# PREVENTION OF CREDIT CARD FRAUD DETECTION USING NOVEL SUPPORT VECTOR MACHINE COMPARED WITH RANDOM FOREST

**B.Sri Sai Chowdary[1], J. Chenni Kumaran[2*]**

**Article History: Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

**Abstract**

**Aim:** The main aim of the research is to detect Credit Card Fraud using Novel Support Vector Machine (SVM) compared with Random Forest (RF).

**Materials and Methods:** When implementing an accurate prediction model it might not be sufficient to just consider one or two parameters. This analysis will be fed to the prediction model. Following the novel SVM algorithm, Random Forest algorithm based on the previously collected datasets, credit card fraud with calculations can be predicted.

**Result:** Comparison is done by using SPSS Software. The Support vector Machine algorithm produces 95.87% whereas Random Forest algorithm produces 94.89% accuracy while detecting credit card fraud on a data set ($p > 0.05$). Hence Support Vector Machine is better than Random forest.

**Conclusion:** After using iterations to get that by using novel SVM algorithm gets 95.87% (0.95) and Random Forest algorithm gets 94.89% (0.94). So it can be stated that by using novel SVM Algorithm performs with more accuracy than RF Algorithm.

**Keywords:** Credit Card Fraud Detection, Novel Support Vector Machine, Random Forest, Machine Learning, fraudulent, Cardholder, Parameter

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

## 1. Introduction

Support Vector Machines can help to reduce risk and improve the quality of service extended to customers in order to succeed in business (Gardos and Cole 1976). The criminals can steal the information of any cardholder and use it for online purchases until the cardholder contacts the bank to block the card (Paruchuri 2017). Researchers have developed machine learning algorithms to predict credit card fraud, although this research has progressed well, it is still challenging in the following areas: how to preprocess imbalanced data, choosing a criterion to judge the performance of different algorithms, and finding an efficient and effective algorithm (Zhang, Bhandari, and Black 2020). Censoring and time varying covariates need to be formally introduced into the analysis through survival methodology to accommodate incomplete data, and proper likelihood functions need to be developed to examine censoring and time varying covariates (Zhang, Bhandari, and Black 2021). The initial task is how to discover beneficial aspects to differentiate scam operators with benevolent operators (Zhang, Bhandari, and Black 2021). Fraud detection is a process of monitoring the transaction behavior of a cardholder in order to detect whether an incoming transaction is done by the cardholder or other parameters(Sudha and Akila 2021).

In the last five years, Science Direct identified almost 42 research articles on Credit card Fraud Detection using machine learning using Novel Support Vector Machine compared with Random Forest.The classification of fraudulent transactions with highly imbalanced classes of data becomes very challenging. Indeed, the number of normal transactions belonging to the normal class is, comparatively, much larger than the number of fraudulent transactions class which contains abnormal cases ("Enhanced Credit Card Fraud Detection Based on SVM-Recursive Feature Elimination and Hyper-Parameters Optimization" 2020). A credit card is a physical medium used for selling goods or services without having cash in hand. Credit Card Fraud Detection (CCFD) is the procedure to identify whether a transaction is normal or abnormal. It can then efficiently discover fraudulent patterns and predict transactions that are probably to be fraudulent. Machine Learning techniques consist in inferring a prediction model on the basis of a set of examples (Tiwari et al. 2021). The unbalanced problem of the data set was solved by using hybrid sampling methods together. On this data set, comparative performance evaluations have been conducted (Fujita et al. 2021). A Fraud Detection System (FDS) should not only detect fraud cases effi-ciently, but also be cost-effective in the sense that the cost invested in transaction screening should not be higher than the loss due to frauds(Venkata Suryanarayana, Balaji, and Venkateswara Rao 2018). A striking case in point is the Ponzi scheme perpetuated by former NASDAQ chairman Bernard Madoff, which has led to the loss of approximately US$50 billion worldwide ("The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature" 2011).Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The research gap identified from the existing system shows poor accuracy. The study is to improve the accuracy of Classification (Zhang, Bhandari, and Black 2020) by incorporating Novel Support Vector Machine and comparing performance with Random Forest. The proposed model improves classifiers to achieve more accuracy for Credit card Fraud Detection.

## 2. Materials and Methods

This study setting was done in the Soft Computing Laboratory, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The number of required samples in research are taken as two groups in which group 1 is Novel Support Vector Machine compared with group 2 of Random Forest Algorithm. The samples were taken from the device and iterated 10 times to get desired accuracy with G power. A dataset consisting of a collection of stocks was downloaded from Kaggle repository (salihfurkansaglam 2021)

**Support Vector Machine**

Support Vector Machine or SVM is one of the most famous Supervised Learning algorithms, which is utilized for Classification as well as Regression issues. Notwithstanding, essentially, it is utilized for Classification issues in Machine Learning.The objective of the Novel Support Vector Machine algorithm is to make the best line or decision boundary that can segregate n-dimensional space into classes so can undoubtedly put the new element in the right classification later on.This best decision boundary is called a hyperplane.

**Algorithm for Novel Support Vector Machine**

Step 1: Load the important libraries

Step 2: Import the dataset and extract the X variables and Y separately.

Step 3: Divide the dataset into train and test

Step 4: Initializing the SVM classifier model
Step 5: Fitting the SVM classifier model
prediction_SVM_all = classifier.predict(X_test_all) #And finally, we predict our data test.
Step 6: Coming up with predictions
cm = confusion_matrix(y_test_all, prediction_SVM_all)
plot_confusion_matrix(cm,class_names)
Step 7: Evaluating the model's performance

**Random Forest**

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

**Algorithm for Random Forest**

Step 1: Load the important libraries
Step 2: pick N irregular records from the informational index.
Step 3: Build a decision tree based on these N records.
Step 4:Pick the quantity in the Random variable algorithm and rehash steps 2 and 3.

```
# importing rfc classifier
from sklearn.ensemble import RandomForestClassifier
RFC = RandomForestClassifier()
RFC.fit(xtrain, ytrain)
ypred = RFC.predict(xtest)
```

Step 5: Fitting the Random forest model

```
n_outliers = len(fraud_)
n_errors = (ypred != ytest).sum()
print("Random Forest Classifier")
```

Step 6: Evaluating the model's performance

Recall that the testing setup includes both hardware and software configuration choices. The laptop has an Intel Core i5 5th generation CPU with 8GB of RAM, an x86-based processor, a 64-bit operating system, and a hard drive. Currently, the software runs on Windows 10 and is programmed in Python. Once the program is finished, the accuracy value will appear. Procedure: Wi-Fi laptop connected. Chrome to Google Collaboratory search Write the code in Python. Run the code. To save the file, upload it to the disc, and create a folder for it. Log in using the ID from the message. Run the code to output the accuracy and graph.

**Statistical Analysis**

SPSS is a software tool used for statistics analysis. The proposed system utilized 10 iterations for each group with predicted accuracy noted and analyzed. Independent samples t-test was done to obtain significance between two groups.

## 3. Results

Table 1 shows the accuracy value of iteration of Novel Support Vector Machine and random forest. Table 2 represents the Group statistics results which depicts Novel Support Vector Machine with mean accuracy of 95.87%, and standard deviation is 1.07. Random forest has a mean accuracy of 94.89% and standard deviation is 1.82. The Proposed Novel Support Vector Machine algorithm provides better performance compared to the random forest algorithm. Table 3 shows the independent samples T-test value for Novel Support Vector Machine and random forest with Mean difference as 0.986, std Error Difference as 0.670. Significance value is observed as 0.92 ($p > 0.05$).

Figure 1 shows the bar graph comparison of mean of accuracy on Novel Support Vector Machine and random forest algorithm. Mean accuracy of Novel Support Vector Machine is 95.87% and random forest is 94.89 %.

## 4. Discussion

In this study the iteration based on the previous historical datasets are considered for getting the accurate value of credit card fraud detection Using Novel Support Vector Machine Algorithm and Random Forest algorithm on the basis of anaconda navigator(Jupituer notebook) and SPSS.After Getting many iterations to compare all the Selective algorithms are verified and decide to get The accurate values of credit card frauds.

The implementation of the machine was to learn about credit card fraud detection. The have a look at makes a speciality of the credit card datasets obtained from diverse portals belonging to some districts of Karnataka inside the country. Datasets ordered in a properly based totally way. The Novel Support Vector Machine set of guidelines is used for the prediction model and its accuracy is received. The destiny is outstanding for the implementation of system studying algorithms inside the vicinity of credit card fraud detection and are hoping to put in force greater advanced algorithms in order that the system becomes more efficient. Hoping to make tool prediction extra sturdy and attain excessive accuracy with the assistance of greater datasets and advanced algorithms.

## 5. Conclusion

In this study, stock price prediction using Novel Support Vector Machine algorithm provides better accuracy than random forest algorithm.

**Declaration**

**Conflict of Interests**

No conflict of interests in this manuscript

**Authors Contribution**

Author SS Chowdary was involved in data collection, data analysis, manuscript writing. Author J Chenni Kumaran was involved in conceptualization, data validation, and critical review of manuscript.

**Acknowledgement**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

**Funding:** Thanks for the following organizations for providing financial support that enabled us to complete the study.

1. Qbec Infosol
2. Saveetha School of Engineering.
3. Saveetha University
4. Saveetha Institute of Medical and Technical Sciences

## 6. References

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

"Enhanced Credit Card Fraud Detection Based on SVM-Recursive Feature Elimination and Hyper-Parameters Optimization." 2020. *Journal of Information Security and Applications* 55 (December): 102596.

Fujita, Hamido, Ali Selamat, Jerry Chun-Wei Lin, and Moonis Ali. 2021. *Advances and Trends in Artificial Intelligence. From Theory to Practice: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26-29, 2021, Proceedings, Part II*. Springer Nature.

Gardos, G., and J. O. Cole. 1976. "Maintenance Antipsychotic Therapy: Is the Cure Worse than the Disease?" *The American Journal of Psychiatry* 133 (1): 32–36.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Paruchuri, Harish. 2017. "Credit Card Fraud Detection Using Machine Learning: A Systematic Literature Review." *ABC Journal of Advanced Research* 6 (2): 113–20.

salihfurkansaglam. 2021. "FraudDetection-Accuracy%99 Recall%99 Precision%99." Kaggle. December 24, 2021.

https://kaggle.com/salihfurkansaglam/fraudde tection-accuracy-99-recall-99-precision-99.

Sudha, C., and D. Akila. 2021. "Credit Card Fraud Detection System Based on Operational & Transaction Features Using SVM and Random Forest Classifiers." In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. IEEE. https://doi.org/10.1109/iccakm50778.2021.93 57709.

"The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature." 2011. *Decision Support Systems* 50 (3): 559–69.

Tiwari, Anurag, Viney Dhiman, Mohamed A. M. Iesa, Haider Alsarhan, Abolfazl Mehbodniya, and Mohammad Shabaz. 2021. "Patient Behavioral Analysis with Smart Healthcare and IoT." *Behavioural Neurology* 2021 (November): 4028761.

Venkata Suryanarayana, S., G. N. Balaji, and G. Venkateswara Rao. 2018. "Machine Learning Approaches for Credit Card Fraud Detection." *International Journal of Engineering & Technology* 7 (2): 917–20.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806 .

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Zhang, Dongfang, Basu Bhandari, and Dennis Black. 2020. "Credit Card Fraud Detection Using Weighted Support Vector Machine." *Applied Mathematics* 11 (12): 1275–91.

———. 2021. "Covariate Selection for Mortgage Default Analysis Using Survival Models." *Mathematical Finance. An International Journal of Mathematics, Statistcs and Financial Economics* 11 (2): 218–33.

**Tables and Figures**

Table 1. Accuracy Values for novel SVM and RF

| S.NO | SVM | RF |
|---|---|---|
| 1 | 96.00 | 92.00 |
| 2 | 97.00 | 93.50 |
| 3 | 94.90 | 94.68 |
| 4 | 97.00 | 96.80 |
| 5 | 96.00 | 95.00 |
| 6 | 95.70 | 96.00 |
| 7 | 96.89 | 92.00 |
| 8 | 94.90 | 95.90 |
| 9 | 96.60 | 96.25 |
| 10 | 93.80 | 96.80 |

Table 2. Group Statistics Results-novel SVM has an mean accuracy (95.87%), std.deviation (1.07), whereas for RF has mean accuracy (94.89%), std.deviation (1.82).

| Group Statistics | | | | | |
|---|---|---|---|---|---|
| | **Groups** | **N** | **Mean** | **Std deviation** | **Std. Error Mean** |
| **Accuracy** | SVM | 10 | 95.8790 | 1.07268 | .33921 |
| | RF | 10 | 94.8930 | 1.82849 | .57822 |

Table 3. Independent Samples T-test - novel SVM seems to be significantly better than RF (p=0.99)

| **Accuracy** | **Independent Samples Test** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Levene's Test for Equality of Variances** | | **T-test for Equality of Means** | | | | | | |
| | **F** | **Sig** | **t** | **df** | **Sig(2-tailed)** | **Mean Difference** | **Std.Error Difference** | **95% Confidence Interval of the Difference** | |
| | | | | | | | | **Lower** | **Upper** |
| **Equal variances assumed** | 3.163 | .092 | 1.47 | 18 | .159 | .98600 | .67037 | -.42240 | 2.39440 |
| **Equal variances not assumed** | | | 1.47 | 14.539 | .163 | .96000 | .67037 | -.44683 | 2.41883 |

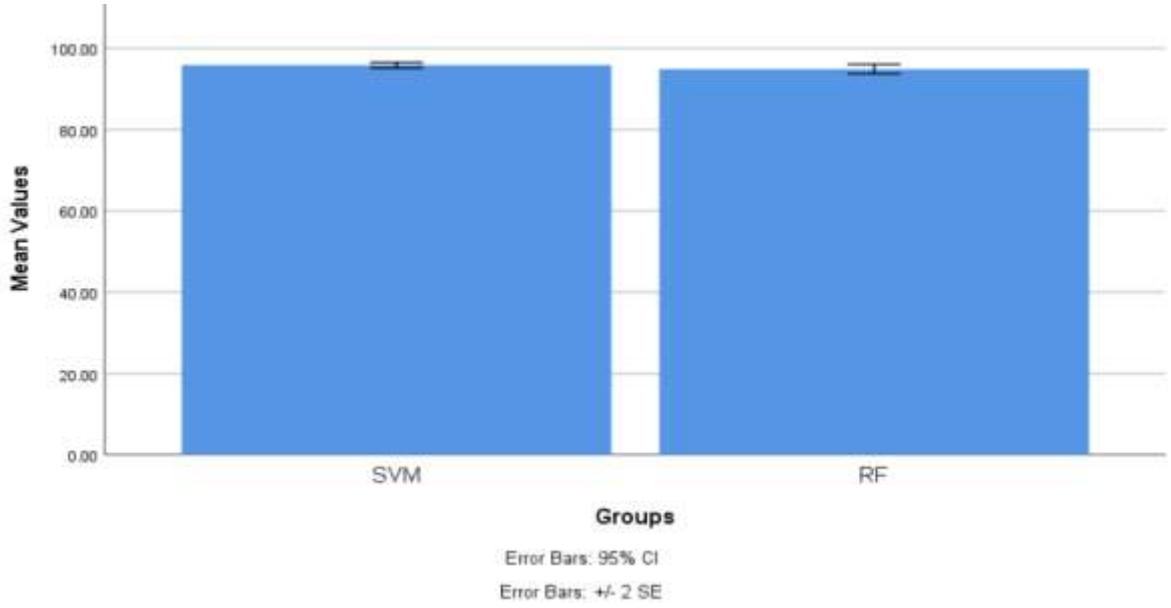


Fig. 1. Bar Graph Comparison on mean accuracy of novel SVM (95.87%) and RF (94.89%). X-axis: novel SVM, RF, Y-axis: Mean Accuracy with ±2SE.