 2*] and it shows variations in grade points among different semesters and their cumulative grade of different students.

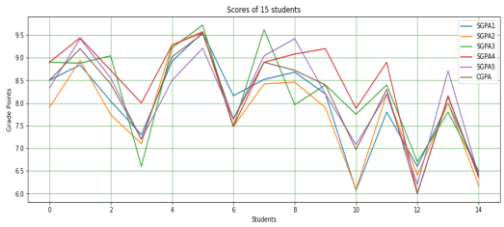

Figure 2: Data of first 15 students

Comparison of the count of each grade point obtained by students in semester I was analyzed as a bar graph. *Fig. 3* represents the same. Similarly, semester II, semester III, semester IV, semester V and the cumulative grade points were also analyzed.

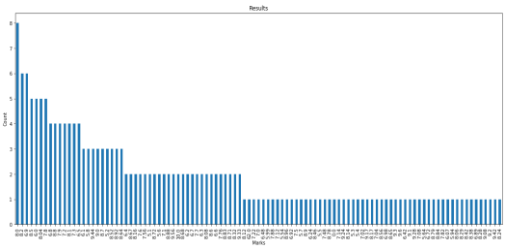

Figure 3: Count of each grade point in Semester I

Data analyzing on basis of gender also plays a huge role. Count of the Male and Female students was made[*Fig. 4*]. Also keeping a track of how many students are either passing or failing [*Fig. 5*] helps in education institutes to know what set of students to concentrate on. Analyzing the data of pass or fail is also done based on gender to know which gender is performing better [*Fig. 6*].

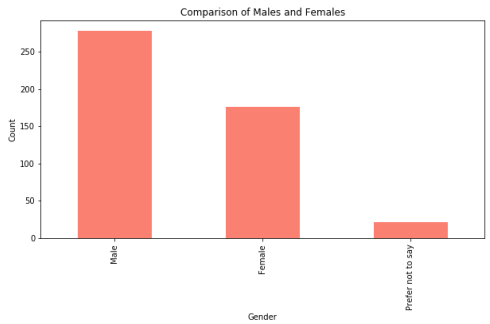

Figure 4: Comparison Males and Females

Published by: The Mattingley Publishing Co., Inc.

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

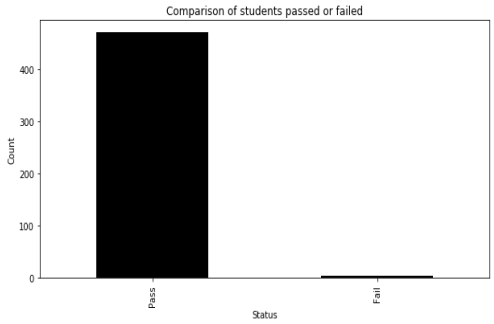

Figure 5: Comparison of students passed or failed

[On basis of gender]

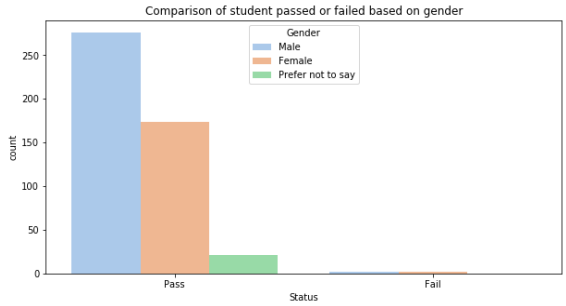

Figure 6: Comparison of students passed or failed with respect to gender

Considering the cumulative grade point (CGPA) of the students, a pie chart [*Fig. 7*] was made to analyze the grade points obtained by converting them into a set of grades, namely O, A+, A, B+, B, C+, C and F. Each grade has a range of grade point. A bar graph [*Fig. 8*] was to have a count of each grade being obtained by students. This analysis was also done based on gender [*Fig. 9*].

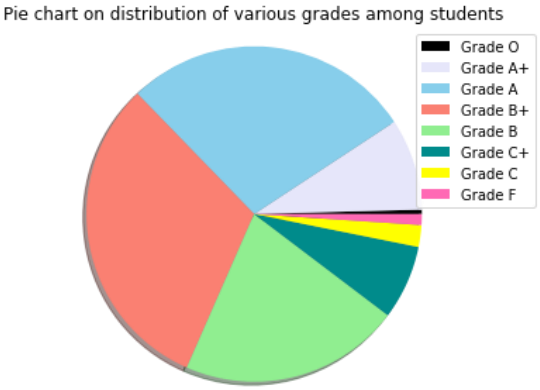

Figure 7: Pie-chart on grades obtained

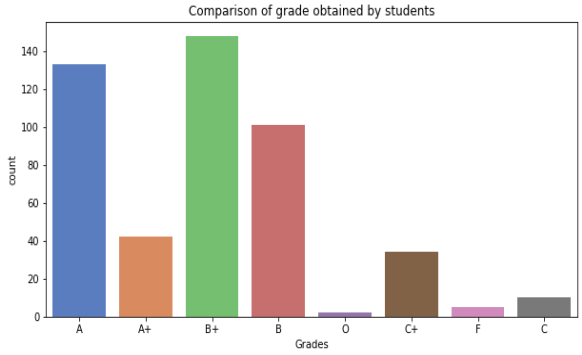

Figure 8: Comparison of grades

[On basis of gender]

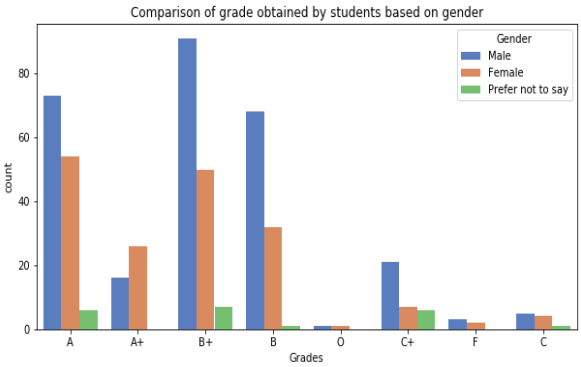

Figure 9: Comparison of grades based on gender

Coding and Execution

At first the dataset of 5 semesters grade point is taken and is split using the train_test_split, which is imported from sklearn. It splits the data into random train and test subsets, in our case random 70% as train subset and 30% as test subset. These subsets are then scaled using a min-max scaler. Our goal is to predict the semester V grade point while taking input of the first 4 semesters grade points. So, our 4 algorithms, namely, Linear Regression, Decision Tree Regressor, Support Vector Regressor models were created and the train and test data were fed through fit() for each model separately. We also took a few particular data entries from our data sets to check if the predicted score was close enough or not. Let us take one student, say ABC, who achieved 8.84, 8.94, 8.88, 9.44 and 9.42 in semesters I, II, III, IV and V respectively, to show to accuracy of each of these algorithms.

First, we implemented theLinear regression algorithm. The subsets were fed into this model and the train and test accuracy was calculated [Table II]. On plotting a scatter plot it was observed that the accuracy was very low [*Fig. 10*]. The mean absolute error, mean squared error and RMSE were calculated [Table II].

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

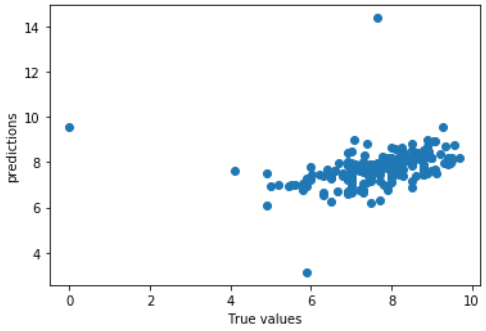

Figure 10: Scatter plot of Linear Regression

On trying to predict the score of the student ABC with this algorithm we got the following result [*Fig. 11*]. As you can see the predicted score is no way close to the actual score obtained by the student (i.e., 9.42).

```
[27]: #Linear Regression
      predict1 = model1.predict([[8.34,8.94,8.88, 9.44]])
      predict1
Out[27]: array([7.96113028])
```

Figure 11: Prediction of grade point, Linear Regression

Next, our subsets were fed into another algorithm, Decision Tree Regressor. It was observed that the train and test accuracy is high when calculated [Table I]. But, the scatter plot diagram did show accuracy but it was not good enough [*Fig. 12*].The mean absolute error, mean squared error and RMSE were also calculated [Table II].
Also, the RMSE of Train and Test errors graph [*Fig. 13*] show that after a point the values are giving a few constant values as output.

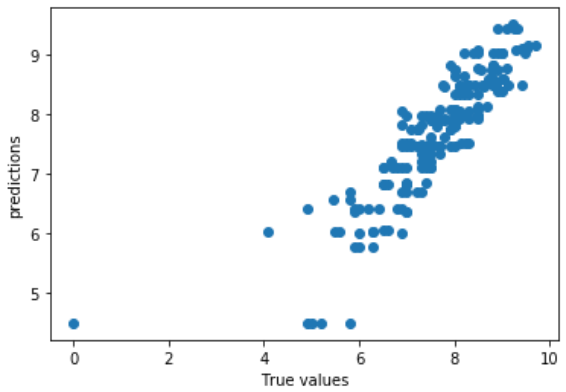

Figure 12: Scatter plot of Decision Tree Regressor

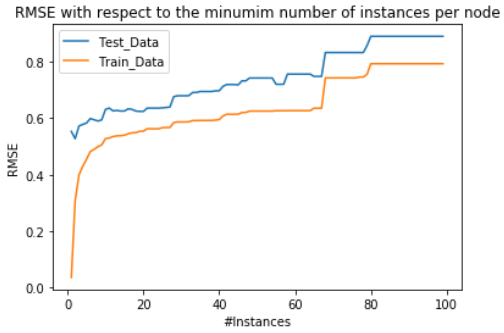

Figure 13: RMSE of Decision Tree Regressor

On trying to predict the score of the student ABC with this algorithm we got the following result [*Fig. 14*]. The predicted score is close to the actual score obtained by the student (i.e., 9.42).

```
[31]: #Decision Tree Regressor
      predict2 = model2.predict([[8.34,8.94,8.88, 9.44]])
      predict2
Out[31]: array([9.09625])
```

Figure 14: Prediction of grade point, Decision Tree Regressor

Next, the Support Vector Regressor was implemented. We fed our train and test subsets into this model and the accuracy [Table I] and errors were calculated [Table II]. The train and test accuracy value was not that good compared to Decision Tree Regressor. It was also observed that error rate is a little high than expected, but a little bit of the error rate is negligible in the case of SVM and SVR. A scatter plot was plotted [*Fig. 15*], and it showed more accuracy compared to Decision Tree Regressor.

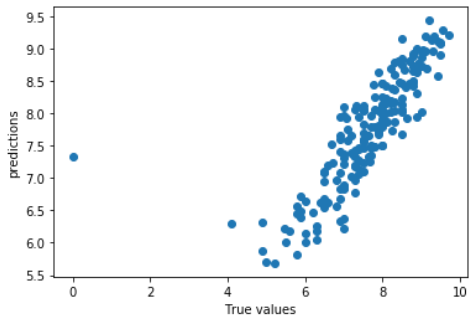

Figure 15: Scatter plot of Support Vector Regressor

On predicting the grade point of student ABC for semester V, it was observed that the predicted score [*Fig. 16*] is very similar to Decision Tree Regressor.

TEST Engineering & Management

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

```
[36]: #Support Vector Regressor
      predict3 = model3.predict([[8.34,8.94,8.88, 9.44]])
      predict3

Out[36]: array([9.09332525])
```

Figure 16: Prediction of grade point, Support Vector Regressor

Next, we implemented our last algorithm model, Random Forest Regressor. The subsets were given to the model. The calculated train and test accuracy was high and similar to Decision Tree Regressor [Table I]. The scatter plot also showed good accuracy [*Fig. 17*]. The mean absolute error, mean squared error and RMSE were calculated [Table II].

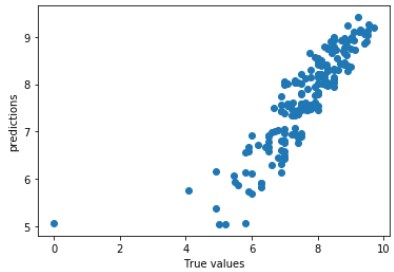

Figure 17: Scatter plot of Random Forest Regressor

The RMSE of train and test data errors showed minimum error rate [*Fig. 18*]. It is observed that, unlike decision trees, there are no constant values which shows that the error rate is pretty low.

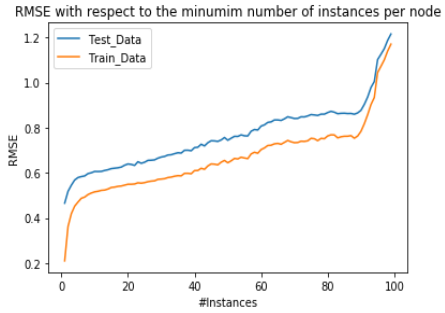

Figure 18: RMSE of Random Forest Regressor

The predicted score for the student ABC, using Random Forest Regressor, was closer the actual value (9.42) compared to Decision Tree and Support Vector Regressors [*Fig. 19*].

```
[40]: #Random Forest Regressor
      predict4 = model4.predict([[8.34,8.94,8.88, 9.44]])
      predict4

Out[40]: array([9.16334389])
```

Figure 19: Prediction of grade point, Random Forest Regressor

Table 1: Comparison of Algorithms (Based On Accuracy)

Algorithms	**Accuracy**	
	Training	***Testing***
Linear Regression	32.6%	7.1%
Decision Tree Regressor	84.89%	76.72%
Support Vector Regressor	78.01%	64.9%
Random Forest Regressor	84.17%	77.13%

Table 2: Comparison Of Algorithms (Based On Errors)

Algorithms	**Errors**		
	Mean absolute Error	***Mean Square Error***	***Root Mean Square Error***
Linear Regression	7.88	15.6	12.5
Decision Tree Regressor	4.06	3.39	5.82
Support Vector Regressor	3.91	5.11	7.15
Random Forest Regressor	3.74	3.38	5.77

On a whole, the following observations were made:

1. Support Vector Regressor and Decision Tree Regressor have low error rates but not as low as Random Forest Regressor.
2. Support Vector Regressor and Decision Tree Regressor predict a similar grade point, even though their test accuracies are not similar. Random Forest Regressor predicts a closer value of the grade point and also has higher test accuracy.
3. Decision Tree Regressor's scatter plot does not show good accuracy when compared to Random Forest Regressor scatter plot. But, Support Vector Regressor scatter plot gives a good level of accuracy.
4. The RMSE graph of Decision Tree Regressor showed constant values, while Random Forest Regressor does not show any constant values.
   So, it was decided to use Random Forest Regressor model as our main model for predictions as it shows better results in every aspect.

4. Result and Conclusion

The success of machine learning in predicting student performance depends on the good use of the data and machine learning algorithms. Picking the correct machine learning method for a problem is essential to achieve the preeminent results. Thus, we concluded on using Random Forest Regressor as it has more accuracy rate, low error rate and good prediction compared to other algorithms used. User can give his input and get his resulting grade point predicted with

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

other aspects [*Fig. 20*]. The student on seeing his predicted score can know how much effort he/she will have to put in to get a better result. Teachers/ Professors can also analyze which student would require more attention in terms of academics.

```
a = input("Enter 1st semester marks: ")
b = input("Enter 2nd semester marks: ")
c = input("Enter 3rd semester marks: ")
d = input("Enter 4th semester marks: ")
e = model4.predict([[float(a),float(b),float(c),float(d)]])
print("Your predicted score is: ", e)
if(e<5):
    print("Fail")
elif(e<7):
    print("Second Class")
elif(e<9):
    print("First Class")
else:
    print("First Class with Distinction")

CGPA = (float(a)+float(b)+float(c)+float(d)+float(e))/5
print("Your CGPA with the predicted score is: ", CGPA)
```

```
Enter 1st semester marks: 8.76
Enter 2nd semester marks: 8.46
Enter 3rd semester marks: 7.89
Enter 4th semester marks: 9.04
Your predicted score is:  [8.79057265]
First Class
Your CGPA with the predicted score is:  8.58811452937422
```

Figure 20: Input taken from user to predict grade point.

Acknowledgment

We wish to express our sincere gratitude to Dr.Shyama Raju, Chancellor and Dr. S Sunil Manvi, H.O.D, Computer and Information Technology Department of REVA University for providing us an opportunity to do our project work on "Predicting Student Academic Performance Using Machine Learning Techniques". This project bears on imprint of many. We sincerely thank our project guide Dr. Gopal Krishna Shyam for his guidance and encouragement in carrying out this synopsis work. Finally, we would like to thank our family and pupils who helped us in completing the Project work successfully.

References

[1] Amandeep Kaur, Nithin Umesh, Barjinder Singh, "Machine Learning Approach to Predict Student Academic Performance", International journal for research in applied science and engineering technology (IJRASET), IC Value: 45.98; SJ Impact Factor: 6.887, Volume No: 6 Issue No: IV, April 2018.

[2] Ammar Almasri, Erbug Celebi, Rami S. Alkhawaldeh, "EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance", Department of Management Information Systems, Institute of Graduate Studies and Research, Cyprus International University, Lefkosa, North Cyprus 99258, Turkey and Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan, 24th of Febraury 2019.

[3] Suja Babu, Aswathy R Kaimal, "Predicting Students Result based on Topic Modelling - LDA", International Research Journal of Engineering and Technology (IRJET), Impact Factor value: 7.211, ISO 9001:2008 Certified Journal, Volume: 06, Issue: 05th of May 2019.

[4] Mushtaq Hussain, Wenhao Zhu, Wu Zhang, Syed Muhammad Raza Abidi, Sadaqat Ali, "Using machine learning to predict student difficulties from learning session data", Article in Artificial Intelligence Review · February 2018, Springer Science+Business Media B.V., part of Springer Nature 2018.

[5] Priyanka Dhamija, Rainu Nandal, Harkesh Sehrawat, "Predicting Student Result On The Basis Of Past Result", International Journal of Engineering and Technology (IJET), ISSN (Print): 2319-8613, ISSN(Online): 0975-4024, 30th of April 2017.

[6] Jabeen Sultana, M. Usha Rani, M.A.H. Farquad, "Student's Performance Prediction Using Deep Learning and Data Mining Methods", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[7] Hany M. Harb and Malaka A. Moustafa, "Selecting Optimal Subset of Features of Student Performance Model", IJCSI International Journal of Computer Science Issue, Vol. 9, Issue 5, No, September 2012.

[8] Carlos Marquez, Cristobal Romero Morales and Sebastian Ventura Soto "Predicting School Failure and Dropout by Using Data Mining Techniques" IEEE Journal Of Latin-American Learning Technologies, Vol. 8, No. 1, February, 2013, pp. 7-14.

[9] Grigorios F. Tzortzis and Aristidis C. Likas, Senior Member, IEEE "The Global Kernel K-Means Algorithm for Clustering in Feature Space" IEEE transactions on neural networks, VOL. 20, NO. 7, July 2009.

[10] Saadat Naziova "Survey on Spam Filtering Techniques", Communication and Network, August 2011, pp. 153-160.

[11] P. Moniza and P. Asha "An Assortment of Spam Detection System", International Conference on Computing, Electronics and Electrical Technologies [ICCEET] 2012, pp.77-83.

[12] Patricia Bellin Ribeiro, Luis Alexandre da Silva and Kelton Augusto Pontara da Costa "Spam Intrusion Detection in Computer Networks Using Intelligent Techniques", IFIP IEEE IM Workshop: 1stInternational Workshop on security for Emerging Distributed Network Technologies (DISSECT), 2015, pp. 304-311.

May-June 2020
ISSN: 0193-4120Page No. 4825-4833

[13] Yen-Liang Chen, Hsiao-Wei Hu and Kwei Tang, "A Novel Decision-Tree Method for Structured Continuous-Label Classification" IEEE Transactions on Cybernetics, 2013, pp. 1734 – 1746.

[14] Qiang Yang, Senior Member, IEEE, Jie Yin, Charles Ling, and Rong Pan, "Extracting Actionable Knowledge from Decision Trees" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 1, JANUARY 2007.

[15] Jasna Soldic-Aleksic , Journal of Economics and Engineering, ISSN.: 2078-0346, Vol. 3. No.1, April 2012, pp. 241-248.

[16] Olaiya Folorunsho, "Comparative study of different data mining techniques performance in knowledge discovery from medical database", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.

[17] Nancy Lekhi and Manish Mahajan, "Outlier Reduction using Hybrid Approach in Data Mining" I.J. Modern Education and Computer Science, 2015, pp. 43-49.

[18] John Jacob, Kavya Jha, Paarth Kotak and Shubha Puthran 'Educational Data Mining Techniques and their Applications' International conference on GreenComputing and Internet of Things, 2015.

[19] Motaz M. H. Khorshid , Tarek H. M. Abou-El-Enien ,Ghada M. A. Soliman, "Hybrid Classification algorithms for terrorism prediction in Middle East and North Africa" International Journal of Emerging Trends & Technology in Computer Science, 2015.

[20] Khan A. R., Amin, H. U., Rehman, Z. U., "Application of Expert System with Fuzzy Logic in Teacher's Performance Evaluation," International Journal of Advanced Computer Science Applications (IJACSA), Vol. 2. No. 2, Feb 2011.