# Fraud Prediction in Movie Theater Credit Card Transactions using Machine Learning

**Areej Alshutayri**
Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
aoalshutayri@uj.edu.sa (corresponding author)

*Received: 12 April 2023 | Revised: 29 April 2023 and 1 May 2023 | Accepted: 2 May 2023*

## ABSTRACT

**This paper highlights how the proliferation of online transactions, especially those involving the use of credit cards, has resulted in the emergence of new security flaws that pose threats to customers and enterprises worldwide. E-commerce and other forms of online monetary transactions have become essential in the manufacturing and service sectors, propelling the global economy. The widespread and dependent connectivity of mobile payment systems using credit card transactions presents chances for fraud, risk, and security breaches. In light of the importance of accurately predicting fraud incidents through payment procedures, this study investigated the credit card payment methods used for movie tickets, using the machine learning logistic regression method to analyze and predict such incidents. This study used a dataset from cinema ticket credit card transactions made in two days of September 2013 by European cardholders, including 284,807 transactions out of which 492 were fraudulent purchases. The results of the proposed method showed a prediction accuracy of 99%, proving its high prediction performance.**

***Keywords-fraud prediction; movie theater; credit card; machine learning***

## I. INTRODUCTION

Credit card frauds have a negative impact not just on customers, but also on businesses [1]. Along with the rise in the popularity of credit card use and online transactions, there is also an increase in fraudulent activity. Therefore, it is vital to detect fraudulent use of credit cards to prevent monetary losses and preserve the personal and financial data of consumers [2]. Identifying fraudulent activity on credit cards can be accomplished using several methods from various angles. Rule-based systems or machine learning algorithms can examine patterns of normal behavior and identify deviations from the norm [3]. Trust is another issue that needs to be addressed in terms of credit card concerns [4]. Payment systems are the engine that drives e-commerce business, and a substantial amount of trust and security in cutting-edge technology is necessary [5]. Furthermore, it is not unusual for mobile users to switch to new solutions, adjust their payment routines, and encounter other types of security challenges [6-7].

This study used a machine learning model to predict fraudulent activity in credit card payment transactions. Today, electronic payment methods are increasingly growing and are increasingly vital and convenient [8-9]. Several studies have shown that phone-based payments are a crucial and necessary addition to the available payment retail [10]. Electronic payment systems are particularly vulnerable to fraud and cyber security breaches since they store sensitive consumer information such as ID, card number, card name, purchase history, and more [11]. Because of all these interconnected factors, it is essential to examine the problem of credit card fraud from a variety of perspectives to determine how it can be resolved. This study also considered other dimensions associated with electronic payment operations to detect unusual or suspicious behaviors, taking into account additional facets related to electronic payment operations, such as the geographic location of transactions. For instance, if a credit card is used unexpectedly multiple times in a short period, this could be indicative of fraudulent activity [12-13]. Another aspect lies in the analysis of user behavior to identify any unique patterns that may emerge. When a user suddenly starts making large transactions or uses his card at irregular times, this may be an indication of fraudulent activity [14]. Machine learning is one of the most important fields in this sector and currently depends on algorithms that can boost an organization's productivity and performance. There are four distinct methodologies of machine learning: supervised, unsupervised, semi-supervised, and reinforced learning [15-18].

This study used Logistic Regression (LR) as a classification algorithm to assign the data to a discrete set of classes. LR resembles the Linear Regression model, however, it uses a more complex cost function called Sigmoid function or logistic function rather than a linear one [19]. LR is a method for making predictions and its outcomes are converted into probability values. To limit the cost function to values between 0 and 1, an LR hypothesis must be taken into account. As a result, linear functions are unable to adequately describe this because they can have a value that is either more than 1 or less than 0, both of which are prohibited by the LR hypothesis [19]. When using a classification strategy, LR is applied to the observations of the discrete classes involved, operating on the principle of probability and employing the predictive analysis on the provided scenarios. To accurately estimate the necessary probabilities, the sigmoid function converts any real values between 0 and 1 by:

$$f(x) = \frac{1}{1+e^{-(x)}} \qquad (1)$$

To validate the hypothesis and ensure its consistency with the expected assumption, the value of the hypothesis must fall somewhere between 0 and 1 by:

$$h\theta(X) = \frac{1}{1+e^{-(\beta 0+\beta 1X)}} \qquad (2)$$

When the prediction function is applied to the classifier, the classifier will provide values based on a set of outputs on probability. This establishes the decision boundaries with a selected threshold value. The accuracy of the optimized model is represented by the cost function after it has been optimized to provide the least amount of inaccuracy:

$$f\theta \frac{1}{2} \sum_{i=1}^{m} (h_\theta)\left(x^{(i)}\right) - (y^{(i)})^2 \qquad (3)$$

Gradient descent is used to reduce the cost values using:

$$\theta_j \coloneqq \theta_j - \propto \sum_{i=1}^{m} (h_\theta)\left(x^{(i)}\right) - (y^{(i)}) x_j{}^{(2)} \qquad (4)$$

Several previous studies used LR for detection and prediction. The need to apply security in e-commerce in a highly effective way creates the concepts of asset protection, security prediction, and vulnerability detection to save end users and resources that can be exploited. LR was a vital component in the development of deep and machine-learning models on the Google Cloud Platform [20]. Similarly, LR was used to forecast student performance in community colleges [21]. In [22], a road crash zone was modeled using LR to determine the nature of the accidents and the types of people that were involved. In a similar vein, a privacy-protecting LR-based diagnosis method for digital healthcare was presented in [23]. The effectiveness of machine-learning RL for tomography processing was shown in [24]. In [25], LR and a survival model were used in Russian exports. In [26], an intelligent categorization of coal structure was presented, using multinomial LR. In [27], LR was effectively used in nursing. In [28], a prediction model for clinical applications was developed using LR, showing its feasibility. In [29], an SK-Tree method to detect malicious events on a portion of a publicly available dataset achieved an AUROC score of 98%. In [19], several machine learning algorithms were investigated to detect and analyze frauds in online transactions with a European credit card dataset, proposing a novel fraud detection method to stream transaction data by analyzing the old transaction details of customers and extracting behavioral patterns. The study in [30] focused on fraud events in real-world transactions, using and evaluating a series of machine learning algorithms, such as Vector Machine, Naive Bayes, K-Nearest Neighbor, and LR. In [31], a credit card fraud detection system was presented using machine learning algorithms that included the modeling of the dataset. The past credit card transactions were modeled with data considered as fraud, and this model was used to define if a new transaction is fraud. This study showed that fraud detection is a classification problem and focused on preprocessing and analyzing datasets using different anomaly detection algorithms, such as local outlier factor and isolation forest, on a Principal Component Analysis (PCA)-transformed credit card transaction dataset. PCA reduced the feature dimensionality [32] and kept the most effective features to improve the prediction process [33].

The above-mentioned studies in the e-commerce security domain used different machine learning algorithms to examine which fit best to their datasets and provide the best accuracy scores. In other words, these studies show the importance of carefully choosing machine learning algorithms that are compatible with the underlying dataset and have the best accuracy performance.

## II. METHODOLOGY

This study focused on applying machine-learning algorithms to predict credit card fraud transactions in movie theaters. Figure 1 shows the flow chart of the proposed method.

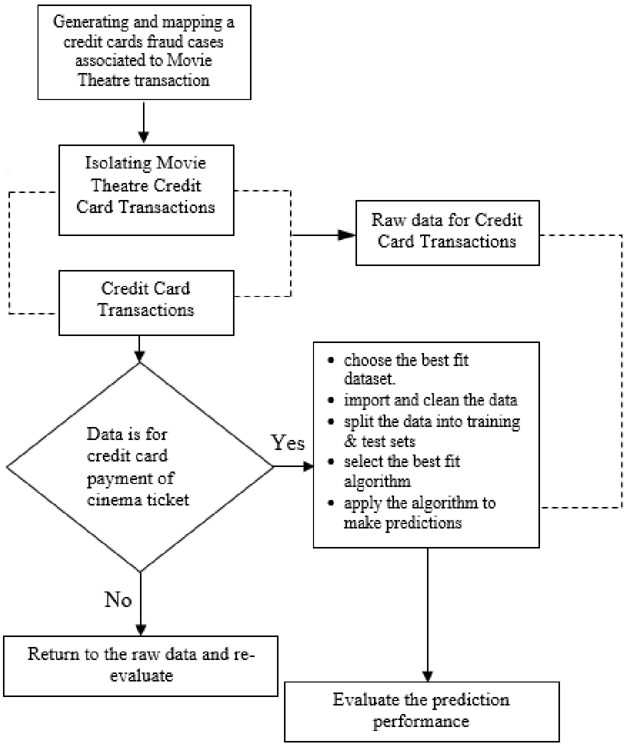

Generating and mapping a credit cards fraud cases associated to Movie Theatre transaction
Isolating Movie Theatre Credit Card Transactions
Raw data for Credit Card Transactions
Credit Card Transactions
Data is for credit card payment of cinema ticket
Yes
No
choose the best fit dataset.
import and clean the data
split the data into training & test sets
select the best fit algorithm
apply the algorithm to make predictions
Return to the raw data and re-evaluate
Evaluate the prediction performance

Fig. 1. The flow diagram of the proposed method.

LR was used to make the predictions, leading to successful results. These two approaches were combined into one. As a consequence, the procedure started with the construction of the model, followed by the collection and preprocessing of the dataset. Then, the model was trained and tested for its prediction accuracy. The most important phase of this method was the generation and mapping of credit card fraud incidents in movie theater transactions. The study began by isolating movie theater credit card transactions, from all the credit card transactions contained in the dataset. In the end, only the dataset for credit card payments of movie tickets was selected for the best-fit test, which was followed by the import and cleaning of the data, their division into training and test sets, and finally using the machine-learning LR algorithm to determine the best-fit for making predictions.

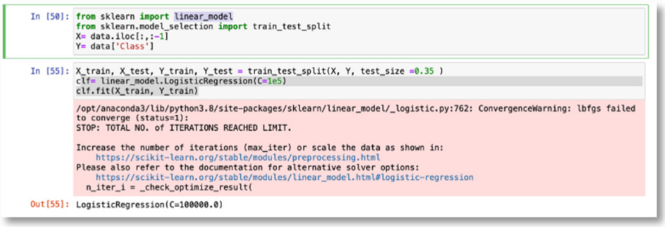
The necessary data were collected from Kaggle [34], as shown in Table I, which contains a total of 284,807 transactions and 492 fraudulent transactions that took place in two days in September 2013 by European cardholders. The dataset has a huge imbalance, as the positive class (frauds) accounts for only 0.172% of all transactions. The features of the dataset were represented by a number value (V1-V28). The PCA method was used, and "Time" and "Amount" were the only unchanged features. The "Time" feature saves the number of seconds that elapsed between the first and each subsequent transaction, for each subsequent transaction in the dataset. The "Class" feature is the response variable, assigned the value of 1 in the event of fraud and 0 in all other cases, whereas "Amount" denotes the total amount of the transaction. This study selected only transactions that involved purchasing a ticket and additional services for a cinema. The data include not just numerical values, but also a description and a "Label".

TABLE I. THE SAMPLE DATASET GENERATED FROM THE CREDIT FRAUD INVESTIGATION

| Time | V1 | V2 | V3 | | V27 | V28 | Amount | Class |
|---|---|---|---|---|---|---|---|---|
| 0 | -1.35981 | -0.07278 | 2.536347 | ... | 0.133558 | -0.02105 | 149.62 | 0 |
| 0 | 1.191857 | 0.266151 | 0.16648 | ... | -0.00898 | 0.014724 | 2.69 | 0 |
| 1 | -1.35835 | -1.34016 | 1.773209 | ... | -0.05535 | -0.05975 | 378.66 | 0 |
| 1 | -0.96627 | -0.18523 | 1.792993 | ... | 0.062723 | 0.061458 | 123.5 | 0 |
| 2 | -1.15823 | 0.877737 | 1.548718 | ... | 0.219422 | 0.215153 | 69.99 | 0 |
| 2 | -0.42597 | 0.960523 | 1.141109 | ... | 0.253844 | 0.08108 | 3.67 | 0 |
| 4 | 1.229658 | 0.141004 | 0.045371 | ... | 0.034507 | 0.005168 | 4.99 | 0 |
| 7 | -0.64427 | 1.417964 | 1.07438 | ... | -1.20692 | -1.08534 | 40.8 | 0 |
| 7 | -0.89429 | 0.286157 | -0.11319 | ... | 0.011747 | 0.142404 | 93.2 | 0 |
| 9 | -0.33826 | 1.119593 | 1.044367 | ... | 0.246219 | 0.083076 | 3.68 | 0 |
| 10 | 1.449044 | -1.17634 | 0.91386 | ... | 0.04285 | 0.016253 | 7.8 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ... | ⋮ | ⋮ | ⋮ | ⋮ |
| 172782 | 0.219529 | 0.881246 | -0.63589 | ... | 0.131507 | 0.081265 | 5.49 | 0 |
| 172783 | -1.77513 | -0.00424 | 1.189786 | ... | 0.181205 | 0.215243 | 24.05 | 0 |
| 172784 | 2.03956 | -0.17523 | -1.19683 | ... | 0.454379 | 0.130308 | 79.99 | 0 |
| 172785 | 0.120316 | 0.931005 | -0.54601 | ... | -0.08083 | -0.07507 | 2.68 | 0 |
| 172786 | -11.8811 | 10.07178 | -9.83478 | ... | 0.21794 | 0.068803 | 2.69 | 0 |
| 172787 | -0.73279 | -0.05508 | 2.03503 | ... | 0.943651 | 0.823731 | 0.77 | 0 |
| 172788 | 1.919565 | -0.30125 | -3.24964 | ... | 0.068472 | -0.05353 | 24.79 | 0 |
| 172788 | -0.24044 | 0.530483 | 0.70251 | ... | 0.004455 | -0.02656 | 67.88 | 0 |
| 172792 | -0.53341 | -0.18973 | 0.703337 | ... | 0.108821 | 0.104533 | 10 | 0 |

## III. EXPERIMENTAL ANALYSIS AND PRESENTATION OF THE RESULTS

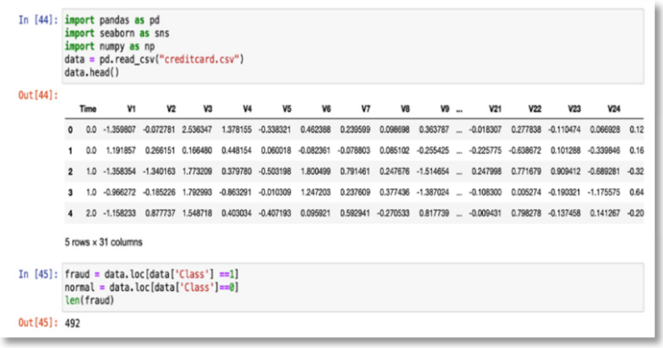
Figure 2 shows the particular processes to import the dataset and initialize the fraud variable.



```
In [50]: from sklearn import linear_model
         from sklearn.model_selection import train_test_split
         X= data.iloc[:,:-1]
         Y= data['Class']

In [55]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size =0.35 )
         clf= linear_model.LogisticRegression(C=1e5)
         clf.fit(X_train, Y_train)

         /opt/anaconda3/lib/python3.8/site-packages/sklearn/linear_model/_logistic.py:762: ConvergenceWarning: lbfgs failed
         to converge (status=1):
         STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

         Increase the number of iterations (max_iter) or scale the data as shown in:
             https://scikit-learn.org/stable/modules/preprocessing.html
         Please also refer to the documentation for alternative solver options:
             https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
           n_iter_i = _check_optimize_result(

Out[55]: LogisticRegression(C=100000.0)
```

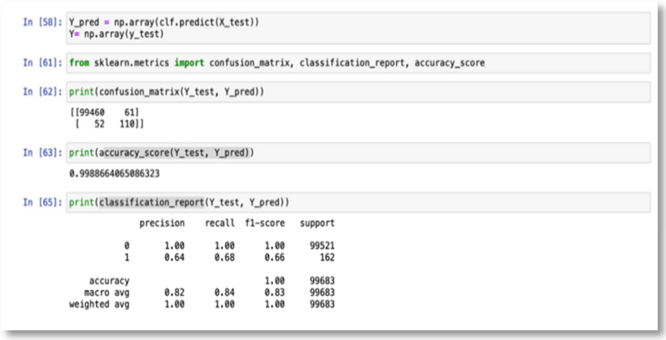
Fig. 2. Experimental import and initialization of the fraud variable.

Python, Pandas, and Scikit-learn were used to develop the proposed method. In the first stage, the dataset was loaded using Pandas, and the value of the fraud variable was initialized to 0 or 1, based on the dataset. In the second step, the test sample was defined, and the algorithm was applied using Scikit-learn, as shown in Figure 3. In the third phase, LR was used to perform the classification, and then the accuracy of the model was assessed, as shown in Figure 4. The variable X is independent and denotes the credit card transactions, while the variable Y is dependent and denotes the class of the transaction or the fraud flag. Figure 4 shows the results of the proposed method, indicating an accuracy of 0.9988.



```
In [44]: import pandas as pd
         import seaborn as sns
         import numpy as np
         data = pd.read_csv("creditcard.csv")
         data.head()

Out[44]:
     Time        V1        V2        V3        V4        V5        V6        V7        V8        V9 ...       V21       V22       V23       V24
  0   0.0 -1.359807 -0.072781  2.536347  1.378155 -0.338321  0.462388  0.239599  0.098698  0.363787 ... -0.018307  0.277838 -0.110474  0.066928  0.12
  1   0.0  1.191857  0.266151  0.166480  0.448154  0.060018 -0.082361 -0.078803  0.085102 -0.255425 ... -0.225775 -0.638672  0.101288 -0.339846  0.16
  2   1.0 -1.358354 -1.340163  1.773209  0.379780 -0.503198  1.800499  0.791461  0.247676 -1.514654 ...  0.247998  0.771679  0.909412 -0.689281 -0.32
  3   1.0 -0.966272 -0.185226  1.792993 -0.863291 -0.010309  1.247203  0.237609  0.377436 -1.387024 ... -0.108300  0.005274 -0.190321 -1.175575  0.64
  4   2.0 -1.158233  0.877737  1.548718  0.403034 -0.407193  0.095921  0.592941 -0.270533  0.817739 ... -0.009431  0.798278 -0.137458  0.141267 -0.20

5 rows x 31 columns

In [45]: fraud = data.loc[data['Class'] ==1]
         normal = data.loc[data['Class']==0]
         len(fraud)

Out[45]: 492
```

Fig. 3. Selecting the SK-learn algorithm, importing the model, and defining the test sample.



```
In [58]: Y_pred = np.array(clf.predict(X_test))
         Y= np.array(y_test)

In [61]: from sklearn.metrics import confusion_matrix, classification_report, accuracy_score

In [62]: print(confusion_matrix(Y_test, Y_pred))

         [[99460    61]
          [   52   110]]

In [63]: print(accuracy_score(Y_test, Y_pred))

         0.9988664065086323

In [65]: print(classification_report(Y_test, Y_pred))

                       precision    recall  f1-score   support

                    0       1.00      1.00      1.00     99521
                    1       0.64      0.68      0.66       162

             accuracy                           1.00     99683
            macro avg       0.82      0.84      0.83     99683
         weighted avg       1.00      1.00      1.00     99683
```

Fig. 4. Calling the predict function, applying classification, and checking accuracy.

The accuracy results of this study were compared with similar studies that used LR to detect fraudulent credit card transactions, as shown in Table II. The study in [35] used a credit card transaction dataset from the UCI Machine Learning Data Repository to detect frauds in credit card transactions, while studies [36-38] used the same dataset as this study. All these studies used the LR method to achieve high accuracy and predictive performance, regardless of their individual approaches. This study achieved the best prediction performance result.

TABLE II. COMPARATIVE PREDICTION PERFORMANCE

| Study | Algorithm | Accuracy |
|---|---|---|
| [35] | Logistic Regression | 77.97% |
| [36] | Logistic Regression | 99.26% |
| [37] | Logistic Regression | 91.20% |
| [38] | Logistic Regression | 74.65% |
| This study | Logistic Regression | 99.88% |

## IV. CONCLUSION

The results of this study indicate that the implementation of a machine learning approach is imperative for the detection and prevention of fraudulent activities within payment operations. This study proposed the use of a logistic regression model to investigate fraudulent credit card transactions, using a machine learning approach to analyze various credit card payment methods for purchasing movie tickets. The raw data were divided into training and testing samples. The use of historical data to accurately predict future outcomes is a significant impetus for the extensive implementation of machine learning in various sectors. This study showcased the impact of the widespread adoption of digital financial transactions, particularly those involving credit cards, on the emergence of novel security risks that pose a threat to businesses and customers worldwide. The emergence of e-commerce and other electronic forms of monetary transactions has a crucial role in the production and service sectors, thereby promoting the advancement of the worldwide economy. The study was initiated based on the acknowledgment that, despite their widespread use, mobile credit card payment systems present opportunities for fraudulent activities, potential hazards, and security violations across all industries due to their ubiquitous and interdependent connectivity.

This study investigated credit card transactions related to cinema ticket purchases to detect potentially fraudulent activities. The application of logistic regression machine learning provided an accurate prediction. The dataset used in this study contained credit card transactions made by European cardholders during September 2013 [38]. The purchase of cinema tickets was segregated and analyzed autonomously from the remaining transactions. The dataset contained 284,807 transactions, where 492 were identified as fraudulent. The data were analyzed using logistic regression, and the findings showed a predictive accuracy of 99.88%, indicating an exceedingly prognostic performance.

## REFERENCES

[1] L. Ni, J. Li, H. Xu, X. Wang, and J. Zhang, "Fraud Feature Boosting Mechanism and Spiral Oversampling Balancing Technique for Credit Card Fraud Detection," *IEEE Transactions on Computational Social Systems*, pp. 1–16, 2023, https://doi.org/10.1109/TCSS.2023.3242149.

[2] N. Shirodkar, P. Mandrekar, R. S. Mandrekar, R. Sakhalkar, K. M. Chaman Kumar, and S. Aswale, "Credit Card Fraud Detection Techniques – A Survey," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, Oct. 2020, pp. 1–7, https://doi.org/10.1109/ic-ETITE47903.2020.112.

[3] R. Van Belle, B. Baesens, and J. De Weerdt, "CATCHM: A novel network-based credit card fraud detection method using node representation learning," *Decision Support Systems*, vol. 164, Jan. 2023, Art. no. 113866, https://doi.org/10.1016/j.dss.2022.113866.

[4] S. Saxena, S. Vyas, B. S. Kumar, and S. Gupta, "Survey on Online Electronic Paymentss Security," in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, Dubai, United Arab Emirates, Oct. 2019, pp. 756–751, https://doi.org/10.1109/AICAI.2019.8701353.

[5] M.-H. Yang, J.-N. Luo, M. Vijayalakshmi, and S. M. Shalinie, "Contactless Credit Cards Payment Fraud Protection by Ambient Authentication," *Sensors*, vol. 22, no. 5, Jan. 2022, Art. no. 1989, https://doi.org/10.3390/s22051989.

[6] E. E.-D. Hemdan and D. H. Manjaiah, "Anomaly Credit Card Fraud Detection Using Deep Learning," in *Deep Learning in Data Analytics: Recent Techniques, Practices and Applications*, D. P. Acharjya, A. Mitra, and N. Zaman, Eds. Cham, Switzerland: Springer International Publishing, 2022, pp. 207–217.

[7] S. Saeed, "A Customer-Centric View of E-Commerce Security and Privacy," *Applied Sciences*, vol. 13, no. 2, Jan. 2023, Art. no. 1020, https://doi.org/10.3390/app13021020.

[8] S. Karnouskos, "Mobile payment: A journey through existing procedures and standardization initiatives," *IEEE Communications Surveys & Tutorials*, vol. 6, no. 4, pp. 44–66, 2004, https://doi.org/10.1109/COMST.2004.5342298.

[9] S. Karnouskos, "Mobile payment: A journey through existing procedures and standardization initiatives," *IEEE Communications Surveys & Tutorials*, vol. 6, no. 4, pp. 44–66, 2004, https://doi.org/10.1109/COMST.2004.5342298.

[10] H. S. Pramanik, M. Kirtania, and A. K. Pani, "Essence of digital transformation—Manifestations at large financial institutions from North America," *Future Generation Computer Systems*, vol. 95, pp. 323–343, Jun. 2019, https://doi.org/10.1016/j.future.2018.12.003.

[11] R. Kumar, R. Singh, K. Kumar, S. Khan, and V. Corvello, "How Does Perceived Risk and Trust Affect Mobile Banking Adoption? Empirical Evidence from India," *Sustainability*, vol. 15, no. 5, Jan. 2023, Art. no. 4053, https://doi.org/10.3390/su15054053.

[12] R. Brown, J. Liñares-Zegarra, and J. O. S. Wilson, "Sticking it on plastic: credit card finance and small and medium-sized enterprises in the UK," *Regional Studies*, vol. 53, no. 5, pp. 630–643, May 2019, https://doi.org/10.1080/00343404.2018.1490016.

[13] S. Carbo-Valverde, H. Pérez Saiz, and H. Xiao, "Geographical and Cultural Proximity in Retail Banking," Bank of Canada, Ottawa,

Canada, Staff Working Paper 2023–2, Jan. 2023. https://doi.org/10.34989/swp-2023-2.

[14] M. F. Rahman and M. S. Hossain, "The impact of website quality on online compulsive buying behavior: evidence from online shopping organizations," *South Asian Journal of Marketing*, vol. 4, no. 1, pp. 1–16, Jan. 2022, https://doi.org/10.1108/SAJM-03-2021-0038.

[15] B. K. Ponukumati, P. Sinha, M. K. Maharana, A. V. P. Kumar, and A. Karthik, "An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 8972–8977, Aug. 2022, https://doi.org/10.48084/etasr.5107.

[16] K. Leavitt, K. Schabram, P. Hariharan, and C. M. Barnes, "Ghost in the Machine: On Organizational Theory in the Age of Machine Learning," *Academy of Management Review*, vol. 46, no. 4, pp. 750–777, Oct. 2021, https://doi.org/10.5465/amr.2019.0247.

[17] H. Saleem, K. B. Muhammad, A. H. Nizamani, S. Saleem, and J. Butt, "Data Science and Machine Learning Approach to Improve e-Commerce Sales Performance on Social Web," *International Journal of Advanced Research in Engineering and Technology*, vol. 12, no. 4, pp. 401–424, Apr. 2021.

[18] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An Introduction to Logistic Regression Analysis and Reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, Sep. 2002, https://doi.org/10.1080/00220670209598786.

[19] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Computer Science*, vol. 165, pp. 631–641, Jan. 2019, https://doi.org/10.1016/j.procs.2020.01.057.

[20] E. Bisong, Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Berkeley, CA, USA: Apress, 2019.

[21] Z. Richards and A. M. Kelly, "Predicting community college astronomy performance through logistic regression," *Physical Review Physics Education Research*, vol. 19, no. 1, Mar. 2023, Art. no. 010119, https://doi.org/10.1103/PhysRevPhysEducRes.19.010119.

[22] A. Vieira, B. Santos, and L. Picado-Santos, "Modelling Road Work Zone Crashes' Nature and Type of Person Involved Using Multinomial Logistic Regression," *Sustainability*, vol. 15, no. 3, Jan. 2023, Art. no. 2674, https://doi.org/10.3390/su15032674.

[23] Y. Zhou *et al.*, "A privacy-preserving logistic regression-based diagnosis scheme for digital healthcare," *Future Generation Computer Systems*, vol. 144, pp. 63–73, Jul. 2023, https://doi.org/10.1016/j.future.2023.02.022.

[24] T. Rymarczyk, E. Kozłowski, G. Kłosowski, and K. Niderla, "Logistic Regression for Machine Learning in Process Tomography," *Sensors*, vol. 19, no. 15, Jan. 2019, Art. no. 3400, https://doi.org/10.3390/s19153400.

[25] K. Malec *et al.*, "Energy Logistic Regression and Survival Model: Case Study of Russian Exports," *International Journal of Environmental Research and Public Health*, vol. 20, no. 1, Jan. 2023, Art. no. 885, https://doi.org/10.3390/ijerph20010885.

[26] Z. Wang, Y. Cai, D. Liu, F. Qiu, F. Sun, and Y. Zhou, "Intelligent classification of coal structure using multinomial logistic regression, random forest and fully connected neural network with multisource geophysical logging data," *International Journal of Coal Geology*, vol. 268, Mar. 2023, Art. no. 104208, https://doi.org/10.1016/j.coal.2023.104208.

[27] L. Connelly, "Logistic Regression," *MEDSURG Nursing*, vol. 29, no. 5, Oct. 2020.

[28] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: an overview," *Journal of Thoracic Disease*, vol. 11, no. Suppl 4, pp. S574–S584, Mar. 2019, https://doi.org/10.21037/jtd.2019.01.25.

[29] T. Cochrane, P. Foster, V. Chhabra, M. Lemercier, T. Lyons, and C. Salvi, "SK-Tree: a systematic malware detection algorithm on streaming trees via the signature kernel," in *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, Rhodes, Greece, Jul. 2021, pp. 35–40, https://doi.org/10.1109/CSR51186.2021.9527933.

[30] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, Jan. 2019, pp. 488–493, https://doi.org/10.1109/CONFLUENCE.2019.8776942.

[31] S. P. Maniraj, A. Saini, S. D. Sarkar, and S. Ahmed, "Credit Card Fraud Detection using Machine Learning and Data Science," *International Journal of Engineering Research*, vol. 8, no. 09, pp. 110–115, Sep. 2019.

[32] A. Rahman and M. N. A. Khan, "A Classification Based Model to Assess Customer Behavior in Banking Sector," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2949–2953, Jun. 2018, https://doi.org/10.48084/etasr.1917.

[33] E. Jamalian and R. Foukerdi, "A Hybrid Data Mining Method for Customer Churn Prediction," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2991–2997, Jun. 2018, https://doi.org/10.48084/etasr.2108.

[34] "Credit Card Fraud Detection using Python." https://kaggle.com/code/renjithmadhavan/credit-card-fraud-detection-using-python.

[35] Y. Kumar, S. Saini, and R. Payal, "Comparative Analysis for Fraud Detection Using Logistic Regression, Random Forest and Support Vector Machine." Rochester, NY, USA, Oct. 18, 2020, https://doi.org/10.2139/ssrn.3751339.

[36] T. Kumar, "Comparison of Logistic Regression and Decision Tree method for Credit Card Fraud Detection," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, no. 5, pp. 680–683, May 2021, https://doi.org/10.22214/ijraset.2021.34241.

[37] F. Itoo, Meenakshi, and S. Singh, "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection," *International Journal of Information Technology*, vol. 13, no. 4, pp. 1503–1511, Aug. 2021, https://doi.org/10.1007/s41870-020-00430-y.

[38] M. V. Krishna and J. Praveenchandar, "Comparative Analysis of Credit Card Fraud Detection using Logistic regression with Random Forest towards an Increase in Accuracy of Prediction," in *2022 International Conference on Edge Computing and Applications (ICECAA)*, Tamilnadu, India, Jul. 2022, pp. 1097–1101, https://doi.org/10.1109/ICECAA55415.2022.9936488.