

DOI: 10.36648/computer-science-engineering-survey.08.01.01

# Predicting the Performance of Primary School Students

**Fahad Ayyaz<sup>1\*</sup>, Amna Asif<sup>2</sup>, Asif Zahoor<sup>3</sup>**<sup>1</sup>Faculty at Allama Iqbal Science College, Bahawalpur, Pakistan<sup>2</sup>Faculty at Islamiya University, Bahawalpur, Pakistan<sup>3</sup> Finance Department Civil Secretariat, Punjab, Pakistan**\*Corresponding author:** Fahad Ayyaz, Allama Iqbal science college, Bahawalpur, Pakistan, Tel: 00923027845344; E-mail: fahadayaz05@gmail.com**Received date:** March 25, 2020; **Accepted date:** April 7, 2020; **Published date:** April 14, 2020**Citation:** Ayyaz F, Asif A, Zahoor A (2020) Predicting the Performance of Primary School Students. American Journal of Computer Science and Engineering Survey Vol.8 No.1: 01. DOI: 10.36648/computer-science-engineering-survey.08.01.01**Copyright:** © 2020 Ayyaz F et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Abstract

Education is important for developing nations. It also needs to be improved in Pakistan. The education system of Pakistan has developed in past years but there is still some needs to improve it further. The measure of student performance is important but it is not possible with on manual manners from some specific personals. Measuring the student performance by analyzing the large amount of accurate results can be done using the machine learning algorithms. From these algorithms the sorted form of data is analyzed and the results are made for making any specific actions. For this purpose, in this research work institute data of different primary schools in Punjab province is gathered. After the collection, raw data is changed into the sorted data and by selecting the eight suitable set of attributes. As each entity has a specific set of attributes and it is depending upon the situation that which attribute is selected for which purpose. From these attributes some are originally taken and some are derived from other set of original attributes. This data is uploaded in the rapid miner analysis tool and analyzed the data in terms of prediction by selecting the retention as special attributes. The algorithms are selected which are more suitable for this data. The attributes are Naive Bayesian, random forest, support vector machine and deep learning. Each of the algorithm has unique nature of classification and produced results in different manners but in the collective sense student and teacher both are involve in the student's performance. The degree of the teacher and age of the teacher are matter a lot while the attendance is another important factor which affects the performance of the student in terms of retention. The reason to select four algorithms is to make more accurate analysis and each algorithm has different level of accuracies. This methodology improves the quality education because through this large amount of student records can be processed and analyzed. Furthermore, this analysis can also be done in future by applying different set of parameters.

**Keywords:** Primary school students; Teachers; Education; Algorithms; Rapid miner tools.

## Introduction

Education is dire need for any society or nation. It becomes more important when the nation is developing. Improvements in other departments of the country is directly related to the education. Giving good quality education means improvements of the country. The quality of the education can also be improved with the help of computer systems or computer-based applications. Some of the computer-based systems for education are available such as distant learning, online test, or computer aided designs etc. But improving quality is related to measuring the performance on large scale which is not possible for humans on manual basis. For measuring performance and checking the factors related to the performance data mining technique and machine learning techniques of artificial intelligence can be used. Through these techniques large of the data can be processed and the results are made on the basis of required algorithms.

## Related Work

Aldowah H et al. proposed their work on the educational mining. It is new field of data mining in which educational sector improvements are being made with the help of data mining and computing algorithm. As each student is important and there are number of students who being taught by the number of teachers. So, it is important to make the automatic decision system with the help of mining tools to categorize the education schemas, study material and the learning capabilities of the students in the educational sector. They have divided their educational mining work into four categories which are computer-based learning analysis, computer-based analysis of behavior, computer based visual analysis and predictive computer-based analysis [1].

Mokhairi et al. worked on improving the education system of Malaysia. They have selected Malaysian school under the Sijil Education crite Arata Suzuki et al. worked on improving the health care system. As the blood pressure is the common issue to most of the person due to their hectic, tensioned but less physical active life. The blood pressure is the pressure of the

blood striking on the walls of heart while supplying the blood or receiving blood. Blood pressure can be observed in two manners systolic and diastolic. The blood pressure is checked using the devices but the conventional devices are time consuming and sometimes error prone. For making the blood pressure gathering mechanism an efficient one, techniques have been introduced and proposed. For implementing the techniques total 84 data sets are selected from these data half are used for teaching and half are used for testing.

Kiumi, Akingbehin proposed his work in the efficient development of the software in terms of quality. The software can be observed in three parameters, such as product, process and project. All these parameters need proper attention so these related to the quality of the software in cost effective. The software quality improvement means the better functioning of the system on which the software is working. Some software operated critical devices such as aircraft, space shuttle, or medical equipment, which should have the high quality but to provide the quality in sophisticated manner is necessary [2].

Khasanah et al. conducted a study to find that high influence attributes may be selected carefully to predict student performance. Feature selection may be used before classification for such job. The student data was from Department of Industrial Engineering Universities Islam Indonesia. They used Bayesian Network and Decision Tree algorithms for classification and prediction of student performance. The Feature Selection methods showed that student's attendance and Grade Point Average in the first semester topped the list of features. When the accuracy rate was considered, the Bayesian Network outperformed the Decision Tree classification in their case [3].

Pedro, Strecht et al. predicted students results (pass/fail) and their grades in their work. They used classification model for the students results and a regression model for the prediction of the grades. They carried out the experiments using the 700 courses student's data who studied at the University of Porto. They used decision trees and SVM for classification while SVM, Random Forest, and AdaBoost.R2 were best suited for regression analysis. The classification model was able to extract useful patterns, but the models for regression were not able to beat a simple baseline. They used Cumulative Grade Point Average (CGPA) for prediction of student's yearly performance. The dataset used was from Bangabandhu Sheikh Mujibur Rahman Science and Technology University students' records. The authors used neural network technique for prediction and it was compared with the real CGPA of the student [4].

Baradwaj, Pal have used Decision Tree algorithm to classify students' dataset obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) to predict students' division on the basis. Not only previous semester result has been chosen as the attributes but the lab work and the seminar performance also have contributed to the findings. Their research also able to identify those students who needed special attention in order to reduce fail ratio. The data collected contained various aspects of students' records including previous academic records, family background and demographics. Three classifiers viz. Decision Tree, Naive Bayes and Rule-Based classifiers are applied to find

the academic performance of students. The experiments showed that Rule Based classifier was the best among the other classifiers and its accuracy was found as 71.3%. The first-year student's level of success was predicted by the model. Developed a data model to predict student's future learning outcomes using senior student dataset. They compared the data mining classification algorithms and found that J48 algorithm was best suited for such job based on their data [5].

Rafique et al. worked on the automatic analysis. For that reason, they have focused on importance of analysis of sentiment and coping the today's needs in a global sense. They have studied the comment and opinions from three selected websites, in which they comments are written using the Urdu or Roman Urdu For understanding and classifying the word for the sentiment analysis they have taken the dataset from these selected websites and then these comments with opinions are then classified by the range of algorithms or classifiers for checking the accuracy and efficiency of each selected classifier in order to use the most suitable algorithm in the future. These classifiers are SVM, Naive Bayesian and LRSGD. These are famous and efficient classifiers but each works well in specific situation and for this situation is best classifier for sentiment analysis of Urdu with the accuracy value of 87.22 percentage [6].

Dewivedi et al. presented the research-based work on all aspects of social media in the daily life. The three main aspects of this research are goodness of social media on the life of the people, the badness of social media and the ugly effect of social media on the people life. The author explains the advantages and disadvantages of social media with different reference and different datasets collected from social media [7].

Khan et al. worked on improving the automobiles industries to proposed the way to analyze the reviews from the customers for automobiles. For this reason, they have selected the automobile websites from the vast range of websites. Then the dataset is collected for the analysis which is written in the Roman Urdu format. After the dataset is collected then it is classified from the set of classification algorithms on the WEKA platform for the efficient execution. With the help of this platform and the presence of classification algorithms the data is used for training as well as for testing. The testing data is ten percent while the rest of the dataset is used for training for better results. The algorithm used in this research work are NB, Deep Neural Networks and Random Forest etc. [8].

Pang et al. elaborated the challenges and the issues related to the opinion-based analysis and the fact-based analysis to most of the sentimental software. This software is compared on the basis of accuracy, efficiency and reliability. Furthermore, new present software is compared with the previously available software for the cost-effectiveness is also done. The dataset is also considered that which kind of data is being analyzed for a specific software. The quality of data is measured in terms of economic factors, privacy and availability. This analysis is suitable for selecting the suitable and relevant software or algorithm for the opinion-based analysis [9].

Mehmood et al. worked on the analysis of sentiment for the Urdu written in Roman style for communication now a days. For

the analysis of sentiment weighting of words method used. This method is helpful for doing the best analysis as it is regardless of structure of written language and also the spelling is defined to the user's site. The analysis is done for weighting method and three suitable weighting methods are considered for this purpose. The weighting methods are "Raw Weighting", "Novel weighting" and "Binary Weighting". They testing is done on t-testing pattern and the results showed that there is a significant amount of change for the improvement by using the weighting methods [10].

Asrofi et al. focused on the long set of textual information. This textual information is difficult to analyze for some sort of sentiment methods because of its large size. But for the high accuracy and more reliability it is adaptive to take the large set of data for training and testing. For the analysis of words, their meanings, sequences and the future usage of the words for understanding the behavior of person is necessary. For the analysis of large data and finding these parameters, they have used Deep Neural Networks for short and long parameter set which works exponentially and the outcomes are efficient. This method is good for large data sets and the large sets are the future needs because of frequent usage. The analysis is basically done for the Roman Urdu in larger perspective. They have also introduced the methods to improve the analysis method for sentiments and set the level of precautions to be avoided while doing the classification in order to train the algorithm. The selection of algorithm is another important thing because some algorithms having limitations for being used in specific situations [11].

Alam et al. focused on the importance of pure Urdu usage in the communication and in most of the websites as language of the nations is necessary to keep the nation recognized. The transmission done in pure Urdu is much difficult because of combining most Urdu words and also Urdu alphabets are more than of English alphabets that is why most of the users find it difficult to use Urdu as the communication. So, in this research work they have first ever proposed the idea to take the sentences in the form of Roman Urdu and converting these sentences in the pure Urdu. The translator is taking the source language and converts the language in to the Urdu. Between this translator a classifier and sentiment analyzer are used which is equipped with the artificial intelligence methods. This translator can also work in parallel manner. There are two main components of this translator: one takes the input as Roman Urdu sentences or sentiments and other proposed the results in Urdu language by considering the meaning, sequence and formation. But using it parallel may degrade the performance so there is a need to accommodate this problem in a sophisticated way [12].

Bilal et al. worked on small datasets of comments written in English as well in the Roman Urdu. The dataset is taken from the blog on the required website and the comments are extracted using the software named as "Web Extractor", the taken data consists of around three hundred comments half of the comments reflects positive meanings and half of the comment reflects the negative meaning of a specific situation. The classifiers are selected to find out the analysis and the

algorithms are selected in this way are NB, K-NN and Decision Tree. These classifiers are suitable for small datasets and when fast classification is required these classifiers are implemented on the WEKA simulation platform and the dataset is given to each of the selected classifiers for training and testing and also these classifiers are compared on the basis of recall, precision, accuracy and F-measure. The results of the simulation show the KNN classifiers is much suitable for small datasets by achieving the highest precision, accuracy and recall [13].

Syed et al. worked on the Urdu lexicon to create a lexicon sentiment of this language. They build a corpus of opinion either positive and negative and made a decision on the basis of weightage. They made a review for electronics by the condition of their products and the usage feasibility. On the other side they have done the review analysis for movies to categories the movies of views [14].

Abdul et al. highlighted the lexical variants for roman Urdu reviews and comments found on internet. They designed phonetic algorithm and map Urdu strings with their phonetic algorithm [15].

Rehman et al. presented the inflection and derivational morphology and how it can be used in various applications. In this paper author stated that no work is reported relating to Urdu stemmer which is beneficial and fulfil the requirement of the Urdu language structure. The author proposed own stemmer "Assas- Band" having accuracy about 91 % and it replaced the old approaches [16].

Riaz et al. presented the problems and challenges faced in developing the application of the Urdu language stemmer. He defined the complexities in understanding of Urdu language into machine readable form due to the less work and the complexity of its morphological structure. The author explains the problems related to stemming and design the prototype of Urdu language as Urdu is quietly different language and diverse nature. Arabic and Farsi have the same script like Urdu but their sentiment cannot be used in the Urdu language [17].

Hridoy et al. developed lexicon-based sentiment analysis for Urdu language. Author develop a lexicon based on negative and positive sentiments and develop an algorithm to calculate the polarity of Urdu text on the basis of developed lexicon. The accuracy was 66%. The author discusses the challenges faced in sentiment analysis of Urdu languages and common problems when mining Urdu language [18].

Javed et al. done the bilingual analysis of tweeters of the national interest. The national interest refers to the general elections or the performance of government. For this analysis they have taken the sentences or comments written in Roman Urdu and English. They have built lexicon and developed the bilingual dictionary for sentiwords on the WordNet [19].

Laudal et al. performed the analysis for sentiment of tweeter's tweets. The tweets are the comments which in form of opinions about a specific topic. The opinions are helpful in proposing the decisions about anything. In their research work they have selected the news and considered they related tweets on the tweeter to find out the best results. The strategies have

been made to concatenate the news with relevant comments either negative or positive and they proper meaningful solution is made. They have compared the exploited statements for check the accuracy and truth of the news on the basis of these comments and weightage values [20].

Peldszus et al. mined the argument-based analysis of the opinion in the formal way of representation and defined the way to improve the arguments by studying the critical analysis of literatures. The critical analysis is related to the deep learning and understanding of the written text with the combinations of arguments and built a decision-making tool for the required needs. The textual data is combination with the arguments is the adaptive approach. In this research work previous mining methods are improved, analysis of the methods is done with RST method of mining spectrum. Analysis on the basis of arguments and comparing the previously available prototyping and mining tools is a good approach to linguistics and computational logics. They solutions are helpful in deciding the textual context meaning in a more sophisticated way then neve be done in the last decade in the field of text mining. The author also recommended these techniques as the standardized task to reduce the cost and time [21].

Li et al. done the review analysis for camera products. Before doing actual analysis and proposing the model for analysis, they have studies the existing text and opinion mining tools and they made a suggestions and improvements from these tools. The existing tools are measures in terms of accuracy that how accurately they performed the analysis of sentiment, the complexity in these tools that a user can found and the mining capability that on to which extent these tools can mine the text. After the analysis they have proposed the mining model for useful reviewing system on automatic basis. They have done this by using SRL and made the feature and sentimental lexicon for the text of these reviews.

Raymond et al. worked on the novel based analysis on the opinion-based mining and the bearing of opinion. They have taken the pairs of opinion from the selected datasets and extracted the required data for the suitable opinion mining. The mining is done on the basis of entropy weightage, which is the calculations of the opinions. The opinions are combined on the basis of part of speech, semantics, position of vectors and part

of speech. This method is efficient for selecting the suitable opinions and targeting the opinions which should be removed [22].

Swaminathan et al. presented an extracting relationships method between the bio-entities such as foods and diseases. Author proposed three level for extraction this type of relation which are weak, medium and strong. In their previous work they have classified the polarities into major categories, which are polarity with no relation, neutral polarity, positive and negative polarities. The also worked on concatenating the semantic and lexicon feature of the datasets to identify lexicon-based features. The new proposed feature is combined with the previously available features to efficiently analyse the opinion-based mining system with the polarity value ( $\sim 0.91$ ). But this is not the surety about the strength and accuracy of the system to consider only polarity value ( $\sim 0.96$ ) [23].

## Methods

### Proposed system

Education data mining is widely being used application to improve the quality of the education. In this research work the data of government primary schools of various districts in Punjab is collected. The data is about the attributes of three entities which are student, teacher and school. The data is collected in raw form but organized in the Microsoft Excel sheets. This collected data is analyzed and sorted on the basis of required attributes. The attributes are arranged and some attributes are derived from two or three originally available attributes [24-28]. The data consists of male and female primary school of either rural areas or urban areas. This data is analyzed by the help of data mining or analysis tool named as Rapid Miner, it is very effective and productive tool that with the short span of time large amount of data can be processed. The data is processed and predicted in by making the retention attribute as special attribute on which results are to be made and prediction with the help of four different algorithms separately to analyze the performance and accuracy of each algorithm [29-35]. The uploaded data with attributes in rapid miner is given in **Tables 1 and 2**.

**Table 1:** Imported data in rapid miner (a).

Roll Number	Name of Student	Gender	Age of Student	Attendance	Retention	marks	T.Gender
366407	ABIDA NOREEN	Female	14	80	Good	68	Female
388730	ADILA SHAHZADI	Female	12	83	Good	67	Female
360399	AFZAL ZATOON	Female	13	84	Good	54	Female
82375	GHAZALA SHAHEEN ZAKAULLAH	Female	11	83	Good	55	Female
369508	Hafiza Beenish	Female	12	84	Good	58	Female
392994	HAFIZA MAMONA MEHRAM	Female	12	85	Good	62	Female
442161	MUHAMMAD ABU BAKAR	Male	11	90	Good	74	Female

456427	MUHAMMAD SHAFI	Male	15	91	Good	85	Female
360067	MUHAMMAD YOUSAF	Male	12	80	Good	59	Female
387742	MUSA YOUSAF	Male	13	45	Good	45	Female
113432	NASRIN FATIMA	Female	7	71	Good	65	Female
393832	RABIA KALSOOM	Female	8	85	Good	62	Female
360513	SAJJAD HAIDER SHAH	Male	5	89	Good	63	Female
82388	SAMINA YASMIN	Female	9	87	Good	41	Male
82379	SHAHNAZ MALIK	Female	5	89	Good	45	Female
412161	Shamsha Mushtaq	Female	12	89	Good	78	Female
126598	SHIRHEEN ARSHAD RAZA	Female	14	89	Good	75	Female
145626	TARIQ MEHMOOD	Male	15	86	Good	74	Male
416058	TUQEER AHMAD	Male	17	83	Good	71	Male

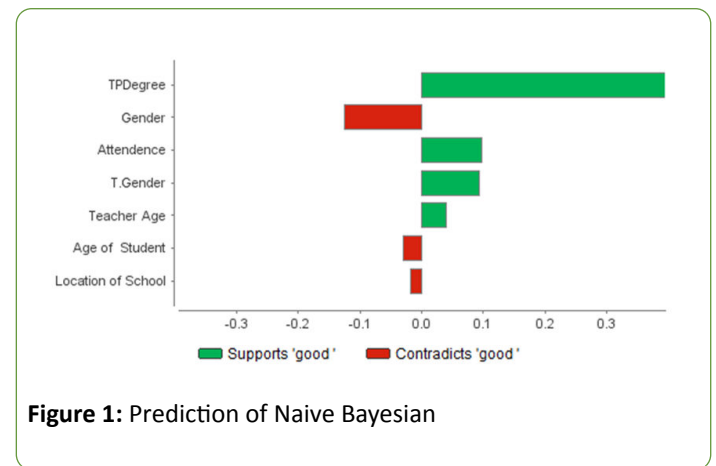
Table 2: Imported data in rapid miner (b).

T.Gender	Teacher Age	TP Degree	Location of School
Female	41	y	rur
Female	35	y	ur
Female	41	n	rur
Female	26	y	ur
Female	55	n	ur
Female	32	n	ur
Female	28	y	rur
Female	25	y	rur
Female	35	y	ur
Female	24	y	rur
Female	56	n	ur
Female	48	n	ur
Female	43	n	ur
Male	36	y	rur
Female	28	y	rur
Female	29	y	ur
Female	36	n	ur
Male	45	n	ur
Male	58	n	ur

## Results and Discussion

After the data uploading process a model for prediction is designed and the required algorithms are selected. The algorithms are Naïve Bayesian, Random forest, support vector machine and deep learning. All of these selected algorithms are suitable for textual mining and prediction. The results are created using the rapid minder tool and results for each algorithm is taken [36-40].

Naive Bayesian: Prediction of Naive Bayesian is shown in Figure 1.



Random Forest: Prediction of Random Forest is shown in Figure 2.

### Important Factors for good

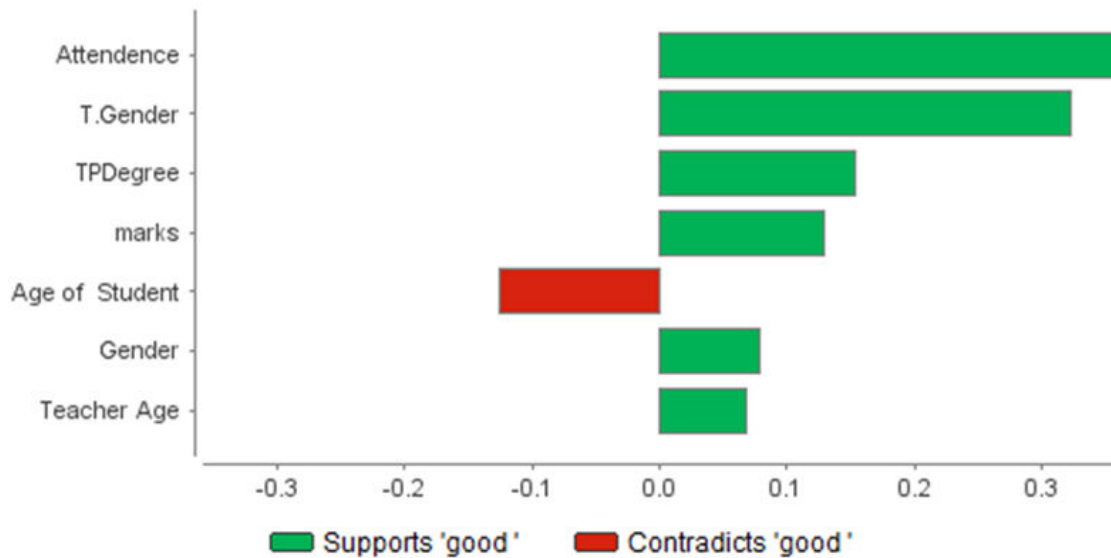


Figure 2: Prediction of Random forest.

Support Vector Machine: Prediction of Support Vector Machine is shown in Figure 3.

### Important Factors for good

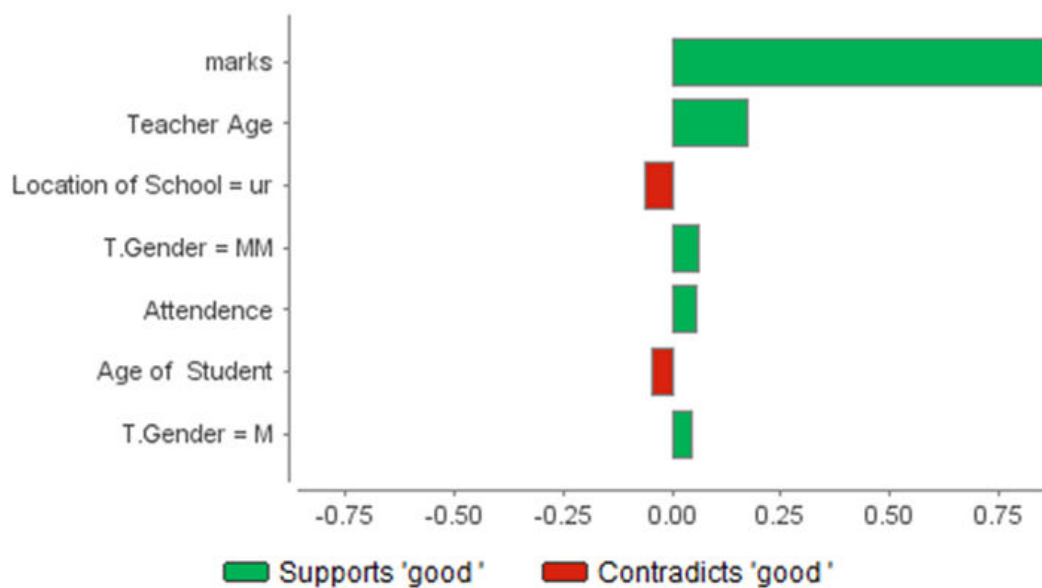
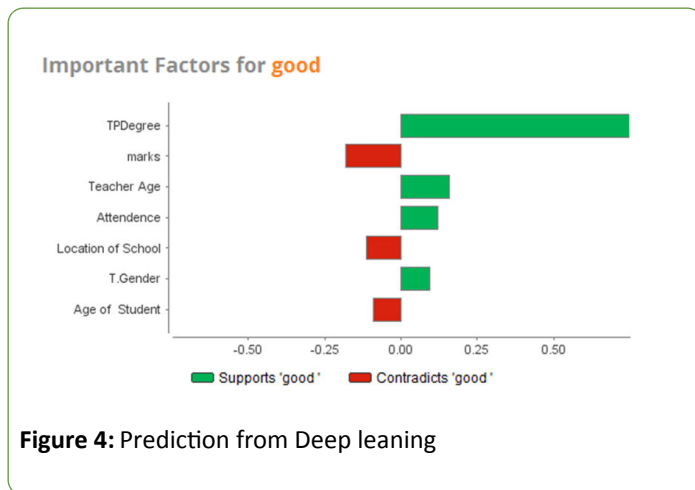


Figure 3: Prediction of SVM

Random Forest: Prediction of Random Forest is shown in Figure 4.



## Comparison and Accuracy of selected Algorithms

Comparison and Accuracy of selected Algorithms is shown in Tables 3 and 4.

**Table 3:** Relevant features in prediction of each algorithms and not supported features

Name of Algorithm	Supported Attributes	Not supported Attributes
Naive Bayesian	T.P Degree, Attendance, T. Gender, T. Age	Gender of student, location of schools, age of student
Random Forest	Attendance, T. Gender, T.P Degree, Marks	Age of the student
Support Vector Machine	Marks, T. Age, T. Gender, Attendance, Gender of student	Location of the school, age of the student
Deep Learning	T.P Degree, T. Age, Attendance, T. Gender	Marks, location of the school, age of the student.

**Table 4:** Accuracy level of each algorithm

Name of Algorithms	Accuracy
Naive Bayesian	75%
Random Forest	67%
Support Vector Machine	73%
Deep Learning	77%

## Conclusion

Education data mining is adaptive application which is widely being used in education sector to improve the quality of education by analyzing and predicting the performance of student on large scale. In this research paper the data of students, teachers and schools of government primary school is collected from various districts of Punjab either in rural areas or urban areas of both male and female schools. The data is sorted and only required features related to the teacher, student and school are selected. The features contain round about eight hundred entries and same entries for the testing. The data is analyzed and predicted using the rapid miner tools by applying the four algorithms. The algorithms are Naive Bayesian, random forest, support vector machine and deep learning. From these algorithms, the results are produced from the results it can be observed that the teacher degree, age and attendance of the student are greatly impact on the performance of the student which is retention in this case. On factors of the institutes can be analyzed such as parent attributes or class room attributes in future as a future work.

## References

1. Aldowah H, Al-Samarraie H, Fauzy MW (2019) Educational Data Mining and Learning Analytics for 21stcentury higher education: A Review and Synthesis. *Telematics and Informatics* 37: 13-49.
2. Akingbehin, Kiumi (2005) Taguchi-based metrics for software quality. Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05). IEEE.
3. Khasanah UA, Harwati (2017) A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques. *IOP Conf Ser Mater Sci Eng* 215:1.
4. Pedro s, Cruz L, Soares C, Moreira JM, Abreu R (2015) A Comparative Study of Classification and Regression Algorithms for Modelling Students Academic Performance. *Int Edu Data Mining Soc* 1-4.
5. Baradwaj, Brijesh Kumar, Pal S (2011) Mining educational data to analyze students' performance. *Int J Adv Computer Sci App* 2: 63-69.
6. Rafique, Ayesha (2019) Sentiment Analysis for Roman Urdu. Mehran University. *ResJ Eng Tech* 38: 463.
7. Dwivedi A (2019) Benefits, of using big data sentiment analysis and soft computing techniques in E-governance. *Int J Recent Technol Eng* 8: 3038-3044.
8. Masud F Kundi, Ahmad S, Khan A, Zubair M Asghar (2014) Detection and scoring of internet slangs for sentiment analysis using SentiWordNet. *Life Sci J* 11: 66-72.
9. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Association for Computational Linguistics 79-86.

10. Mehmood, Asad, S. Palli A, Khan M. N. A (2014) A study of sentiment and trend analysis techniques for social media content. *I.J. Modern Education and Computer Science. Int J Modern Edu Computer Sci* 12: 47-54.
11. Buntoro, Asrofi GB, Bharata Adji T, Erna A Permanasari (2016) Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute. *Int J Computer App* 136: 23-29.
12. Alam MA, Daiyan M, Kumar M, Tiwari SK (2015) A literature review on opinion mining and sentiment analysis. *Int JEmerging Technol Advanced Engineering* 5: 262-280.
13. Bilal M, Israr H, Shahid M, Khan A (2016) Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *J King Saud University Computer Information Sci* 28: 330-344.
14. Syed AZ, Aslam M, Martinez-Enriquez AM (2010) Lexicon based sentiment analysis of Urdu text using SentiUnits. *Mexican International Conference on Artificial Intelligence. Springer, Berlin, Heidelberg* 4: 32-43.
15. Korayem M, Crandall D, Abdul-Mageed M (2012) Subjectivity and sentiment analysis of arabic: A survey. *International conference on advanced machine learning technologies and applications. Springer, Berlin, Heidelberg* 128-139.
16. Rehman ZU, Bajwa IS (2016) Lexicon-based sentiment analysis for Urdu language. *2016 sixth international conference on innovative computing technology (INTECH) IEEE*.
17. Ain QT, Ali M, Riaz A, Noureen A, Kamran M (2017) Sentiment analysis using deep learning techniques: a review. *Int J Adv Comput Sci Appl* 8: 424-433.
18. Anwar AS Hridoy, Tahmid M Ekram, Samiul M Islam, Ahmed F, Rahman M Rashedur (2015) Localized twitter opinion mining using sentiment analysis. *Decis Anal* 2: 1-19.
19. Khan HF, Qamar U, Younus M Javed (2014) Sentiview: A visual sentiment analysis framework. *International conference on information technology (i-society 2014)* 5: 291-296.
20. Laudal A Olav, Piene R (2004) *The Legacy of Niels Henrik Abel. The Legacy of Niels Henrik Abel. Springer, Berlin, Heidelberg* 1-3.
21. Peldszus A, Stede M, Potsdam U (2013) From argument diagrams to argumentation mining in texts: A survey. *Int J Cognitive Informatics Natural Intelligence (IJCINI)* 7: 1-31.
22. Raymond YK Lau, Li C, Stephen SY Liao (2014) Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis. *Decision Support Systems* 65: 80-94.
23. Shankar S Dasgupta, Swaminathan N, Kaipa KK, Kumar S Bhattacharjee, Viswanathan A (2015) Sentiment analysis of Facebook data using Hadoop based open source technologies. *International Conference on Data Science and Advanced Analytics (DSAA) IEEE*.
24. Samuel A, Ahmad NBB, Shamsuddin SM (2019) Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods. *IOP Conf Ser Mater Sci Eng* 10:1-5.
25. George S, Kyatanavar DN (2012) Applications of fuzzy logic in sugar industries: A review. *Int J Engineering Innovative Technol* 1: 226-231.
26. Chen J Chien, Hsu LI, Chih-Hao W, Wei-Liang S, Yi-Hsiang H et al. (2005) Biomarkers of exposure, effect, and susceptibility of arsenic-induced health hazards in Tai wan. *Toxicol applied pharmacol* 206: 198-206.
27. Xiang W, Kong J, Chen S, Zheng W, Liu W (2005) Cardiac hypertrophy in vitamin D receptor knockout mice: role of the systemic and cardiac renin-angiotensin systems. *Am J Physiol Endocrinol Metab* 288: E125-E132.
28. Quiceno-Manrique AF, Godino-Llorente JI, Velasco MB, Dominguez GC (2010) Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic signals. *Anna Biomed Eng* 38: 118-137.
29. Mondal A, Bhattacharya P, Saha G (2013) An automated tool for localization of heart sound components S1, S2, S3 and S4 in pulmonary sounds using Hilbert transform and Heron’s formula. *SpringerPlus* 2: 512.
30. Pedrosa J, Castro A, Vinhoza TTV (2014) Automatic heart sound segmentation and murmur detection in pediatric phonocardiograms. *Conf Proc IEEE Eng Med Biol Soc* 10:2294-2297.
31. Jeyarani AD, Jaya ST (2010) Analysis of noise reduction techniques on QRS ECG waveform-by applying different filters. *Recent Advances in Space Technology Services and Climate Change RSTS & CC-2010 IEEE* 149-152.
32. Ganguly A, Sharma M (2017) Detection of pathological heart murmurs by feature extraction of phonocardiogram signals. *J Appl Adv Res* 2: 200-205.
33. Xue Y (2011) Reengineering legacy software products into software product line based on automatic variability analysis. *Proceedings of the 33rd International Conference on Software Engineering* 1114-1117.
34. Kumar AD, Radhika V (2014) A survey on predicting student performance. *Int J Computer Sci Information Technol* 5: 6147-6149.
35. Rao, Srinivasa K (1991) Computers in Education. *Physics Education (India)* 7: 315-19.
36. Wayne JW, Bhargava A, Gramopadhye AK, Madathil KC, Babu SV (2016) Examining the effects of interaction fidelity on task performance and learning in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 24: 1418 – 1427.
37. Gardner J, Brooks C (2018) Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction* 28: 127-203.
38. Sajid A, Ahmad T, Khalid T (2015) Stress in medical undergraduates; its association with academic performance. *Bangladesh Journal of Medical Science* 14: 135-141.
39. Hussain S, Dahan NA, Ba-Alwib FM, Ribata N (2018) Educational data mining and analysis of students’ academic performance using WEKA. *Indonesian J Elec Eng & Comp Sci* 9: 447-459.
40. Vinodhini G, Chandrasekaran RM (2012) Sentiment analysis and opinion mining: a survey. *Int J Advanced Res Computer Scif Software Engg* 2: 282-292.