

# Systematic Review of Predicting Student's Performance in Academics

Mukesh Kumar, Yass Khudheir Salal

**Abstract:** Data mining (DM) gaining popularity due to its advantages in the educational environment. Most of the educational institution, now a day applied these techniques to make improvement in their education system. By using these techniques, academic performance of the student is analyzed and if find anything wrong with the student performance then timely help will be provided to that student. In our education system, we lack in finding those factor which mostly affects the student performance in academics. Therefore, a systematic review of all the authors work done in this field is required to understand the data mining application in education and how it helps to improve and predict the student academic performance. In this article, the main focus moves around two important factors: Firstly, to find the most critical factors which affect the student performance used by the most researcher and secondly to find the algorithm which is mostly used.

**Index Terms:** Academic Performance, Educational Data Mining, Prediction, Classification.

## I. DATA MINING INTRODUCTION IN EDUCATION

Education is a critical factor for the development of any nation. So, proper attention for its implementation should be taken into consideration. Every institution collected lots of data from the student in the form of session test, assignment marks, internal and external laboratory evaluation, personal attributes. But all these information used to make final result of the student whether pass or fail the particular examination. If student data is analyzed timely then definitely it will help the student to improve their performance in academics. Now a day's educational system is not limited to any formal teaching within a classroom but it goes beyond that like online MOOC course, Web-based education system (WBS), an intelligent tutorial system (ITS), Project based learning (PBL), webinar, Seminar, Workshops and many more. All these teaching methods make the education system more attractive and provide lots of knowledge and learning through a different source. If all these methods do not properly check and evaluated then it adversely affect the education system. So for making any education system to have a success, then a proper evaluation check should be maintained. All these teaching methodologies discussed above generate a lot of data related to the student progress in academics, students marks and student attendance and if that data is not properly

analyzed then all the collected data is going to be wasted and no further use of that data happens. So, now a day's every educational institution applies an assessment or evaluation criteria to observe the performance of the students. In this modern era of education, there are lots of evaluations or assessment tools are available, which are helpful to observe the academic performance of the students. So at present DM is one of the best assessment or analysis tools, which is used to analyze the student data. The span of DM techniques is not only limited to the education system, but it is going to covers almost entire organisations where data are generated. The important application areas of DM are retail, banking, telecommunication, marketing, tourism, production, hospitality, hospital system etc. These entire listed organisations are taking the benefits of implementing DM to increase their performance in future by analyzing historical data. The DM techniques are used to analyze the historical data of any organization by using an inbuilt algorithm, which further finds the hidden information from that data which is not possible to find manually. By periodically evaluating the student progress, timely help and support are provided to those students who are on the risk of failure or dropout, so that they can improve their performance in future. It also helps the management and administrative staff of the organization to take suitable action for making improvement in the system. By taking all these factors into mind, a detail literature survey is performed. In this survey report, the main focus is to find the factors which affect the student performance, mostly used DM algorithm, Accuracy of the used DM algorithm.

## II. THE METHODOLOGY USED FOR REVIEW

Now a day, many young researchers working in the field of prediction academic performance of students, predicting dropout student, predicting student campus placement etc. So for making an effective literature review a systematic work plan is a must. So for that, we divide our studies into different parts. The first part of this section is related to building research questions formulation for the review process. The second part of this section is related to Search Strategy used for search some good research article/papers. The third part of this section is related to a number of research papers studied in this systematic review process and the last part of this section is related to the detailed classification of EDM Research.

### A. Research Questions formulation:

To start any study, it should be better to set all the objectives for your study. All the study work is carried out successfully and competed on time if followed by a work plan.

Manuscript published on 28 February 2019.

\* Correspondence Author (s)

Mr. Mukesh Kumar\*, Assistant Professor, Chitkara University School of Engineering and Technology, Chitkara University, Himachal Pradesh, India.

Mr. Yass Khudheir Salal, Department of System Programming, South Ural State University (National Research University) (Chelyabinsk, Russian Federation).

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Systematic Review of Predicting Student's Performance in Academics

So for better implementation of this literature review, we proposed some objectives, which further restrict our review process to some specific points. Below is a list of proposed objectives:

- Identify student's attributes influencing the performance in academics.
- Identify different DM techniques for predicting student's performance in academics.
- Identify existing gaps in the current research work

As the objectives for the review process is formulated, so for better understanding of this study, Main intention behind this literature study is to consider the suitability of the considered objectives in hand.

### B. Search Strategy used for searching some good research article/papers:

As the research in EDM increased since last ten years, so to find the best research papers is not an easy task. A lot of world-class database of research papers are available today. So to find some good research papers, selection of good research journal or database are the most important task. For that purpose, we finalize some research journal or research papers database to find some good research papers, technical notes, technical article, review papers, conference proceeding etc. Below is a table-1 in which searched database, search items, publication period and search strategies are mentioned:

**Table-1: Literature study contains the following terms for search on EDM work**

Terms used for search	Possible Values used
Databases	IEEE Xplore, Springer Link, Science Direct, Reserachgate, Google Scholar, ACM Digital Library
Search Articles	Journal, Workshops, Abstract, Technical, Book Chapters, Conference
Search applied on	Full text, Chapters, Article
Publication period	2012, 2013, 2014, 2015, 2016, 2017
Search Strategies	EDM, Student academic prediction, Student performance analysis, Student performance prediction in academic, DM application in Education

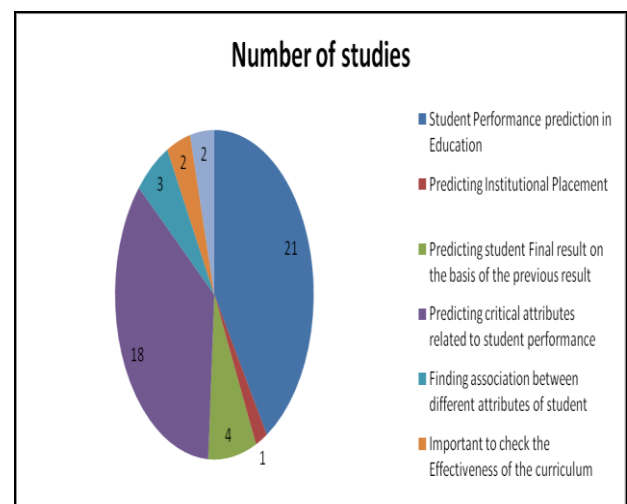
Table-1 shows us that, we follow some standard while finding some good research papers. For finding these papers we are searching on IEEE Xplore, Springer database, Science Direct, ACM, Reserachgate and Google scholar. In papers categories, journal articles, workshops papers, conference proceeding, Technical article are searched. Here, we consider research papers from 2012-17 for our review process. In this, we considered maximum papers (46 research papers, review papers, technical articles etc.) of the year 2012-2017, which is further known as the latest years of publication of the research article. All these papers are taken from world-class research database with good citation, conference proceeding papers, research articles, technical papers, IEEE papers, review papers etc for consideration. In Table 2, we tried to introduce only those research objectives/questions in which the young researcher are working since last few years

for the betterment of the education system. In our review, we found that "Student Performance prediction in Education" and "Predicting critical attributes related to student performance" is the hottest topic of research. Around 60% of the research scholar are worked on these topics and found some good result. Below is the list of all the research topic of the EDM with their numbers of studies, which are considered in this review papers:

**Table 2: Detailed classification of EDM Research**

Classification of EDM Research work	Number of studies
Student Performance prediction in Education	28
Predicting Institutional Placement	1
Predicting student Final result on the basis of the previous result	4
Predicting critical attributes related to student performance	18
Finding an association between different attributes of the student	3
Important to check the Effectiveness of the curriculum	2
Clustering similar attributes student in a class	2

In this particular section, our main focused was to develop a good research question for our review process. So only three main objectives are formed which is very important for further research point of view. Secondly, we follow a good search strategy used for searching good research article/papers. Thirdly, we are selecting 38 odd research papers in this systematic review. In the end, we divided all the research area into different subtopics. Below is the graphical representation of data represented in table 2 for better understanding.



**Fig 1: Graphical representation of numbers of studies in the area of EDM**



### III. IMPORTANT FACTORS FOR PREDICTING STUDENT'S PERFORMANCE

Use either SI (From conclusion of section 2, we found the predicting student performance in academic is most researched articles. This section will intensively discuss the significant factor in predicting student's academic performance in education. Mainly two factors affect the student's performance in academics, which are student attributes and prediction techniques. Table 4 shows the entire common student attributes and Table 5 shows the prediction techniques used by different research scholar in their research papers. Section 3.1 will be focused on the significant student attributes used by the different researcher for predicting student academic performance and section 3.2 will be focused on prediction techniques used in predicting students academic performance.

#### A. Important factors in predicting the performance of student's in academics

This section of literature is mainly to identify those factors which affect the student's academic performance prediction result. Student attributes are mainly classified into a different group like students academic attributes, family attributes, social attributes, personal attributes, institutional attributes etc. In this process, we almost review 47 odd research papers. The student attributes which most of the researcher are frequently used are cumulative grade point average (CGPA) and assessment marks in the academics. In his study, Mukesh Kumar and Prof. A. J Singh found that academic attributes and family attributes are the deciding factors of student academic performance. They are not considered personal attributes of the student into consideration because they feel that at the time of data collection student may not fill their correct information which further affects the prediction result [37]. In another study authors found that it is possible to predict the graduation performance in the fourth year at university using only pre-university marks and marks in first and second year's course. They were not considered socio-economic or demographic features of the students. In pre-university marks, they considered higher or senior secondary marks of the students [17]. In another study, authors found that while predicting the student performance in academics, attendance marks and cumulative grade point average (CGPA) are the most influencing factors. They considered these factors are important because these marks are deciding factor in there higher education. They further given more stress on CGPA because it will result whether a student is an educational dropout or the student will complete their education [20]. M. Kumar, S. Shambhu & P. Aggarwal, in their study found that, while predicting student performance attendance (ATND), internal grade of the student (INT\_GRD) is the deciding factors. They further gave stress on the medium of instruction, school location, school types are also are the deciding factor for the performance prediction [28]. Edin Osmanbegović and Mirza Suljic in their study found that university entrance examination (URK), GPA (PO), family (BCD), Material (MAT) are the attributes which affect the student academic performance the most [6]. In [23] authors found that the midterm-1 grade and midterm-2 is the strong predictor for the

final failing grade. They found 91% accuracy model to predict the failing students earlier before their final exam. Table 3 list all the attributes which are considered by the different researcher for their prediction. These entire attributes are further divided into five different categories like academic, personal, family, social and school-related attributes [23]. They are using only the female student's dataset of one-year duration, but the more accuracy of the result they further suggested to include male student's data also.

**Table 3: Critical factors for predicting the student's performance in academics**

Student Attributes	Possible Values Used in all research papers while implementing DM algorithm
Academic Attributes	Internal and External Assessment, Lab Marks, Sessional Marks, Attendance, CGPA, semester marks, grade, seminar performance, assignment, attendance, schools marks etc.
Personal Attributes	Age, Gender, height, weight, Student interest, Level of motivation, sports person, hobbies etc.
Family Attributes	Qualification, Occupation, Income, Status, Support, Siblings, Responsibilities etc.
Social Attributes	A number of friends, Social networking, Girls/boys friends, Movies, Travel Outing, friends parties etc.
School Attributes	Teaching Medium, Accommodation, Infrastructure, Water & Toilet facilities, Transportation system, Class Size, School Reputation, School Status, Class Size, School Type, Teaching Methodology etc.

**B. Algorithm predicting academic performance of student's**  
In EDM, predicting student academic performance usually, we use predictive modelling. To build a predictive modelling, lots of tasks are taken into consideration such as classification, clustering, regression, association rule mining etc. But in our study, we found that classification algorithms are the most popular algorithm used for prediction.

#### Decision Tree algorithm used for prediction:

A decision tree is frequently implemented classification algorithm used for prediction. It is used decision trees, where nodes represent a feature/attribute, branches represent a decision and leaf of the tree represents an outcome. This classification algorithm belongs to the categories of supervised learning algorithms. Here, while building a tree, the most influencing factor of the student's dataset is at the root level. There are different methods to find the root node like Gini Index, Information Gain. As shown in Table 5, approximately 20 research papers are used decision tree algorithm for prediction. In [27] authors used academic and demographic student's factors for making academic prediction and found 98.86% accuracy. In [29] authors used CGPA, Attendance, 12th marks, Engineering Cut-off, Education Medium, Board Type are influencing factors for prediction and found maximum accuracy of up to 97.27%.

## Systematic Review of Predicting Student's Performance in Academics

In [41] authors (2013) used internal assessment data as the main predictor and found the accuracy up to 56.25% which is not considered as a good prediction model [41].

**Table 5: Result accuracy using the Decision Tree algorithm**

Student Attributes used for Prediction	Author's Name	Year	Accuracy
midterm-I examination result	Mashaal A. et al [23]	2015	70.88%
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al [17]	2015	73.08%
finance level, motivation level, gender and grades	Kolo David Kolo et.al [24]	2015	66.80%
gender, race, hometown, GPA, family income, Grade Point Average (GPA), sem-1, sem-2 grades	Fadhilah Ahmad et. al. [22]	2015	68.8%
academic and demographic information	C. McGuinness et. al [19]	2014	65.00%
Internal Grade, Internet access and Attendance	Mrinal Pandey et. al.[27]	2016	98.86%
10th, 12th, graduation marks, parent's qualification	Mukesh Kumar et. al.[28]	2016	69.34%
pre-university, 1st and 2nd-year courses marks	Mukesh Kumar et. al.[37]	2017	61.40%
Int. assessment, CGPA, Extra-curricular activities	Raheela Asif, et. al.[33]	2017	74.78%
Internal, Demographic, Extra-curricular	M. Mayilvaganan et. al [47]	2014	66.00%
External, CGPA, Demographic, Extra-curricular	N. J. Aarthi et. al. [13]	2014	90.00%
Psychometric, Extra-curricular activities, soft skills	Natek, Zwilling et. al. [16]	2014	90.00%
Demographic attribute, High school background	T. Mishra et al. [15]	2014	88.00%
External assessment of student	P. Parkavi et. al. [12]	2013	65.00%
Entrance examination, GPA, Materials, Family	U. K. Singh et. al. [2]	2012	85.00%
CGPA, Attendance, 12th marks, Eng. Cut-off, Education Medium, Board	E. Osmanbegović et. al. [6]	2012	73.93%
Sex, race, home, income, university entry mode, and GPA	E.S.V. Kumar et. al.[39]	2016	97.27%
	AA Aziz, N H Ismail et. al. [40]	2014	68.80%

internal assessment data dCentrality, nMessages, nReplies, dPrestige and aEvaluation A. Priyam et. al. [41] M.I. López, J.M Luna et. al.[42] 2013 56.25% 2012 86.80%

*Neural Network algorithm used for prediction:*

Neural Network algorithms are frequently used learning algorithm used in Machine Learning. It is another algorithm which is implemented in EDM prediction. It is used to detect all the possible interactions between predictors attribute. D. M. S. Anupama Kumar (2012) used student's attribute like Internal and External assessment as the most influencing factors for making a prediction in self-regulated learning in higher education with maximum accuracy 98.00% [1]. P. M. Arsal, N. Buniyamin, J.-l. A. Manan (2013) used External assessment of student for student's performance prediction with 97.00% accuracy [13]. In [33] authors (2015) used pre-university, 1st and 2nd-year courses marks as main attributes for predicting academic performance at degree level with accuracy up to 62.50% [17]. As shown in Table 6, approximately 8 research papers are used Neural Network algorithm for prediction.

**Table 6: Result accuracy using Neural Network algorithm**

Student Attributes used for Prediction	Author's Name	Year	Accuracy
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al [17]	2015	62.50%
Grade Point Average (GPA), sem-1, sem-2 grades	C. McGuinness et. al [19]	2014	69.00%
CGPA, Quiz, Midterm, Lab. and Atted, Final grade	S T Jishan et. al. [20]	2015	75.00%
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al.[33]	2017	70.43%
Demographic attribute, High school background	P. Parkavi et. al. [12]	2013	72.00%
External assessment	N. Buniyamin et. al. [13]	2013	97.00%
Internal assessments, External assessment	Anupama Kumar et. al. [1]	2012	98.00%
Entrance examination, GPA, Materials, Family	E. Osmanbegović et. al. [6]	2012	71.20%

*Naive Bayes algorithm used for prediction:*

Naive Bayes algorithm is also a simple but powerful algorithm for making a prediction. It is easy to build and useful for very large student datasets. This algorithm would then be able to classify an unlabelled example based on the information learned from the labeled examples. This algorithm is the second frequently used machine learning algorithm for building predictive modeling. As shown in Table 7, approximately 16 research papers are used Naive Bayes algorithm for prediction. Mrinal Pandey and S. Taruna (2016) used academic and demographic attributes for student's performance prediction and got 91.57% accuracy [27]. V. Ramesh, P. Parkavi, K. Ramar (2013) used Demographic and High school background attributes for predicting student's performance with maximum accuracy 50.00%, which is not considered as a good predictive model and need some improvement.

**Table 7: Result accuracy using Naive Bayes algorithm**

Student Attributes used for Prediction	Author's Name	Year	Accuracy
midterm-I examination result	Mashaal A. et. al. [23]	2015	81.01%
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al [17]	2015	83.65%
gender, race, hometown, GPA, family income, university mode entry, and SPM grades, subjects	Fadhilah Ahmad et. al.[22]	2015	67.0%
Grade Point Average (GPA), sem-1, sem-2 grades	C. McGuinness et. al [19]	2014	76.51%
CGPA, Quiz, Midterm, Lab. and Atted, Final grade	S T Jishan et. al. [20]	2015	75.00%
Qualification, Courses, class, Date_Of_Work	A.M. El-Halees et.al. [21]	2015	77.46%
academic and demographic information	Mrinal Pandey et. al.[27]	2016	91.57%
Internal Grade, Internet access and Attendance	Mukesh Kumar et. al.[28]	2016	69.34%
10th, 12th, graduation marks, parent's qualification	Mukesh Kumar et. al.[37]	2017	61.40%
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al.[33]	2017	75.65%
Int. assessment, CGPA, Extra-curricular activities	M. Mayilvaganan et. al [47]	2014	73.00%
Demographic attribute, High school background	P. Parkavi et. al. [12]	2013	50.00%
Entrance examination, GPA, Materials, Family	E. Osmanbegović et. al.	2012	76.65%

CGPA, Attendance, 12th marks, Eng. Cut-off, Education Medium, Board	E.S. Vinoth Kumar et. al.[39]	2016	85.92%
Sex, race, home, income, university entry mode, and GPA	AA Aziz, N H Ismail et. al. [40]	2014	63.30%
dCentrality, nMessages, nReplies, nWords, dPrestige, aEvaluation	M.I. López, J.M Luna et. al.[42]	2012	89.40%

*K-Nearest Neighbor algorithm used for prediction*

KNN algorithm used for classification and regression prediction. It is commonly used for its ease of interpretation and low calculation time. M. Mayilvaganan, D. Kalpanadevi (2014) used Internal, Extra-curricular performance and CGPA for prediction the performance with maximum accuracy 83.00% [47]. In [19] authors (2014) used Grade Point Average (GPA), sem-1, sem-2 grades for predicting learner progression in tertiary education with accuracy up to 69.00%. As shown in Table 8, approximately 4 research papers are used K-Nearest Neighbor algorithm for prediction.

**Table 8: Result accuracy using K-Nearest Neighbor algorithm**

Student Attributes used for Prediction	Author's Name	Year	Accuracy
pre-university, 1st and 2nd-year courses marks	Raheela Asif, et. al [17]	2015	74.04%
Grade Point Average (GPA), sem-1, sem-2 grades	C. McGuinness et. al [19]	2014	69.00%
Qualification, Courses, class, Date_Of_Work	A.M. El-Halees et.al. [21]	2015	77.46%
Int. assessment, CGPA, Extra-curricular activities	M. Mayilvaganan et. al [47]	2014	83.00%

*Support Vector Machine algorithm used for prediction:*

In machine learning, SVM is a supervised learning that analyzes the dataset used for classification and regression analysis. There are few works that have used SVM as techniques to predict student's academic performance. S.A. Oloruntoba, J. L. Akinode (2016) used the Academic result, semester-1 and semester-2 grades for student academic performance prediction using support vector machine with the accuracy of 98.00% [43]. Mrs. M. Jamuna, Mrs. S. A. Shoba (2017) used students' University Admission Score, Number of Failures at the first-year university examination attributes for Students Performance Prediction [44]. K. B. Eashwar, R. Venkatesan (2017) used psychological, social-behavior, family-circumstances and parental-care parameters for Students Performance Prediction [46]. As shown in Table 9, approximately 6 research papers are used Support Vector Machine algorithm for prediction.



**Table 9: Result accuracy using a Support Vector**

Machine algorithm	Author's Name	Year	Accuracy
Student Attributes used for Prediction	C. McGuinnes et. al [19]	2014	81.22%
Grade Point Average (GPA), sem-1, sem-2 grades	M. Mayilvagan et. al [47]	2014	80.00%
Int. assessment, CGPA, Extra-curricular activities	S.A. Oloruntoba et. al. [43]	2016	98.00%
Academic result, sem-1, sem-2 grades	Mrs M. Jamuna et. al. [44]	2017	97.62%
students' University Admission Score, Number of Failures at the first-year university examination academic details and conducting online test scores	A Kadamban de et. al. [45]	2017	90.00%
psychological, social-behaviour, family-circumstances and parental-care parameters	K. B. Eashwar et. al. [46]	2017	96.70%

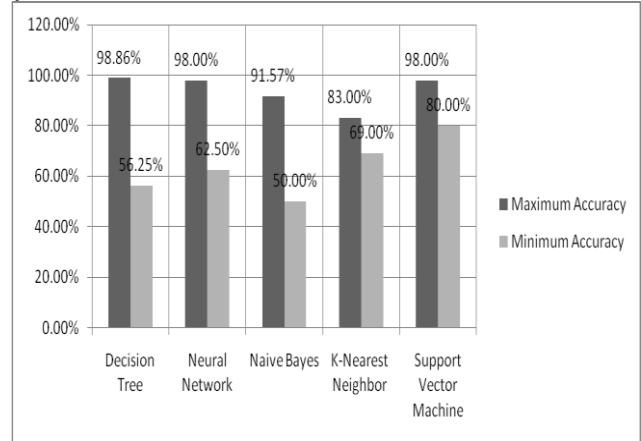
**IV. RESULT ANALYSIS**

Here, we are trying to discuss the different classification algorithms with their maximum and minimum accuracy. From section 3.2, we discuss DT, NN, NB, K-NN and SVM in detail. In Table 5, we analyze twenty papers and found that the decision tree algorithm has maximum accuracy is around 98.86% and minimum accuracy is 56.25%. In Table 6, we analyze eight papers and found that the Neural Network (NN) algorithm have maximum accuracy is around 98.00% and minimum accuracy is 62.50%. In Table 7, we analyze sixteen papers and found that the Naive Bayes (NB) algorithm have maximum accuracy is around 91.57% and minimum accuracy is 50.00%. In Table 8, we analyze four papers and found that the K-Nearest Neighbor (KNN) algorithm have maximum accuracy is around 83.00% and minimum accuracy is 69.00%. In Table 9, we analyze six papers and found that the Support Vector Machine (SVM) algorithm have maximum accuracy is around 98.00% and minimum accuracy is 80.00%. In Table 10, show the detail of different algorithm with their accuracy.

**Table 10: DM Classification algorithms with their maximum and minimum accuracy**

DM Classification Algorithm	Maximum Accuracy	Minimum Accuracy
Decision Tree Algorithm	98.86%	56.25%
Neural Network Algorithm	98.00%	62.50%
Naive Bayes Algorithm	91.57%	50.00%
K- Nearest Neighbor algorithm	83.00%	69.00%
Support Vector Machine Algorithm	98.00%	80.00%

Fig. 2 shows the student's performance prediction accuracy that uses classification algorithm from 2012 to 2017. Decision Tree algorithm has a maximum accuracy (98.86%) followed by Neural Network (98.00%), Support Vector Machine (98.00%), Naive Bayes (91.57%) and K-Nearest Neighbor algorithm (83.00%). In the end, the algorithm that has lower prediction accuracy is the Naive Bayes algorithm by (50.00%).



**Fig 2: Graphical representation of Classification algorithms with their max and min accuracy**

There are lots of software tool available in the market like WEKA, RapidMiner, MATLAB, KNIME, Rattle GUI, Orange, Apache Mahout, R, ML-Flex, Natural Language Toolkit etc. But in our analysis, WEKA is the frequently used software tool by the researcher for prediction. RapidMiner is the second most frequently used software tool. These tools are freeware and easily available for use. These software tools are supporting a lot of classification, clustering and association rule mining algorithm for use on different types of datasets.

**V. CONCLUSION**

EDM is one of the most critical areas of research today. Analyzing almost 46 papers, we found that, predicting student's performance in academic is a useful topic of research which helps educator, academician, policy makers and management for improving the teaching and learning process throughout the world. The student's attributes taken by different researchers are categories as academic, psychological, social behavior, family attributes, and personal attributes. In most of the reviewed papers, researchers used CGPA and internal and external assessment marks for their prediction. While analyzing prediction algorithms we found that Classification, Clustering, Linear regression, Association Rule mining are used for prediction. But in most of the cases, Classification algorithms are frequently used for making student's academic prediction. Under the Classification algorithms, DT, NN, NB, K-NN and SVM are frequently used by many researchers. In the end, we enthused to accomplish our research on student's academic prediction in our educational system which helps students, educators, management and policymaker to improve the educational system by continues monitoring of students.



## REFERENCES

- D. M. S. Anupama Kumar, "Appraising the significance of self-regulated learning in higher education using neural networks", *International Journal of Engineering Research and Development* Volume 1 (Issue 1) (2012) 09–15.
- K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students using classification, in *Wireless and Optical Communications Networks (WOCN)*, 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.
- S. Parack, Z. Zahid, F. Merchant, "Application of data mining in educational databases for predicting academic trends and patterns", in *Technology Enhanced Education (ICTEE)*, 2012 IEEE International Conference on, IEEE, 2012, pp. 1–4.
- Anuwatvisit S, Tungkasthan A, Premchaiswadi W., "Bottleneck mining and Petri net simulation in education situations. In: *ICT and Knowledge Engineering (ICT & Knowledge Engineering)*, 10th International Conference. IEEE; Bangkok, Thailand; 2012, 244–251.
- Mohammed M. Abu Tair, Alaa M. El-Halees, Mining Educational Data to Improve Students' Performance: A Case Study, *International Journal of Information and Communication Technology Research*, ISSN 2223-4985, Volume 2 No. 2, February 2012.
- Edin Osmanbegović and Mirza Suljic, Data Mining Approach For Predicting Student Performance, *Economic Review – Journal of Economics and Business*, Vol. X, Issue 1, May 2012.
- M. of Education Malaysia, National higher education strategic plan (2015). URL <http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025>
- U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: *Engineering Education (ICEED)*, 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
- D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)* 1 (1) (2013) p12–16.
- K. F. Li, D. Rusk, F. Song, Predicting student academic performance, in *Complex, Intelligent, and Software Intensive Systems (CISIS)*, 2013 Seventh International Conference on, IEEE, 2013, pp. 27–33.
- C. Romero, M.-I. L'opez, J.-M. Luna, S. Ventura, Predicting students' final performance from participation in on-line discussion forums, *Computers & Education* 68 (2013) 458–472.
- V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, *International Journal of Computer Applications* 63 (8) (2013) 35–39.
- P. M. Arsad, N. Buniyamin, J.-I. A. Manan, A neural network students' performance prediction model (nnsppm), in *Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2013 IEEE International Conference on, IEEE, 2013, pp. 1–5.
- G. Elakia, N. J. Aarthi, Application of data mining in the educational database for predicting behavioural patterns of the students, *Elakia et al.(IJCSIT) International Journal of Computer Science and Information Technologies* 5 (3) (2014) 4649–4652.
- T. Mishra, D. Kumar, S. Gupta, Mining students' data for prediction performance, in *Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT '14*, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262. doi:10.1109/ACCT.2014.105. URL <http://dx.doi.org/10.1109/ACCT.2014.105>
- S. Natek, M. Zwilling, Student data mining solution–knowledge management system related to higher education institutions, *Expert systems with applications* 41 (14) (2014) 6400–6407.
- Raheela Asif, Agathe Merceron, Mahmood K. Pathan, Predicting Student Academic Performance at Degree Level: A Case Study, *IJ. Intelligent Systems and Applications*, 2015, 01, 49-61 Published Online December 2014 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijisa.2015.01.05.
- A. Bogar'in, C. Romero, R. Cerezo, M. S'anchez-Santill'an Clustering for improving educational process mining, in *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, ACM, 2014, pp. 11–15.
- G. Gray, C. McGuinness, P. Owende, An application of classification models to predict learner progression in tertiary education, in *Advance Computing Conference (IACC)*, 2014 IEEE International, IEEE, 2014, pp. 549–554.
- S. T. Jishan, R. I. Rashu, N. Haque, R. M. Rahman, Improving accuracy of student's final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, *Decision Analytics* 2 (1) (2015) 1–25.
- Randa Kh. Hemaïd and Alaa M. El-Halees, Improving Teacher Performance using Data Mining, *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 4, Issue 2, February 2015.
- Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, The Prediction of Students' Academic Performance Using Classification Data Mining Techniques, *Applied Mathematical Sciences*, Vol. 9, 2015, no. 129, 6415 - 6426. HIKARI Ltd, [www.m-hikari.com](http://www.m-hikari.com) <http://dx.doi.org/10.12988/ams.2015.53289>.
- Mashaël A. Al-Barrak and Mona S. Al-Razgan, predicting students' performance through classification: a case study, *Journal of Theoretical and Applied Information Technology* 20th May 2015. Vol.75. No.2.
- Kolo David Kolo, Solomon A. Adepoju, John Kolo Alhassan, A Decision Tree Approach for Predicting Students Academic Performance, *I.J. Education and Management Engineering*, 2015, 5, 12-19 Published Online October 2015 in MECS (<http://www.mecs-press.net>) DOI: 10.5815/ijeme.2015.05.02.
- Mohammed I. Al-Twijri and Amin Y. Noaman, A New Data Mining Model Adopted for Higher Institutions, *Procedia Computer Science* 65 (2015) 836 – 844, doi: 10.1016/j.procs.2015.09.037.
- Mihai Dascalu and Elvira Popescu et. al., Predicting Academic Performance Based on Students' Blog and Microblog Posts, Springer International Publishing Switzerland 2016 K. Verbert et al. (Eds.): EC-TEL 2016, LNCS 9891, pp. 370–376, 2016. DOI: 10.1007/978-3-319-45153-4\_29.
- Mrinal Pandey and S. Taruna, Towards the integration of multiple classifiers pertaining to the Student's performance prediction, <http://dx.doi.org/10.1016/j.pisc.2016.04.076> 2213-0209/© 2016 Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).
- Mukesh Kumar, Shankar Shambhu & Punam Aggarwal, "Recognition of Slow Learners Using Classification Data Mining Techniques" *Imperial Journal of Interdisciplinary Research (IJIR)* Vol-2, Issue-12, 2016 ISSN: 2454-1362, <http://www.onlinejournal.in>
- Doleck T, Jarrell A, Poitras EG, Chouachi M, Lajoie SP. Examining diagnosis paths: a process mining approach. In: *Computational Intelligence & Communication Technology (CICT)*, Second International Conference. IEEE, Ghaziabad, India, 2016, 663–667.
- Vidal JC, Vázquez-Barreiros B, Lama M, Mucientes M. Recompiling learning processes from event logs. *Knowledge-Based Syst* 2016, 100:160–174.
- Julia Rudnitckaia. *Process Mining: Data Science in Action*. Berlin, Germany: Springer; 2016.
- Romero C, Cerezo R, Bogarín A, Sánchez-Santillán M. Educational process mining: a tutorial and case study using moodle data sets. In: *Data Mining and Learning Analytics: Applications in Educational Research*. Hoboken, NJ: John Wiley & Sons; 2016, 1–28.
- Raheela Asif, Saman Hina and Saba, "Predicting Student Academic Performance using Data Mining Methods", *International Journal of Computer Science and Network Security (IJCSNS)*, VOL.17, 2017.
- Ricardo Mendes And Joao P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications", IEEE, 2017.
- Schulte J, Fernandez de Mendonca P, Martinez- Maldonado R, Buckingham Shum S. Large-scale predictive process mining and analytics of university degree course data. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM; Vancouver, Canada; 2017, 538–539.
- Mukesh Kumar, A.J. Singh, Disha Handa, "Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques", *International Journal of Education and Management Engineering(IJEME)*, Vol.7, No.6, pp.40-49, 2017.DOI: 10.5815/ijeme.2017.06.05
- Mukesh Kumar, A.J. Singh, "Evaluation of Data Mining Techniques for Predicting Student's Performance", *International Journal of Modern Education and Computer Science (IJMECS)*, Vol.9, No.8, pp.25-31, 2017.DOI: 10.5815/ijme.2017.08.04
- Febrianti Widyahastuti, Viany Utami Tjhin, "Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron", IEEE, 2017.

39. R. Sumitha and E.S. Vinoth Kumar, Prediction of Students Outcome Using Data Mining Techniques, International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-6, June 2016 ISSN: 2395-3470.
40. Azwa Abdul Aziz, Nor Hafieza Ismailand Fadhilah Ahmad, First Semester Computer Science Student's Academic Performances Analysis by Using Data Mining Classification Algorithms, Proceeding of the International Conference on Artificial Intelligence and Computer Science(AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA. (e-ISBN978-967-11768-8-7).
41. Priyam, A., Gupta, R., Rathee, A., & Srivastava, S. (2013). Comparative Analysis of Decision Tree Classification Algorithms, 334–337.
42. López, M. I., Luna, J. M., Romero, C., & Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. International Educational Data Mining Society. 148-151.
43. S.A. Oloruntoba, J. L. Akinode, "Student Academic Performance Prediction Using Support Vector Machine" International Journal of Engineering Sciences & Research Technology, 2016, DOI: 10.5281/zenodo.1130905, ISSN: 2277-9655.
44. Mrs M. Jamuna, Mrs S. A. Shoba, "Educational Data Mining & Students Performance Prediction Using SVM Techniques", International Research Journal of Engineering and Technology (IRJET), 2017, e-ISSN: 2395-0056.
45. A Kadambande, S Thakur et. al., "Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology, e-ISSN: 2395 -0056, 2017.
46. K. B. Eashwar, R. Venkatesan, " Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET), Volume 8, Issue 11, November 2017, pp. 649–662, Article ID: IJMET\_08\_11\_066
47. M. Mayilvaganan, D. Kalpana Devi, Comparison of classification techniques for predicting the performance of student's academic environment, in Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, 2014, pp. 113–118.

### Authors Profile



**Mukesh Kumar** is presently working as an Assistant Professor in Chitkara University School of Engineering and Technology, Chitkara University, Himachal Pradesh, India. He completed his M.Tech in Computer Science and Engineering from HPU Shimla in 2008. He is currently pursuing PhD degree in the department of Computer Science, Himachal Pradesh University, Summer hill, Shimla, India. His research interest includes Machine learning, Artificial intelligence, Information security, Educational Data Mining. He has 10 years of teaching experience and published 15 research papers in different international journals.



**Yass Khudheir Salal** was born in Al-Qadisiyah, Iraq on Sep. 06, 1983. He received the B.E. degree in Computer Science from Qadisiyah University, Faculty of Science, Al-Qadisiyah, Iraq, in 2004. He received the M.E. degree from Belarus State University of Electronic and radio communications, security department of information and data protection, Belarus, Minsk, in 2014, He is currently pursuing PhD degree in the department of system programming in Graduate School of Electronics and Computer Science, South Ural State University (SUSU), national research university, Chelyabinsk, Russia. His research interest includes Big Data, Machine learning.