



## TRANSACTION DATA SIMULATOR FOR CREDIT CARD FRAUD DETECTION USING NOVEL LOGISTIC REGRESSION ALGORITHM AND COMPARE THE ACCURACY RATE WITH SUPPORT VECTOR MACHINES ALGORITHM

G. Satya Sai Vivek<sup>1</sup>, T. Rajesh Kumar<sup>2\*</sup>

---

**Article History:** Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

---

### Abstract

**Aim:** To Predict the Transaction data simulator for Credit Card Fraud detection using Novel Logistic Regression Algorithm and Compare the accuracy rate with Support Vector Machines Algorithm.

**Materials and Methods:** Using the gradient boosting builds on tree and Support vector Machine algorithm, this study implements the process to detect the fraudulent and to get the best accuracy by comparing the algorithms. Here the G-power test analysis was carried out with a confidence interval of 80% and the sample size for the two groups are 20.

**Result:** Novel Logistic Regression algorithm was done using Data Quality issues and Support Vector Machine Algorithm with the Accuracy of 94.20% and 76.00% respectively. There is a statistical 2-tailed significant difference in accuracy for two algorithms is 0.002 ( $p < 0.05$ ) by performing independent samples t-tests.

**Conclusion:** Novel Logistic Regression performs significantly better than the Support Vector Machines in Credit Card Fraudulent detection.

**Keywords:** Machine Learning, Support Vector Machine, Data Simulator, Fraud Detection, Credit Card, Novel Logistic Regression.

---

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

<sup>2\*</sup>Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

## 1. Introduction

In recent years, the predominant information mining concerns individuals with charge card misrepresentation location models in light of information mining. Since our concern is drawn nearer as an arrangement issue, traditional information mining calculations are not straightforwardly applicable (Georgieva, Markova, and Pavlov 2019). This project is to propose a Credit card misrepresentation identification framework utilizing supervised learning calculation. regulated calculations are developmental calculations that target acquiring better arrangements as time progresses (Padvekar et al. 2016). A credit card is the most well-known method of installment in supervised techniques (Lacruz and Saniie 2021). As the quantity of Mastercard clients is rising around the world, wholesale fraud precision is expanded, and cheats are likewise increasing. In the virtual card buy, just the card data is required, for example, data simulator, card number, termination date, secure code, and so on. Such buys are regularly done on the Internet or over the phone (C. and C. 2020). To submit extortion in these sorts of buys, supervised techniques are used for individuals just to realize the card subtleties (Dorronsoro et al. 1997). The method of installment for online buy is for the most part done by credit card. The subtleties of Credit cards to be kept hidden (Montague 2004). To get Credit card protection, the subtleties ought not to be spilled (Prusti and Rath 2019).

There are 200 research articles published on fraudulent detection in sciencedirect and 240 papers on Google Scholar and 50 research articles were published in IEEE xplore. John O. Awoyemi proposed an examination through which the exhibitions of a few calculations were assessed when they were applied on charge card misrepresentation information that is exceptionally slanted (Bodepudi 2021). The European cardholders' 284,807 exchanges were utilized as a source to produce the dataset of Visa exchanges (Porkess and Mason 2012). On the slanted information, a half breed approach of under-testing and oversampling is performed (Bianchini, Maggini, and Scarselli 2009a). On crude and preprocessed information, there are three unique strategies applied in Python (Padvekar et al. 2016). In view of specific boundaries like accuracy, responsiveness, precision, adjusted arrangement rate, etc, the exhibitions of these strategies are assessed (Dorronsoro et al. 1997). It is seen through the accomplished outcomes that in contrast with strategic relapse and backing vector machines draws near, the presentation of calculated relapse is better.

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). The research gap identified from the survey is that there are many methods proposed for credit card fraud detection in the crowd but most of the methods which are proposed have less accuracy rate. The main aim of this study is to recognize credit card fraud detection based on transaction data simulator by using Novel logistic regression algorithm and support vector machine algorithm to attain better accuracy (Prusti and Rath 2019).

## 2. Materials and Methods

The research work was performed in the laboratory of Deep learning, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The sample size taken for conducting the experiment was 20. Two groups are considered as classifiers algorithms in order to classify the prediction of the fare amount, machine learning classification algorithms are used. Group 1 is the Novel Logistic Regression algorithm and Support Vector Machines Algorithm is group 2 and they are compared for more by using a dataset to calculate the Support Vector Machines Algorithm using a clinical calculator. The application of algorithm for Data quality issues and Novel Logistic Regression algorithm with the CREDIT CARD dataset testing the algorithm with the accuracy 80%, alpha and beta values with 0.05, 0.2 respectively and threshold value is 0.002 (Azhan and Meraj 2020).

The data collection is taken from the open source access website IEEE-dataport.org that is used for Support Vector machine algorithms using Transaction Data Simulator and Novel Logistic Regression technique. First of all, we obtained our dataset from Kaggle, a data analysis website which provides datasets. Inside this dataset, there are 31 columns out of which 28 are named as v1-v28 to protect sensitive data. The Jupyter software with windows 10.1 system has been used to develop this project. The proposed system uses two groups Novel Logistic Regression technique and Support Vector Machines Algorithm where this algorithms are fitted into the dataset which is then tested and trained for the process of estimating the frauds happened with user's credit card where the cost estimation and the Time estimation is known, the sample size of the dataset is 74.

### Novel Logistic Regression Algorithm

This algorithm is used in various fields, including machine learning, most medical fields, and social sciences. Novel Logistic Regression may be used to predict the risk of developing a given software / data simulator based on observed characteristics of the product. It is also used in marketing applications such as prediction of a customer's propensity to purchase a product or halt a subscription.

**Step 1:** In the node we first compute the linear part.

**Step 2:**  $Z=WX+B$

**Step 3:** Notice that the result is only a single value.

**Step 4:**  $A=1/1+e$

**Step 5:** First, always inspect the columns and data types.

**Step 6:** Now that we have prepared the data we can create a model in Tensor flow.

**Step 7:** In Numpy, and in math in general  $XY=Xi, jY, k$ .

**Step 8:** print accuracy.

**Step 9:** End.

### Support Vector Machines Algorithm

Support vector machines are a set of supervised learning methods used for classification, regression, and outlier detection based on transaction data simulators. All of these are common tasks in machine learning. You can use them to detect cancerous cells based on millions of images or you can use them to predict future driving routes with a well-fitted regression model.

**Step 1:** Import the dataset

**Step 2:** Explore the data to figure out what they look like

**Step 3:** Pre-process the data

**Step 4:** Split the data into attributes and labels

**Step 5:** Divide the data into training and testing sets

**Step 6:** Train the SVM algorithm

**Step 7:** Make some predictions

**Step 8:** Evaluate the results of the algorithm

### Statistical Analysis

The analysis was done by IBM SPSS version 21. Independent variables project, team\_exp, year\_end etc and dependent variables are id, length, effort (Porkess and Mason 2012). For both proposed and existing algorithms iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analysing accuracy. With value obtained from The iteration Independent Sample T-test was performed.

### 3. Result

In Table 1, it was observed that the Novel Logistic Regression algorithm is significantly better than the Support Vector Machines Algorithm. In the CREDITCARD dataset, it is observed that accuracy and performance of Novel Logistic Regression was significantly better than Support Vector Machines Algorithm.

In Table 2, the Novel Logistic Regression algorithm achieved a mean of 93.1500, Standard Deviation of 0.81802 and the Standard Error Mean of 0.25868. For the Support Vector Machines Algorithm. The mean is 71.5900, Standard Deviation is 3.05158 and standard error mean is 0.96499. In Table 3, The 2-tailed significance value smaller than 0.000 ( $p<0.005$ ) showed that our hypothesis holds good.

Figure 1 represents the mean accuracy of the prediction of Credit card fraud for Novel Logistic Regression Algorithm and Support Vector Machines Algorithm. The Novel Logistic Regression model obtained 94.20% accuracy and the Support Vector Machines Algorithm obtained 76.00% accuracy. The Novel Logistic Regression algorithm technique achieved better performance than the Support Vector Machines Algorithm.

### 4. Discussion

Based on the above study it is observed that the Novel Logistic Regression algorithm has better accuracy of 92.45% than the Support Vector Machines Algorithm which has 71.02% in prediction of credit card fraud. There is a statistical 2-tailed significant difference in accuracy for two algorithms is 0.002 ( $p<0.05$ ) by performing independent samples t-tests

In the existing systems the accuracy for the Novel Logistic regression and random forest are 84.40% and 73.46% respectively. This analysis paper makes use of machine learning to predict the hassle of knowing all the certain parameters (Ryman-Tubb and Krause 2018). Two styles of models of cost estimation are there; one is algorithmic and the alternative is non algorithmic (Bianchini, Maggini, and Scarselli 2009b). The challenge with project estimation in software development, unlike other industries, is that it's often done with partial data and sometimes with incorrect data, too (Pumplun et al. 2021). Factors affecting the algorithms are sample size of the dataset and test size of dataset (Patel 2019). Based on the above finding, the proposed algorithm was chosen to improve the accuracy.

The limitations of the Novel Logistic regression algorithm is that on high dimensional datasets, this may lead to the model being over-fit on the training set, which means overstating the accuracy of predictions on the training set and thus the model

may not be able to predict accurate results on the test set. So on high dimensional datasets, Regularization techniques should be considered to avoid overfitting (Goyal and Sharma 2020). The work can be done in an enhanced manner for effort estimation by considering many attributes for the Novel Logistic regression algorithm in future so that it can work effectively and improve the prediction accuracy. Attributes like TeamExp, Entities, PointAdjus can improve the accuracy.

## 5. Conclusion

In this research, the aim was to predict Credit card Fraud with accurate values. Here we proposed a method for Credit card Fraud Detection using machine-learning techniques, Accuracy of Novel Logistic Regression algorithm (92.45%) performs better than the accuracy of Support Vector Machines Algorithm (71.02%) in predicting Credit card fraud detection.

### Declarations

### Conflicts of Interest

The authors of this paper declare no conflicts of interest.

### Author Contribution

Author DV was involved in data collection, data analysis, manuscript writing. Author TRK was involved in conceptualization, guidance and critical review of manuscript.

### Acknowledgment

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

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Maso Technologies, Bangalore.
2. Saveetha University, Chennai.
3. Saveetha Institute of Medical and Technical Sciences, Chennai.
4. Saveetha School of Engineering, Chennai.

## 6. References

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture

Environment and Enhancing the Yield." *Sustainable Energy Technologies and Assessments*.

<https://doi.org/10.1016/j.seta.2022.102142>.

Azhan, Mohammed, and Shazli Meraj. 2020. "Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques." *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. <https://doi.org/10.1109/iciss49785.2020.9316002>.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." *Bioresource Technology* 351 (May): 126956.

Bianchini, Monica, Marco Maggini, and Franco Scarselli. 2009a. *Innovations in Neural Information Paradigms and Applications*. Springer Science & Business Media.

———. 2009b. *Innovations in Neural Information Paradigms and Applications*. Springer Science & Business Media.

Bodepudi, Hariteja. 2021. "Credit Card Fraud Detection Using Unsupervised Machine Learning Algorithms." *International Journal of Computer Trends and Technology*. <https://doi.org/10.14445/22312803/ijctt-v69i8p101>.

C., Dr Victoria Priscill, and Victoria Priscill C. 2020. "Analysis of Performance on Classification Algorithms for Credit Card Fraud Detection." *Journal of Advanced Research in Dynamical and Control Systems*. <https://doi.org/10.5373/jardcs/v12sp3/20201391>.

Dorransoro, J. R., F. Ginel, C. Sgnchez, and C. S. Cruz. 1997. "Neural Fraud Detection in Credit Card Operations." *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council* 8 (4): 827–34.

Georgieva, Sevdalina, Maya Markova, and Velizar Pavlov. 2019. "Using Neural Network for Credit Card Fraud Detection." *RENEWABLE ENERGY SOURCES AND TECHNOLOGIES*. <https://doi.org/10.1063/1.5127478>.

Goyal, Yogita, and Anand Sharma. 2020. *Credit Card Fraud Detection and Analysis Through Machine Learning*.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO<sub>2</sub> Smart Carriers for Self-Healing High Corrosion Protective

- Coatings.” *Environmental Research*. <https://doi.org/10.1016/j.envres.2022.113095>.
- Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. “Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application.” *Materials Chemistry and Physics*. <https://doi.org/10.1016/j.matchemphys.2022.125900>.
- Lacruz, Francisco, and Jafar Saniie. 2021. “Applications of Machine Learning in Fintech Credit Card Fraud Detection.” 2021 *IEEE International Conference on Electro Information Technology (EIT)*. <https://doi.org/10.1109/eit51626.2021.9491903>.
- Montague, David A. 2004. *Fraud Prevention Techniques for Credit Card Fraud*. Trafford Publishing.
- Padvekar, Suchita Anant, Atharva College of Engineering Department of Computer Science University of Mumbai, Pragati Madan Kangane, and Komal Vikas Jadhav. 2016. “Credit Card Fraud Detection System.” *International Journal Of Engineering And Computer Science*. <https://doi.org/10.18535/ijecs/v5i4.22>.
- Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. “A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications.” *Journal of Energy Storage*. <https://doi.org/10.1016/j.est.2022.104422>.
- Patel, Ankur A. 2019. *Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data*. “O’Reilly Media, Inc.”
- Porkess, Roger, and Stephen Mason. 2012. “Looking at Debit and Credit Card Fraud.” *Teaching Statistics*. <https://doi.org/10.1111/j.1467-9639.2010.00437.x>.
- Prusti, Debachudamani, and Santanu Kumar Rath. 2019. “Web Service Based Credit Card Fraud Detection by Applying Machine Learning Techniques.” *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. <https://doi.org/10.1109/tencon.2019.8929372>.
- Pumplun, Luisa, Mariska Fecho, Nihal Wahl, Felix Peters, and Peter Buxmann. 2021. “Adoption of Machine Learning Systems for Medical Diagnostics in Clinics: Qualitative Interview Study.” *Journal of Medical Internet Research* 23 (10): e29301.
- Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. “Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications.” *Sustainable Energy Technologies and Assessments*. <https://doi.org/10.1016/j.seta.2022.102102>.
- Ryman-Tubb, Nick, and Paul Krause. 2018. *Machine Learning Advances in Payment Card Fraud Detection*. Academic Press.
- Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. “Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System.” *Computational Intelligence and Neuroscience* 2022 (February): 5906797.
- Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarathanan, Natchimuthu Karmegam, and Woong Kim. 2022. “The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?” *Fuel*. <https://doi.org/10.1016/j.fuel.2022.123494>.
- Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. “Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review.” *International Journal of Biological Macromolecules* 209 (Pt A): 951–62.
- Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. “Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity.” *Fuel*. <https://doi.org/10.1016/j.fuel.2022.123814>.

## Tables and Figures

Table 1. Comparison of prediction accuracy between Novel Logistic Regression and Support Vector Machines Algorithm The Novel Logistic Regression clustering obtained accuracy of 92.45% compared to Support Vector Machines Algorithm having 71.02%.

Execution	Novel Logistic Regression	Support Vector Machines
1	93.00	76.00
2	92.60	75.20
3	92.00	74.00
4	94.02	73.40
5	94.00	72.00
6	92.30	71.00
7	93.01	70.00
8	94.00	69.30
9	94.12	68.00
10	92.45	67.00

Table 2. Mean of the Novel Logistic Regression algorithm is 93.1500 and Support Vector Machines Algorithm mean value is 71.5900. The below table will show the Novel Logistic Regression algorithm obtained standard deviation (0.81916) and standard error means(0.25868).

	Group	N	Mean	std.deviation	std.Error mean
Accuracy	Novel Logistic Regression	10	93.1500	0.81916	0.25868
	Support Vector Machines	10	71.5900	3.05158	0.96499

Table 3. Independent Samples Test. The accuracy increases and the error rate decreases. The 2-tailed significance is less than 0.002.

	Levene's Test for Equality of variances		T-test for Equality of Means						
	f	sig	t	df	sig(2-tailed)	mean diff	Std. Error difference	95% Confidence Interval of the Difference	
								Lower	Upper

<b>Equal Variances assumed</b>	14.357	0.001	21.580	18	0.000	21.56000	0.99906	19.46105	23.65895
<b>Equal Variances not assumed</b>			21.580	10.287	0.000	21.56000	0.99906	19.34233	23.7777

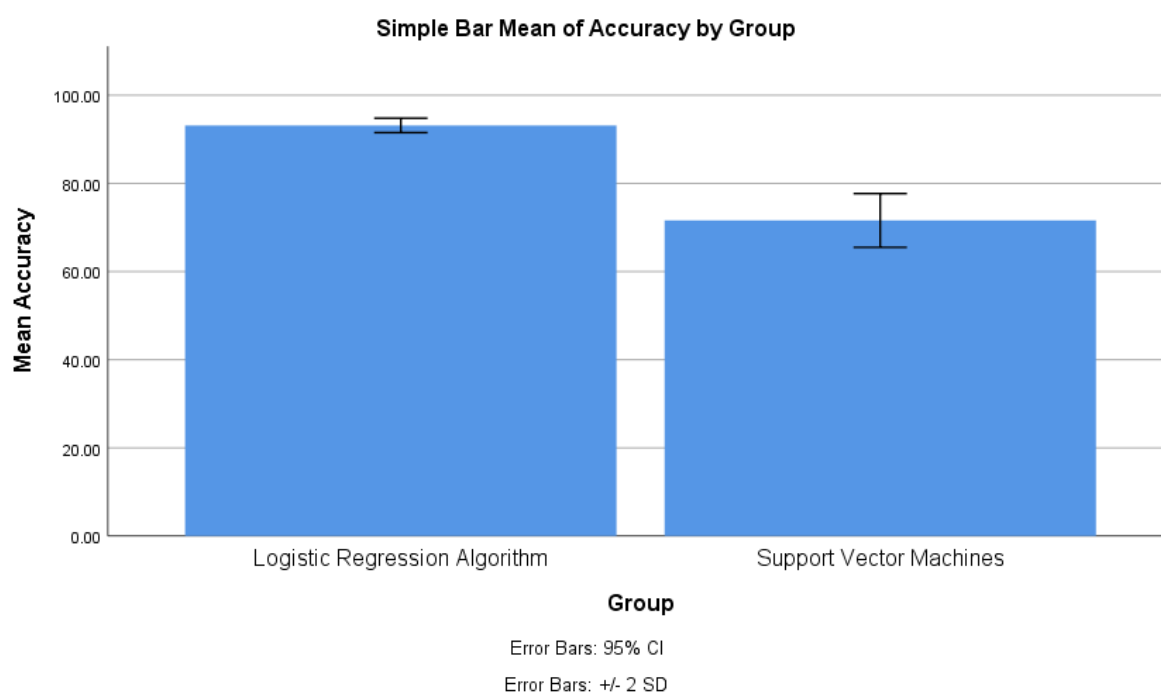


Fig. 1. Prediction accuracy for the two algorithms. The accuracy of the Novel Logistic Regression algorithm is better than the accuracy of the Support Vector Machines Algorithm. X Axis: Novel Logistic Regression vs Support Vector Machines Algorithm Y Axis: Mean accuracy of detection +/-2SD.