

Classification and Prediction of Student Academic Performance using Machine Learning: A Review

Zeba Parveen^{1*}, Mohatesham Pasha Quadri²

^{1,2}Dept. of Computer Science and Information Technology, Maulana Azad National Urdu University, Hyderabad, India

*Corresponding Author: zebapar618@gmail.com,

DOI: <https://doi.org/10.26438/ijcse/v7i3.607614> | Available online at: www.ijcseonline.org

Accepted: 12/Mar/2019, Published: 31/Mar/2019

Abstract— Today, as education is very important for all human being, so it is necessary to analyze and improve the education system as technologies growing day by day, so use of latest technologies is very crucial to enhance the education system and academic performance of the student. Many researchers have been worked on predicting student performance and built predictive models to measure and predict students' performance and found interesting results. This classification presents a review of works previously done by different authors on student performance by using different techniques. The aim of this work is to review the available study, to compare different models developed by different authors accordingly and to find out the best model from it. This study shows how different techniques used and produces result and which is best suitable technique. The various factors identified with the representation of machine learning algorithms based on methods and tools followed by their attributes and results respectively. This can help students, faculties, and institutions to increase the performance.

Keywords—Student Performance, Machine learning algorithms, Tools

I. INTRODUCTION

Machine Learning is very popular and most used in every field. It is given more concentration in educational field. Machine learning is broad area of research. Many researches have used machine learning techniques for measuring, predicting and improving students' performance in academic year.

To measure and predict performance of higher education students is very challenging task for researchers. Many researches have proposed and going on in this field because it is a field where improvement is needed time by time, for this it is crucial to analyze previous performance of the students so that their performance can be predicted in future. For this purpose, researchers have utilized the student data from university database in their study based upon identified attributes/ factors that affect a student performance and by using different techniques of machine learning and data mining, they have predicted students' performance in upcoming year which may be useful and helpful for the students as well as faculties to improve student grades in academics.

There are different factors identified by researchers that affect a student performance like-personal, psychological, cognitive, personality, economic, demographic, institutional

etc. they have applied different machine learning and data mining techniques such as- naïve Bayes, SVM, Random forest, Neural network, etc. It has been stated that the Naïve Bayes machine learning algorithm to predict and identify slow learners is most accurate [16, 23].

Rest of the paper is divided into three sections in which section-II represents literature survey of related works, section-III shows the comparative study of previous researches and section-IV defines conclusion and future work of this study.

II. RELATED WORK

Amjad Abu Saa conducted a research on a group of students enrolled in different courses. He constructed a survey and collected data from students using google form and their daily class. He used different classification techniques to measure students' performance in previous semester and predicted their grade at the end of semester. He used WEKA and Rapid Miner tools for implementation. He compared Naïve Bayes and decision tree algorithms and found that CART DT algorithm has best accuracy than others [1].

Brijesh Kumar Bhardwaj and Saurabh Pal conducted a study on BCA students for improvement of their performance. They collected data from 300 students, in which 226 were male students and 74 were female students, through

questionnaire and university database. They used Naïve Bayes classification algorithm to design prediction model and MATLAB tool to implement it. At last they found that student performance does not only depend on their academic results but it is also affected by other factors such as- social, personal and environmental [2].

Augusto Sandoval compared three different models for predicting students' performance. He collected information from 21314 UG students of three semesters from second semester of 2013 to second semester of 2014. After comparing the three models Random Forest (RF), Linear Regression (LR) and Robust Linear Regression models it was found that Random Forest gives best performance over others [3].

Joannis E. Livieris used semi-supervised ML techniques for predicting students' performance. They had collected information of 5 years from 3716 students studying Mathematics at secondary schools. The code for implementation was written in JAVA using WEKA ML toolkit [4].

Evandro B. Costa presented a comparative study on effectiveness of educational DM techniques. The purpose of this study was to early detect those students who were going to fail with their precision. For this purpose, they collected students' data from two independent sources like-distance education and on-campus. In this study, Pentaho Data Integration Tool was used for preprocessing the data and WEKA tool was used to apply EDM techniques [5].

Parneet Kaur et al. predicted slow learners using DM techniques. They have taken 152 students of high school as their dataset. They compared Multilayer Perception (MLP), Naïve Bayes (NB), SMO, J48 and REPTree algorithms using WEKA tool. At the end the result shown that MLP technique performs best [6].

Shiwani Rana and Roopali Garg had implemented a modified Naïve Bayes algorithm to predict slow learners. They conducted their study on 60 students of BE (3rd semester) of Punjab University. They have compared Naïve Bayes (NB) and Modified Naïve Bayes (MNB) algorithms and found that the MNB algorithm gives more accuracy [7].

A. T. M. Shakil Ahamed et al. made a comparative study to predict students' grade. They have taken 423 students of HSC as their data sample. They compared three ML algorithms namely- ANN, k-NN and SVM in Rapid Miner. At the end, it was observed that ANN had best accuracy over others [8].

Ahmed Mueen et al. did a study and predicted students' performance using Data Mining (DM) techniques. They collected data from undergraduate students of PF and AOS

courses during period August 2014 - May 2015. In this study, they compared three different classification models NB, NN, and DT in WEKA tool. It was observed that Naïve Bayes performed better on other two [9].

Fadhilah Ahmad et al. proposed a model for predicting performance of first year degree students of computer science (CS). They collected students' data of 8 years (July 2006-07 till July 2013-14). The classification was done to the students' data based on the following different algorithms- Naïve Bayes (NB) Decision Tree (DT), and Rule-Based (RB), in which, Rule-Based was best model among the other techniques. The implementation was performed in WEKA [10].

Jie Xu et al. conducted a research on student performance. They proposed a novel ML method for predicting student performance in undergraduate courses. They collected data for their study from mechanical and aerospace engineering students of three years (2013, 2014, 2015). They compared four different techniques of machine learning for prediction that are- linear regression, logistic regression, random forest and k-NN, from which they came across with the result that random forest performed better in most cases [11].

M. Mayilvaganan and D. Kalpanadevi applied some classification techniques on the student data to predict their performance. They compared different techniques of machine learning like- C4.5, AODE, naïve Bayes and multi labeled k-Nearest Neighbor. They collected 197 students' data from the department of commerce and department of computer science and engineering. They found that multi labeled k-nearest neighbor had best accuracy over all other techniques used in the study. The experimentation was done in WEKA tool [12].

M. Durairaj and C. Vijitha used k-means clustering algorithm for student performance prediction. They have used 38 students record from 300 records of Bharathidasan university for their study and compared two techniques-naïve Bayes and decision tree algorithms and found that NB algorithm produced best and accurate result than the other. The implementation work in this study was done in WEKA tool [13].

S.Taruna and Mrinal Pandey made a comparative study on academic performance prediction and compared five different classification algorithms named as- Decision Tree(DT), Naïve Bayes(NB), Naïve Bayes Tree (NB Tree), k-Nearest Neighbor (k-NN) and Bayesian Network (Bayes Net) algorithms using WEKA toolkit. They have collected students' data from an engineering university consisting 1000 instances and 18 attributes from which 7 attributes were removed and only 11 attributes were used for this study. In conclusion it was observed that J48 decision tree has improved its accuracy in this study than previous researches

and is the best classifier for predicting student grade in future [14].

Olugbenga Adejo and Thomas Connolly proposed a framework for early detection and prediction of slow performer students of higher education institutions. They have identified six most influencing variable domains such as- cognitive, psychological, economical, personality, demographic and institutional domains. They concluded that this conceptual framework will provide an improved and enhanced model that will be able to predict and detect slow performers with high accuracy, efficiency and robustness [15].

Maria Koutina and Katia Lida Kermanidis conducted a study on PG students of Ionian University Informatics to predict grades in final year. They have applied six different classification techniques on students' data to find out best classifier. They collected the demographic, in-term performance and in-class behaviors of 117 students of different subjects. The machine learning techniques used for this study were- J48, k-NN, Naive Bayes, Random Forest, SMO and J-Rip. The implementation was performed in WEKA tool. At the end they observed that the other attributes did not affect student performance but the presence in the class highly influence and can improve their final grade [16].

Ashkan et al. introduced a model for predicting SAP of engineering students. They gathered information for their study from university database of UIC (University of Illinois at Chicago). The size of the dataset was 300 and UG (engineering) students of the university were samples for their study. The goal of this study was to predict students' grades in three different courses and they found the result that the proposed model improved the grades of the students. The ML technique used in this study was Bayesian Network. They compared three proposed model with other previous models and concluded that BN performed well and was more efficient and accurate to predict student grades than other models [17].

Paulo Cortez and Alice Silva performed a study on secondary school students of Portugal to determine the student achievement, dropout rate and failing ratio in two core subjects- mathematics and Portuguese. They used four data mining (DM) techniques- Decision Tree (DT), Neural Network (NN), Random Forest (RF) and Support Vector Machine (SVM) to build and evaluate the model. They considered data during the year 2005-2006 of two public schools of Portugal. By applying data mining techniques, they predicted student grades in secondary schools and also found that previous performances of the students highly affect the student achievement progress [18].

Qasem et al. worked for the improvement of the quality of education in higher education students. This study was done on UG students of computer science of Yarmouk university and data was collected from students through questionnaire. Three classification algorithms like- ID3, C4.5 and Naive Bayes (NB) were applied on collected data in WEKA tool. The classification accuracy of this model was very low because of very small dataset. It was noticed that the classification accuracy can be improved using large sample size [19].

M. Ramaswami and R. Bhaskaran identified different factors affecting a student performance and developed a model to predict their performance using CHAID algorithm. A total 1000 records of students were collected from five different schools. The CHAID algorithm was applied on students' data and performance of the students was predicted. The result obtained from this model was satisfactory but not enough good. It seems that by applying another technique, the accuracy can be improved. The prediction accuracy of this model was 44.69% [20].

Ali Daud et al. presented a study on graduate and undergraduate students of engineering to forecast their performance using advanced learning analytics. They collected information from different universities of Pakistan of the year 2004 to 2011. This study was conducted with the aim that the student will complete his/her degree course or drop. For this purpose, they applied five different classification algorithms that are- BN, NB, SVM, C4.5 and CART and concluded that their proposed features (i.e; student personal and his family expenditure information) have effect on student performance and SVM is more suitable for this model [21].

Shaobo Huang and Ning Fang constructed four different mathematical models for predicting students' average academic performance. The data was collected from undergraduate students of four semesters. The purpose of this study was to find out low performer students and to improve their performance. This study is helpful for poor performer students to improve their learning and also for the good performer students to encourage them to enhance their learning and understanding capabilities. Among the four predictive models developed, the SVM model was best achieving accuracy 90.1% [22].

S. Kotsiantis et al. predicted student performance in distance mode (DM) education. They used student demographic information and marks in assignments as attributes and gathered information from Hellenic Open University. The experiment in this study was done using six machine learning algorithms that are- C4.5, back propagation (BP), NB, 3NN, logistic regression and SMO. The all six algorithms were tested and compared and it was found that there were no significant differences among the four algorithms i.e; NB,

logistic regression, BP and SMO and it was observed that NB was most appropriate algorithm having best accuracy over others. The aim of this study was to find the more appropriate and suitable algorithm for predicting student performance as well as providing supporting tool for educational tutors [23].

Reynold A. Rustia conducted a study on students appearing in board examination. They used various data mining techniques to build the prediction model to measure student performance in licensure examination for teachers (LET). They have used several DM techniques like- NN, SVM, C4.5, NB and logistic regression and observed that C4.5 was most suitable algorithm. This model can help to identify the students expecting to fail in LET and they will be given higher priority in their mock board review so that they can be able to perform better in board examination [24].

III. COMPARATIVE STUDY

A. Algorithms

This section describes comparative study of various model, techniques and tools used in previous researches to determine and forecast learner’s performance. The most widely used data mining and machine learning techniques used in this field are naïve Bayes (NB), decision tree (DT), neural network (NN), support vector machine (SVM) and k-nearest neighbor (k-NN) algorithms. The objective of all the studies is to measure students’ performance in previous courses and based upon that their grades/performance in upcoming courses can be predicted and necessary decisions can be made to improve their learning capabilities. There are techniques too that were applied by other researchers in their studies like- linear regression [11], [22], logistic regression [11], [23], [24], random forest [11], [16], rule-based classification [10], Bayesian network [14], [17], [21], radial basis function (RBF) [22] that also produce satisfactory results.

The algorithms depicted through the following diagram are used by different authors and some of them that are mostly used and produce best models are defined through tables.

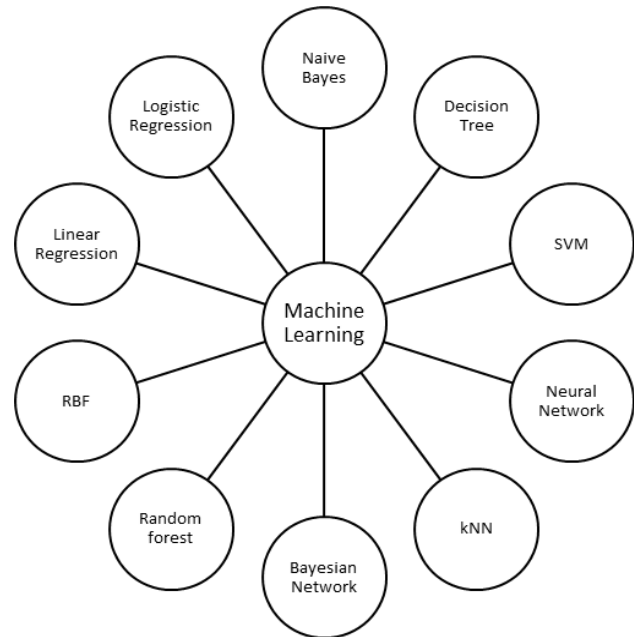


Figure 1. Machine Learning algorithms

Tabular representation of algorithms

The most used algorithms applied by various authors are defined in below tables.

1) Naïve Bayes Algorithm

Table1. Percentage accuracy of naïve Bayes algorithm

Method	Attributes	Results	Reference
Naive Bayes	Personal, Social and Academic	36.40%	(Saa, 2016) [1]
	Personal, Academic	65.13%	(Kaur, 2015) [6]
	General, Forum and Academic info.	85.7%	(Mueen, 2016) [9]
	Personal, Academic and Course related	67.0%	(Ahmad, 2015) [10]
	Academic and Demographic	84.37%	(S.Taruna, 2014) [14]
	Personal, Family related	84.8%	(Daud, 2017) [21]
	Demographic, Marks in	72.48%	(Kotsiantis, 2004) [23]

	Assignments		
	Subject Areas and LET result	62.98%	(Rustia, 2018) [24]

Naive Bayes algorithm is used by many authors to build the predictive model and to forecast student performance. The models defined in the above table shows that there are three models that has given satisfactory accuracy and performance which are 84.37% (S.Taruna, 2014), 85.7% (Mueen, 2016) and 84.8% (Daud, 2017) among which the best model and best accuracy is 85.7%.

2) Decision Tree Algorithm

Table 2. Percentage accuracy of decision tree algorithm

Method	Attributes	Results	Reference
Decision Tree	Personal, Academic	69.73%	(Kaur, 2015) [6]
	General, Forum and Academic info.	80.5% (C4.5)	(Mueen, 2016) [9]
	Academic and Demographic	95.49%	(S.Taruna, 2014) [14]
	Personal, Social, Psychological and Academic	44.69% (CHAID)	(Ramaswami, 2010) [20]
	Personal, Family related	76.6% (C4.5) 71% (CART)	(Daud, 2017) [21]
	Demographic, Marks in Assignments	69.99% (C4.5)	(Kotsiantis, 2004) [23]
	Subject Areas and LET result	73.10% (C4.5)	(Rustia, 2018) [24]

The above table shows the decision tree model for performance prediction. The accuracies 80.5% (Mueen,2016) and 95.49% (S.Taruna, 2014) are higher than other models. Other models also produce good results but 95.49% accuracy shows the best accuracy.

3) Neural Network

Table 3. Percentage accuracy of neural network algorithm

Method	Attributes	Results	Reference
Neural Network	Personal, Academic	75% (MLP)	(Kaur, 2015) [6]
	Demographic, Psychological and Academic	86.11% (ANN)	(A.T.M.Shakil Ahmad, 2017) [8]
	General, Forum and Academic info.	81.4% (MLP)	(Mueen, 2016) [9]
	Mid Exam Scores, Final Exam Score	89.6% (MLP)	(Huang, 2013) [22]
	Demographic, Marks in Assignments	72.26% (BP)	(Kotsiantis, 2004) [23]
	Subject Areas and LET result	65.67%	(Rustia, 2018) [24]
	SS Grade, Living Location, Med. of Teaching	70%	(Agrawal, 2015) [26]

The above table presents neural network model developed by various authors. The higher accuracies of models are- 89.6% (Huang, 2013), 81.4% (Mueen, 2016) and 86.11% (A.T.M.Shakil Ahmad, 2017). Form the result it can be observed that the model constructed by Huang with scores in the exam is the best model having accuracy of 89.6%.

4) Support Vector Machine

Table 4. Percentage accuracy of SVM algorithm

Method	Attributes	Results	Reference
Support Vector Machine	Demographic, Psychological and Academic	77.88%	(A.T.M.Shakil Ahmad, 2017) [8]
	Personal, Family related	86.7%	(Daud, 2017) [21]
	Mid Exam Scores, Final Exam Score	90.1%	(Huang, 2013) [22]

	Demographic, Marks in Assignments	72.17% (SMO)	(Kotsiantis,2004) [23]
	Subject Areas and LET result	61.89%	(Rustia, 2018) [24]

In the above table SVM model of performance prediction and their accuracies are defined. Result shows that 90.1% (Huang,2013) accuracy is the best accuracy over all other models defined in the table in which examination scores were used as attributes.

5) *k*-Nearest Neighbor Algorithm

Table 5. Percentage accuracy of *k*-NN algorithm

Method	Attributes	Results	Reference
K-Nearest Neighbor	Demographic, Psychological and Academic	69.31%	(A.T.M.Shakil Ahmad, 2017) [8]
	Demographic, Marks in Assignments	66.93% (3NN)	(Kotsiantis, 2004) [23]

This table represents *k*-NN model for prediction in which we find that *k*-NN model don't produce satisfactory results. The accuracy of the model is only 69.31% which is not enough good. It needs more improvement.

Algorithms with best accuracy

Table 6. The best percentage accuracy of above algorithms

Methods	Results	Reference
Naïve Bayes	85.7%	(Mueen, 2016) [9]
Decision Tree	95.49%	(S.Taruna, 2014) [14]
Neural Network	89.6% (MLP)	(Huang, 2013) [22]
Support Vector Machine	90.1%	(Huang, 2013) [22]
k-Nearest Neighbor	69.31%	(A.T.M.Shakil Ahmad, 2017) [8]

In the above table we have compared best accuracies of all the above models and it is found that decision tree (DT) model presented by S. Taruna in 2014 is best prediction model achieving the best accuracy of 95.49%, and in which academic and demographic information of the students are used which was implemented in WEKA toolkit.

B. Attributes

In this section, the most used and influencing attributes on the students' academic performance are identified by comparing several research studies. The below attributes are identified as mostly and commonly used attributes in previous studies-

- Age
- Sex
- Parent's education
- Parent's occupation
- Family income
- Time spent on study
- Attendance
- Resources (Internet, Lab, Library)
- Examination grade

The above defined attributes are mostly used by many researchers and are influencing factors in academics. As the student age and sex affect their performance in the way that male students don't feel comfortable with female students for combined studies and vice versa. Also, parent's education, parent's occupation and family monthly/annual income are important factors to affect students' performance because if the parents are literate and are not busy in their jobs then they can give time to their children and can help them in their studies. The student how much time spent in learning, how much classes he attends and what resources are available at his institution are also the factors that influence a student learning. And the last and highly influencing factor is his examination grades, according to the grade scored in the examination he can make preparation for next exams for improving his grade.

C. Tools

The tools used in previous studies are described in this section. There are four important tools that were used which are-WEKA, MATLAB, Python and RapidMiner. Weka is the tool which was used in most of the studies discussed above and produced satisfactory results. MATLAB [2] is rarely

used because it is not free. Python [7] is also used and its performance accuracy is about 70% which not enough good and the accuracy found by RapidMiner [1], [8] is about 75%-85% that is good but also can be to be improved.

D. Observation and Analysis

From the above reviews it is observed that the students' performance can be enhanced by applying different machine learning techniques on their data. It was found that the factors that highly influence their performance is the student own effort in academics. Academic data can include student personal inadequacy, Fear of examination, Classroom environment, Time management, Student-Teacher relationship, Teaching-Learning methods, Inadequate study facilities etc.

IV. CONCLUSION AND FUTURE WORK

This review paper presents comparison of different prediction models for academic performance developed by various authors. The aim of the present study is to find important factors/attributes that influence a student performance, to compare different methods and draw the best method from them, to discover best prediction model for student performance.

At the end of the study we found interesting results that can be helpful for students, teachers as well as institutions for the improvement of student performance in academics. It is observed that the best algorithm for building prediction model is decision tree and best suited tool is WEKA that is freely and easily available for students and best attributes are student personal, family-related and academic info. In conclusion we can say that by using the above techniques, tools and attributes we can build a best prediction model to increase performance and to enhance learning and education systems.

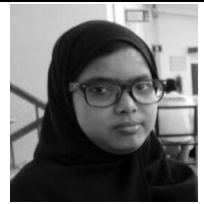
REFERENCES

- [1] Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications(IJACSA), Vol. 7, No. 5, 2016.
- [2] Brijesh Kumar Bhardwaj and Saurabh Pal, "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security(IJCSIS), Vol. 9, No. 4, 2011.
- [3] Augusto Sandoval, Carlos Gonzalez, Rosa Alarcon, Karim Pichara, Maximiliano Montenegro, "Centralized student performance prediction in large courses based on low-cost variables in an institutional context", The Internet and Higher Education, Vol. 37, pp.76-89, 2018, DOI:10.1016/j.iheduc.2018.02.002.
- [4] Ioannis E. Livieris, Konstantina Drakopoulou, Vassilis T. Tampakas, Tassos A. Mikropoulos, Panagiotis Pintelas, "Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach", Journal of Educational Computing Research, 2018, DOI: 10.1177/0735633117752614.
- [5] Evandro B. Costa, Balduino Fonseca, Marcelo Almeida Santana, Fabr sia Ferreira de Ara jo, Joilson Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses", Computers in Human Behavior, Vol.73, pp. 247-256, 2017, doi:10.1016/j.chb.2017.01.047.
- [6] Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector", 3rd International Conference on Recent Trends in Computing(ICRTC-2015), Vol.57, pp. 500-508, 2015.
- [7] Shiwani Rana, Roopali Garg, "Slow Learner Prediction using Multi-Variate Na ve Bayes Classification Algorithm", International Journal of Engineering and Technology Innovation, Vol. 7, No. 1, pp. 11-23, 2017.
- [8] A. T. M. Shakil Ahamed, Navid Tanzeem Mahmood, Rashedur M Rahman, "An intelligent system to predict academic performance based on different factors during adolescence", Journal of Information and Telecommunication, Vol.1, No.2, pp.155-175, 2017.
- [9] Ahmed Mueen, Bassam Zafar, Umar Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques", I.J. Modern Education and Computer Science, Vol.11, pp. 36-42, 2016.
- [10] Fadhilah Ahmad, Nur Hafieza Ismail, Azwa Abdul Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques", Applied Mathematical Sciences, Vol. 9, No. 129, pp. 6415 - 6426, 2015.
- [11] Jie Xu, Kyeong Ho Moon, Mihaela van der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs", IEEE Journal of Selected Topics in Signal Processing, Vol.11, No.5, pp. 742-753, 2017.
- [12] M. Mayilvaganan, D. Kalpanadevi, "Comparison of Classification Techniques for predicting the performance of Students Academic Environment", International Conference on Communication and Network Technologies (ICCNT). IEEE, pp.113-118, 2014, doi:10.1109/CNT.2014.7062736.
- [13] M. Durairaj, C. Vijitha, "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, pp.5987-5991, 2014.
- [14] S.Taruna, Mrinal Pandey, "An Empirical Analysis of Classification Techniques for Predicting Academic Performance", 2014 IEEE International Advance Computing Conference (IACC), pp. 523-528, 2014, doi: 10.1109/IAdCC.2014.6779379.
- [15] Olugbenga Adejo, Thomas Connolly, "An Integrated System Framework For Predicting Students' Academic Performance In Higher Educational Institutions", International Journal of Computer Science & Information Technology (IJCSIT), Vol 9, No. 3, 2017.
- [16] Maria Koutina, Katia Lida Kermanidis, "Predicting Postgraduate Students' Performance Using Machine Learning Techniques", Springer, Berlin, Heidelberg, Vol. 364, pp.159-168, 2011.
- [17] Ashkan Sharabiani, Fazle Karim, Anoo Shiravan Sharabiani, Mariya Atanasov, Houshang Darabi, "An Enhanced Bayesian Network Model for Prediction of Students' Academic Performance in Engineering Programs", IEEE Global Engineering Education Conference (EDUCON), pp.832-837, 2014, doi: 10.1109/EDUCON.2014.6826192.
- [18] Paulo Cortez, Alice Silva, "Using Data Mining To Predict Secondary School Student Performance", EUROSIS, 2008.

- [19] Qasem A. et al., "Mining Student Data Using Decision Trees", The International Arab Conference on Information Technology (ACIT'2006), Nov. 2006, Jordan.
- [20] M. Ramaswami, R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues(IJCSI), Vol. 7, Issue 1, No. 1, January 2010.
- [21] Ali Daud, Naif Radi Aljohani, Rabeeh Ayaz Abbasi, Miltiadis D. Lytras, Farhat Abbas, Jalal S. Alowibdi, "Predicting Student Performance using Advanced Learning Analytics", 2017 International World Wide Web Conference Committee (IW3C2), pp. 415-421, 2017.
- [22] Shaobo Huang, Ning Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models", Computers & Education Vol.61, pp.133-145, 2013.
- [23] S. Kotsiantis et al., "Predicting Students' Performance In Distance Learning Using Machine Learning Techniques", Applied Artificial Intelligence, Vol.18, No.5, pp.411-426, 2004.
- [24] Reynold A. Rustia et al., "Predicting Student's Board Examination Performance using Classification Algorithms", in the proceedings of the 2018 7th International Conference on Software and Computer Applications (ICSCA), pp.233-237, 2018, doi: 10.1145/3185089.3185101
- [25] J.F. Superby et al., "Determination of factors influencing the achievement of the first-year university students using data mining methods" 8th international conference on intelligent tutoring systems, Educational Data Mining Workshop, Jhongli, Taiwan, 2006.
- [26] Havan Agrawal, Harshil Mavani, "Student Performance Prediction using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Vol. 4, Issue 03, March-2015.
- [27] Amirah Mohamed Shahiri et al., "A Review on Predicting Student's Performance using Data Mining Techniques", The Third Information Systems International Conference, Procedia Computer Science, Vol.72, pp. 414-422, 2015.

Authors Profile

Ms. Zeba Parveen completed Bachelor of Technology(B.Tech) from Maulana Azad National Urdu University(MANUU), Hyderabad, India, in 2017 and currently pursuing Master of Technology(M.Tech) from the same university that is MANUU. She presented a paper in national conference on "EMERGING TRENDS AND ISSUES IN INFORMATION TECHNOLOGY & COMMUNICATION" held in MANUU in 2018 and attended workshops. Her research interest is in machine learning, Data Mining and Big Data.



Mr. Mohatesham Pasha Quadri was born in Hyderabad, India, in 1987. He received the M.Tech degree in Computer Science & Engineering from the JNTU, Hyderabad, India, in 2014 and the MCA and B.Sc.(Mathematics, Statistics & Computer Science) degrees from Osmania University, Hyderabad, India, in 2011 and 2008, respectively.



In 2014, he joined the Department of Computer Science and Information Technology, Maulana Azad National Urdu University, as an Assistant Professor. His current research interests include Cloud Computing, Web Semantics, machine learning and Big Data.