DEEP LEARNING MODEL FOR CREDIT CARD FRAUD DETECTION

Hemalatha M[1], Deepa N[2]

[1]*Computer Science Engineering, RVS College Of Engineering, hemamithu6@gmail.com*
[2]*Computer Science Engineering, RVS College Of Engineering,deepanatrayan@gmail.com*

Abstract— Detecting Credit card Fraud detection is an at most need currently. This is viewed in two perspectives in this work, first as a classification problem and second as the problem of finding the outliers. As a classification problem, various machine learning models will be employed and the results will be compared. In the perspective of finding the outlier, a deep learning model will be designed to detect fraudulent data. The experimentation will be made and analyzed with a dataset that has the details of the applicants who have applied for availing the credit card.

Keywords— Machine Learning, Deep Learning, Fraud Detection, Logistic Regression, Support Vector Classifier, Decision Tree and Random Forest.

I.INTRODUCTION

Fraud detection has been one of the major challenges for most organizations particularly those in banking, finance, retail, and e-commerce. This goes without saying that any frauds negatively affects an organization's bottom line, its reputation and deter future prospects and current customers alike to transact withit.

More often than not, for any fraud detected, the organization ends up paying for the losses. Additionally, it takes the good customers away from them while attracting more fraudsters. Given the scale and reach of most of these vulnerable organizations, it has become indispensable for them to stop these frauds from happening or even predict all suspicious actions beforehand at all times. Frauds can range from really small like non- payment for e-commerce orders to threatening (to organization's existence) like public exposure of customers' credit carddetails.

Machine learning comes to the rescue here. On setting up automated data science processes with deep learning algorithms, organizations can greatly reduce the risk of their exposure to most of such frauds.

1.1 Dataset

The dataset employed in the project is downloaded from UCI Machine Learning repository. The dataset is called as statlog – Australian credit card application dataset. This dataset holds credit card applications. Basically, the bank issues credit cards to customers and they are trying to figure out did any faulty application got approved by mistake. Our goal is to detect possible frauds so they can be further investigated by the bank. This way, the bank can protect itself from possible losses in the future. In the dataset, all attribute names and valueshave been changed to meaningless symbols to protect the confidentiality of the data.This dataset has a good mix of attributes – continuous, nominal with small numbers of values, and nominal with larger numbers of values, this makes the problem harder.

1.2 Problem statement

The problem is to identify the fraudulent data from the given data. It can be considered either as a problem of classification or as a problem of finding the outliers. The first phases of the work consider it as the classification problem and apply various machine learning algorithms to find the fraudulent data. the machine learning algorithms that are applied are as follows.

II. SYSTEMARCHITECTURE

From the dataset it is observed that the features have no semantics, except the final feature which is indicating had credit card been issued to the customer or not.The Architecture of the model is given below.

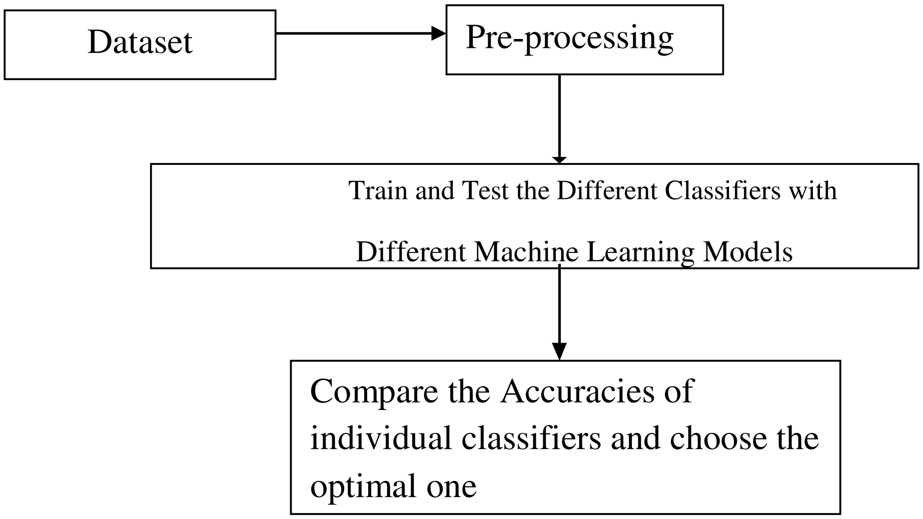
Dataset
Pre-processing
Train and Test the Different Classifiers with
Different Machine Learning Models
Compare the Accuracies of individual classifiers and choose the optimal one

Fig 1. System Architecture

III. SYSTEMMODEL

The module includes pre-processing followed by the implementation of Machine Learning Models. The architecture and the modules are explained in this chapter.

3.1 Modules

The Modules include

- FeatureScaling
- Implementation of the Machine Learning Model

3.1.1Featurescaling

Feature scaling is the process in which all the values in the data set are normalized.Normalization is the process in which the values are converted to a range between 0 to1. This is done with the scalar function and this is done to avoid the influence of a feature that has larger values than the other features. The formula used for normalization is

xnew= (x-xmin)/(xmax-xmin)

where

x is the value to be normalized

xmaxand xmin are the maximum and the minimum values

3.1.2Implementation of Machine Learning Model

The following machine learning models are built for detecting the fradulent applications

- LogisticRegression
- Support VectorMachine
- DecisionTree
- RandomForest

The uniqueness of the approach is that efforts were taken to find the optimal parameters for each model and the model is built as shown in the following figure

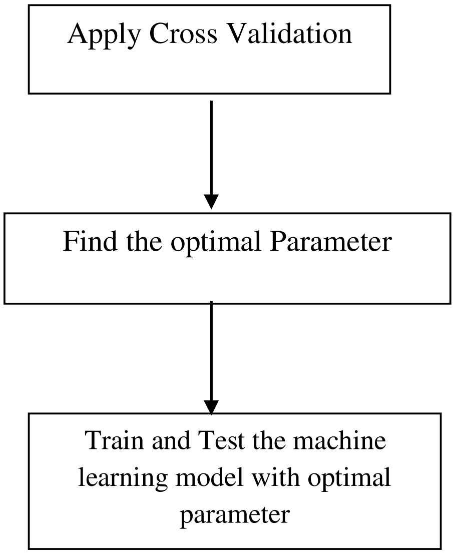
Apply Cross Validation
Find the optimal Parameter
Train and Test the machine learning model with optimal parameter

Fig 2.overview of the system

K-Fold cross validation mechanism is employed and the optimal parameters are identified and the best scores are used in the model. The steps are as follows

For Each Model

Identify the parameters ($P_1,P_2\ldots P_3$) that has high impact on the accuracy. For Each parameter P_i List out the possible values $v_1,v_2\ldots v_n$ of the parameter For

each value v_i

Perform K-fold cross validation and find the best mean score Savethe value of the parameter that produces the best result. .

Inverse of regularization: Inverse of regularization is the parameter that is used for

The over fitting problem that arises in the data set with small amount of data and in

we have only 150 Rows. A good value of regularization would reduce the error value by decreasing

Value .

. IV.RESULTS

The performance parameters that are considered and evaluated for the considered machine learning models are as follows.

- AUC
- Accuracy

- Recall

The performance metrics are calculated as follows

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision}=\frac{TP}{TP+FP}$$

$$\text{Recall}=\frac{TP}{TP+FN}$$

$$\text{Specificity}=\frac{TN}{TN+FP}$$

Area under the curve is the region that is below the receiver operating characteristic curve which is drawn with the true positive rate and the false positive rate. Recall is known as the true positive rate and False positive rate can be defined as follows

$$\text{False positive rate}=\frac{FP}{FP+TN}$$

Experimentation has been conducted in i3 system with 4 GB RAM. The implementation is done with the python libraries. The values achieved for these parameters are given in the followingfigures.

The following Figure represents the best training accuracy obtained with the best parameters identified with K-Fold cross validation.

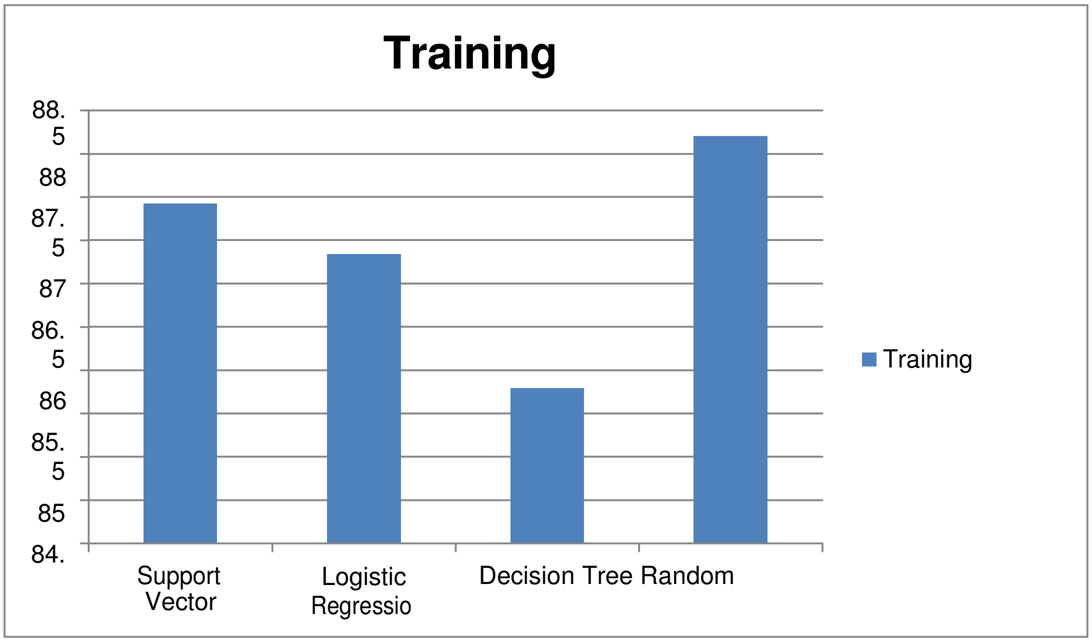

Fig 3. Best Training Accuracy of various Machine learning model

The Following figure represents the Comparison of the Test accuracy of the different machine learning models.

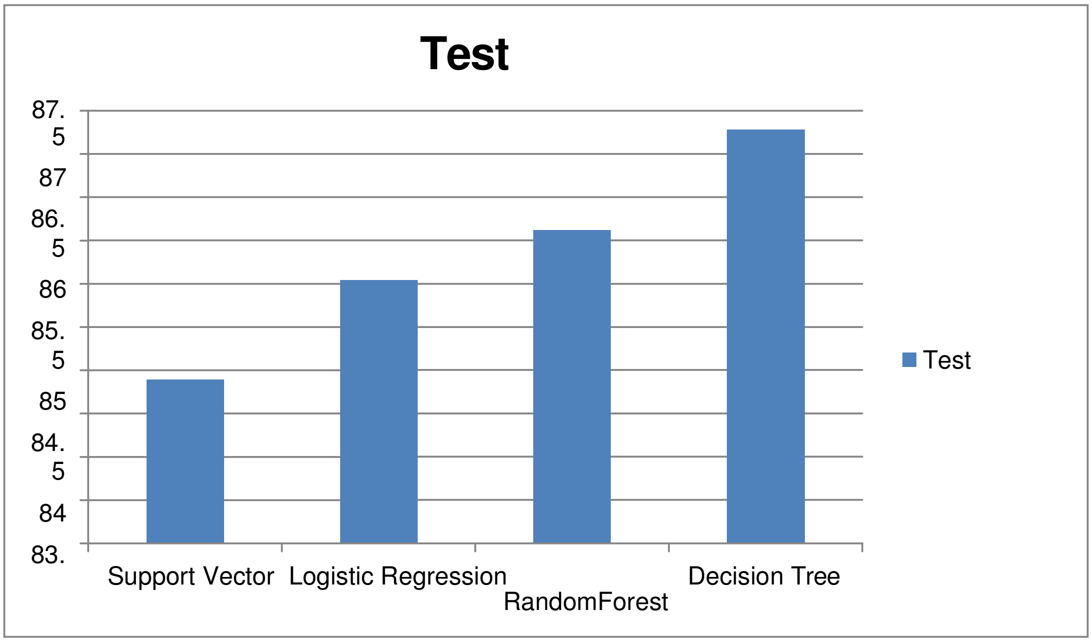

Fig 4.comparison of Test Accuracy of various Machine learning model

The Following figure represents the Comparison of the Recall of the different machine learning models**.**

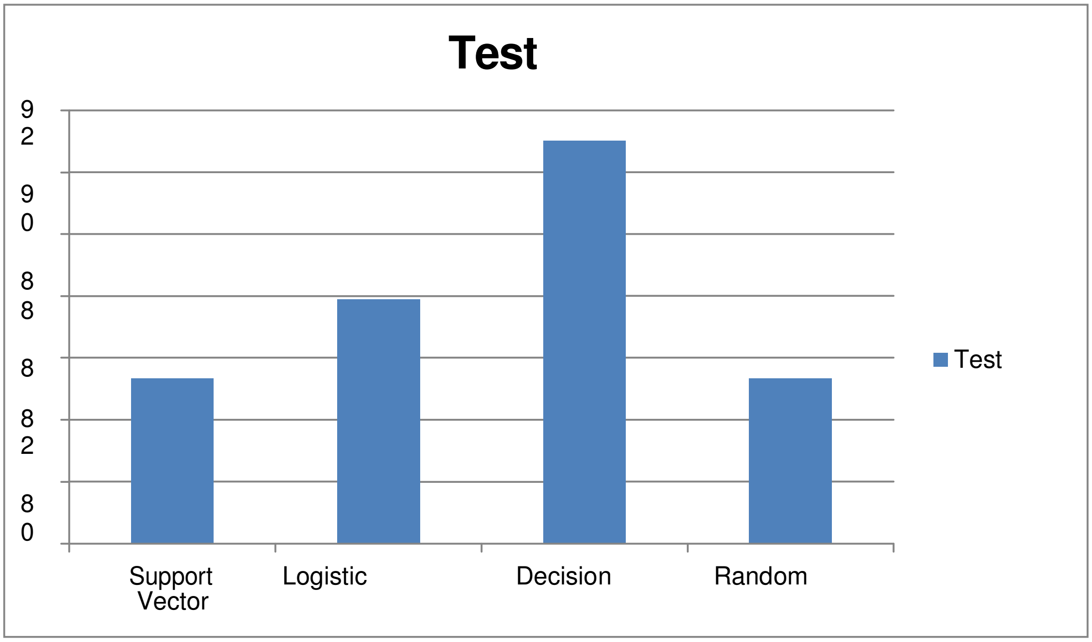

Fig 5.comparison of Test Recall of various Machine learning models

The Following figure represents the Comparison of the Area under the Curve of the different machine learning models.

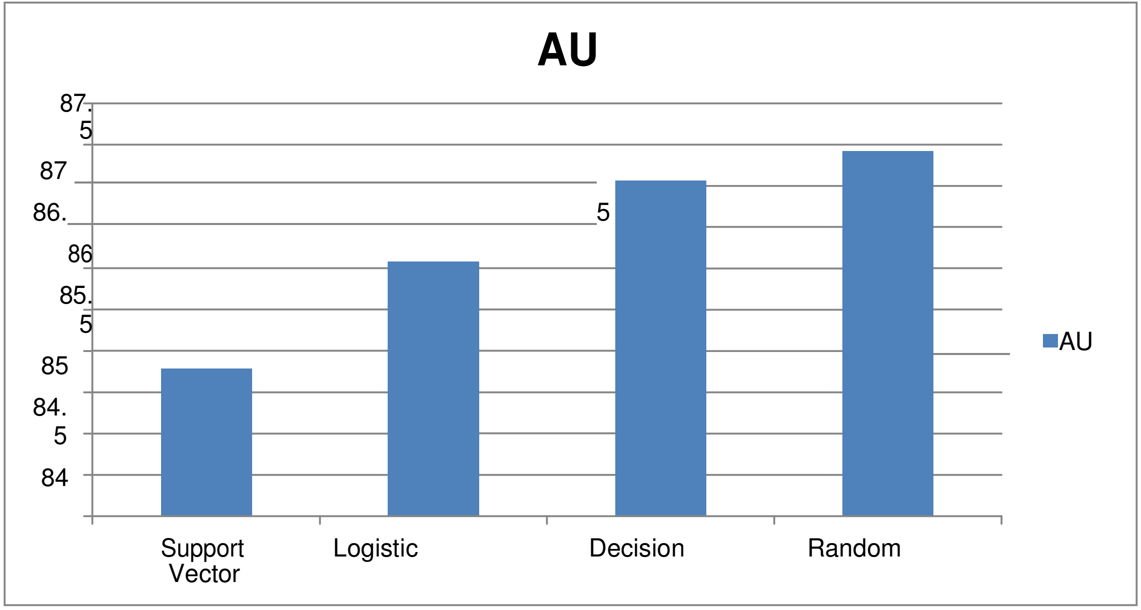

Fig 6.comparison of Test AUC of various Machine learning model.

V.CONCLUSION

Various Machine Learning Algorithms are implemented for finding the fraudulent applications among the set of data of the given credit card applications. It is considered as the classification problem. The machine learning algorithms employed are Support vector classifier, Logistic Regression, Decision Tree and Random Forest. The performance parameters considered for evaluating the model are accuracy, precision, recall and Area under the curve. The future work would include considering this as the problem of finding the outliers. The outliers will be identified with the help of deep learning model.

REFERENCES

[1] SumitMisra, Soumyadeep Thakur, ManosijGhosh, Sanjoy Kumar Saha, ,"An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction", Procedia Computer Science, Volume 167, 2020, Pages 254-262,https://doi.org/10.1016/j.procs.2020.03.219.

[2] SiddhantBagga, AnishGoyal, Namita Gupta, ArvindGoyal, "Credit Card Fraud Detection using Pipeling and Ensemble Learning", Procedia ComputerScience,Volumem173,2020,

[3]Darwish, S.M. A bio-inspired credit card fraud detection model based on user behavior analysis suitable for business management in electronic banking. *J Ambient Intell Human Comput* (2020). https://doi.org/10.1007/s12652-020-01759-9

[4]Ajeet Singh &Anurag Jain (2020) Cost-sensitive metaheuristic technique for credit card fraud detection, Journal of Information and Optimization Sciences, 41:6, 1319-1331, DOI: 10.1080/02522667.2020.1809090[5]S. C. Dubey, K. S. Mundhe and A. A. Kadam, "Credit Card Fraud Detection using Artificial Neural Network and BackPropagation," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 268-273, doi:10.1109/ICICCS48265.2020.9120957.

[6]R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1264-1270, doi:10.1109/ICICCS48265.2020.9121114.

[7]Dawei Cheng, et.al ," Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection", proceedings of the AAAI conference on Artificial Intelligence. Vol 34 No 01: AAAI-20 ,https://doi.org/10.1609/aaai.v34i01.5371

[8]Z. Li, G. Liu and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," in IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 569-579, April 2020, doi: 10.1109/TCSS.2020.2970805.

[9] Yvan Lucas, Pierre-EdouardPortier, LéaLaporte, Liyun He-Guelton, Olivier Caelen, Michael Granitzer, Sylvie Calabretto, "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs".

[10]NaoufalRtayli, NourddineEnneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization" , Journal of InformationSecurityandApplications,Volume55,2020,https://doi.org/10.1016/j.jisa.2020.102596.

[11]A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," in IEEE Access, vol. 8, pp. 25579-25587, 2020, doi: 10.1109/ACCESS.2020.2971354.

[12] Honghao Zhu, Guanjun Liu, Mengchu Zhou, Yu Xie, Abdullah Abusorrah, Qi Kang, "Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to creditcard fraud detection", Neurocomputing.

[13] ToluwaseAyobamiOlowookere, Olumide Sunday Adewale, "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach", Scientific African, Volume 8, 2020,https://doi.org/10.1016/j.sciaf.2020.e00464.