



An approach to Predict Student's marks using Personalized analytics in Covid-19 situation

Ayush Ranjan Srivastava, Prachi Sharma, Ashok Kumar Rai

Vandana Tiwari, Varun Kumar Singh

1. Computer Science and Engineering, Buddha Institute of Technology, India

ABSTRACT

This paper focuses on predicting student performance using personalized analytics based on educational databases. Predicting student performance becomes more challenging due to the huge amount of databases. We present two different approaches to improving student achievements. Both approaches are validated on one course which was offered to students of the school of computing between the years of 2019 and 2020. The first approach is based on regression algorithms to predict student performance. Regression is a data mining function that predicts a number. The main goal is to find how well a student can perform in the programming language by predicting grades based on their school background and performance in semester exams. In the model build (training) process, a regression algorithm estimates the value of the dependent variable as a function of the predictors in the build data based on the independent variables. These relationships between predictors and targets are summarized in a model, which can then be applied to a different dataset in which the target values are unknown. The second approach is to find the error rate of regression algorithms by using root mean square error. The obtained results reveal that the school background also plays a major role in predicting grades. Finally, we can identify the students who are at risk and provide better additional training for the weak students.

Measuring the academic performance of students is more challenging since the International Journal of Pure and Applied Mathematics students " s academic performance depends on diverse factors like personal, socio-economic, psychological, and other environmental variables. The scope of this paper is to predict the student grades using the best algorithms with high accuracy. There is a critical need to develop innovative approaches that ensure students become graduates in a timely fashion and are well trained and workforce ready in their field of study. From the paper, Graduation is important in a student" s life because it is one of the main factors that are going to decide the future of the student based on their mark in the higher secondary examination, they are going to get a college education and the field of information technology based on their knowledge in programming. We present methods that draw on techniques from recommender systems to accurately predict students" programming course grades. Data mining provides many tasks that are used to predict the student's performance. In this paper, the regression task is used to evaluate the performance of a student and as many approaches are used for data regression, the multi-linear regression and SVM, random forest, and regression tree methods were used here. For this study, recent real-world data has been collected. Information like the medium of study, syllabus, intermediate background, etc., were collected. mathematics marks are also collected to know the logical ability of the student and English marks to know the communication and understanding level of the student This study is more useful for identifying weak students in the programming at the beginning of the semester and the identified students can be assisted by the educators so that their performance is better in future. This study investigates the accuracy of some regression techniques for predicting the performance of a student.

Keywords: Marks Prediction, algorithm, linear regression, Machine Learning



1. INTRODUCTION

Over the past 35 years, a vast amount of knowledge has been accumulated on text mining for Information Retrieval (IR). Using automated text mining algorithms to discover knowledge from natural language texts provides numerous challenges but also offers unique possibilities. One of the most natural forms of storing information is in the form of natural language texts. This can be easily interpreted by a human but it is still a great challenge for computers to derive meaning from this data. However, computers do offer an important advantage over human capabilities: computing power. This means that computers can find patterns, which are non-trivial recurrences, within data faster and more accurately than their human counterpart, but this can only be done if the structure of the data is known. Natural language does contain implicit grammatical structures, but these structures are deeply complex and vary across different languages. The main aim of this project is to use data mining methodologies to study students' performance in the courses. Data mining provides many tasks that could be used to study student performance. In this research, the classification task is used to evaluate students' performance and as many approaches are used for data classification, the decision tree method is used here. Information like Attendance, Class test, Seminar, and Assignment marks was collected from the student's management system, to predict the performance at the end of the semester. This paper investigates the accuracy of data mining classification methods for predicting student performance.

Education is a very important issue regarding the development of a country. The main objective of educational institutions is to provide high-quality education to their students. One way to accomplish this is by predicting student's academic performance and thereby taking early steps to improve student's performance and teaching quality. This system aims to predict student's marks using linear regression. The idea behind this analysis is to predict the marks of students by their studying hours. Through this project we can determine: How many hours need to do the study to get 99% marks If I will do study $x()$ hours per day so how much marks I will get Through these points the school can determine the performance of the student. To implement this model we are using Jupyter Notebook which is an open web source application. The model is deployed on the python framework called Flask. The data set is taken from Kaggle, which provides data for free.

The Main Objectives of this Study:

- Predict the student's success or failure
- Predict the final grade

The problem of the student final grade prediction in a particular course has recently been addressed using data mining techniques. Researchers usually examine study-related records, e.g. the age, gender, and the field of study because of their easy availability in university information systems. The most typical way how to obtain such data is to conduct questionnaires but it tends to have a lower response rate. Therefore, only the data originated from the college are considered for our experiments. As stated previously, student's low academic performance in the engineering dynamics course has been a long-standing problem. Before designing and implementing any pedagogical and instructional interventions to improve student learning in engineering dynamics, it is important to develop an effective model to predict student academic performance in this course so the instructor can know how well or how poorly the students in the class will perform. This study focused on developing and validating mathematical models that can be employed to predict student academic performance in engineering dynamics. The goal of this study is to develop a validated set of mathematical models to predict student academic performance in engineering dynamics, which will be used to 5 identify the academically-at-risk students. The predicted results were compared to the actual values to evaluate the accuracy of the models. The three objectives of the proposed research are as follows: Identify and select appropriate mathematical (i.e., statistical and data mining) techniques for developing predictive models. Identify and select appropriate predictor variables/independent variables that can be used as the inputs of predictive models. Validate the developed models using the data collected in four semesters and identify academically at-risk students.

Three research questions have been designed to address each research objective of the study. These three research questions include:

How accurate will predictions be if different statistical/data mining techniques such as multiple linear regression (MLR), multilayer perceptron (MLP) networks, radial basis function (RBF) networks, and support vector



machine (SVM) are used? What combination of predictor/independent variables yields the highest prediction accuracy? What is the percentage of academically at-risk students that can be correctly identified by the model?

The rest of our work are stated below

2. RELATED WORK

Samrat Singh, Dr. Vikesh Kumar [1]. Data Mining is a powerful tool for academic performance. Educational Data Mining is concerned with developing new methods to discover knowledge from the educational database and can be used for decision-making in the educational system.

M. Goyal and R. Vohra [2]. Data analysis plays an important role for decision support irrespective of the type of industry like any manufacturing unit and educations system. If data mining techniques such as clustering, decision tree, and association are applied to higher education processes, it would help to improve student's performance, their life cycle management, selection of courses, to measure their retention rate, and the grant fund management of an institution.

Jason Brownlee [3]. After you have found a well-performing machine learning model and tuned Sample output to test PDF Combine only P a g e | 7 it, you must finalize your model so that you can make predictions on new data.

Neelam Naik & Seema Purohit [4] . Quality higher education is required for the growth and development of a country. Professional education is one of the pillars of higher education. Data mining techniques aim to discover hidden knowledge in existing educational data, predict future trends and use it for the betterment of higher educational institutes as well as students.

Alaa M.El-Halees, Mohammed M. Abu Tair. [5] Educational data mining concerns developing methods for discovering knowledge from data that come from the educational domain. In this paper, we used educational data mining to improve graduate students' performance and overcome the problem of low grades of graduate students.

B.K. Bharadwaj and S. Pal [6]. Nowadays the amount of data stored in educational databases increasing rapidly. These databases contain hidden information for the improvement of students' performance. The performance is higher education in India is a turning point in academics for all students. This academic performance is influenced by many factors, therefore it is essential to develop a predictive data mining model for students' performance so as to identify the difference between high learners and slow learners students. In the present investigation, an experimental methodology was adopted to generate a database.

Suchita Borkar, K. Rajeswari [7]. Education Data Mining is a promising discipline that has an imperative impact on predicting students' academic performance. In this paper, student's performance is evaluated using an association rule mining algorithm. Research has been done on assessing student's performance based on various attributes. In our study, important rules are generated to measure the correlation among various attributes which will help to improve the student's academic performance.

Randhir Singh, M.Tiwari, Neeraj Vimal [8]. Educational institutions are important parts of our society and playing a vital role in the growth and development of the nation and prediction of student's performance in educational environments is also important as well. A student's academic performance is based upon various factors like personal, social, psychological, etc.

D.Magdalene Delight Angeline [9]. The objective of the educational institution that is producing good results in their academic exams can be achieved by using the data mining techniques which can be applied to predict the performance of the students and to impart the quality of education in the educational institutions. Data mining is used to extract meaningful information and to develop relationships among variables stored in a large data set.

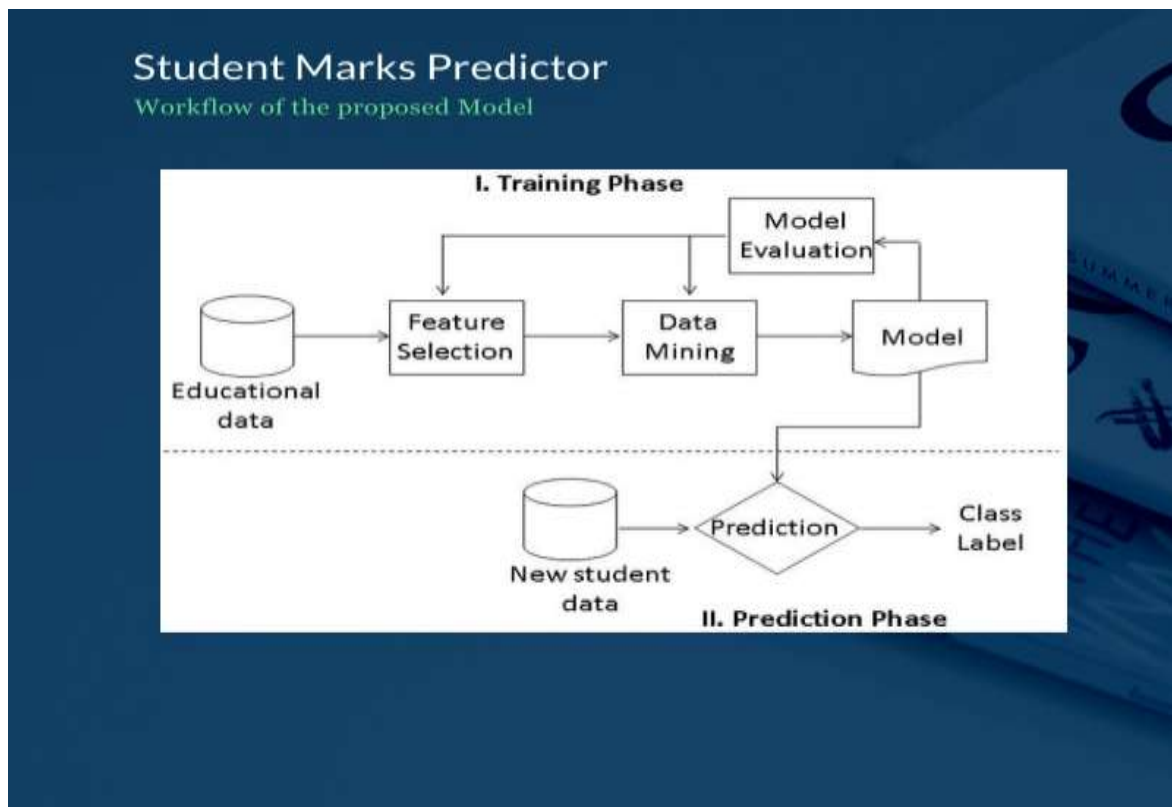
Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas [10]. In recent years, the analysis and evaluation of students' performance and retaining the standard of education is a very important problem in all educational institutions.



The most important goal of the paper is to analyze and evaluate the school students' performance by applying data mining classification algorithms in WEKA tool.

S. Anupama Kumar and Dr. Vijayalakshmi M.N [11]. Educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. Classification methods like decision trees, rule mining, Bayesian network, etc can be applied to the educational data for predicting the student's behavior, performance in the examination, etc.

3. METHODOLOGY



Student Characteristics Regression is the most often used technique for student performance prediction. Researchers usually examined study-related (SR) data. Our study-related data contained attributes such as higher secondary school background, the medium of study, syllabus, mathematics, and English marks. We built a multi-linear regression for the programming course based on the training set and evaluated the results using various models(SVM, RF, Decision tree). The method that achieved the best results was subsequently validated on the test set. Grade prediction regression is a commonly used technique for student grade prediction. And also SVM Reg., Random Forest, Decision Tree are used to validate the models. The baseline model predicts the programming marks of the training set of a given instance of attributes. In addition to accurately predicting students' performance, the multi-regression model can be used to analyze how the different features contribute to the predicted grades and thus gain some insights about the student's performance. For a proper analysis of the estimated model parameters, it is more convenient that all the attributes have non-negative values which will make all the model's components contribute additively to the predicted grades.

3.1 IMPLEMENTATION:



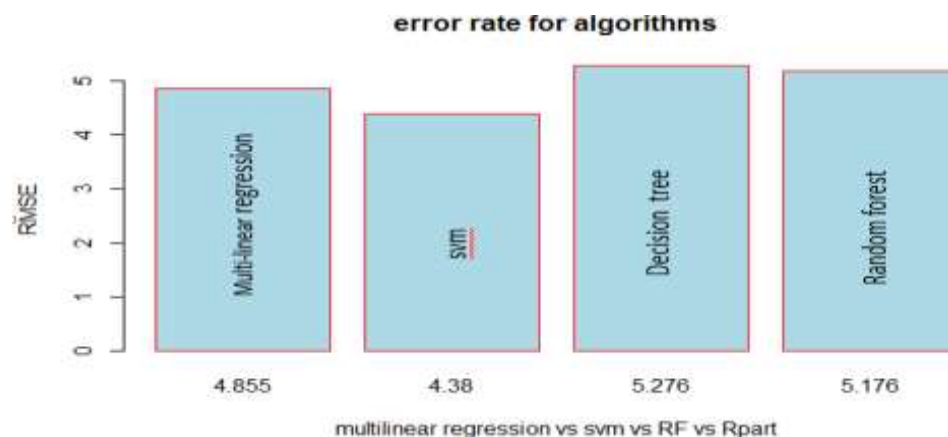
3.1.1. Multi-linear regression model:

$$Y = 0.0267(m) + 0.00444(e) + (-0.25824(sicse)/0.59825(sstate)) - 0.49231(llocal) + 0.56192(bckmaths) - 0.59242(csyas) + 1.79781(int)$$

lm(formula = c ~ m + e + s + l + bck + cs + int, data = trai)

3.1.2. COMPARING APPROACHES:

In comparison with the method using all grades, both approaches had positive effects on the number of calculations. RMSE gives the standard deviation of the model prediction error. A smaller value indicates better model performance. In this case, residuals of models are considered and RMSE is calculated for both multilinear regression and SVM has almost the same error rate so any of the models can be used to test the data and predict the values. Fig represents the RMSE rate for algorithms



3.2 ROOT MEAN SQUARE ERROR

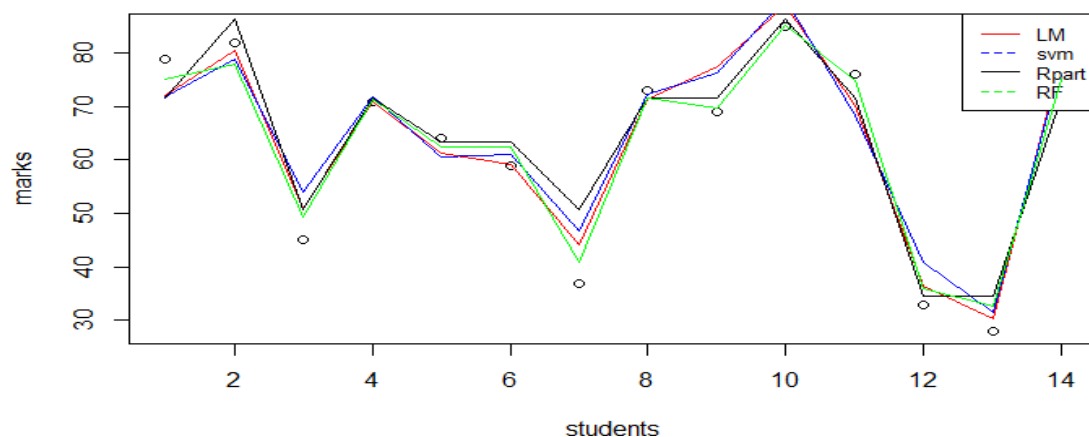
Root mean square is calculated by identifying residuals and squaring residuals and finding the mean of the squared residuals and finally calculating the root of the mean squared gives RMSE. By using RMSE we can identify which algorithm gives high accuracy results for the given datasets. In this case, we observe that SVM, linear model has the least RMSE. Experimental results represent the graphical output for four algorithms in test data. The goal of this study was to develop a validated set of statistical and data mining models to predict student academic performance in an engineering dynamics course. This chapter describes how the predictive models were developed using six combinations of predictors and four modeling techniques (MLR, MLP network, RBF network, and SVM). The models were developed and validated based on the quantitative data of student academic performance collected during four semesters from 2008 to 2011. The criteria used to evaluate and compare the models are also defined. The three objectives of this research were as follows: Identify and select appropriate mathematical (i.e., statistical and data mining) techniques for constructing predictive models. Identify and select appropriate predictor variables (i.e., independent variables) that can be used as inputs for predictive models. Validate the developed models using the data collected during multiple semesters to identify academically at-risk students.

Three research questions were designed to address each research objective: How accurate will predictions be if different statistical and data mining modeling techniques such as traditional multiple linear regression, MLP networks, RBF networks, and SVM are used? What particular combination of predictor variables will yield the highest prediction accuracy? What is the percentage of academically-at-risk students that can be correctly identified by the models?



MODEL	Root Mean Square Error	Mean Absolute Error	Explained Variance Score	R2 Score	Predicted Weight
Baseline Model	1.344	9.86	1.11*e-16	0	6.834
Linear and Ridge Regression	1.056	0.830	0.3687	0.3685	7.4877
K-Nearest Neighbor(k=30)	1.298899	1.013725	0.132261	0.096206	7.0625
Decision Tree(max depth=7)	1.022807	0.795126	0.406359	0.406292	7.5971
Ada Boost (estimators=30)	1.033814	0.805893	0.431822	0.429056	7.7089
Random Forest (depth= 10)	0.992277	0.768459	0.451359	0.451355	7.6648

LM vs SVM vs RF vs RPART





```
In [15]: # y = m * x + c
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(X_train,y_train)#Linear regression

Out[15]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

In [16]: lr.coef_

Out[16]: array([[3.93571882]])

In [17]: y_pred = lr.predict(X_test)

y_pred

In [18]: lr.intercept_

Out[18]: array([50.44735504])

In [19]: pd.DataFrame(np.c_[X_test, y_test, y_pred], columns = ["study_hours", "student_marks_original", "student_marks_predicted"])

Out[19]:
```

	study_hours	student_marks_original	student_marks_predicted
0	8.309000	82.02	83.113815
1	7.230000	77.55	78.902595
2	8.670000	84.19	84.570030
3	8.900000	85.46	85.829460
4	8.710000	84.03	84.727459
5	7.709000	80.81	80.752384
6	5.680000	73.61	72.841591
7	5.390000	70.90	71.668875
8	5.790000	73.14	73.235162

4. CONCLUSION

Predicting student's performance would boost the results of student's grades and gives teachers a better approach for teaching the students who are at risk of failure. Regression models, tree-based models and created to make the best predictions with high accuracy. The basic idea is to increase the efficiency of the prediction results using various algorithms. Thus by finding the RMSE, we observed that SV, linear model gives optimum results.

The following observations are made:

In internal validation, SVM models have relatively low APA, but relatively high PAP.

RBF models yield the lowest average PAP among the four types of models in internal validation.

Although MLP models generate good APA in external validation, RBF and SVM models outperform MLP models in terms of PAP.

RBF and SVM 91 models have nearly the same level of performance in terms of APA and PAP.

The MLP models have the lowest performance among the four types of models based on the data collected in this study.