I.J. Modern Education and Computer Science, **2023, 1, 46-57**
Published Online on February 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2023.01.04

Analysis of Student's Academic Performance based on their Time Spent on Extra-Curricular Activities using Machine Learning Techniques

Neeta Sharma*
Associate Professor, Department of CS, Noida International University, Noida, India
Email: neeta.msit@gmail.com
ORCID iD: https://orcid.org/0000-0002-9288-3239
*Corresponding Author

Shanmuganathan Appukutti
Professor & Principal, Northern Institute of Engineering and Technology, Alwar, India
Email: profshan.sai@gmail.com
ORCID iD: https://orcid.org/0000-0002-7759-5343

Umang Garg
Assistant Professor, Department of CSE, Graphic Era Hill University, Dehradun, India
Email: umangarg@gmail.com
ORCID iD: https://orcid.org/0000-0002-1815-5794

Jayati Mukherjee
Assistant Professor, Department of CSE, Noida International University, Noida, India
Email: jayati.uit93@gmail.com
ORCID iD: https://orcid.org/ 0000-0003-3433-8141

Sneha Mishra
Assistant Professor, Department of CSE, Noida International University, Noida, India
Email: snesh_shukla@live.in
ORCID iD: https://orcid.org/0000-0002-6234-7670

Received: 23 October, 2022; Revised: 19 December, 2022; Accepted: 12 January, 2023; Published: 08 February, 2023

Abstract: The foundational tenet of any nation's prosperity, character, and progress is education. Thus, a lot of emphasis is laid on quality of education and education delivery system in India with current financial year (2022-23) education budget outlay of Rs. 1,04,277.72 crores. This research contributes in analyzing how students perform in academics depending upon the time spent on their extracurricular activities with the help of three Machine Learning prediction algorithms namely Decision Tree, Random Forest and KNN. Additionally, in order to comprehend the underlying causes of the shortcomings in each machine learning technique, comparisons of the prediction outcomes obtained by these various techniques are made. On our dataset, the Decision Tree outscored all other algorithms, achieving F1 84 and an accuracy of 85%. The research, which is at an introductory level, is meant to open the door for more complexes, specialised, and in-depth studies in the area of predicting the performance in academics.

Index Terms: Algorithms, Extra-curricular activities, KNN, Decision Tree, Random Forest, Machine Learning, Prediction.

1. Introduction

It is impossible to overstate the significance of quality education for students' career development, universities' performance advancement, and of course for country building. Today's fiercely competitive and disruptive corporate world makes it even more important to assess educational quality. There are numerous approaches to track students' academic progress and learning habits using tried-and-true questionnaire-based analysis methods to identify the variables influencing academic achievement. However, the majority of these studies only present a post-mortem analysis of the issue or remedy. What if we could use machine learning to forecast the students who could drop out in a specific semester. Such knowledge would provide us the opportunity to act quickly to improve pupils' academic performance [1,2].

Due to the Corona pandemic, student performance in academics has declined and drop-out numbers have increased which calls to taking an active approach for finding a solution(s) that not only uses postmortem methods to analyse how/why many students failed or dropped out, but also uses machine learning (ML) to forecast and find in advance that the students who are likely to perform below-par.

1.1. Classifiers and Predictors:

Numerous predictors may have an impact on a student's academic achievement. Being a pilot study, we only included one predictor—the amount of time spent participating in extracurricular activities. In order to recommend the best prediction algorithm to educational institutions, the research also compares the accuracy levels of several prediction algorithms. Three classifiers—Decision Tree, Random Forest Tree, and KNN—are employed because they are the simplest and most effective supervised learning algorithms. When classifying a new case based on similarities discovered, KNN bases its work on the similarity between fresh data and existing cases. Therefore, the KNN algorithm can quickly classify fresh data that appears at a university into the appropriate category. Decision Tree and Random Forest Tree being supervised learning model works by solving regression and classification problems.

Furthermore, depending on F1 score and accuracy, all deployed models are contrasted. The F1 Score formula is as follows:

$$F1 = 2 * \frac{(presion * recall)}{(presion + recall)} \tag{1}$$

In order to compare the models' predictions, the F1 Score and Accuracy parameters have been used as measures of model performance. Because it is a reliable indicator of an algorithm's effectiveness. Future studies could include the prediction of academic performance based on more characteristics.

1.2. Contribution:

In this paper, the contribution lies in:

1. Extracurricular activities have huge impact on students' academic performance, which we have tried to investigate in this article. As far as we are aware, no solid paper has yet been proposed for this type of research work
2. The application of machine learning techniques on education analysis has been less explored, which we have try taken up in our article. We have used three popular algorithms for supervised machine learning namely Random Forest Decision Tree, and K-NN on our dataset.
3. We have reduced the manual intervene by automating the parameter tuning using advance Python library.
4. We have done the analysis in the detailed manner with the help of Confusion Matrix, RoC Curve and Precision Table. Finally, the results are compared, and best algorithm is found on the basis of F1 Score and accuracy.

In out paper, the authors have defined the objectives of the study, methodologies, viewing how different ML algorithms, such as namely Decision Tree, Random Forest, and K-NN perform. The paper is structured and described starting with Introduction than thorough recent Literature review, Methodology, Implementation of algorithms, results and discussion, and finally conclusion.

2. Literature Review

In the current year (2022), Neeta Sharma et. el. examines the value of analysing academic performance in the year 2022. Additionally using extracurricular activities as a predictor, the authors' study. On their selected predictor, they use Logistic Regression (LR) and the K-NN model. In comparison to KNN, the accuracy and F1 Score of logistic regression are the highest. As part of a future project, the authors advise implementing SVM, MLP, Decision Trees, ANN, NB (Naive Bayes) etc. [9].

In the current year (2022), Qiu et. el. highlighted the use of information technology in the eLearning process in terms of modern education to improve quality of educational that may be attained through real-time monitoring and feedback. As a result of his research, the author developed the BCEP (Behaviour Classification based eLearning Performance) prediction Framework that found to be more effective than conventional classification techniques [3].

In the current year (2022), Sudais et. el. examined the transcript students' data, that contained their Grades and CGPA transversely all courses in the dataset. The performance of the pupils was predicted using the NN, NB classifier, Decision tree, and SVM techniques. The most accurate method is found to be Nave Bayes. The overall accuracy and reliability of all machine learning methods deployed are not sufficient [4].

In the one year back (2021), Lonia Masangu et al. pointed out that most students receive some instruction on how to use online learning environments so they may comprehend the method of precise outcome prediction. According to their methodology, students those do not turn-in an valuation or do not attend a requisite forum will receive a grade of 0% for that particular assessment because all students are expected to receive the same amount of course information.

The authors' criteria for predicting academic performance include student demographics and data related to online logging (concentrating on students' engagement and how frequently they check their messages, how many numbers of hands are raised, how many forums are accessed, and how many resources are accessed to predict student success). Decision trees, SVM, Perceptron classifiers, Random Forest and Logistic regression classifier are some of the techniques employed. The SVM algorithm was shown to be the best algorithm for envisaging students' academic accomplishment in this paper [5].

In the one year back (2021), Giannakas et al. explored the potential of Deep Neural Network (DNN) categorization in a few particular computer science courses in 2021 by predicting the academic students' performance. The Deep Neural Network models with the Relu and Sigmoid activation methods' two hidden layers. Using learning efficiency of 80.76% and 86.57%, respectively [6].

In the one year back (2021), Mehmet Kokoc et. el. has investigated the degree to which a learner's involvement with the learning dashboards can be utilised to forecast and/or guide their academic success is one of the outcomes of an online learning experience that can be predicted. Data mining techniques have been used to analyse log data as well as student academic achievement. Based on user engagement by the unbending erudition dashboard, cluster analysis revealed that students were divided into mainly four categories based on their behavioural propensities.

According to the study's predictive analysis, engagement with prescriptive learning dashboards had some substantial influence on learners' academic achievement, and ANN algorithm produced the greatest results for forecasting the performance related to academics. The findings suggest that strict learning dashboards can be used in online courses as a teaching tool to enhance student performance as well as instructional design in eLearning environments [7].

In the one year back (2021), Ansar et al. Forecasting student's academic performance is vital in cultivating the values of education which was to study to find the key variables which have a substantial impact on 10^{th} standard schoolchildren's performance and to evolve an operative classification model which can be used for the prediction of academics' achievement by joining solo as well as ensemble-based classifiers. First, three independent one classifiers—a MLP (Multilayer Perceptron), a J48, and a PART—as well as three familiar ensemble algorithms— Voting (VT), Bagging (BAG) and MultiBoost (MB)—were studied. Nine additional models were created by combining solo as well as ensemble-based classifiers in order to improve the classifiers' performance as described above. The evaluation findings showed that MB with MLP outpaced the others as reaching 98.6% precision, 98.7% accuracy, F-score and recall. According to the research study, the anticipated model may be effective for early identification of secondary level students' academic performance so as to enhance learning conclusions [8].

3. Methodology

The methodology used starts with the gathering and integrating of student data, followed by conversion and normalisation in the .csv (Comma Separated Value) file, pattern extraction using classifier methods, and compared display of the results. The following are the goals established for this work:

1. To forecast how well pupils will succeed academically.
2. To compare as well as select the top algorithm for predicting students' academics' performance.

The collected dataset will be processed via the phases of conversion to numerical values from string so that the algorithms may be implemented, scaling of features have been done on different data characteristics to let our algorithms toward unite at a fast rate. Lastly normalisation has been completed. The prepared dataset has been used to implement the chosen classification algorithms such as; Random Forest, Decision Tree, and K-NN [8]. A 70:30 ratio between the training and test datasets has been chosen. Then, utilising the Precision matrix, RoC (Receiver Operating Characteristics) curve, and Confusion matrix, all models were assessed.

The graphical representation of RoC has been used to investigative binary-classifier model ability, while comparing classifiers has been done using the confusion matrix. As mentioned in the research's objectives, the comparison of results in relation to accuracy of period using various matrices have been then evaluated to offer the algorithm for optimum classification for academic performance prediction the as result of time spent on extracurricular activities. The following is a list of some methodology steps:

3.1. Collection and Processing of Data

We discovered after a thorough investigation that there are few universities in India that maintain track of students' extracurricular activities, hence we were unable to obtain a bigdata set for our predictor. For the purpose of gathering information about the students, we have turned to questionnaires. To provide a quicker turn-around of responses, the survey was administered in physical mode to 500 students. Only 395 of the 415 responses we received were legitimate and used for additional research.

The picture below illustrates many phases in data processing:

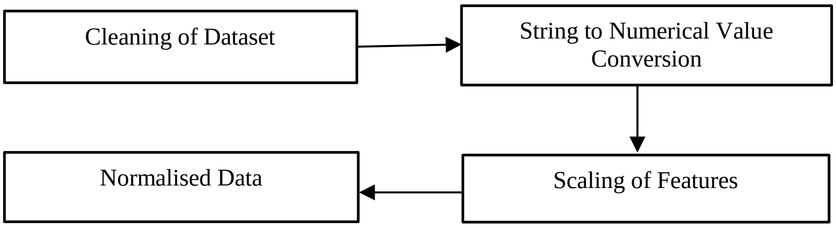

Fig. 1. Data Processing Steps

a. **Cleaning of Dataset:** In the absence of instances from the dataset are deleted at this point, and inadequate records are not added for further processing. Additionally, at this point, the features in the dataset which are not pertinent to the study are removed.
b. **String to numeric values Conversion**: In order to be used for model training, text values that is fail/pass and/or female/male like that be transformed into numeric or binary values.
c. **Scaling of Features:** As a matter of best practise, it is a technique to have characteristics in the number form to improve categorization outcomes. When working with enormous datasets, it is beneficial.
d. **Normalised Data**: The data has been cleaned, normalised and consistent for the use by the training model at this point in the processing procedure.

3.2. Algorithms Implementation

As, discussed earlier, three major ML algorithms are implemented and tested on the dataset. We will detail our algorithms' implementation in this part.

3.2.1 Decision Tree Algorithms

Decision Tree algorithm is widely used for supervised ML algorithms which comes under the classification and regression tasks. There are two popular attribute selection measures.

Steps to Compute Gini Index

Step 1: Split the data for training and testing (70:30) for Gini Index.

Step 2: Choose the max. depth=3 and attribute for Gini Index (to be consider as a root node at each level of Gini Method to split points where it operates with the definite/ categorical target variable as "Passed" or "Failed".

- Compute Gini for sub-nodes, using the given below formula

$$\text{Gini index}=1-n\sum i=1(pi)2 \tag{2}$$

that is, the sum of the square of probability (pi) for success and failure.

- Compute Gini index for splitting by the use of weighted Gini score each node's that split.

Step 3: Visualise the data using Confusion matrix and RoC Curve with Gini Index

Step 4: Visualise the Precision Table for Gini Index

Steps to Compute Entropy Index

Step 1: Split the data for training and testing (70:30) for Entropy.

Step 2: Choose the max_depth=3 and attribute Entropy (Information Gain) which we study by way of the root node at each level.

(Information-gain as a criterion, in general we estimate the information confined by attributes where Entropy measures the impurity in the given dataset).

Step 3: Visualise the data using Confusion matrix and RoC Curve with Entropy

Step 4: Visualise the Precision Table for Entropy

3.2.2 Random Forest

Random forest algorithm is the supervised ML algorithm which is used for both Regression as well as Classification problems. It creates decision trees based on diverse samples or subsets of data and choose their common vote for the classification decision and the average if there is regression. Random-forest Algorithm is well known for its feature which can be handle the dataset that contains continuous variables as in case of regression problems and categorical/definite variables as in case of classification problems. It provides improved results for classification categories. Below are the steps to implement the Random Forest Classifier.

Step 1: Split the train and test data (70:30)
Step 2: Train your splitting data keeping default parameters.
Step 3: Visualise the data after implementation in the form of Confusion Matrix and RoC Curve
Step 4: Precision matrix based on Confusion Matrix.

3.2.3 K-Nearest Neighbors Algorithm (KNN)

This is the most straightforward supervised machine learning methods (labelled dataset). It uses similarity to classify the data and is effective even with noisy datasets. The following objectives are what we are working for in this section:

- **KNN algorithm with its parameters**- The two characteristics listed below serve to define the KNN algorithm, a kind of supervised ML algorithm:

    - **Lazy learning algorithm:** As it doesn't have the dedicated training phase as well as uses all of the data during categorization for training.
    - **Non-parametric learning algorithm**: Because KNN not makes any assumptions inn respect to the primary data also known as a non-parametric learning.

  K-NN algorithm can be used for both regression and as well as classification predictive problem with its major application in classification predictive problems in industry.

- **KNN hyperparameters tuning:** Before dividing the data into smaller pieces and fitting it to the K-NN method, possibly a wise move would be to modify the ideal random state. Due of this, we'll pick a method that returns the "ideal state," which optimises both F1 score and accuracy for the specified number of rounds.

    - Using a variety of techniques to improve accuracy.
    - Find the best model with the highest level of accuracy.

The following steps have been taken in the implementation:
Step 1: Split the data and use it to the algorithms KNN
Step 2: Hyper tune the parameters associated to KNN Algorithms with the help of different methods

i. **Using visualization: Plotting** the ROC curve
ii. Using Grid Search CV Method: There is no way to know in advance what the ideal hyperparameter values should be. The manual method takes a lot of time, so we automated tweaking of hyperparameters using GridSearch CV; So that find the Best K for three purposes as given below:
  a. Training (13
  b. Testing (11)
  c. For all dataset (11)
iii. Find and use the value of K to K and search for Best Metric(s) (distance) based on accuracy and time.

Step 3: Optimal random state (1st step) and hyper parameter tunning (2nd Step) to find the improved accuracy by choosing ***K=13***
Step 4: Evaluating the KNN model with the following as given below:
  i. Confusion matrix
  ii. Precision Matrix
  iii. RoC Curve

4. Result & Discussions

This section will cover the result after the implementation of algorithms. We have also compared the results and found the best result on the basis of results in each implemented algorithm. An error analysis is also presented in this section.

4.1 Decision Tree

After the implementation of decision tree with Gini index according to the step three we find the Confusion matrix, RoC Curve and Precision Table as given below:

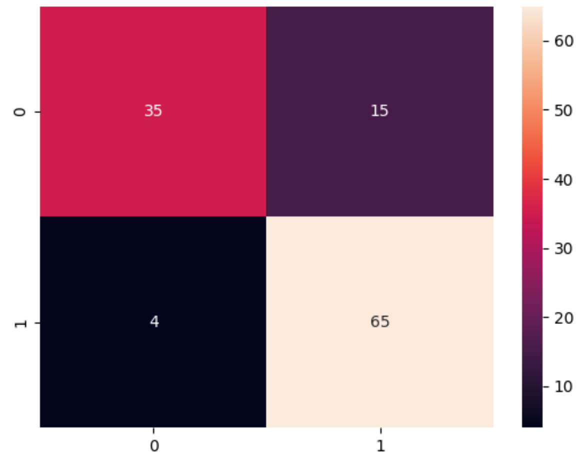

Fig. 2. Confusion Matrix Visualization for Gini Index

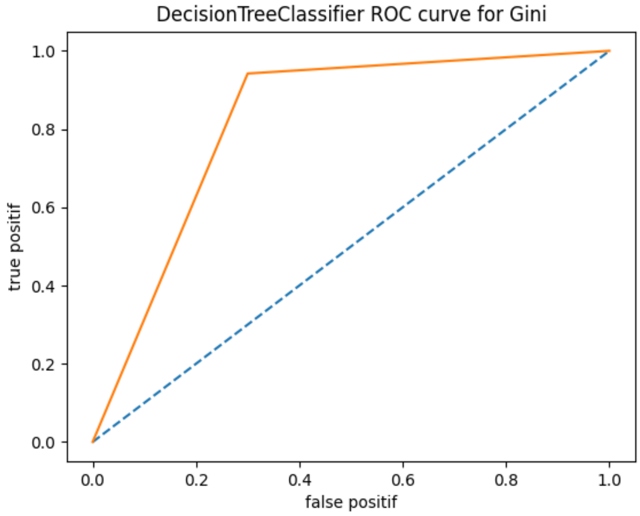

Fig. 3. RoC Curve Visualization for Gini Index

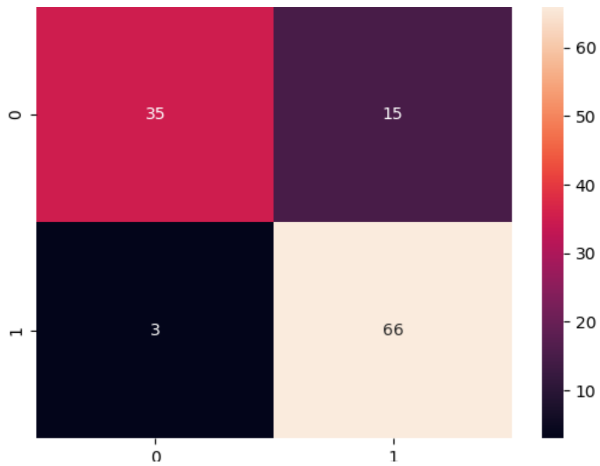

Fig. 4. Confusion matrix with entropy

Table 1. Precision Table for Decision Tree- Gini Index

	Precision	Recall	F1 Score	Support
0.0	90%	70%	79%	50
1.0	81%	94%	87%	69
Accuracy			84%	119
Macro avg	86%	82%	83%	119
Weighted avg	86%	84%	84%	119

After the implementation of decision tree with **Entropy** index according to the step three we find the Confusion matrix, RoC Curve and Precision Table as given abrove:

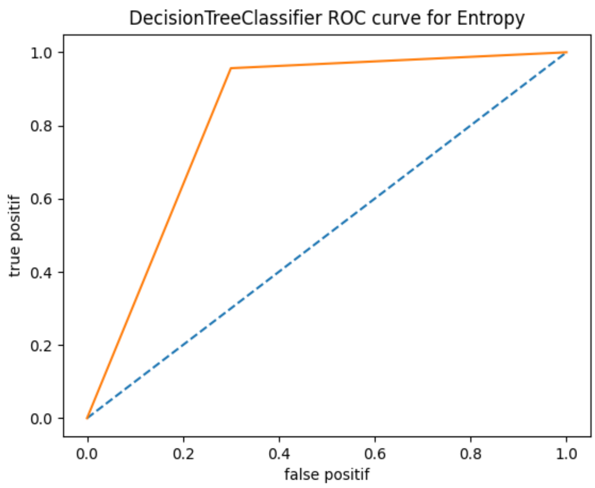

Fig. 5. RoC Curve with Entropy

Table 2. Precision Table for Decision Tree- Entropy Index

	Precision	Recall	F1 Score	Support
0.0	92%	70%	80%	50
1.0	81%	96%	88%	69
Accuracy			85%	119
Macro avg	87%	83%	84%	119
Weighted avg	86%	83%	84%	119

Conclusion of Decision Tree Algorithm: Here, we found two values F1 Score and accuracy, one gained with the training-set besides other with the test-set. As per the above table we achieved high accuracy and F1 score in Entropy method for Decision Tree. However, there is very miner difference between both methods as given in the below:

Table 3. Accuracy & F1 score result of Decision Tree

	Accuracy	**F1 Score**
Gini Index	84	83
Entropy	85	84

4.2 Random Forest

After the implementation of Random Forest according to the step three we find the Confusion matrix, RoC Curve and Precision Table as given below:

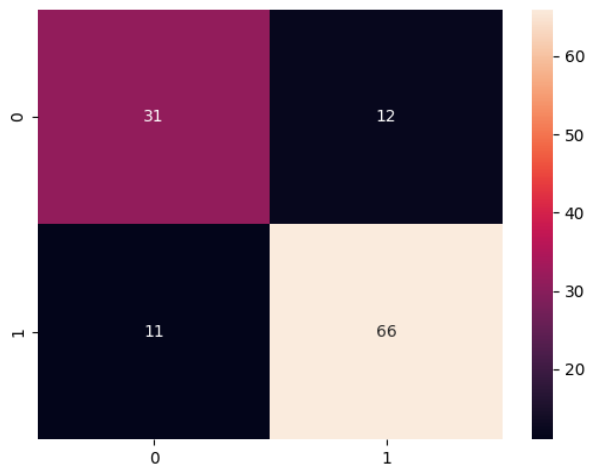

Fig. 6. Confusion Matrix for Random Forest

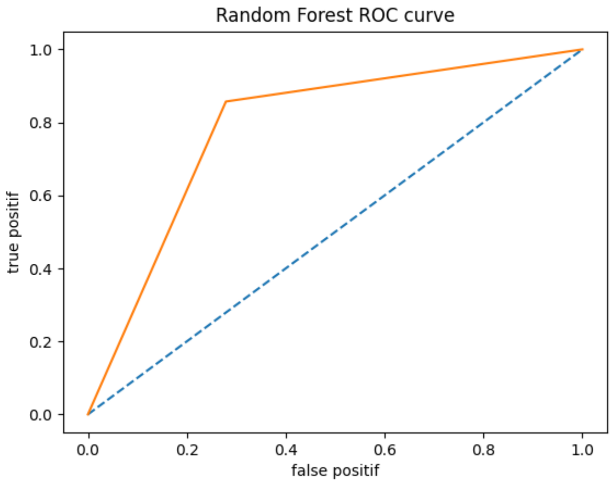

Fig. 7. RoC Curve (Random Forest)

Table 4. Precision Table for Random Forest

	Precision	Recall	F1 Score	Support
0.0	%74	72%	73%	43
1.0	%85	86%	85%	77
Accuracy			81%	120
Macro avg	%79	79%	79%	120
Weighted avg	%81	81%	81%	120

Conclusion: In the implementation of Random Forest, we achieve the Accuracy test 81% and 79% F1 Score. We did simple implementation of Random Forest and did not implement hyperparameter tunning.

4.3 K Nearest Neighbour

After the implementation of KNN according to the step two and three we find the Confusion matrix, RoC Curves (KNN and **varying number of neighbours)** and output table for all the matrices such as Euclidean Manhattan and Chebyshev are as given below:

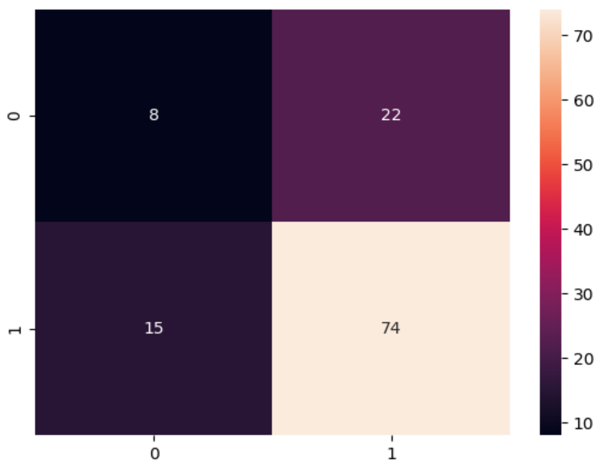

Fig. 8. KNN Confusion matrix (without hypertunning)

It can be seen that the we get the best random state is 71027464, that has F1 score 55% and an accuracy of 65%.

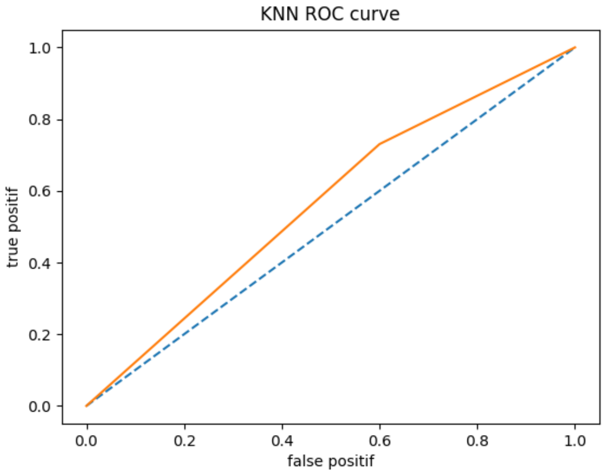

Fig. 9. RoC Curve for KNN

Table 5. Result of all Chosen Matrices in KNN

Model Name	Execution Time	Accuracy
Euclidean, N-neighbors: 11	0.015ms	0.64
Euclidean, N-neighbors: 13	0.017ms	0.64
Manhattan, N-neighbors: 11	0.0ms	0.64
Manhattan, N-neighbors: 13	0.011ms	0.64
Chebyshev, N-neighbors: 11	0.015ms	0.64
Chebyshev, N-neighbors: 13	0.00ms	0.64

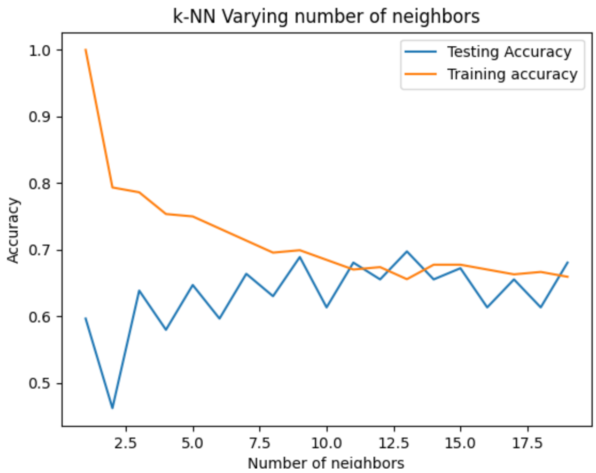

Fig. 10. K-NN Mutable Number of neighbours

Below we have provided the step three result of RoC for, Confusion matrix and precision table after hyper tunning. Evaluate the KNN Model Using Confusion Matrix.

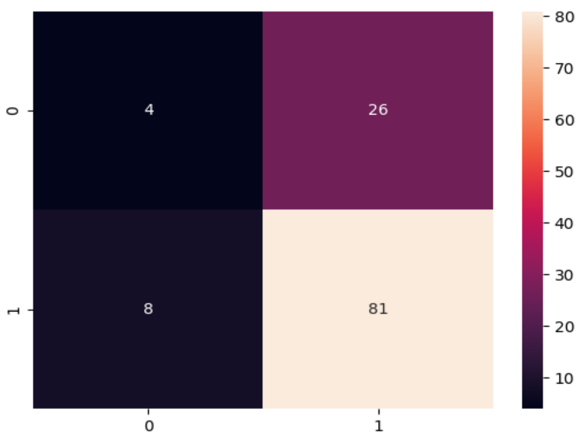

Fig. 11. Evaluating KNN Model using Confusion Matrix

Precision Matrix Evaluation

Table 6. Precision Matrix

	Precision	Recall	F1 Score	Support
0.0	33%	13%	19%	30
1.0	76%	91%	83%	89
Accuracy			71%	119
Macro avg	55%	52%	51%	119
Weighted avg	65%	71%	67%	119

RoC Curve Evaluation

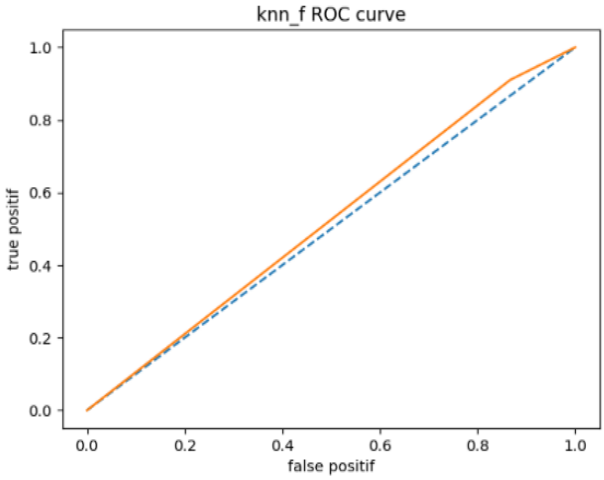

Fig. 12. Evaluation of Model using RoC Curve

As we can see, accuracy increases from 64% to 71% when the best KNN parameters and optimal state are combined. It is obvious that adjusting the hyperparameters improves the outcome, however the most significant parameters are random state.

KNN: Understanding the effects of K-NN hyper parameter adjustment is the main goal of this section. The ideal random state is tuned using a model that we first implement without defining the KNN parameters, then we test the model with the an of accuracy of 71%.

4.4 Cumulative Result

From the table 7, it is evident that on comparison of F1 score and accuracy of the three algorithms i.e., Decision Tree, Random Forest and KNN (shown as in table 7 below), the Decision Tree (Entropy) is superior than Random Forest and KNN model in terms of accuracy and efficiency for predicting student academic performance.

Table 7. Comparison of models' F1 Score and Accuracy

Model Name		**Accuracy**	**F1 Score**
Decision Tree	**Gini**	84	83
	Entropy	85	84
Random Forest		81	79
K-NN Algorithm		71	51

4.5 Error Analysis

The error which we have in the result of algorithm implementation may be due to in the erroneous data supplied by the recipients. Additionally, the database size has great impact on machine learning algorithms. We have taken small dataset as we were trying to perform a pilot study. The results can be differed in the form of accuracy and F1 score if the dataset will be large. In future we will extend the study on large dataset with more features.

5. Conclusion & Future Work

This study used machine learning (ML) models to predict students' academic achievement based on the amount of time they spent participating in extracurricular activities. Three Algorithms namely, Random Forest, K-NN and

Decision Tree have been implemented in this research work and Based on accuracy and F1 scores, it was found that the Decision Tree with Gini Index is a better algorithm for accurately predicting students' academic success.

Future study on this topic is advised to use more machine learning (ML) prediction models, namely, SVM, MLP-ANN, and Nave Bayes (NB), among others, for comparably better outcomes and to develop the best algorithm for predicting students' academic success.

References

[1] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers in Education, 113, 177–194.

[2] Xu J., Han Y., Marcu D., van der Schaar M.: Progressive prediction of student performance in college programs. In: AAAI pp. 1604–1610 2017.

[3] Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., ...& Chen, P. K. (2022). Predicting students' performance in e-learning using learning process and behaviour data. Scientific Reports, 12(1), 1-15

[4] Sudais, M., Safwan, M., Khalid, M. A., & Ahmed, S. (2022). Students' Academic Performance Prediction Model Using Machine Learning. Research Square; 2022. DOI: 10.21203/rs.3.rs-1296035/v1.

[5] Masangu, Lonia, Ashwini Jadhav, and Ritesh Ajoodha. "Predicting Student Academic Performance Using Data Mining Techniques." Advances in Science, Technology and Engineering Systems Journal 6.1 (2021): 153-163.

[6] Giannakas, F., Troussas, C., Voyiatzis, I., &Sgouropoulou, C. (2021). A deep learning classification framework for early prediction of team- based academic performance. Applied Soft Computing, 106, 107355.

[7] Koko ç, M. & Altun, A. discuss the effects of learner interaction with learning dashboards on academic performance in an e-learning environment. Behav. Inf. Technol. 40, 161–175. https://doi.org/10.1080/0144929X.2019.1680731 (2021).

[8] Siddique, A., Jan, A., Majeed, F., Qahmash, A. I., Quadri, N. N., & Wahab, M. O. A. (2021). Predicting academic performance using an efficient model based on fusion of classifiers. Applied Sciences, 11(24), 11845.

[9] N. Sharma and M. Yadav, "A Comparative Analysis of Students' Academic Performance using Prediction Algorithms Based on Their Time Spent on Extra-Curricular Activities," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), 2022, pp. 745-750, doi: 10.1109/ICICICT54557.2022.9917606.

Authors' Profiles

Dr. Neeta Sharma is serving as an Associate Professor in the Department of Computer Science in Noida International University since Feb 2013. She is having more than 13 years of Excellence in the Academics. She is completed her Doctorate in Cloud Computing (CSE) in 2016. She has authors forty plus research papers in the reputed journal and the National/ International conferences including Scopus, IEEE and Book chapter in Apple Academic Press. She has published four Patent in her name.

Prof. (Dr.) A.Shanmuganathan is a committed academician, researcher and speaker, has vast professional experiences over 26 years in Academics and in Industries. As a professor in the Dept. of Mechanical Engg. He has published patents and a number of research papers in leading international journals and conferences including Elsevier and Scopus indexed journals. He has worked abroad in Government Universities for two decades and worked as Dean, Principal and Director-Academics during his tenure emphasizing collaborated-working-model in Higher Educational Institutions (HEIs). He has extended his consultancy services to industries and local communities ensuring excellence of University-Industry-Government Linkage. He did research works contributing to "University-Industry-Government" linkage in King Khalid University, Saudi Arabia. He did mechanical projects where students are involved to improve their technical skills in making them Industry-Ready

Mr. Umang Garg received the master degree from AKTU, Lucknow, India and perusing Ph.d from Graphic Era Hill University, Dehradun. He is working as an Assistant Professor in Graphic Era Hill University, Dehradun since 2017. His area of Interest includes IoT Security, Intrusion Detection System for IoT, and Machine Learning. He is reviewer of various Journals and Conference Proceeding. He has published many research papers in the reputed journal and the National/ International conferences including Scopus, IEEE and many Patent in his name.

Ms. Jayati Mukherjee, currently working as an Assistant Professor in Noida International University, has done her B.E. from the University of Burdwan, WB in 2015. She completed her M.Tech from BIT, Mesra in 2017. Currently she is pursuing her Ph.D. in Pattern Recognition from Visva-Bharati University. She has worked as a PLP in the Indian Statistical Institute. Her research interests include Machine Learning, Pattern Recognition and Image Processing. She has published in multiple national and international reputed journals and conferences. She has 2 patents in the IPO.

Ms. Sneha Mishra serving as an Assistant Professor in Noida International University and pursuing her PhD from Galgotias University, Greater Noida. She has done her Bachelor's degree in CSE from Rishi Institute of Engineering and Technology affiliated from U.P.T.U U.P, in 2012 and her M.Tech. From Amity University, Noida,U.P. in 2014. Her current research interests include image processing, neural networks and Computer Vision. She has five Patents in the IPO. She has various research papers in international journals.

How to cite this paper: Neeta Sharma, Shanmuganathan Appukutti, Umang Garg, Jayati Mukherjee, Sneha Mishra, "Analysis of Student's Academic Performance based on their Time Spent on Extra-Curricular Activities using Machine Learning Techniques", International Journal of Modern Education and Computer Science(IJMECS), Vol.15, No.1, pp. 46-57, 2023. DOI:10.5815/ijmecs.2023.01.04