ANALYSIS OF STUDENT ACADEMIC PERFORMANCE USING MACHINE LEARNING ALGORITHMS:– A STUDY

VRATIKA GUPTA*

College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad Uttar Pradesh, India. *Corresponding Author Email: vratika1997@gmail.com

PRIYANK SINGHAL

College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad Uttar Pradesh, India.

VIPIN KHATTRI

College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad Uttar Pradesh, India.

Abstract

Student academic performance is the great value of institutes, universities and colleges. All colleges majorly focus on the career development of students. The academic performance of students plays a vital role in the establishment of a bright career. On the basis of better academic performance, the placement of the students will be better and the same will be reflected in the form of better admission and future. Machine learning can be deployed for the prediction of student performance. Various algorithms are playing an important role in the prediction of the accuracy of various machine learning models. These articles discuss various algorithms that can be helpful to deploy for predicting student academic performance. The article discusses various methods, predictive features and the accuracy of machine learning algorithms. The primary factors used for predicting students performanceare academic institution,sessional marks,semester progress, family occupation, methods and algorithms. The accuracy level of various machine learning algorithms is discussed in this article.

Keywords: Algorithms, Accuracy, Prediction, Student Performance.

1. INTRODUCTION

Nowadays, all university requires enhancing quality of education in order to achieve the better results in student performance. It is beneficial for students; with the help of prediction students improve their performance. We need different kinds of techniques to measure students' strengths and weaknesses in the subjects. Many researchers have worked on prediction techniques to measure student performance using machine learning algorithms.

Prediction is the most challenging factor in various fields as stock market, education system, sales forecast, budget growth, weather forecast, etc. In this article, previous predictions of student performance were reviewed using machine learning algorithms. Various kinds of machine learning processes are used such as regression, classification, decision tree, clustering, ANN and K Means to predict student performance. These predictions are helpful for students and faculty members to easily see the student's weaknesses in the subjects. Many researchers have used WEKA tool, Python and R languages to predict the accuracy of machine learning models.

The next section focuses to the methods, search strategies and Research questions. Section 3 focuses a detailed background of the previous paper, while Section 4 discusses the types of algorithms and result of previous papers. Lastly, section 5 and section 6 are defining the result and conclusion.

2. MATERIALS AND METHODS

This article's primary goal is to define machine learning algorithms.The most significant factors are:

A) Set of Research Questions

It is necessary to formulate the appropriate research question to determine the key studies related to the prediction of student performance. To choose the right research question, we used the PICO framework. The PICO framework is is utilized to display the Population, Intervention, Context and outcome.

Table 1 represented the PICO framework in detail:

According, this piece of work is conducted to answer the following research questions:

Research Question

1. What type of algorithm is applied to measure the performance of students?
2. What model is the greatest fit to measure student performance?
3. What are the accuracy and outcomes of these ml algorithms?
4. Which tools and methods are used in measuring in ml algorithms?

B) Study Preference and Search Procedure

A systematic search was required to answer the set of research question in this paper. For this objective, many online databases were searched from Jan 2015 – Jan 2023, including Research Gate, Google Scholar, IEEE Xplore, Scopus and EBSCO host. This database was very helpful to understand the current working of student academic performance. The search of the paper was 150. The important papers of 50 provide good details of student academic performance. The 30 papers used abstract title screening. The 20 papers used to study over all papers and results.

C) Eligibility Criteria

We included these studies 1) Student academic performance based on machine learning and deep learning. 2). Publish papers between 2015 to 2023, 3). Research Papers are collected from reputed Journals and International conferences. 4). At any education level 5). There is no empirical or experimental data. 6). Define tools and techniques used to measure student performance.

3. LITERATURE REVIEW

Numerous studies have been carried out in the past. Study on machine learning, deep learning, and educational data mining is still quite popular. This section is defining the study of previous research articles.

Alfiani, A. P et al.(2015) used a data mining approach to map the students' performance. In this article, the researcher used the K-mean clustering algorithm to predict students' performance using variables (gender, GPA, and the grade of certain courses). **Akinrotimi et al. (2018)** defined the Random tree and the c4.5 Algorithm to forecast student performance using CGPA (Cumulative Grade Point Average), SSCE grade score, and UME. The result was showing the accuracy and Confusion Matrix. In future work on large data sets and used the SVM classifier algorithm. **Jadhav, P (2019)** defined the decision tree Algorithm. It showed how to improve the performance of students which is helpful to the growth of institutes. Several factors affect the performance of students like internal marks, external marks, attendance, and assignments. **Birgili,B et al.(2021)** used flipped technique in 2012 & 2018 data. Flipped learning helps to increase student performance. **Hooda, M (2022)** defined ANN and FCN algorithms. Learning Analytics is added to the fully connected neural network to improve student quality evaluation in higher education. They defined the ANN model accuracy was low and FCN model accuracy was high.

4. STUDENT ACADEMIC PERFORMANCE& PREDICTION TECHNIQUES

Researchers have been defined different algorithms, applications and tools to measure the performance of students. The different types of algorithms are Naïve Bayes 4.1, Artificial Neural Network **Error! Reference source not found.**, Support Vector Machine, Decision Tree 4.5,Regression, KNN 4.6, Educational Data Mining, Random forest and 4.7 Deep Learning. Applications and tools are Python Jupiter notebook, R studio WEKA, Orange tool etc. *To predict student performance uses different parameters like Marks, CGPA, Attendance, and Courses.*

In the following section provide the brief details of machine learning algorithms, average accuracy of student performance.

4.1 Naïve Bayes

Naïve Bayes is part of   classification algorithms. It is used for supervised algorithms such as spam or not spam, face detection, and character Recognition. **Table 4.1** is defining the deep study of research articles in 2017 – 2022 to predict the student performance using Naïve Bayes algorithms. **Table 4.1** is defining that variable, parameters and result. The result is defining the accuracy and F1-Score. F1-score is defining the precision and recall into the matrix. Accuracy is used to define the overall result. The highest accuracy of model using naïve Bayes technique was 94% and the lowest accuracy was 80%. The working of naïve Bayes technique in student performance is used to calculate the probability using multiple variables. This is the formula which is applying in Naïve Bayes algorithms to measure the probability.

$$P(A|B) = \frac{P(B|A).P(B)}{P(A)}$$

4.2 Artificial Neural Network

In this work, our search is defining the ANN techniques to predict the student's performance. After studying the several research paper of 2017 -2022, Table 4.2 is defining the parameters, predictive features and result. The highest accuracy of the model was 80% and lowest accuracy was 69%.

4.3 Support Vector Machine

SVM is applied to regression and classification problem. It offers the information in best decision boundary. **Table 4.3** is defining the research articles of 2019-2021 in SVM algorithms. In **Table 4.3** is describing the parameters, accuracy and tools which are used as a performance indicator. The highest accuracy of the model was 90% and the lowest accuracy of the model was 64%.

4.4 K Means Clustering Algorithm

K means clustering algorithm is a component of unsupervised algorithm. The working of k Mean clustering algorithm is first select then centroid point randomly. Group the data to form k clusters. Each cluster has used to identify patterns. **Table 4.4** defined the parameters, methods to measure the performance of students.

4.5 Decision Tree

Decision tree is top down approach. It works on a tree-based model. It defined root node, leaf node and splitting node. There are three type of decision tree ID3, C4.5 and J48. ID3 is used for top down approach and greedy approach for small data sets. C4.5 is better algorithm to ID3 ; it works on large data set. J48 is extended version of decision tree. It works on classification data set. These algorithms are used to prediction. After study 2015 -2023papers we are identify the decision tree provide good result of student performance.

It used to predict student performance using different parameters. The different parameters are defined in **Table 4.5**. Weka tool is open-source software. It used for data preparation, classification, regression, Clustering, association rules mining, and visualization. Many research articles used WEKA tool to predict student performance. It is also helpful to create a decision tree for j48 algorithm.

The decision tree is used for the if–else condition. A decision tree example **Fig 1** *defines in university data set Graduation, Post-Graduation and Doctorate students work like a decision node. Where GS = Graduate Students, PS = Post Graduate Students and Doc = Doctorate Students.*

Decision tree Algorithm (IF-ELSE condition)

R1 if GS > =50%, Print "PASS successfully& interested to take admission in post-graduation" then else GS <= 40 %   Print "FAIL"

R2   if PS >= 50% then print "pass & take admission in Ph.D." then else PS <= 40 % "FAIL"

R3 if Doc >=50 % PRINT "PASS successfully complete then else PS <= 40% "FAIL".
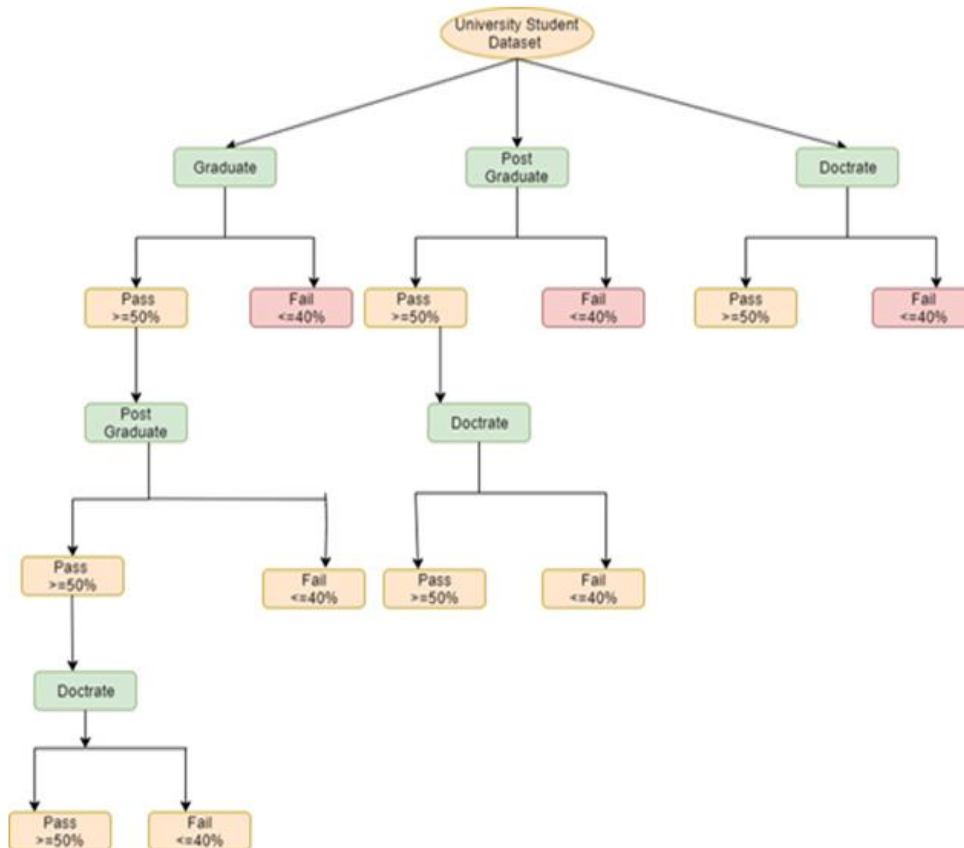

Fig.1: Decision Tree algorithm using student academic Performance

4.6 K-Nearest Neighbor

KNN is a method of supervised machine learning algorithm. It is called non parametric because it doesn't assume anything about the underlying data and lazy learning algorithm. It is worked on both classification and algorithm. It is called non parametric because it doesn't assume anything about the underlying data and lazy. The method is mainly solved categorization difficulties. **Table 4.6** is defining the parameters and predictive features. The highest accuracy of the student performance using KNN algorithm is 94% and the lowest accuracy of the KNN algorithm is 62%.

4.7 Random Forest

Random forest is used for both classifier and regression methods. It is a supervised learning algorithm. It works on continuous and categorical variables. It performs better in classification algorithms for categorization. **Table 4.7** is defining the accuracy and those parameters are which used to random forest method. The random forest algorithm

provides the good accuracy to measure student performance. The highest accuracy of the model was 97% and the lowest accuracy of the model was 75%.

4.8 Deep Learning

Deep learning is a part of machine learning. It is based on artificial neural networks. Like example neural network is working to mimic the human brain same deep learning is also used to mimic human brain. In today's time, it works on fast manner in the education field and large amount of data sets. It is also working to measure the performance of the students, Biometric Attendance, Finger Print and

scholarship. **Table 4.8** used to defining the attributes, parameters, tools, algorithms and accuracy of the model using deep learning techniques. Different type of deep learning algorithms discussed in **Table 4.8**. The highest accuracy of the model was 97.16% and the lowest accuracy of the model was 81.64%.

5. RESULT

The article discusses the result of algorithms in parameters, tools and average accuracy into the table. On the basis of, article shows the highest and the lowest count of the publish paper in student performance. The highest count of the paper is decision tree and support vector machine. The articles are discussing in the result of the published paper in year 2015-2023 in student performance. The best count of paper is 2015 -2023.
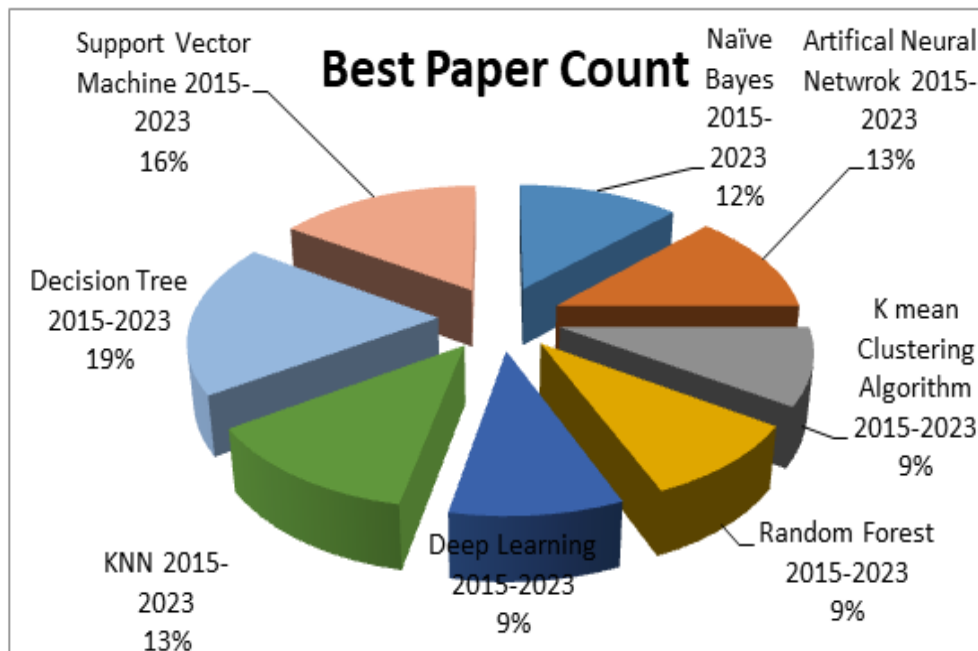

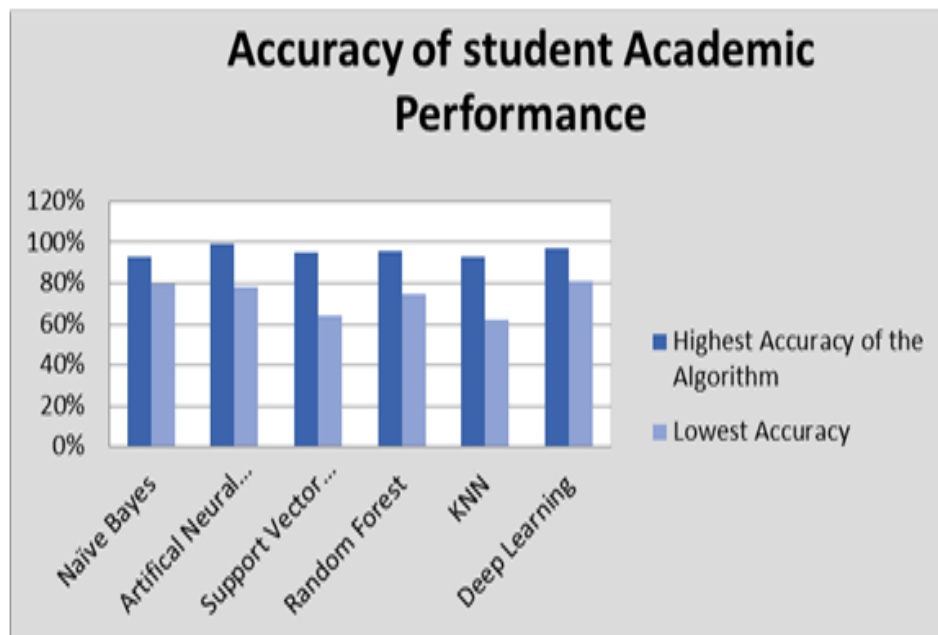**Fig.2: Highest Count of paper 2015 to 2023**

Fig 3: Prediction accuracy define for student academic performance 2015 to 2023

Table 1: Define the pico criteria context of student academic performance

Pico Criteria	Study
Population	Male/female students, above 18 years; all institution resources.
Intervention	Machine learning and Deep Learning Algorithms
Context	Academic institutions; university; college; high school
Outcome	Get the knowledge of which model is best for algorithms, Accuracy, ROC, AUC curve, F-score, Predictive Features, Highest and Lowest accuracy of the model

3.1: Table for Literature Review

Sr. No	Year	Paper Name	Keywords	Findings
	-2015	Mapping Student's Performance Based on Data Mining Approach	Performance, student, clustering, K-mean, pattern	Meta-Analysis is helpful to further predict student performance.
2.	-2018	Student Performance Prediction Using Random Tree and C4.5 Algorithm	Decision Tree, Machine Learning, Encryption, Prediction, Data Mining.	This paper used a random tree and c4.5 Algorithm and the result shown on the basis of accuracy, Confusion matrix, and further work on the SVM classifier
	-2019	A review paper on student academic performance using decision tree Algorithms	Performance of the students, decision tree, data mining, educational data mining	This paper predicts the performance of students using the decision tree Algos

	-2020	Clustering-Based EMT Model for Predicting Student Performance	Educational data mining · Clustering-based EMT, Wrapper-based feature selection, Machine learning	In this paper used Cluster based supervised classifier model & accuracy is 96.25%
5.	-2021	The trends and outcomes of flipped learning research between 2012 and 2018: A descriptive content analysis	Active learning, Descriptive content analysis, Flipped learning, Student performance	In this method used flipped approach 2012 & 2018. Flipped learning increase student performance.

Table 4: Average accuracy of ANN, DT, SVM, NB, DL, KNN and RF algorithm

Algorithm	Average accuracy (%)	Study
Naïve Bayes (NB)	85%	[(2017),(2022),(2022)]
Artificial neural network (ANN)	79%	[((2021)(2022),(2023)]
Decision tree (DT)	87%	[(2016),(2022),(2022)(2022)]
Support vector machine (SVM)	83.4	[(2019)]
K-nearest neighbor (KNN)	82.25%	[(2021)(2019)(2022)]
K-Mean clustering Algorithm (KMC)	83%	[2022)(2019)(2019)]
Random Forest (RF)	91.75%	[(2022),(2021)]
Deep Learning (DL)	91.2%	[(2022), (2021)(2019)]

Table 4.1: Accuracy of Naïve Bayes algorithm using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2017)	English, Education, History, Mathematics, Additional Mathematics, Physics, Chemistry, Category -Average, Good, Excellence, Poor	Naïve Bayes, Data set – 488, *Cross* validation-10	Accuracy = 80.94%
(2022)	Independent Variables (Gender, Address, Family Type, Family Annual Income, Instruction Material (Book, Encyclopedia), Extra-Curricular, Internet Access, Medium of instruction, In Relationship, Weekly study time),Dependent Variable(GWA)	Naïve Bayes Classifier, cross validation - 10 WEKA tool	Accuracy = 93.48%

Table 4.2: Accuracy of ANN algorithm using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2021)	Gender, Age, Address, father's and Mother's job, Family size, Travel and Study time, and First, Second, and Final Grades.(comparison paper)	ANN Data size- 480 students attributes of columns – 16 Data collection – Kaggle	Data set DS 2 (Accuracy = 79.17%, Recall= 0.79 Precision = 0.79 F- Measure = 0.79)
(2022)	Student Registration, Courses, Assessments And Student Assessments	ANN	Accuracy = 78%, Recall= 0.88,F1-score= 0.76, precision = 0.81
(2023)	Individual, University, Graduation, Academic satisfaction, Education service, Socio cultural and Perceptions	ANN AND SEM (Structural Equation Modeling) ANN	Data size = 420, SEM Accuracy = 69% and ANN Accuracy = 99%

Table 4.3: Accuracy of Support Vector algorithm using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2019)	High, Low, Average	SVM Linear and SVM Radial	Accuracy =64%, Accuracy = 90%
(2020)	Age, Study time, Failures, Famrel, Free time,Go out, Dalc, Walc Health, Absences G1 , G2,G3	SVM	Accuracy = 93%
(2021)	Gender, , Place of Birth, Stage ID, Grade ID, Section ID, Nationality, Topic, Semester, Relation, Raise hand, Visited Resources, Student Absence Day, Class, Failing percentage (0% to 69%), Low passing (70% - 89%), High percentage (90%-100%)	SVM	Accuracy = 70.8% Precision=78.8% Recall=78.8% F-Measure=78.7% Roc Area= 86%
(2022	Midterm, Final, stdID, Faculty, Department	Orange machine learning software, SVM	Accuracy=74% AUC value= 80%
(2022)	Core, Delta, Bifurcation and Ridge Ending	SVM	Accuracy = 95.02%, MSE= 0.66

Table 4.4: Accuracy of K Mean Clustering algorithm using student academic Performance

References	Parameters & Predictive Features	Tools, Methods and algorithm	Result
(2022)	Sample size – 80, Cluster – 3, High, Medium And Low, Student's Name, Skills, Skill Competence, Extracurricular, Discipline, Knowledge	Web based, K Mean Clustering algorithm	WCV (With in cluster Variation) = 360.97 BCV (Between Cluster Variation) = 7.357
(2019)	Students data – 3lakh, 200 MOOCS	K Mean Clustering	-
(2019)	Age, Medu, Fedu, Qualtiy of family relationship, Health, Abasences, G1, G2 and G3 Final Year Grade	Elbow Method, K Mean Clustering and MLP	Elbow curve Clustering data set into 2 clusters

Table 4.5: Accuracy of Decision Tree algorithm using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2016)	Roll No, Internal, Sessional, Adm score	Decision Tree	-
(2019)	Lab Exercise projects, Quizzes, Pass failed and Conditional, Midterms	J48	Accuracy = 91.47%
(2020)	M1,M2,M3,SGPA and CGPA, Result (Fail or Pass)	C4.5	Accuracy = 77.12%
(2022)	WSM, DSM, Living-area, Romantic–status, parent-Edu Technology Desire-higher-education	Decision Tree	Accuracy = 90%
(2022)	Scholarship, Job of father, Income of mother, Job of mother, Domicile, Transportation, Income of father	ID3	Accuracy =89% and F1 score= 72%

Table 4.6: Accuracy of KNN algorithm using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2022)	Race, Gender, Home Language, Age at first year, Home province, Rural or Urban, Home country, School, Quantile, English FAL, Mathematics major, Computers, Additional Maths, Year started, Probability of streams, Plan Description, Aggregate, Number of years in degree	KNN	Accuracy = 87%
(2021)	Reg no, Name, Attendance, Class Performance, Fees Payment, Punctuality, Parents Involvement, Result	KNN AND K values = 5	Accuracy =%, Kappa = 0.25%

| (2019) | Grade Points(0-4), GPA(0-4), hometown(1: city close from campus, 0: city near from campus), type_of_school(1: public school 0: private school), major (1: computer/informatics 2: science major 3: others), parents job(civil servant, employee , entrepreneur, farmer, fisherman , others) AKTIF (1 = active, 0 = non-active) | KNN, R studio Data size – 1530 rows and 7 attributes, Data split – 75% and 25%, | Accuracy = 93.81% Active (True = 94% And False = 6%) Non active (True = 85% And False = 15%) |

Table 4.7: Accuracy of Random Forest using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2020)	Age, Study time, Failures, Famrel, Free time,Go out, Dalc, Walc Health, Absences G1 , G2,G3	Random Forest, Training and testing data split – 70:30, Dataset -649 rows	Dataset 2 Accuracy = 93%
(2022)	Session, Student, Exercise, Activity, Time begin, Time end, Idle time, Mouse wheel, Mouse Wheel click, Mouse click left, Mouse click right, Mouse movement, Keystrokes	Random Forest, Training and Testing data set - 80:20	Accuracy=97.4%, Precision =0.97, Recall =0.97, F1-score=0.97, ROC=0.98
(2022)	Independent variables Gender (Male, Female), Age (≤ 25, 26 – 35,36 – 45,46 – 55, ≥ 56) Education, Marital status (Single, Married), Job, Initial registration year, Number of registrations, Credits, and GPA, Dependent Variables Status (Not Active, Active)	Random Forest and CART DATA SIZE = 1493, Training data set = 70% And Testing data set= 30%	Accuracy = 96.28% And CART = 96% Training (AUC values = 0.98% CART = 0.95) Testing (AUC values = .98%, CART = 0.95)

Table 4.8: Accuracy of Deep Learning using student academic Performance

References	Attributes and predictive Features	Tools, Methods and algorithm	Result
(2019)	Exam (BA, B.SC), Sub (English, Physics), IN_Sem1, Sem2, Sem3, Sem4, Sem5 and Sem 6(Internal Assessment Marks1, Marks 2, Marks 3, Marks 4, Marks 5, Marks 6), Pc(Percentage), Result(Pass, Fail)	Deep Learning and RNN Data set – 10140	Accuracy = 95.34% Precision – 96% Recall- 99% F-score- 98% Kappa Statistics- 26%
(2019)	Exam (BA, B.SC), Sub (English, Physics), IN_Sem1, Sem2, Sem3, Sem4, Sem5 and Sem 6(Internal Assessment Marks1, Marks 2, Marks 3, Marks 4, Marks 5, Marks 6), Pc(Percentage), Result(Pass, Fail)	Artificial Immune Recognition	Accuracy - 93.18% Precision – 92.6% Recall-93.2% F-score- 92.6% Kappa Statistics – 0.21

(2021)	School_f,Sex_f,Age_f,Address_f ,Famsize_f,Pstatus_f,Medu_f,Fedu_f,Mjob_f,Fjob_f,Reason_f,Guardian_f,Traveltime_f,Studytime_f,Failures_f,Schoolsup_f,Famsup_f,Paid_F,Activities_f,Nursery_f ,Higher_f,Internet_f,Romantic_f, Famrel_f,Freetime_f,Gout_f,Dalc _f,Walc_f,Health_f,Absences_f, G1_f,G2 _f,G3_f (target)	BILSTM (Bidirectional long short-term memory) (With Feature Selection), without Feature Selection Split Data set = 70:30	(Accuracy = 90.16%, P= 90%, R=90%, F-Score = 90%) (Accuracy =88.46 % Precision = 88% Recall= 88% F-Score = 88%)
(2021)	School_f,Sex_f,Age_f,Address_f ,Famsize_fPstatus_f,Medu_f,Fedu_f,Mjob_f,Fjob_f,Reason_f,Guardian_f,Traveltime_f, Studytime_f,Failures_f,Schoolsup_f,Famsup_f,Paid_F,Activities_f ,Nursery_f,Higher_f,Internet_f,Romantic_f,Famrel_f,Freetime_f, Gout_f,Dalc_f,Walc_f,Health_f,Absences_f,G1_f,G2 _f,G3_f (target)	RNN (Recurrent Neural Network), Data set = 1044 records Data set contains = 33 attributes	Accuracy = 81.64%, P= 0.85% ,R= 0.82% ,F-Score = 0.81%
(2022)	Core, Delta, Bifurcation and Ridge Ending	Image Processing Method, Dataset content = 500 finger Print, Images = taken 100 different fingers Take 5 Times One layer Perceptron, Two Layer Perceptron	Accuracy = 97.75%, DPI = 300, Size = 256* 256, One layer Accuracy = 97.50%, MSE = 0.28 Two Layer Acc = 97.75%, MSE = 0.20

6. CONCLUSION

This paper is helpful to study, those techniques which are used to predict student academic performance. The article provides the brief detail of how to measure the performance of students. The main aim of this article defines the new approaches of machine learning algorithms and better understanding of new parameters of student's performance. Study the different algorithms like Decision Tree, Artificial Neural networks, Support Vector Machines, and Naïve Bayes. To achieve the best model, we required different variables like CGPA, Attendance, marks, semester marks, CT, and total marks.

References

1) Aghaei, S., Shahbazi, Y., Pirbabaei, M., &Beyti, H. (2023). A hybrid SEM-neural network method for modeling the academic satisfaction factors of architecture students. *Computers and Education: Artificial Intelligence,* 10012 .[Cross Ref] [Google **Scholar**] **[Publisher Lin**k]

2) Akinrotimi, AkinyemiOmololu & Aremu, Dayo Reuben &Aremu, Dayo Reuben Aremu, DayoReuben(2018) -"Student Performance Prediction Using Random Student Performance Prediction Using Random tree and C4.5 Algorithm" Journal of Digital Innovations & Contemp Res. In Sc., Eng & Tech. Vol. 6, No. 3. Pp 23-34. [Cross Ref] [Google **Scholar**] [Publisher Link]

3) Alfiani, A. P., &Wulandari, F. A. (2015). Mapping student'sTheir performance based on data mining approach (a case study). Agriculture and Agricultural Science Procedia, 3, 173-177. [Cross Ref] [Google **Scholar**] [Publisher Link]

4) Almasri, A., Alkhawaldeh, R. S., &Çelebi, E.(2020). Clustering-based EMT model for predicting student performance. Arabian Journal for Science and Engineering, 45(12), 10067-10078. [Cross Ref] [Google **Scholar**] [Publisher Link]

5) Al-Sahaf, H., Bi, Y., Chen, Q., Lensen, A., Mei, Y., Sun, Y.,& Zhang, M. (2019). A survey on evolutionary machine learning. *Journal of the Royal Society of New Zealand*, *49*(2), 205-228. [Cross Ref] [Google **Scholar**] [Publisher Link]

6) Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*, *2022*. [Cross Ref] [Google **Scholar**] [Publisher Link]

7) Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., &Gani, A. (2022). Data Mining Techniques to Analyze the Impact of Social Media on Academic Performance of High School Students. *Wireless Communications and Mobile Computing*, *2022*. [Cross Ref] [Google **Scholar**] [Publisher Link]

8) Apolinar-Gotardo, M. (2019). Using decision tree algorithm to predict student performance. *Indian Journal of Science and Technology*, *12*, 5. [Cross Ref] [Google **Scholar**] [Publisher Link]

9) Aremu, D. R., Awotunde, J. B., &Ogbuji, E. (2022). Predicting Students Performance in Examination Using Supervised Data Mining Techniques. In *Informatics and Intelligent Applications: First International Conference, ICIIA 2021, Ota, Nigeria, November 25-27,* 2021*: Revised Selected Papers* (p. 63). Springer Nature. [Cross Ref] [Google **Scholar**] [Publisher Link]

10) Asiah, M., Zulkarnaen, K. N., Safaai, D., Hafzan, M. Y. N. N., Saberi, M. M., &Syuhaida, S. S. (2019). A review on predictive modeling technique for student academic performance monitoring. In MATEC Web of Conferences (Vol. 255, p. 03004). EDP Sciences. [Cross Ref] [Google **Scholar**] [Publisher Link]

11) Birgili, B., Seggie, F. N., &Oǧuz, E. (2021). The trends and outcomes of flipped learning research between 2012 and 2018: A descriptive content analysis. Journal of Computers in Education, 8(3), 365-394. [Cross Ref] [Google **Scholar**] [Publisher Link]

12) Brahim, G. B. (2022). Predicting student performance from online engagement activities using novel statistical features. *Arabian Journal for Science and Engineering*, 1-19. [Cross Ref] [Google **Scholar**] [Publisher Link]

13) Burman, I., & Som, S. (2019, February). Predicting students academic performance using support vector machine. In 2019 Amity international conference on artificial intelligence (AICAI) (pp. 756-759). IEEE. [Cross Ref] [Google **Scholar**] [Publisher Link]

14) Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., &Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. Decision Support Systems, 135, 113325. [Cross Ref] [Google Scholar] [Publisher Link]

15) Crivei, L. M., Czibula, G., Ciubotariu, G., &Dindelegan, M. (2020, May). Unsupervised learning based mining of academic data sets for students' performance analysis. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 000011-000016). IEEE. [Cross Ref] [Google Scholar][Publisher Link]

16) Gupta, V., Mishra, V. K., Singhal, P., & Kumar, A. (2022, December). An Overview of Supervised Machine Learning Algorithm. In *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 87-92). IEEE. [Cross Ref] [Google **Scholar**] [Publisher Link]

17) H. Alamri, L., S. Almuslim, R., S. Alotibi, M., K. Alkadi, D., Ullah Khan, I., & Aslam, N. (2020, December). Predicting student academic performance using support vector machine and random forest. In *2020 3rd International Conference on Education Technology Management* (pp. 100-107). [Cross Ref] [Google **Scholar**] [Publisher Link]

18) Hamsa, H., Indiradevi, S., &Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia Technology*, *25*, 326-332. [Cross Ref] [Google **Scholar**] [Publisher Link]

19) HassanA.El-Sabagh,"Adaptive-learning environment based on learning styles and its impact on development students 'engagement(2021) "International Journal of Higher educational technology in Higher Education. [Cross Ref] [Google **Scholar**] [Publisher Link]

20) Hooda, M., Rana, C., Dahiya, O., Shet, J. P., & Singh, B. K. (2022). Integrating LA and EDM for improving students Success in higher Education using FCN algorithm. *Mathematical Problems in Engineering*, *2022*. [Cross Ref] [Google **Scholar**] [Publisher Link]

21) https://www.javatpoint.com/unsupervised-machine-learning.

22) Hussain, S., Muhsin, Z., Salal, Y., Theodorou, P., Kurtoğlu, F., & Hazarika, G. (2019). Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, *14*(8). 10.3991/ijet.v14i08.10001 [Cross Ref] [Google **Scholar**] [Publisher Link]

23) Jadhav, P., (2019). A review paper on student performance using Decision tree algorithm, JETIR March 2019, Volume 6, Issue 3 (ISSN-2349-5162). [Cross Ref] [Publisher Link]

24) Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, *2*(1), 1-25. [Cross Ref] [Google **Scholar**] [Publisher Link]

25) Karo, I. M. K., Fajari, M. Y., Fadhilah, N. U., &Wardani, W. Y. (2022). Benchmarking Naïve Bayes and ID3 Algorithm for Prediction Student Scholarship. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1232, No. 1, p. 012002). IOP Publishing. [Cross Ref] [Google **Scholar**] [Publisher Link]

26) Kausar, S., Oyelere, S., Salal, Y., Hussain, S., Cifci, M., Hilcenko, S.,.&Huahu, X. (2020). Mining smart learning analytics data using ensemble classifiers. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(12), 81-102. [Cross Ref] [Google **Scholar**] [Publisher Link]

27) Kaushik, A., Khan, G., &Singhal, P. (2022, December). Cloud Energy-Efficient Load Balancing: A Green Cloud Survey. In *2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 581-585). IEEE. [Cross Ref] [Google **Scholar**] [Publisher Link]

28) Liza, J. C., & Gerardo, B. D.(2022)A Naïve Bayes Students' Performance Prediction Model for Decision Support System. [Cross Ref] [Google **Scholar**] [Publisher Link]

29) Maheswari, K., Priya, A., Balamurugan, A., & Ramkumar, S. (2021). Analyzing student performance factors using KNN algorithm. *Materials Today: Proceedings*. [Cross Ref] [Google **Scholar**] [Publisher Link]

30) Makhtar, M., Nawang, H., & WAN SHAMSUDDIN, S. N. (2017). ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE BAYES CLASSIFIER. Journal of Theoretical & Applied Information Technology, 95(16). [Cross Ref] [Google **Scholar**] [Publisher Link]

31) Masangu, L., Jadhav, A., &Ajoodha, R. (2021). Predicting Student Academic Performance Using Data Mining Techniques. *Advances in Science, Technology and Engineering Systems Journal*, *6*(1), 153-163. [Cross Ref] [Google **Scholar**] [Publisher Link]

32) Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM, 42(11), 30-36.

[Cross Ref] [Google **Scholar**] [Publisher Link]

33) Mendy Nefale, M., & Ajoodha, R. (2023, April). Using Numerous Biographical and Enrolling Observations to Predict Student Performance. In *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications: ICAIAA 2022* (pp. 649-660). Singapore: Springer Nature Singapore. [Cross Ref] [Google Scholar][Publisher Link]

34) Nguyen, H. T., Afanasiev, A. D., & Nguyen, T. V. (2022). Automatic identification fingerprint based on machine learning method. *Journal of the Operations Research Society of China*, *10*(4), 849-860. [Cross Ref] [Google **Scholar**] [Publisher Link] .

35) Omolewa, O. T., Oladele, A. T., Adeyinka, A. A., & Oluwaseun, O. R. (2019). Prediction of student's academic performance using k-means clustering and multiple linear regressions. *Journal of Engineering and Applied Sciences*, *14*(22), 8254-8260. [Cross Ref] [Google Scholar][Publisher Link]

36) Lestari, A. F. APPLICATION OF DECISION TREE AND NAIVE BAYES ON STUDENT PERFORMANCE DATASETS. [Cross Ref] [Google **Scholar**] [Publisher Link]

37) Ragab, M., Abdel Aal, A. M., Jifri, A. O., &Omran, N. F. (2021). Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. *Wireless Communications and Mobile Computing*, *2021*. [Cross Ref] [Google **Scholar**] [Publisher Link]

38) Smirani, L. K., Yamani, H. A., Menzli, L. J., &Boulahia, J. A. (2022). Using ensemble learning algorithms to predict Student failure and enabling.customized educational paths. *Scientific Programming*, *2022*. [Cross Ref] [Google **Scholar**] [Publisher Link]

39) Syahputra, Y. H., &Hutagalung, J.(2022). Superior class to improve student achievement using the K-means algorithm. *Sinkron: jurnaldanpenelitianteknikinformatika*, *7*(3), 891-899. [Cross Ref] [Google **Scholar**] [Publisher Link]

40) Thai, H. T. (2022, April). Machine learning for structural engineering: A state-of-the-art review. In *Structures* (Vol. 38, pp. 448-491). Elsevier. [Cross Ref] [Google **Scholar**] [Publisher Link]

41) Troussas, C., Giannakas, F., Sgouropoulou, C., &Voyiatzis, I. (2020) . Collaborative activities

42) Wiering, M. A., & Van Otterlo, M. (2012). Reinforcement learning. *Adaptation, learning, and optimization*, *12*(3), 729. [Cross Ref] [Google Scholar] [Publisher Link]

43) Wiyono, S., & Abidin, T. (2019). Comparative study of machine learning KNN, SVM, and decision tree algorithm to predict student's performance. *International Journal of Research-Granthaalayah*, *7*(1), 190-196. [Cross Ref] [Google **Scholar**]

44) Xing, W. (2019). Exploring the influences of MOOC design features on student performance and persistence. *Distance Education*, *40*(1), 98-113. [Cross Ref] [Google Scholar] [Publisher Link]

45) Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 1-19. [Cross Ref] [Google **Scholar**] [Publisher Link]

46) Yousafzai, B. K., Khan, S. A., Rahman, T., Khan, I., Ullah, I., Ur Rehman, A., &Cheikhrouhou, O. (2021). Student-performulator: student academic performance using hybrid deep neural network. *Sustainability*, *13*(17), 9775. [Cross Ref] [Google Scholar] [Publisher Link]

47) Kumar, A., Gupta, V., Chaudhary, S., Gupta, V. K., & Kumar, P. (2023). Robust Dynamic Clustering System: An Architecture for Scalable Web Services. [Cross Ref] [Google Scholar]

48) V. Gupta, P. Singhal and V. Khattri, "Student Performance Using Antlion Optimization Algorithm and ANN Regression," *2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART)*, Moradabad, India, 2023, pp. 468-471, doi: 10.1109/SMART59791.2023.10428404.