



# International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

**COPY RIGHT**



**ELSEVIER**  
**SSRN**

**2022 IJIEMR.** Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 25th Jun 2022. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05](http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05)

**DOI: 10.48047/IJIEMR/V11/SPL ISSUE 05/26**

Title **STUDENT PERFORMANCE ANAYSIS BASED ON MACHINE LEARNING**

Volume 11, SPL ISSUE 05, Pages: 170-176

Paper Authors

**Mr. P. V. Kishore Kumar, M.V.S. Bhagyasri, N. Kalyan Naidu, T. Mahima, V. Nikhilendra**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

## STUDENT PERFORMANCE ANALYSIS BASED ON MACHINE LEARNING

Mr. P. V. Kishore Kumar<sup>1</sup>, M.V.S. Bhagyasri<sup>2</sup>, N. Kalyan Naidu<sup>3</sup>, T. Mahima<sup>4</sup>, V. Nikhilendra<sup>5</sup>

<sup>1</sup>Assistant professor, Dept. of C.S.E., <sup>2</sup>18ME1A0570, <sup>3</sup>18ME1A0573, <sup>4</sup>18ME1A05A8, <sup>5</sup>18ME1A05B8

RAMACHANDRA COLLEGE OF ENGINEERING, ELURU

[pvkishorekumar@rcee.ac.in](mailto:pvkishorekumar@rcee.ac.in), [mvsbhagyasri@gmail.com](mailto:mvsbhagyasri@gmail.com),

[nalagopukalyannaidu8280@gmail.com](mailto:nalagopukalyannaidu8280@gmail.com), [tadimahimaprasad@gmail.com](mailto:tadimahimaprasad@gmail.com),

[nikhilendra.05@gmail.com](mailto:nikhilendra.05@gmail.com)

### Abstract:

The goal is to be expecting the scholar overall performance primarily based totally on sure attributes of the scholar which includes semester marks, attendance, dating with family, buddy ship. Academic success is critical for the a success improvement of younger human beings in society. We additionally evaluate the KNN, SVM, Logistic regression which one is extra correct in locating the output. Based on attributes we will perceive the scholar dropouts and the scholar who want the unique interest from instructor to offer counselling which improves the overall performance in students. By the usage of gadget mastering we examine the scholar overall performance.

### Introduction

Education is an critical detail of the society, each authorities and usa withinside the global paintings so tough to enhance this sector. With the corona-virus outbreak that has disrupted lifestyles around the world in 2020, the instructional structures had been affected in lots of methods; research display that student's overall performance has reduced given that then, which highlights the want to cope with this hassle extra severely and attempt to locate powerful solutions, in addition to the influencing factors. The academic structures want, at this unique time, revolutionary methods to enhance exceptional of training to acquire the nice consequences and reduce the failure rate.

As college students of the IT branch who studied withinside the ultimate month a bit approximately device learning, we recognize that during order for an institute to offer exceptional training to

inexperienced persons, deep evaluation of preceding facts of the inexperienced persons can play a essential role, and desired to paintings in this difficult task.

### Literature Survey

Creators in [6] thought of managed AI methods for the Botnet Detection in a specific organization. They have utilized the CNN, ANN organizations. To look at the outcomes they've utilized Metric which are decided to assess the exhibition of the strategies. While carrying out calculations they've looked at test\_precision, test\_review, test\_f1. In this paper, we've given a complete audit of generally utilized machine learning strategies to decide the exhibition of AI procedures to find some wide certifiable digital assaults. They have contrasted the precision that got and the Machine Learning Algorithms with the

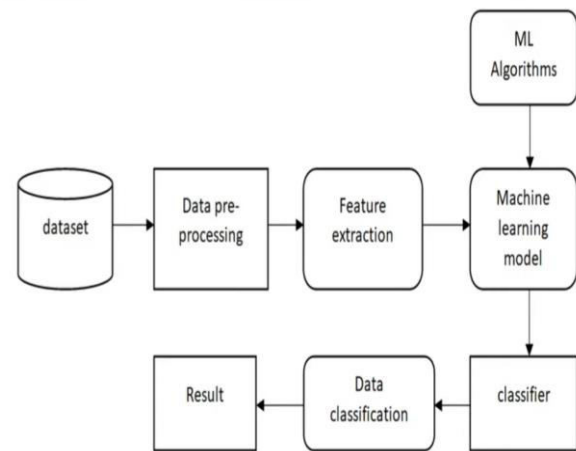
Dense Neural Network and closed which strategy gives the best exactness. The preceding predictive fashions simplest centered on the usage of the scholars demographic records lifestyles age, gender, own circle of relatives status, own circle of relatives profits and classifications. In addition to this associated attributes inclusive of the house paintings and observe hours in addition to the preceding achievements and grades

Creators in [7] thought of an applied number related semi-managed AI procedure for interruption identification in robot cell phones. the ascent in information traffic additionally will make to cybercrimes. Thus, to protect robot cell phones against cutting edge cybercrimes, extra high level AI strategies are expected to be created to find malevolent exercises. Creators in [8] have utilized AI methods to find a few wide world-renowned cybercrimes. They examined 3 wide utilized AI methods, specifically: call tree, profound conviction organization, and backing vector machine. they need to consider 3 significant digital dangers. Interruption location, spam identification, and malware detection. While carrying out calculations they've looked at test\_precision, test\_review, test\_f1. They have contrasted the precision that got and the Machine Learning Algorithms with the Dense Neural Network and closed which strategy gives the best exactness. In this paper, we've given a complete audit of generally utilized machine learning strategies to decide the exhibition of AI procedures to find some wide certifiable digital assaults. we've dissected 3 wide utilized AI procedures, specifically: Logistic Regression, choice tree, Random Forest, and MLP Classifier. The majority of the audit articles exclusively designated

a particular danger. Be that as it may, we've considered 2 significant digital assault sorts. DoS (Denial of Service) assault and Probe assault are considered for this review. we've given a thorough correlation with determine the exhibition of each and every equation with the help of datasets.

The proposed framework principally follows a bunch of steps. We can accomplish our objective assuming we follow those means. The calculation's key advances are recorded underneath. Each dataset ought to be standardized. Partition the first dataset into testing and preparing datasets. Create IDS models utilizing Logistic Regression, Decision Tree, Random Forest, and MLP. Evaluate the presentation of each model.

### System architecture:



### Existing System

The preceding predictive fashions simplest centered on the usage of the scholars demographic records lifestyles age, gender, own circle of relatives status, own circle of relatives profits and classifications. In addition to this associated attributes inclusive of the house paintings and observe hours in addition to the preceding achievements and grades.

## Proposed System

The proposed body paintings first of all cognizance on merging the demographic and observe associated attributes with the academic psychology fields, via way of means of including the scholars mental traits to the formerly used information sets. Comparing the KNN, SVM, Logistic regression which we offers the correct output.

## Dataset processing

We need to map every string to a numerical fee in order that it may be used for version training.

Let`s take an instance to make this clear :  
The mom process column incorporate 5 values : 'teacher',

'health','services', 'at\_home' and 'other'.

Our process then is to map every of those string to numerical values, and that is how it is done :

```
df['Mjob'] = df['Mjob'].map()
```

1. We need to carry out function scaling :
2. Feature scaling is a way used to normalize the variety of unbiased variables
3. or capabilities of statistics. In statistics processing, it's also called statistics normalization
4. and is usually achieved throughout the statistics
5. **pre processing step.**
6. This will permit our mastering algorithms to converge very

1. quickly. The operation calls for
2. to take every column, let`s say 'col', and update it by :
3. But this isn't the best scaling you could do, you could strive also :  
  
, wherein std refers to traditional deviation.

## Dataset Visualization :

After having our dataset processed, we flow to the subsequent step that's dataset visualization; in order that patterns, tendencies and correlations that won't in any other case be detected may be now exposed.

This subject will deliver us a few insights into our information and assist us recognize the dataset via way of means of putting it in a visible context the usage of python libraries such as: matplotlib and seaborn. The preceding predictive fashions simplest centered on the usage of the scholars demographic records lifestyles age, gender, own circle of relatives status, own circle of relatives profits and classifications. In addition to this associated attributes inclusive of the house paintings and observe hours in addition to the preceding achievements and grades There are such a lot of approaches to visualise a dataset. In this undertaking we selected to:

#1- Plot distribution histograms in order that we will see the range of samples that arise in every unique category.

E.g.  
Internet accessibility at domestic distribution indicates that our dataset includes greater than three hundred college students with domestic net accessibility, at the same time as there are approximately 50 college students who've no get entry to the net at domestic.

#2- Plot “Boxplots” to peer to peer how the scholars fame is shipped consistent with every variable.

#3- Plot the correlation output that ought to listing all of the functions and their correlations to the goal variable.

So that we've an concept approximately the maximum impactful factors on the scholars fame.

### Model evaluation (metrics) :

Before starting training our classifiers let us define the metrics that we will use to compare between our three classifiers :

1) Confusion matrix :

2) F1 score :

TP = quantity of real positives

FP = quantity of fake positives

FN = quantity of fake negatives

3) The roc curve : A receiver working function curve, or ROC curve, is a graphical plot that illustrates the diagnostic cap potential of a binary classifier device as its discrimination threshold is varied.

4) ROC score : it`s honestly the fee of the location below the roc curve.

The high-quality fee is 1 due to the fact the location of 1x1 rectangular is 1.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

### Logistic regression

#### KNN

#### 1) Introduction to KNN :

K-Nearest Neighbour is one of the best Machine Learning algorithms primarily based totally on Supervised Learning technique. Its assumes the similarity among the brand new case facts and to be had instances and positioned the brand new case into the class this is maximum just like the to be had categories. K-NN set of rules shops all of the to be had facts and classifies a brand new facts factor primarily based totally at the similarity. This way while new facts seems then it could be without problems categorised right into a nicely suite class via way of means of the usage of K- NN set of rules.

#### Python implementation of the KNN algorithm

In this step we are able to put into effect knn for our case look at with the aid of using following this step:

- a) Data preprocessing step
- b) Hyperparameters tuning
- c) Fitting the K-NN set of rules to the Training set
- d) Predicting the take a look at result
- e) Test accuracy of the result

\*Let`s inspect every step one by one

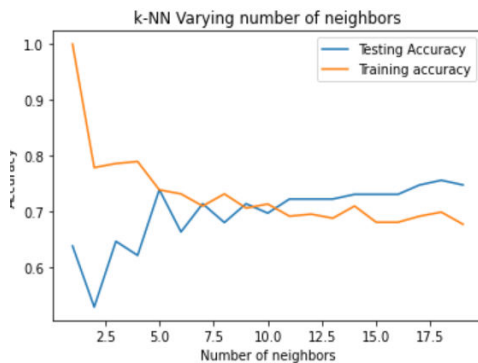
a) Data preprocessing step:

we have to inspect preceding segment

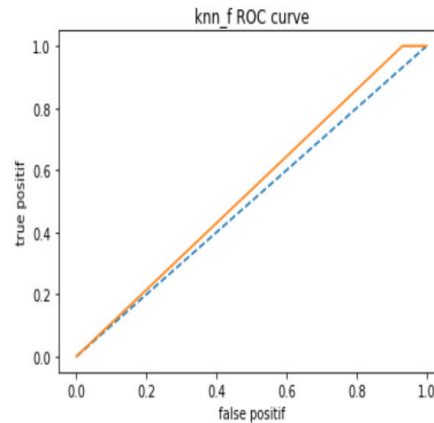
b) Hyperparameters tuning:

In this step we appearance after 2 techniques to song the quality parameters for our version

First method: tuning the best k for better test accuracy and training accuracy using KNN Varying numbers



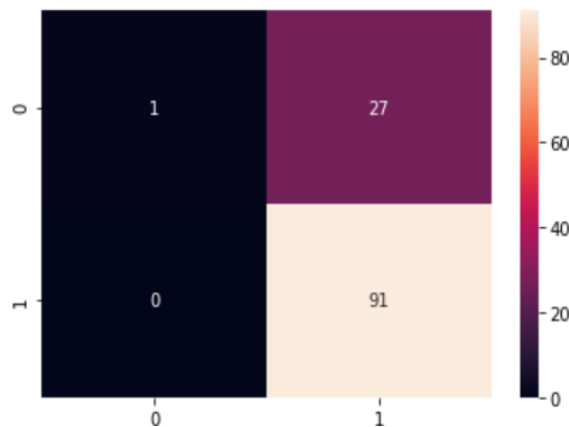
the ROC curve:



After prediction we will evaluate the model using various methods:

### 1) Confusion\_matrix:

Accuracy is: 0.773109243697479



### 2) Classification\_report :

2)classification\_report

```
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.0	1.00	0.04	28
1	1.0	0.77	1.00	91
accuracy			0.77	119
macro avg	0.89	0.52	0.47	119
weighted avg	0.83	0.77	0.68	119

### 3) Plotting Roc curv:

## SVM

Now we are able to use Support Vector Machine set of rules and notice how it's going to act on our records. But, let's outline what's svm set of rules

In gadget studying, support-vector machines are supervised studying fashions with related studying algorithms that examine records for type and regression analysis.

It makes use of a method referred to as the kernel trick to convert your records after which primarily based totally on those ameliorations it unearths an foremost boundary among the viable outputs. We will use 3 kernel : Linear, polynomial and gaussian kernel.

1) Linear kernel : Linear Kernel is used whilst the facts is Linearly separable, that is, it is able to be separated the use of a unmarried Line. It is one of the maximum not unusualplace kernels to be used.

It is broadly speaking used whilst there are a Large wide variety of functions in a selected Data Set.

2) The polynomial kernel is a kernel feature typically used with assist vector machines and different kernelized

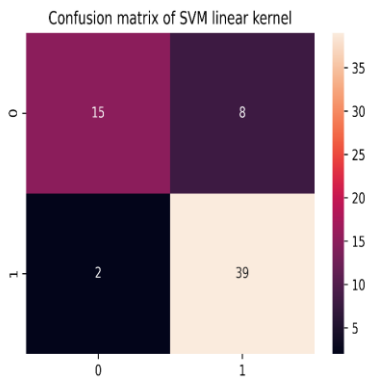
fashions, that represents the similarity of vectors in a characteristic area over polynomials of the unique variables, permitting studying of non-linear fashions.

3) Gaussian RBF(Radial Basis Function) is every other famous Kernel approach utilized in SVM fashions for more.

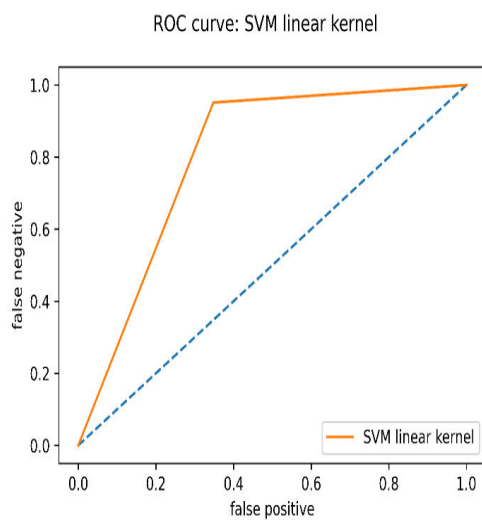
RBF kernel is a feature whose fee relies upon on the gap from the beginning or from a few point.

Gaussian Kernel is of the subsequent format:

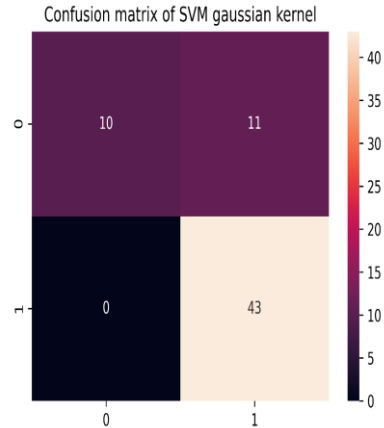
### Confusion matrix :



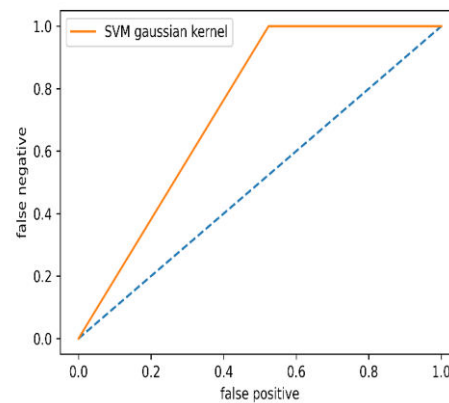
### ROC Curve :



### Confusion matrix :

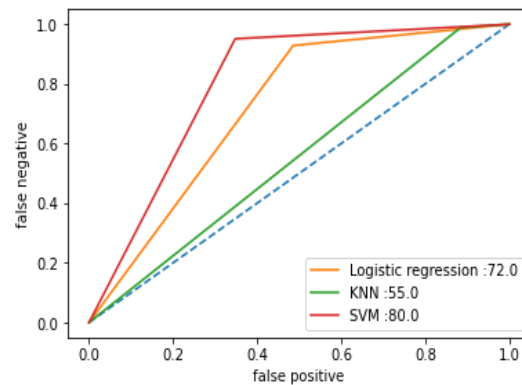


ROC curve: SVM gaussian kernel



### Result:

Comparison of three ROC curves



### Conclusion:

Improving the training device is a huge problem, as an engineering scholar we will assist acquire this intention through the usage of technology and examine sources like device gaining knowledge of

materials, to provide you with an revolutionary option to assist the scholar in need, in particular college students who stay in hard situations. situations (demographic, social and academic issues). In this mission, we got here up with the concept of making a version that predicts the popularity of college students primarily based totally on one of a kind functionalities. Our primary demanding situations had been to outline the excellent class set of rules and pick out the maximum influential elements for the instructional popularity of college students to offer them with a precis or valedictorian of the excellent situations for college kids to acquire excessive educational popularity and keep away from failures. For this mission entitled "Prediction of scholar overall performance and difficulty" ", we used numerous class strategies which include logistic regression, KNN and SVM and we compare this version the usage of one of a kind metrics like f1 score, roc curve and the confusion matrix and subsequently we were given a winner with SVM with a precision of 80%as compared to different set of rules . Before taking over our primary demanding situations, there had been numerous steps to take: - statistics processing-statistics visualization -Implementation of models -contrast of three algorithms.

## References:

[1] U. Fayyad, G. Piatesky-Shapiro and P. Smyth, "From information mining to know-how discovery databases.," *AI Magazine*, pp. 37-53, 1996.

[2] "ÍndiceBrasscom de Convergência Digital," 2015. [Online]. Available: [www.brasscom.org.br](http://www.brasscom.org.br). [Accessed: 14-May-2016].

[3] H. Witten, E. Frank and M. A. Hall, *Data mining: Practical Machine Learning*

*Tools and Techniques*. Burlington: Morgan Kaufmann, 2011.

[4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA information mining software: an update," in *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10-18, Jun. 2009.

[5] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students` Performance," in *International Journal of Advanced Computer Science and Applications – IJACSA*, vol. 2, no. 6, pp. 63-69, 2011.

[6] J. R. Quinlan, *C4.5*, San Francisco: Morgan Kaufmann Publishers Inc., 1993.

[7] E. L. Ferreira, H. Rausch, S. Campos, A. Faria-Campos, E. Pietra and L. da Silva Santos, "Medical information mining: A case have a look at of a Paracoccidioidomycosis patient's database," *e-Health Networking, Applications and Services (Healthcom)*, 2014 IEEE sixteenth International Conference on, Natal, 2014, pp. 275-280.

[8] F. H. D. Araujo, A. M. Santana and P. A. S. Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process," in *IEEE Latin America Transactions*, vol. 13, no. 1, pp. 299306, Jan. 2015.