

## **Credit Card Fraud Detection Using Machine Learning Techniques: Incorporating Data Analytics and Data Modelling**

<sup>1</sup>Arushi Srivastava  
BCA 6th Semester  
Kalinga University, Raipur  
[srivastava.arushi0612@gmail.com](mailto:srivastava.arushi0612@gmail.com)  
m

<sup>2</sup>Anish Anand  
BCA 6th Semester  
Kalinga University,  
Raipur  
[loianesptg24@gmail.com](mailto:loianesptg24@gmail.com)  
m

<sup>3</sup>Pawan Kumar  
Assistant Professor,  
Faculty of CS & IT  
Kalinga University, Raipur  
[pawan@kalingauniversity.ac.in](mailto:pawan@kalingauniversity.ac.in)  
n

<sup>4</sup>Prachi Kushwaha  
BCA 6th Semester  
Kalinga University, Raipur  
[kushwahaprachi1312@gmail.com](mailto:kushwahaprachi1312@gmail.com)

<sup>5</sup>S Sanskar Verma  
BCA 6th Semester  
Kalinga University, Raipur  
[sanskarverma511@gmail.com](mailto:sanskarverma511@gmail.com)

### **Abstract**

Credit card swindling is a growing problem in today's financial markets. The degree of frauds has increased rapidly in recent years, causing serious financial consequences that harm many organizations, businesses and government agencies. The aim of this work is to identify fraudulent credit card transactions using machine learning techniques, i.e., logistic regression model, on a dataset of credit card transactions to prevent unauthorized access to customer accounts by fraudsters.

**Keywords:** Credit card, Fraud identification, Binary variables, Logistic regression, Neural network, Bayesian network.

### **Introduction**

E-commerce has evolved significantly since its inception. It has become an essential tool for most organizations, businesses and administrations to promote international trade. A key factor in the success of e-commerce is the ease of online payments. Every time we discuss a financial exchange, we must also consider financial fraud. Financial fraud is an intentional crime in which the perpetrator profits by depriving the victim of his rights or by obtaining a financial advantage. Due to credit card transactions becoming the popular mode of payment in recent years, fraud has

increased significantly. The research aims to detect fraudulent credit card transactions by carefully analyzing the data set and performing logistic regression modeling to create a reliable transaction algorithm.

## Literature Review

Zareapoor and his research team used various methods to identify the most effective model to detect fraudulent transactions, evaluate the model based on accuracy, detection rate and cost. Applied models included neural networks, Bayesian networks, SVM, KNN, among others. The neural network performed well and provided fast detection with reasonable accuracy. KNN showed good speed with moderate accuracy, while SVM scored one lower due to its slowness and moderate accuracy. In terms of cost, all the models built were expensive (Zareapoor et al., 2012).

Awoyemi's team compared the use of three ML techniques for credit card fraud detection, the first being KNN, the second Naive Bayes, and the third logistic regression. They chose from different distributions to observe different results. The highest accuracy for the 10:90 distribution is Naive Bayes 97.5%, followed by KNN 97.1%, logistic regression performed poorly with a precision of 36.4%. The second division studied was 34:66. In this case, KNN led with a low accuracy of 97.9%, followed by Naive Bayes at 97.6%.

## Methodology

### Data Collection

The data set comes from the open access platform Kaggle.com. It consists of information on businesses done using credit cards in Europe in just two days in 2013. The data consists of 29 columns, 280,800 rows. 25 properties are numeric variables. For confidentiality and privacy, customer data has been transformed by PCA (Principal Component Analysis) transformation, the other three attributes are "Time" which contains seconds elapsed between the initial transaction and subsequent transactions, "Amount" is each transaction and the final attribute "Class" which includes binary variables where "1" is a fraudulent event and "0" is a legitimate event.

### Data Cleaning

The credit card details were loaded into Python with the import command and the data cleaning process was completed. This process involved two main tasks: 1. removing nulls and missing values, and 2. Deviation management. The data contains a total of 1,048,575 events and no null or missing values were found.

## **Data Preprocessing**

Data preprocessing is necessary before implementing a machine learning algorithm because different models have different requirements for predictors and training data can affect prediction performance. The purpose of data processing is to clean and prepare the data to a state that minimizes bias, controls for missing values, and reduces variability. The dataset contains both numerical and categorical data, so categorical data must be coded before they can be used for modeling.

## **Data Modelling**

A logistic regression model is a statistical model in which estimates are formed based on the relationship between dependent qualitative variables or at least three variables. For multinomial logistic regression, there are one or more independent explanatory variables, which can be either qualitative or quantitative. The Logistic Regression model managed to obtain an accuracy of 93.8026 percent.

## **Results**

We obtained an accurate credit card fraud detection rate of 93.8026% using the logistic regression model. Compared to existing models, this proposed model is suitable for a wider dataset and provides more accurate results. The logistic regression model provides better performance with a lot of training data, but the speed during testing and running still suffers. Adding pre-processing methods would also help.

## **Conclusion**

To conclude, the project focusses solely on identifying fraud related to e-transactions done using credit cards. This was achieved using a logistic regression model. The model will help to decrease credit card frauds and increase customer satisfaction by providing them with enhanced security.

## **References**

[1] Adepoju, O., Wosowei, J., Lawte, S. and Jaiman, H. (2019). A comparative evaluation of credit card fraud detection using machine learning techniques. 2019 Global Conference on Technology Development (GCAT). <https://doi.org/10.1109/gcat47503.2019.8978372>

[2] Alenzi, H. Z. and Aljehane, N. O. (2020). Credit card fraud detection using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). <https://doi.org/10.14569/ijacsa.2020.0111265>

[3] Awoyemi, J. O., Adetunmbi, A.O and Oluwadare, S.A. (2017). Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computer Networking and Informatics (ICCNI). <https://doi.org/10.1109/iccni.2017.8123782>

[4] Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., and Rohith, R. (2020). Credit card fraud detection using machine learning algorithms. diary.