Stochastic Modelling and Computational Sciences

PERFORMANCE EVALUATION OF DIFFERENT MACHINE LEARNING ALGORITHMS ON STUDENT DATASET CLUSTERED BY K-MEANS ALGORITHM

Gautam Appasaheb Kudale and Dr. Sandeep Singh Rajpoot
Computer Science, Dr. A.P.J. Abdul Kalam University, Indore, M.P., India
gaukudale@gmail.com and sandeepraj413@gmail.com

ABSTRACT
This research paper presents an innovative approach to cluster the dataset and apply the different machine learning algorithms. For clustering, K-Means is used. Before applying K-Means, estimation of value of K is done by using the Elbow method followed by Silhouette method. For the entire experimental work, the Kaggle dataset has been taken into account. K-Means clustering used over here offers five different clusters. Here, each cluster shows identification of distinct student. Subsequently, four prominent machine learning algorithms—K-Nearest Neighbor (KNN), Neural Network (NN), Random Forest (RF), and Support Vector Machine (SVM) are applied and the performance metrics are measured. The comparative analysis of the machine learning algorithms reveals varying levels of accuracy, precision, recall, and F1-score across different clusters. The outcomes highlight the algorithm that exhibits superior performance in this specific context. Here the highest accuracy i.e. 92.00% is achieved with Random Forest algorithm.

Keywords: K-means clustering, K-Nearest Neighbors, Random Forest, Support Vector Machine, Neural Network

1. INTRODUCTION
In today's era of artificial intelligence, Machine Learning algorithms are become more and more popular. Day by day, by varying certain parameters, different moderations and inventions are being implemented. For every kind of estimation and classification task, concerned machine learning algorithms are available.

This paper tries to perform an experiment on the students dataset obtained from Kaggle. The experimental work starts with cleaning and preprocessing the said dataset. Further, K-Means algorithm gives different kinds of clusters depending upon the students characteristics.

Four different machine learning algorithms have been taken into consideration so as to apply on the clusters yielded by K-Means. These four machine learning algorithms are – a) Support Vector Machine, b) Random Forest, c) K-Nearest Neighbors and d) Neural Network. The result of the experiment shows the graph from where the most efficient machine learning algorithm having the highest accuracy on student dataset has been sorted out.

2. RELATED WORK
Jehad Ali et al. (2012) worked on twenty versatile datasets with two machine learning algorithms random forest and J48 (a decision tree algorithm). The study finds that Random Forest performs better for datasets with a larger number of instances, while J48 is more effective for datasets with fewer instances. Specifically, Random Forest shows higher accuracy when the number of instances increases, indicating its ability to handle large datasets effectively. [39]

Cherry D. Casuat et al. (2019), predicted students' employability using machine learning algorithms i.e. Decision Trees, Random Forest, and Support Vector Machine, they observed that Support Vector Machine (SVM) obtained the highest accuracy of 91.22%. The authors concluded that the SVM model was not overfit or underfit.[35]

Slamet Wiyono et al. (2019) predicted student performance to prevent inactive students and increase the number of students graduating on time. After comparing KNN, SVM and DT algorithms on student's data, they concluded that the SVM algorithm got the highest accuracy, achieving 95% precision in predicting student performance. The Decision Tree algorithm has shown an accuracy of 93%, and the KNN algorithm has shown accuracy of 92%. [26]

(Seyhmus Aydogdu, 2020) In his research focused on predicting the performance of university students in an e-learning environment using artificial neural networks. In his study he observed that artificial neural network achieved an accuracy of 80.47% in predicting student performance. The study also examined the contribution of input variables to the prediction of the output variable. [41]

Leena H. Alamri et al. (2020) in their research on datasets which consisted of records for mathematics and Portuguese language courses, used classification algorithms SVM and RF which has shown both SVM and RF algorithms applied to the datasets gave high accuracy in binary classification, reaching 93% accurate prediction. [15]

Muhammed Berke YILDIZ et al. (2020) focused on ninth-grade students' data, utilizing data mining techniques to address a classification problem, specifically predicting academic success. Their study has compared various supervised classification algorithms like ADTree, JRip, NN, Naive Bayes, and J48, with the Neural Network algorithm achieving the highest accuracy (98.6%). [40]

3. METHODOLOGY

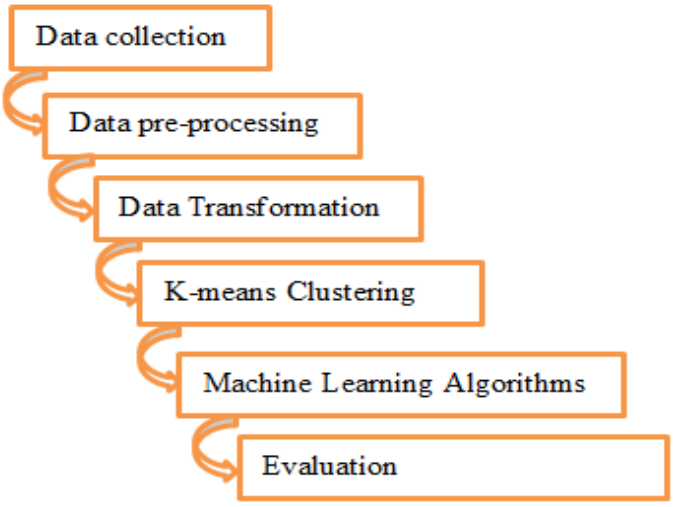

Fig. 1: Methodology of Research Work

The dataset utilized in this research is obtained from open source repository Kaggle, (https://www.kaggle.com/datasets/spscientist/students-performance-in-exams) which includes a comprehensive set of attributes related to student profiles and their academic history. The dataset is having one thousand records and total eight attributers. The dataset has the attributes "gender", "race/ethnicity", “parental level of education”, “lunch”, "test preparation course", "maths score", "reading score" and "writing score".

Before proceeding to further experiments, data pre-processing is achieved by using Vector Assembler and Standard Scalar. After this, Principal Component Analysis algorithm is applied to reduce the un-necessary dimensions.

3.1 K-MEANS CLUSTERING

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process.

Estimation of value of K - Elbow method and Silhouette method:

The dataset used for this experimental work contain heterogeneous rows. To segregate the records, we have used the K-Means algorithm. The K-Means is a clustering technique which is used as a preprocessing step to categorize the students having similar characteristics into distinct clusters.

First of all, we have estimated the value of K using Elbow plot and Silhouette plot. We have fed thousand records to plot the Elbow plot where a graph of K versus WCSS. Here WCSS is called as within Cluster Sum of Squares which is defined by the formula.

$$WCSS = \sum_{C_k}^{C_n}\left(\sum_{d_i in C_i}^{d_m} distance(d_i, C_k))^2\right) \quad (1)$$

Where C is the cluster centroid and d is the data point in each cluster. *Ci* indicates the k clusters from i=1 to k and *di* is the data point in each cluster.

Similarly, Silhouette plot is a plot of K versus Silhouette Score. The equation to calculate the Silhouette coefficient for a particular data point is given by the formula.

$$S(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (2)$$

Where-

S(i) is the silhouette coefficient of the data point i.

a(i) is the average distance between i and all the other data points in the cluster to which i belongs.

b(i) is the average distance from i to all clusters to which i does not belong.

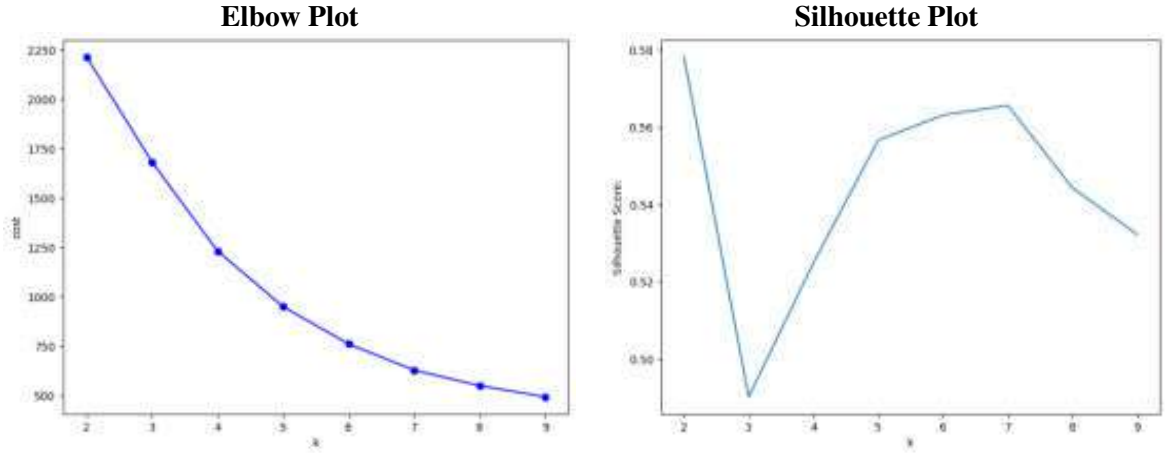

Fig. 2: Elbow Plot and Silhouette Score Plot for PCA features

The Fig. 2. shows the two plots Viz Elbow plot and Silhouette plot respectively. Traditionally, from the elbow plot, the correct value of k is estimated by observing the first or most significant turning point of the curve [4]. In our experimental work, we have achieved multiple significant turning points for the values of k like k=4, k=5, k=6, k=7 or any of these. Still here, estimation of correct value of k remains a challenge.

To estimate the exact value of k, we have calculated Silhouette Score and plotted the Silhouette plot. The Fig 2 also shows the Silhouette graph where global maximum or optimal value of k is clearly seen. Here, we have obtained the value k=5, which is also present in the list of our elbow plot. So, we have fixed value of k = 5 as a number of clusters.

3.2 ALGORITHM K-MEANS (K, D)

Step l: Identify the k data points as the initial centroids (cluster centers).

Step 2: Repeat

a) For every data point x ∈ D do

b) Calculate the distance from x to the centroid.

Assign x to the closest centroid (a centroid represents a cluster).

Re-calculate the centroids using the current cluster memberships.

Until centroids do not alter:

Stopping criteria

i. Centroids of newly formed clusters do not vary.
ii. Points should remain in the same cluster.
iii. Maximum numbers of iterations or cycles are reached.

In short, the iteration will be continued until the convergence so as to form the clusters. [5, 8, 18, 25, 27, 28].

In this study, the entire experimental work is carried out using Python programming language with the help of open source tool Jupyter. The data set is divided in to 80:20 for training and testing respectively. For k-means clustering pyspark python library is used.

4. ADVANCED MACHINE LEARNING ALGORITHMS

Four advanced machine learning algorithms are taken into consideration whose accuracy has to be evaluated with respect to student dataset. These are Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN) and Neural Network (NN).

4.1 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks.

Here we have used the Non-Linear SVM that classifies the data within each cluster. For this purpose it uses the corresponding kernel functions. SVMs are adaptable and efficient in a variety of applications because they can manage high-dimensional data and nonlinear relationships. [30, 31, 32, 33, 34, 35, 36]

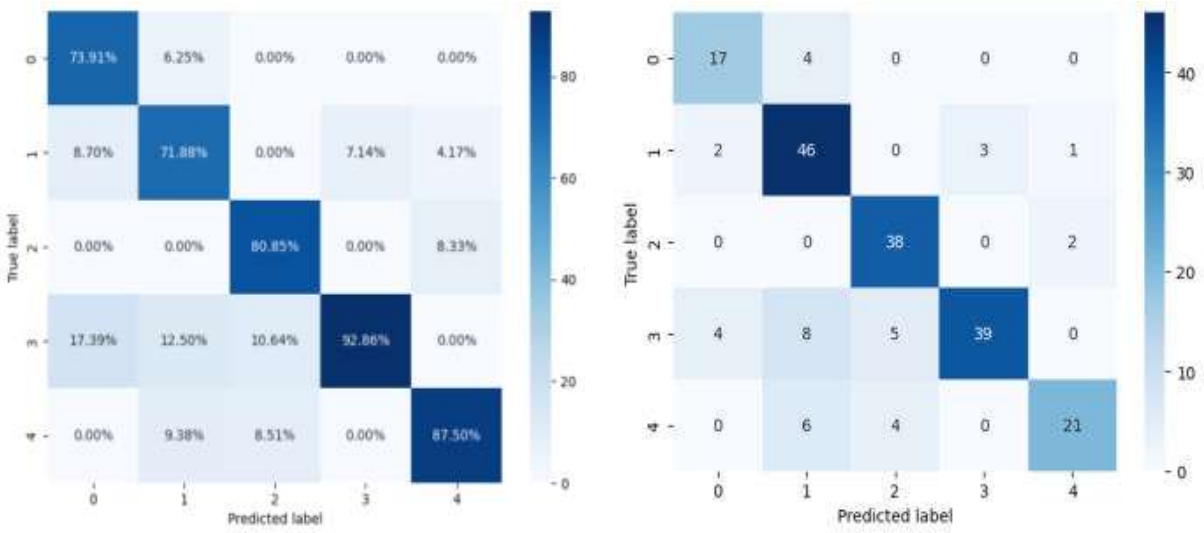

Fig. 3: Confusion Matrix obtained after applying the SVM algorithm

Table. 1: Classification Report of Support Vector Machine (SVM)

	Cluster	Precision	Recall	F1-Score	Support
	0	0.74	0.81	0.77	21
	1	0.72	0.88	0.79	52

	2	0.81	0.95	0.87	40
	3	0.93	0.7	0.8	56
	4	0.88	0.68	0.76	31
Accuracy				0.81	200
Macro Average		0.81	0.8	0.8	200
Weighted Average		0.82	0.81	0.8	200

The above table shows the classification report of SVM for testing dataset which is comprised of 20% of the dataset. Here, we have got an 81% accuracy for testing. Moreover, the metrics such as precision, recall, f1score and support values are calculated and shown in the table for each cluster.

4.2 RANDOM FOREST (RF)

This is the second machine learning model considered for the experimental work. It is used for classification and regression tasks. [15, 16, 17, 18, 19, 20, 21]. The purpose behind using this algorithm is just classification.

Here, we have used Random forest algorithm to classify the testing dataset comprising of 200 records which is 20% of total dataset. This algorithm is effective for complex and high-dimensional datasets. Random Forest combines multiple decision trees to create a robust and accurate prediction model.

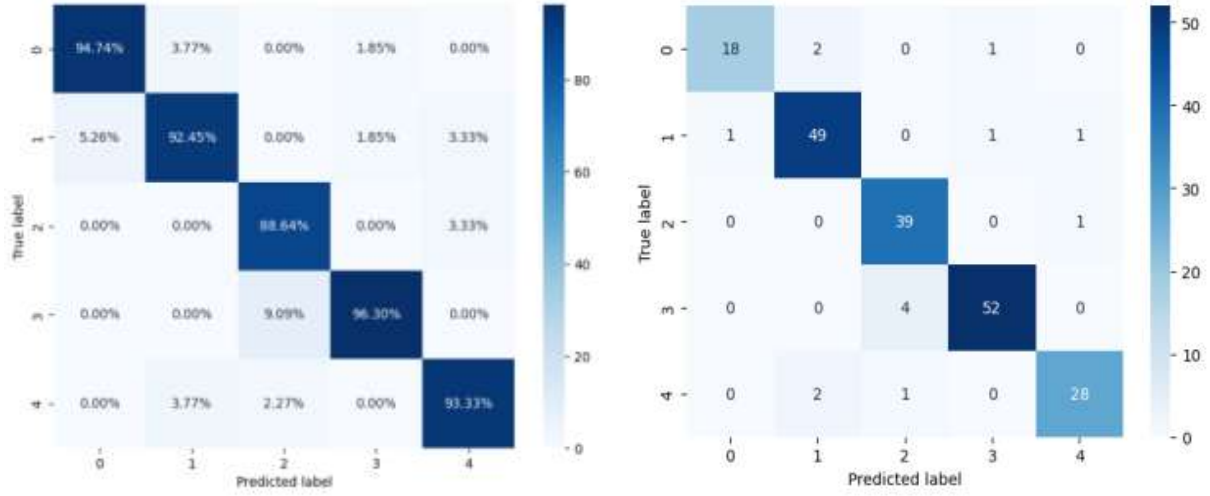

Fig. 4: Confusion Matrix obtained after applying the RF algorithm

Table. 2: Classification Report of Random Forest (RF)

	Cluster	Precision	Recall	F1-Score	Support
	0	0.94	0.81	0.87	21
	1	0.9	0.9	0.9	52
	2	0.93	0.97	0.95	40
	3	0.93	0.95	0.94	56
	4	0.9	0.9	0.9	31
Accuracy				0.92	200
Macro Average		0.92	0.91	0.91	200
Weighted Average		0.92	0.92	0.92	200

This model yields testing accuracy of 92%. For this also, we have calculated the metrics such as precision, recall, f1-score and support parameters.

4.3 K-NEAREST NEIGHBOR (KNN)

We have also taken into consideration the KNN Machine learning model to apply on all of the clusters. K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning. In this K-NN, to find the correct neighbor, the algorithm uses Euclidean distance as a distance metric.

Euclidean distance is calculated by using the below formula. [22, 24, 27]

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

Where (x1, y1) and (x2, y2) are two points in a two-dimensional plane.

KNN algorithm, at its training phase, just stores the dataset, when it gets new data; it classifies that data into a category that is much similar to the new data. [22, 23, 24, 25, 26, 27]

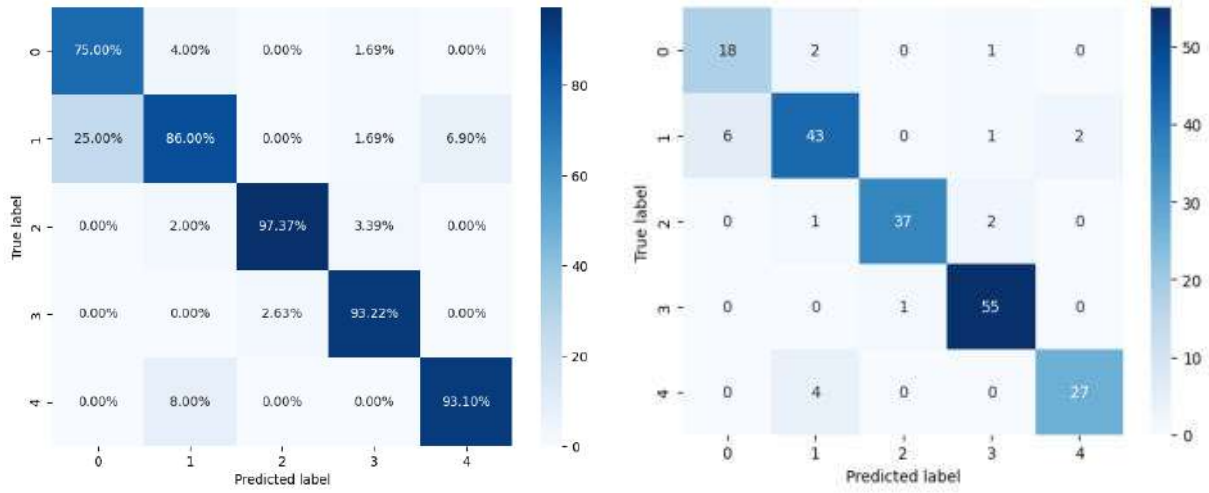

Fig. 5: Confusion Matrix obtained after applying the KNN algorithm

Table. 3: Classification Report of K Nearest Neighbors (KNN)

	Cluster	Precision	Recall	F1-Score	Support
	0	0.85	0.81	0.83	21
	1	0.79	0.85	0.81	52
	2	0.83	0.88	0.85	40
	3	0.88	0.79	0.83	56
	4	0.78	0.81	0.79	31
Accuracy				0.82	200
Macro Average		0.83	0.82	0.82	200
Weighted Average		0.83	0.82	0.83	200

From the above table, it is clear that obtained accuracy for testing is 82%. Also here the performance metrics such as precision, recall, f1-score and support are calculated for each cluster.

4.4 NEURAL NETWORK (NN)

The next classifier used over here is Multi-layer Perceptron. This MLP Classifier trains iteratively since at each time step, the partial derivatives of the loss function with respect to model the parameters are computed to update the parameters.

Neural networks symbolize the convergence of artificial intelligence and brain-inspired architecture, driving a paradigm shift in contemporary computing. [32, 35]

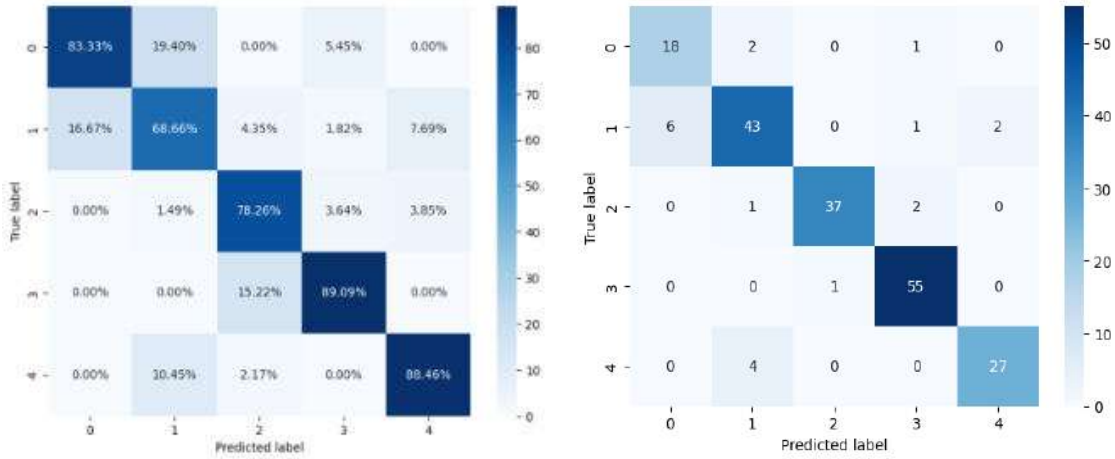

Fig. 6: Confusion Matrix obtained after applying NN algorithm.

Table. 4: Classification Report of Neural Network (NN)

	Cluster	Precision	Recall	F1-Score	Support
	0	0.66	0.9	0.76	21
	1	0.8	0.71	0.76	52
	2	0.95	0.88	0.91	40
	3	0.89	1	0.94	56
	4	0.92	0.74	0.82	31
Accuracy				0.85	200
Macro Average		0.84	0.85	0.84	200
Weighted Average		0.86	0.85	0.85	200

This Multi-layer perceptron model has given testing accuracy of 85%. For this also, here the performance metrics such as precision, recall, f1-score and support parameters are calculated for each cluster.

5. PERFORMANCE METRICS USED IN THE EXPERIMENTAL WORKS

Evaluation metrics such as accuracy, precision, recall, support and F1-score are employed to gauge the algorithms' effectiveness in predicting academic performance across different student clusters. Macro average and weighted average g is also calculated. These metrics offer insights into both the overall predictive accuracy and the algorithms' ability to correctly identify the students at risk of underperformance or excelling.

5.1 ACCURACY

Accuracy is a fundamental performance metric used in machine learning to measure the overall correctness of a model's predictions, especially in classification tasks. It quantifies the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances in the dataset. In other words, accuracy tells you how accurate or correct your model's predictions are compared to the ground truth. [18, 20, 33, 35, 36]

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Instances} \quad (4)$$

5.2 PRECISION

Precision is a performance metric used in machine learning algorithms, especially in the context of classification tasks. It measures the accuracy of the positive predictions made by the model. In other words, precision focuses on the quality of the positive predictions, specifically the ratio of correctly predicted positive instances to the total instances predicted as positive [18, 20, 23, 35, 36]

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (5)$$

5.3 RECALL

Recall also known as Sensitivity or True Positive Rate, is another important performance metric used in machine learning algorithms, particularly in classification tasks. Recall measures the ability of a model to correctly identify all relevant instances from the total actual positive instances. In simpler terms, it quantifies the model's ability to capture all positive instances. [18, 20, 23, 33, 35, 36]

$$Recall\ or\ Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (6)$$

5.4 F1-SCORE

F1-Score is a widely used performance metric in machine learning, especially in classification tasks, that balances both precision and recall into a single value. It is particularly useful when there is a need to consider both false positives and false negatives, and there is a desire to find a balance between them. The F1-score is the harmonic mean of precision and recall, combining both metrics into a single value that takes into accounts the trade-off between them. [18, 20, 23, 35, 36]

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

In the context of machine learning and data analysis, "support" usually refers to the number of occurrences or instances of a particular class or category within a dataset. It is particularly relevant in classification tasks, where you are trying to categorize data points into different classes or labels.

Conventions used in the above expressions 1 to 4:

- True Positives (TP) are the instances that are correctly predicted as positive by the model.
- False Positives (FP) are the instances that are actually negative but are incorrectly predicted as positive by the model.
- True Negatives (TN) are the instances that are correctly predicted as negative by the model.
- False Negatives (FN) are the instances that are actually positive but are incorrectly predicted as negative by the model.
- Total Instances is the sum of true positives, true negatives, false positives, and false negatives.
- Precision is the ratio of true positives to the sum of true positives and false positives.
- Recall is the ratio of true positives to the sum of true positives and false negatives

6. RESULT AND DISCUSSION

Table. 5: Algorithms and their Accuracy Percentage

Algorithm	Accuracy
KNN	0.82
NN	0.85
RF	0.92
SVM	0.81

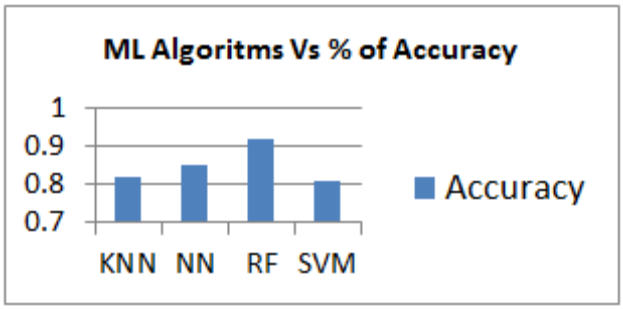

Fig. 7: Column Chart of ML Algoritms Vs % of Accuracy

We have made the Kaggle dataset ready for our experiment by cleaning and pre-processing. After this, we estimated the correct value of K by applying Elbow method followed by Silhouette method where we got the value of K=5.

Subsequently, we applied K-means algorithm on the dataset where we got 5 discrete clusters. On each cluster, we applied different Machine Learning Algorithms Viz – SVM, RF, KNN and NN. The Performance Metrics for all of the Machine Learning Algorithms are calculated and exhibited in the form of table as well as a comparative graph. The table 5 above shows the accuracy of each machine learning algorithm that is used throughout the experiments.

7. CONCLUSION

The accuracy obtained with KNN machine learning algorithm on the cluster is 82%. The accuracy obtained with Neural Network algorithm is 85%. The accuracy with Random Forest algorithm is 92%. Also the accuracy with Support Vector Machines is 81%. This clearly shows that maximum accuracy is obtained with Random Forest machine learning model. This is how, by calculating the accuracy, we have evaluated the performance of each machine learning model used over here. So, we can conclude that the Random Forest is well suited and efficient machine learning algorithm for such kind of classification of student dataset.

8. FUTURE WORK

Considering the strengths of different algorithms, future research could explore hybrid models that combine the benefits of diverse algorithms. This approach might yield even more accurate and robust predictions.

REFERENCES

[1] Margaret H. Dunhan, “Data Mining Introductory and Advanced Topics”, Pearson

[2] Ian H.witten, Eibe Frank, Mark A. Hall, “Data Mining Practical Machine Learning Tools and Techniques”, 3rd Edition

[3] Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, “Mining of Massive Datasets”, 2nd Edition

[4] Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining, Concepts and Techniques”, 3rd Edition

[5] Alvaro Fuentes, “Hands-on Predictive Analytics with Python”

[6] Richard O. Duda, Peter E. Hart, David G. Stork, “Pattern Classification”, 2nd Edition

[7] Satish Kumar, “Neural Networks A classroom Approach”, 2nd Edition

[8] Prof. Prashant Sahai Saxena, Prof. M. C. Govil, “Prediction of Student’s Academic Performance using Clustering,” Special Conference Issue: National Conference on Cloud Computing & Big Data

[9] Oyelade, O. J, Oladipupo, O. O., Obagbuwa, I. C., “Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance,” (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010

[10] Rakesh Kumar Arora, Dr. Dharmendra Badal, "Evaluating Student's Performance Using k-Means Clustering," International Journal of Computer Science And Technology, IJCST Vol. 4, Issue 2, April - June 2013, ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

[11] Mr. Shashikant Pradip Borgavakar, Mr. Amit Shrivastava, "Evaluating student's performance using k-means clustering," International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 6 Issue 05, May – 2017

[12] Dr. G. Rajitha Devi, "Prediction of student academic performance using clustering," International Journal of Current Research in Multidisciplinary (IJCRM) ISSN: 2456-0979 Vol. 5, No. 6, (June'20), pp. 01-05

[13] Yann Ling Goh, Yeh Huann Goh, Chun-Chieh Yip, Chen Hunt Ting, Raymond Ling Leh Bin, Kah Pin Chen, "Prediction of students' academic performance by k-means clustering," Peer-review under responsibility of 4th Asia International Multidisciplinary Conference 2020 Scientific Committee

[14] Revathi Vankayalapati, Kalyani Balaso Ghutugade, Rekha Vannapuram, Bejjanki Pooja Sree Prasanna, "K-means algorithm for clustering of learners performance levels using machine learning techniques," Revue d'Intelligence Artificielle Vol. 35, No. 1, February, 2021, pp. 99-104

[15] Leena H. Alamri, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, Nida Aslam ICETM 2020, December 17–19, 2020, London, United Kingdom Predicting Student Academic Performance using Support Vector Machine and Random Forest.

[16] Gorinkala Hemasri, Kalla.Kiran International Journal for Research Trends and Innovation © 2022 IJRTI | Volume 7, Issue 12 | ISSN: 2456-3315 Students Performance Prediction Using Random Forest Algorithm

[17] Saba Batool , Junaid Rashid, 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), A Random Forest Students' Performance Prediction (RFSPP) Model Based on Students' Demographic Features.

[18] Sujith Jayaprakash, Sangeetha Krishnan, Jaiganesh V, 2020 International Conference on Emerging Smart Computing and Informatics (ESCI) AISSMS Institute of Information Technology, Pune, India. Mar 12-14, 2020 , Predicting Students Academic Performance using an Improved Random Forest Classifier.

[19] Vrushali Y Kulkarni, Pradeep K Sinha, International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, May 2014, Effective Learning and Classification using Random Forest Algorithm.

[20] Shaohai Huang1 and Junjie Wei, Scientific Programming Volume 2022, Article ID 9340434, 9 pages https://doi.org/10.1155/2022/9340434, Student Performance Prediction in Mathematics Course Based on the Random Forest and Simulated Annealing

[21] P.Ajay, M.Pranati, M.Ajay, P.Reena, T. BalaKrishna International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 08 | Aug 2020 Prediction Of Student Performance Using Random Forest Classification Technique.

[22] Moohanad Jawthari and Veronika Stoffov An International Journal for Engineering and Information Sciences DOI: 10.1556/606.2021.00374 © 2021 Predicting students' academic performance using a modified KNN algorithm

[23] Tuomas Tanner and Hannu Toivonen Predicting and preventing student failure – using the k-nearest neighbour method to predict student performance in an online course environment

[24] Thomas Asril, Sani M. Isa International Journal of Emerging Trends in Engineering Research Emerging Trends in Engineering Res ISSN 2347 – 3983 https://doi.org/10.30534/ijeter/2020/60862020 Prediction of Students Study Period using K-Nearest Neighbor Algorithm

[25] Ihsan A. Abu Amra, Ashraf Y. A. Maghari 2017 8th International Conference on Information Technology (ICIT) Students Performance Prediction Using KNN and Naïve Bayesian

[26] Slamet Wiyono, Dega Surono Wibowo, M. Fikri Hidayatullah and Dairoh International Journal Of Computing Science And Applied Mathematics, Vol. 6, No. 2, August 2020 Comparative Study of KNN, SVM and Decision Tree Algorithm for Student's Performance Prediction

[27] A Seetharam Nagesh, Ch V S Satyamurty and K Akhila CVR Journal of Science and Technology, Volume 13, December 2017 ISSN 2277 – 3916 Predicting Student Performance using KNN Classification in Bigdata Environment

[28] DIKSHA A. BANSOD, PROF. AMIT D. SHAH 2021 IJCRT International Journal of Creative Research Thoughts (IJCRT) | Volume 9, Issue 8 August 2021 | ISSN: 2320-2882 A Review Of Student Performance Prediction Techniques In Virtual Learning Environment

[29] J. Dhilipan, N.Vijayalakshmi, S.Suriya, Arockiya Christopher IOP Conf. Series: Materials Science and Engineering 1055 (2021) 012122 IOP Publishing doi:10.1088/1757-899X/1055/1/012122 Prediction of Students Performance using Machine learning

[30] Slamet Wiyono , Taufiq Abidin ©International Journal of Research – GRANTHAALAYAH Vol.7 (Iss.1): January 2019 ISSN- 2350-0530(O), ISSN- 2394-3629(P) DOI: 10.5281/zenodo.2550651 Comparative Study of Machine Learning KNN, SVM, and Decision Tree Algorithm to Predict Student's Performance.

[31] Tatiana A. Cardonaa, Elizabeth a. Cudneya 25th International Conference on Production Research Manufacturing Innovation: Cyber Physical Manufacturing August 9-14, 2019 | Chicago, Illinois (USA Predicting Student Retention Using Support Vector Machines.

[32] Engr. Sana Bhutto Dr. Qasim Ali Arain Dr. Isma Farah Siddiqui Maleeha Anwar 2020 International Conference on Information Science and Communication Technology Predicting Students' Academic Performance Through Supervised Machine Learning.

[33] Iti Burman, Subhranil Som 978-1-5386-9346-9/19/$31.00 ©2019 IEEE Predicting Students Academic Performance Using Support Vector Machine.

[34] Yulei Pang, Nicolas Judd, Joseph O'Brien, Michael Ben-Avie 978-1-5090-5920-1/17/$31.00 ©2017 IEEE Predicting Students' Graduation Outcomes through Support Vector Machines.

[35] Cherry D. Casuat, Enrique D. Festijo 2019 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS) Predicting Students' Employability using Machine Learning Approach.

[36] Pooja Kumari, Praphula Kumar Jain , Rajendra Pamula 4th Int'l Conf. on Recent Advances in Information Technology | RAIT-2018 |An Efficient use of Ensemble Methods to Predict Students Academic Performance.

[37] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua et al., ,Neurocomputing, https://doi.org/10.1016/j.neucom.2019.10.118. A comprehensive survey on support vector machine classification: Applications, challenges and trends.

[38] Sri. Udaya Damuluri , Khondkar Islam, Pouyan Ahmadi, Namra Qureshi Emerging Science Journal Vol. 4, No. 4, August, 2020 Analyzing Navigational Data and Predicting Student Grades Using Support Vector Machine.

[39] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012 ISSN (Online): 1694-0814 Random Forests and Decision Trees.

[40] Muhammed Berke YILDIZ, Caner BÖREKCİ Journal of Educational Technology & Online Learning Volume 3 │Issue 3│2020 https://dergipark.org.tr/tr/pub/jetol Predicting Academic Achievement with Machine Learning Algorithms.

[41] Şeyhmus Aydoğdu Education and Information Technologies (2020) 25:1913–1927 https://doi.org/10.1007/s10639-019-10053-x Predicting Student Final Performance Using Artificial Neural Networks in Online Learning Environments.