MACHINE LEARNING ALGORITHMS TO PREDICT STUDENT'S ACADEMIC PERFORMANCE.

BY :
[1] BASHRU ALIYU SANI, BADAMASI HARUNA[2]
Department of Computer Science
Abdu Gusau Polytechnic,
Talata Mafara, Zamfara State
EMAIL: baliyusani@gmail.com

Abstract
The greatest aim of every educational setup is giving the best educational experience and knowledge to the students. Discovering the students who need extra support and guidance so as to carry out the necessary actions to enhance their performance plays an important role in achieving that aim. In this research work, three machine learning algorithms have been used to build a classifier that can predict the performance of the students in higher institutions considering three institutions which are Federal University Dutsinma, Katsina State, Abdu Gusau Polytechnic TalataMafaraand college of Education Maru, Zamfara State. The machine learning algorithms include Support Vector Machine, Linear Regression and Stochastic Gradient descent algorithms. The models have been compared using the Mean Absolute Error, Mean Square Error and Root Mean Square Error classification accuracy. The dataset used to build the models is collected based on a survey given to the students and the students' grade book. The support vector machine model achieved the best performance that is equal to 99.09%.

Keywords: support vector machine, Linear Regression and Schocastic algorithm, Students' Performance Prediction

INTRODUCTION

With the wide usage of computers and internet, there are recently been a huge increase in publicly available data that can be analyzed. Be it online sales information, website traffic, or user habits, data is generated every day. Such a large amount of data presents both a problem and an opportunity. The problem is that it is difficult for humans to analyze such large data. The opportunity is that this type of data is ideal for computers to process, because it is stored digitally in a well-formatted way, and computers can process data much faster than humans.

The concept of machine learning is something born out of this environment. Computers can analyze digital data to find patterns and laws in ways that is too complex for a human to do. The basic idea of machine learning is that a computer can automatically learn from experience, (Alpaydin, 2004). Although machine learning applications vary, its general function is similar throughout its applications. The computer analyzes a large amount of data, and finds patterns and rules hidden in the data. These patterns and rules are mathematical in nature, and they can be easily defined and processed by a computer.

The computer can then use those rules to meaningfully characterize new data. The creation of rules from data is an automatic process, and it is something that continuously improves with newly presented data.

Applications of machine learning cover a wide range of areas. Search engines use machine learning to better construct relations between search phrases and web pages. By analyzing the content of the websites, search engines can define which words and phrases are the most important in defining a certain web page, and they can use this information to return the most relevant results for a given search phrase (Witten et al., 2016). Image recognition technologies also use machine learning to identify particular objects in an image, such as faces. First, the machine learning algorithm analyzes images that contain a certain object. If given enough images to process, the algorithm is able to determine whether an image contains that object or not (Watt et al., 2016). In addition, machine learning can be used to understand the kind of products a customer might be interested in. By analyzing the past products that a user has bought, the computer can make suggestions about the new products that the customer might want to buy (Witten et al., 2016). All these examples have the same basic principle. The computer processes data and learns to identify this data, and then uses this knowledge to make decisions about future data. The increase in data has made these applications more effective and thus more common in use.

Depending on the type of input data, machine learning algorithms can be divided into supervised and unsupervised learning. In supervised learning, input data comes with a known class structure (Mohri et al., 2012). This input data is known as training data. The algorithm is usually tasked with creating a model that can predict one of the properties by using other properties. After a model is created, it is used to process data that has the same class structure as input data. In unsupervised learning, input data does not have a known class structure, and the task of the algorithm is to reveal a structure in the data (Sugiyama, 2015).

This proposed research focuses on supervised learning, more specifically predictive analytics, which is the process of using machine learning to predict future outcomes (Nyce, 2007). Predictive analytics has a wide range of applications, such as fraud detection, analyzing population trends, or understanding user behavior (Sas, 2017).

The specific focus of this thesis is education. The aim is to predict student performance. Data about students is used to create a model that can predict whether the student is successful or not, based on other properties. First, the training data set is taken as input. There are different data sets, containing different types of information like certain information about a student, such as age, gender, family background or medical information. The algorithms create a model, which is a function that outputs success or failure of the student, using other variables as input.

This proposed research evaluates the effectiveness of three different machine learning algorithms and enhanced the performance of the best which are Support Vector Machine, Linear Regression and Stochastic Gradient descent algorithms.

To determine which algorithm provides the best results, Mean Absolute Error, Mean Square Error and Root Mean Square Error were used and the implementation was done on Python.

LITERATURE REVIEW

Many researches have been done in the aspect of predicting students' performance so as to guide the lecturers on how to guide the student at risk, below are some of the literatures reviewed:

Havan, (2015) compared five algorithms, viz. Decision Trees (C4.5), Naive Bayes algorithm (Bayesian networks), 3-NN (kNN), RIPPER (Rule Learning) and WINNOW (Perceptron based neural networks). This study was composed of two experimental stages, training and testing. During these stages, number of attributes was increased step-by-step. For example, while only demographic data was included in the first step, performance attributes were added in the next step. Five algorithms were tested for each these subsequent steps and then they were compared. However, classification of data into binary groups seems insufficient. The primary goal of this study was only detecting at-risk students instead of determining performance levels of students. Classifying students according to their performances in different levels (e.g. poor, average, good, excellent, etc.)

Harshil, (2015) conducted a study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India.

By means of Bayesian classification method on 17 attributes, it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

Meier et al, (2015) worked onPaper Features Dataset Size Machine Learning Algorithms Grades which used New algorithm proposed (NAP), KNN Linear regression, logistic regression, SVM and the New algorithm proposed(NAP perform best.

Guleria et al, (2014) proposed Class Performance, Attendance, Assignment, Lab Work, Sessional Performance using decision tree machine learning algorithm for prediction.

Xu et al, (2017)used Grades and Backgrounds of student as parameters while Linear Regression, Logistic Regression, RF, kNN, and Progressive Prediction algorithm where the progressive predictive algorithm give large percentage accuracy.

Arsad et al ,(2013) uses only Grade as the dataset and the algorithms used was ANN
Dekker et al., (2009) studied evaluating the effectiveness of machine learning for dropout prediction was done at the Eindhoven University of Technology. Basic methodology was to build multiple prediction models using different machine learning methods, such as CART, BayesNet, and Logit. Then, prediction results of different models were compared in terms of their effectiveness. Most successful model was built by using the J48 classifier.

Yadav et al., (2012) made a similar study which was made by researchers from three different universities in India. A data set of university students was analyzed by different algorithms, after which precision and recall values of the predictions were compared. The ADT decision tree model provided the most accurate results.

Kalles and Pierrakeas, (2006) made in the Hellenic Open University, analyzed the usage of machine learning in distance education. Genetic algorithms and decision trees were used to build a predictive model, and the results were compared in terms of accuracy. Most accurate results were provided by the GATREE (genetically evolved decision trees) model

Amrieh et al., (2016) studied about performance prediction which was made at the University of Jordan. A data set of students from different countries was used.

In addition to using individual machine learning methods, the researchers also applied ensemble methods, and compared the results between them. Decision trees provided the best results. Another area that the researchers focused on was behavioral features. A model was built with and without these features. It was found that the inclusion of behavioral features improved the prediction results.

Cortez and Silva, (2008) proposed a research which was done at the University of Minho, Portugal. The data set contained information about whether the student had passed the exam in the subjects of math and Portuguese language. Decision trees, random forest, neural networks, and support vector machines were used. These methods were compared in terms of accuracy. Another comparison was made between a data set that included the past exam results and the one that did not. Inclusion of the past grades resulted in an improved performance.

Li et al, (2013)also considered Grades with the use of ANN, SVM and PCA, and PCA performed well better.

Gray et al, (2014) Aptitude, Personality, Motivation Learning strategies Naïve Bayes, Decision Tree, Logistic Regression, SVM, ANN, KNN were used and SVM, KNN, Naïve Bayes gives best results Buniyamin et al, (2016)used Grades and in-built features to feed the machine using Neuro-Fuzzy classification.

Alharbi et al, (2016)considered student demographics, general performance, students' modules 1789 dataset were used for Testing and 898 were used for Training, so he fed them into machine that used Logistic regression, ANN, DL, BN, DA, DT , And Ensemble approach and it was concluded that there was no overall winner.

Livieris et al, (2012)used only Grades by considering 279 students data from ATYI College, Gambia using ANN, DT, NB, Rule- Learning, and SVM where ANN and SVM were the best. Hamsa et al, (2016)proposed the collection of Internal grades, sessional grades and admission score of 168 students and fed them into Fuzzy Genetic Algorithm and DT, FGA model is less strict than DT. Arsad et al, (2014)used grades of 896 candidates and the machine learning algorithm used were ANN, Linear Regression and both performed well concurrently

Sarker et al, (2014) Personal and demographics information, student satisfaction and integration of 149 students were collected for preparation, they used ANN and Logistic Regression where logistic regression has the highest percentage of 49% Huang et al, (2011)from a University used GPA and Grades of 239 students and considered Linear Regression, ANN, Radial Basis Function NN, SVM. But all in all SVM performed more better.

Elaf, (2016)proposed a model for the prediction of students' performance using data mining approaches with some few characteristics referred as student's behavioural features. This model was evaluated with three classifiers which are; Artificial Neural Network, Decision tree and Naïve Bayesian. Bagging, Random Forest, and Boosting were ensemble methods used in order to enhance the performance of the classifiers. The model achieved up to 22.1% more in accuracy as compared to when the behavioural features were not considered. The model improved to 25.8% accuracy on using the ensemble methods.

METHODOLOGY

To address the common issues of above review literatures such as class imbalance, data hi-dimensionality and classification errors, this study has proposed a model which have following phases. Fig 1 shows the main steps of proposed methodology.

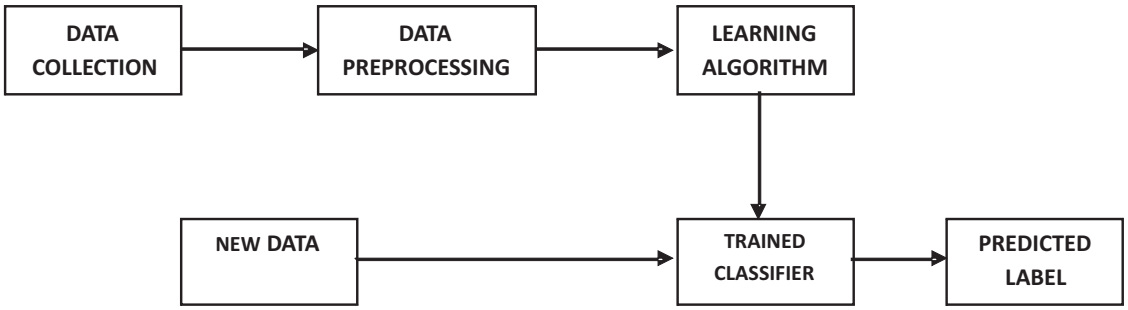


Fig1: steps involved in the proposed model

Data collection

The dataset used in this research was collected from the Computer Science department of the Federal University Dutsinma, Katsina State, Abdu Gusau Polytechnic TalataMafaraand college of Education Maru, Zamfara State dur**ing the 2015 and 2019 academic years. Three data sources have been used, survey collected from the students and the students' grades data records.**

Data preprocessing

Pre-processing plays an important in data mining. Its purpose is to convert raw data into a suitable form which can be used by mining algorithms. Following tasks are performed in this phase. The parameters used are Field of study, Mothers education, Family support, Study time, Internet usage, Health status, Missing classes, Time spent on sport and games, Time spent on social media and Frequent attending of social gatherings. And the following were used for the preprocessing:

- **Data integration**

Data Integration refers to gathering of the data from the multiple sources into single repository. Redundancy is one of the common problems that arose when integrating data. The dataset consists of two comma separated values files which were taken from Machine learning repository. These files contained the performance data of various courses which were studied by the Students of **Federal University Dutsinma, Katsina State,** Abdu Gusau Polytechnic TalataMafaraand college of EducationMaru, Zamfara State. In this step, multiple files are integrated into one file.

- **Data cleaning**

This phase deals with the handling of missing and noisy data to achieve data consistency. The dataset occupied by this study must not have any missing and outliers etc. for effective prediction.

- **Discretization**

The discretization mechanism is a technique used to transform the needed data from numerical values into nominal values. Some classifiers are not applicable on continues data. That's why target attribute which is Grade has been converted into nominal.

- **Class balancing**

This phase deals with data balancing; data balancing technique is applied after data pre-processing for solving the class imbalance problem. The class imbalanced problem arises when the number of instances in one class is much smaller than the number of instances in another class or other classes. Traditional classification algorithms provide high accuracy for majority classes when data is un-balance because during classification, they have much intension towards majority class instances and have less intension for minority class instances.

The adjustment of the ratio of two class samples can improve the machine's learning performance. Therefore, we employed a class balancing method known as re-sampling in this phase. After re-sampling on the training set, 50% PASS and 50% FAIL students are obtained.

- **Feature selection**

The student performance dataset contains several attributes, which may be inappropriate for classification sake. The problem of data high dimensionality arises when included large amounts of student's characteristics which can influence student performance such as educational background, social, demographics, family, socioeconomic status etc. This issue can be resolve by focusing on the important features from the dataset. The purpose of feature selection is to select an appropriate subset of features which can efficiently describe the input data, which reduces the dimensionality of feature space and removes irrelevant data. Feature selection methods are of two types which are wrapper-based and filter-based methods. Filter method means searching for the minimum set of relevant features while leaving the rest. It uses variable ranking techniques to rank the features where the highly ranked features are selected and applied to the learning algorithm. This study applied filter method using information gain-based selection algorithm to evaluate the feature ranks. It's checking which features are most important to build students' performance model. During feature selection, a rank value is assigned to each feature according to their influence on data classification. Fig 3 depicts the level of each of the parameters both the dependents and the independent (GPA).

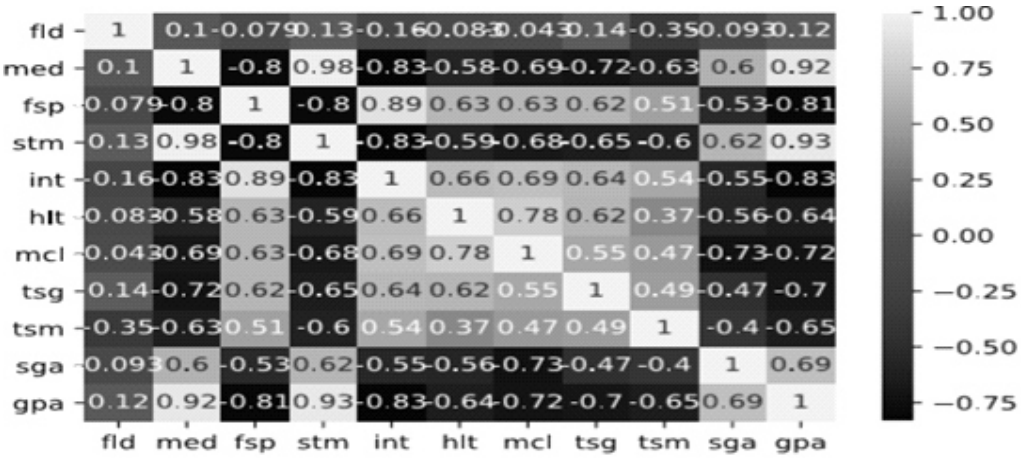


Fig 3.0: Measure of each parameter

The literature review recommends that in general there is no single classifier that works best in all contexts to provide good prediction and those that has good prediction didn't improve the better algorithm.

In this paper, student's performance prediction model is build using ensemble method. Ensemble method is a learning approach that combines multiple models to reduce their classification errors as well as to improve the accuracy of better classifier. The predictions made by ensembles are usually more accurate than predictions made by a single model. Generally, Ensemble has two types of approaches:

The **Homogeneous ensembles**, in this a combination of one ensemble learning (Meta model) such as Bagging, Boosting and one base model is applied.

Heterogeneous ensembles: is an ensemble that merges two or more different base methods. This study has employed heterogeneous ensemble approaches. In the model building for this, three classifiers were applied on the clean dataset. Then a **Meta classifier Realadaboost** has been applied to improve the accuracy of the best classifier by reducing its classification errors. The classifiers that were used are support vector machine, Naïve Bayes and Stochastic Gradient descent.

Stochastic Gradient Descent binary classifiers

Stochastic Gradient Descent (SGD) has gained popularity for solving large scale supervised machine learning problems. It provides a rapid method for minimizing a number of loss functions and is applicable to Support Vector Machine (SVM) and Logistic optimizations. However SGD does not provide a convenient stopping criterion. Generally an optimal number of iterations over the data may be determined using held out data. SGD has proved to be a very effective method of training machine learning algorithms. It has generally been found to confer a significant decrease in training time without sacrificing accuracy. SGD can be applied to standard convex loss functions with regularization terms with good effect, but was later suggested using SGD without the usual regularization term and performing the regularization with early stopping. This has become a widely practiced approach and is implemented by dividing the training set into disjoint pieces consisting of a new training set and a validation set. The advantage of Stochastic Gradient Descent is that it is efficient and easy to implement compared to other machine learning models like deep learning. However, the demerits of Stochastic Gradient Descent are that; requires a number of hyper parameters such as the regularization parameter and the number of iterations and is sensitive to feature scaling.

Support Vector Machine

Support Vector Regression is the prediction type of Support Vector Machine which assigns support vectors in order to separate features. SVMs are described as a set of related supervised learning techniques used for classification and regression. They are member of a family of generalized linear classification. An important property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. Thus, SVM is also known as a Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyperplane that separate the data. The separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. In classification phase, increment of classes may cause the reduction of success rate in SVM. However, it can be used effectively for 2-class problems.

Linear Regression

Linear Regression is an attractive model because it is a linear model, i.e. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x). The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

For example, in a simple regression problem (a single x and a single y), the form of the model would be:

$y = B0 + B1*x$

In higher dimensions when we have more than one input (x), the line is called a plane or a hyper-plane. The representation therefore is the form of the equation and the specific values used for the coefficients (e.g. B0 and B1 in the above example).

It is common to talk about the complexity of a regression model like linear regression. This refers to the number of coefficients used in the model. When a coefficient becomes zero, it effectively removes the influence of the input variable on the model and therefore from the prediction made from the model ($0 * x = 0$).

This becomes relevant if you look at regularization methods that change the learning algorithm to reduce the complexity of regression models by putting pressure on the absolute size of the coefficients, driving some to zero. Machine learning, more specifically the field of predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explain ability.

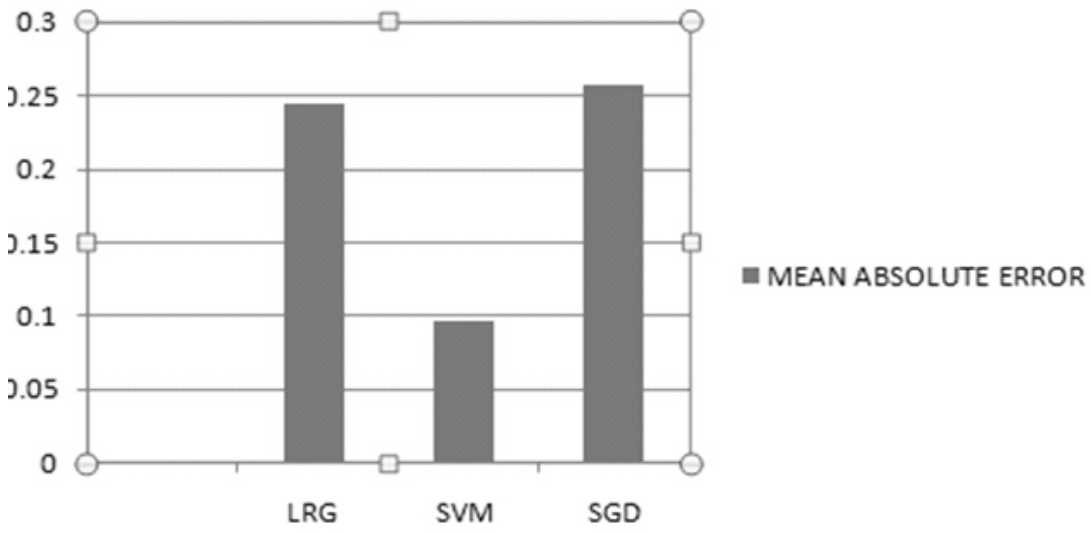
RESULTS AND DISCUSSION

Model evaluation

For our experiments, three classifiers Support Vector Machine, Linear Regression and Stochastic

```
n [5]: #division of datasets in 1:4 (i.e. 80% as training dataset while 20% as te
       X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.8,
       X_test.fillna(X_train.mean(), inplace=True)
```

Gradient descent algorithms have evaluated using 100 folds cross validation technique. This technique divides the data set into 100 subsets of equal size; 80% of the subsets are used for training, while 20% is left out and used for testing. The final result is estimated as the average error rate on test. Fig 3.1 shows the percentage for training and testing.



Absolute Error for the classifiers

It was observed the Mean absolute error for Linear Regression is 0.24, for Support Vector Machine its 0.009 and SGD Regression shows 2.26, therefore, it means SVM has the lowest number of mean absolute error which indicates that the accuracy will be higher that the two other classsifiers

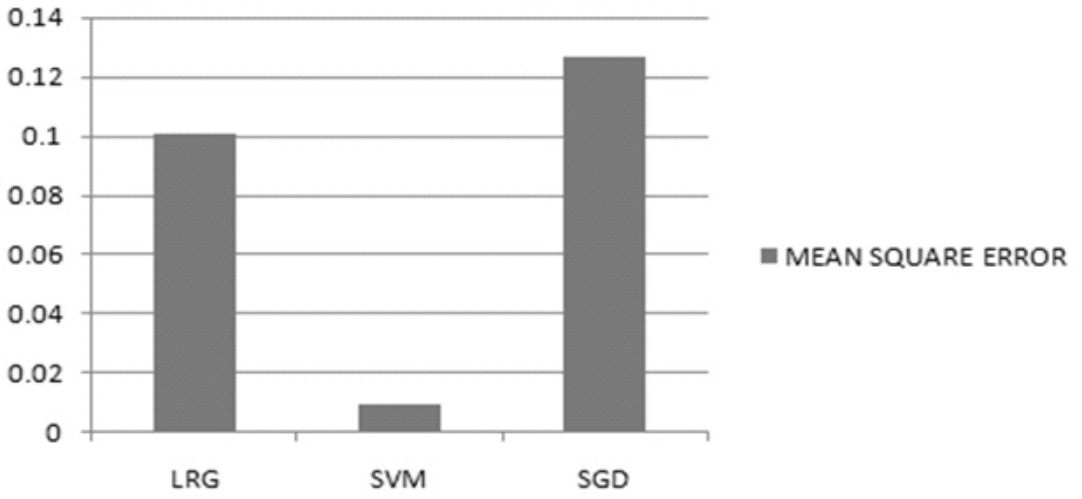


Fig 3.3: Mean Square Error on the classifiers

From Fig 3.3, Linear Regression shows 0.100, SVM shows 0.001 while SGD regression shows 0.126, by indication SVM shows the lower error in mean square error, which means its own accurate prediction will be higher and be the best.

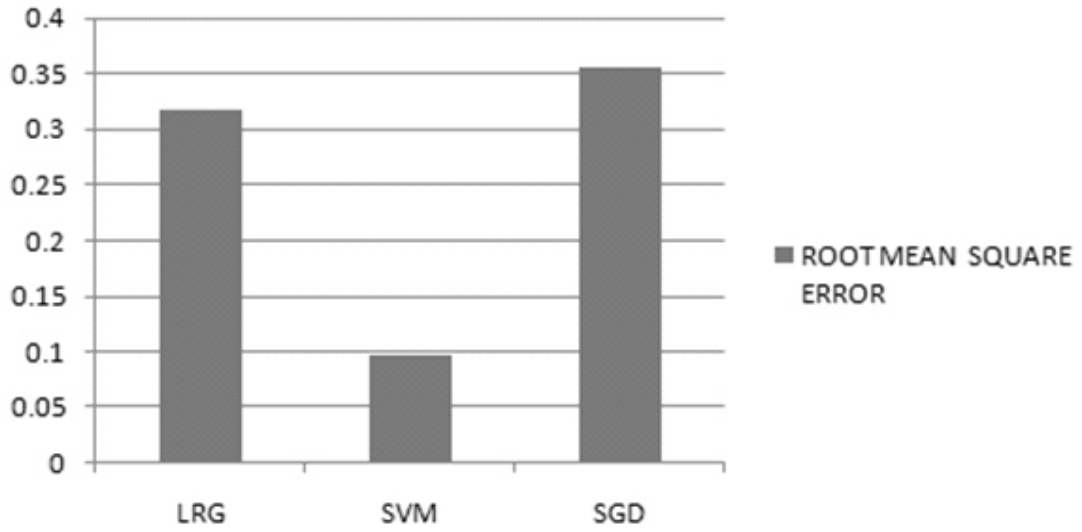


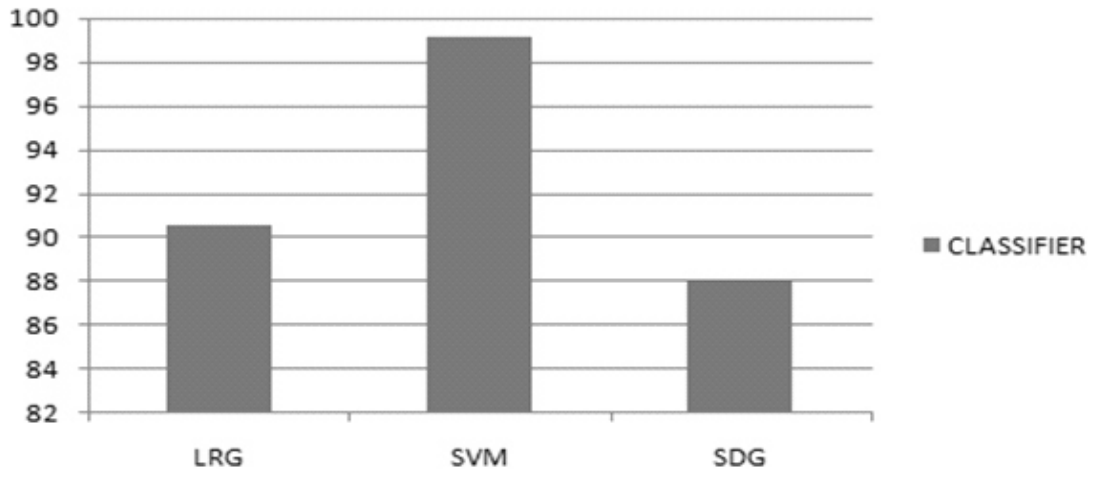
Fig 3.4: Root Mean Square Error for the classifiers

Fig 3.4 indicates that Linear Regression has 0.32, SVM has 0.09 and SGD has 0.36, this implies that SGD has the highest root mean square error, followed by Linear Regression, while SVM show the least,that means that the prediction is better.

Result analysis

In the experiment, three classification algorithms (Linear Regression, Support Vector Machine and Stoichastic Gradient Descent) are simply executed all together on dataset.

In the light of results shown in graphical representation of Fig 3.5, it has been observed that the highest accuracy achieved by Support Vector Machine which was not enough but was improved and 99.1% was realized as the accuracy as compared to previous studies and lowest accuracy has achieved by Stochastic Gradient Descent Regression which is 88.1%.



Accuracy comparism

Fig 3.6 shows the output of the classifiers compared to the actual value. And this shows that Support vector machine predict better after the improvement

Out[17]:

	Actual	linearR_Prediction	supportVM_Prediction	stochasticGD_Prediction
788	2.55	2.804363	2.650293	3.002955
440	2.33	2.765376	2.430158	2.824360
139	2.87	2.003657	2.769936	1.806641
695	4.70	4.566729	4.600461	4.524137
66	1.96	1.792276	1.859942	1.886638
...	...	...	...	...
617	3.01	2.772933	2.909822	2.925005
449	3.66	2.720499	3.559692	2.780741
331	4.62	4.447525	4.520102	4.677777

classifiers prediction outputs

CONCLUSION AND FUTURE WORK

To predict student's performance in advance is a very important issue. We concluded after deep studies that various datasets of student provides different results with different attributes. This is the reason that the results are vary with different evaluation measures like accuracy, precision and geometric mean. It was concluded after these studies that every algorithm results are varied according to the dataset and variable attribute used for prediction. However, if we use the Linear Regression, Support Vector Machine and Stochastic Gradient Descent, according to our requirements these algorithms provide extra ordinary accurate results for future prediction and help in the betterment of education system. However, Support Vector Machine algorithm performed better and was improved toperform more better prediction. For future research, other classifiers can be used for prediction.

Reference

Alharbi, Z., Cornford, J, Dolder, L. and De La Iglesia, (2016) "Using data mining techniques to predict students at risk of poor performance," Proc. 2016 SAI Comput. Conf. SAI 2016, pp. 523–531

Arsad, P., Buniyamin, N., and Manan, L.A, (2013) "A neural network students' performance prediction model (NNSPPM)," 2013 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2013, pp. 26–27

Buniyamin, N., Bin U. M. and Arshad, P.M. (2016) "Educational data mining for prediction and classification of engineering students achievement," IEEE 7th Int. Conf. Eng. Educ. ICEED 2015, pp. 49–53.

Elaf A. A., Thair, H. and Ibrahim, A. (2016) "Mining educational data to predict student's academic performance using ensemble methods". International Journal of Database Theory and Application 9(8), 119-136.

Ethem, A. (2004). "Introduction to Machine Learning". Cambridge, MA

Gray, G., McGuinness, C. and Owende, P. (2014) "An application of classification models to predict learner progression in tertiary education," in Souvenir of the 2014 IEEE International Advance Computing Conference, IACC.

Guleria, P., Thakur, N., and Sood, M. (2015) "Predicting student performance using decision tree classifiers and information gain," Proc. 2014 3rd Int. Conf. Parallel, Distrib. Grid Comput. pp. 126–129

Hamsa, H., Indiradevi, S., and Kizhakkethottam, J. (2016) "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," Procedia Technol.

Havan, A., Harshil, M. (2015) "Student Performance Prediction using Machine Learning" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 4 Issue 03.

Huang, S. and Fang, N. "Work in progress: Early prediction of students' academic performance in an introductory engineering course through different mathematical modeling techniques," Proc. - Front. Educ. Conf. FIE, vol. 1, pp. 3–4, 2012.

Kalles D. and Pierrakeas, C. (2006), "Analyzing student performance in distance learning with genetic algorithms and decision trees".Applied Artificial Intelligence 20(8), 655- 674.

Li, K., Rusk, D. and Song, F.(2013) "Predicting student academic performance," Proc. - 2013 7th Int. Conf. Complex, Intelligent, Softw. Intensive Syst., pp. 27–33.

Livieris, I.E., Drakopoulou, K. and Pintelas, P.(2012) "Predicting students' performance using artificial neural networks," Proc. 8th Pan-Hellenic Conf. "Information Commun. Technol. Educ., pp. 28–30.

Meier, Y., Xu, J., Atan, O. and Van Der Schaar, M (2016) "Predicting grades," IEEE Trans. Signal Process., vol. 64, no. 4, pp. 959–972

Mohd, P., Arsad, N. Buniyamin, and AbManan, J. (2014) "Neural Network and Linear Regression methods for prediction of students' academic achievement," IEEE Glob.Eng. Educ. Conf. EDUCON, no. April, pp. 916–921.

Mohri, M., Rostamizadeh, A. and Talwalkar.A. (2012)."Foundations of Machine Learning (Adaptive Computation and Machine Learning Series)".MIT Press.

Nyce, C. and CPCU.(2007) "A. Predictive analytics white paper".American Institute for CPCU. Insurance Institute of America, 9-10

Sarker, F., Tiropanis, T., and Davis, H. (2014) "Linked data, data mining and external open data for better prediction of at-risk students," in Proceedings - 2014 International Conference on Control, Decision and Information Technologies.

Sas, S. (2017) "Predictive Analytics: What it is and why it matters", https://www.sas.com/en_us/insights/analytics/predictive-analytics.html. Retrieved April 24, 2017.

Sugiyama, M, (2015) "Introduction to Statistical Machine Learning". Morgan Kaufmann.

Watt, J., Borhani, R and Aggelos K. (2016). "Machine Learning Refined: Foundations, Algorithms, and Applications". Cambridge University Press.

Witten, H. I., Eibe F., Mark A. Hall, and Christopher J. Pal. (2016). "Data Mining: Practical Machine Learning Tools and Techniques". Morgan Kaufmann.

Xu, J. Moon, K and Van M. Der Schaar, (2017) "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol. 11, no. 5, pp. 742–753

Yadav, S.K., Brijesh B, and Saurabh P. (2012) "Mining education data to predict student's retention: A comparative study". International Journal of Computer Science and Information Security 10(2), pp. 113-117.