International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Credit Card Fraud Detection Using Machine Learning

Sachin Acharya
Dept. of Computer Science & Engineering
ACS College of Engineering
Bangalore, India
acharya4023@gmail.com

Muhammad Owais A
Assistant Professor
Dept. of Computer Science & Engineering
ACS College of Engineering
Bangalore, India
owais8825811761@gmail.com

R Ram
Dept. of Computer Science & Engineering
ACS College of Engineering
Bangalore, India
ramrreddy2482@gmail.com

Rajasekar G
Dept. of Computer Science & Engineering
ACS College of Engineering
Bangalore, India
rajeshseemalamudi2003@gmail.com

Raghavendra Biradar
Dept. of Computer Science & Engineering
ACS College of Engineering
Bangalore, India
rpbiradar27@gmail.com

ABSTRACT

Credit card fraud has become a main issue in banking, causing big losses and diminishing trust among customers. Conventional methods of detection fail to catch up with new tactics used by fraudsters. The use of machine learning strategies in credit card fraud detection has been discussed here. We apply complex algorithms such as logistic regression, decision trees, random forests, and neural networks to analyze transaction data in order to identify patterns characteristic of fraud patterns. The study focuses on the feature engineering, model estimation, and real-time detection mechanisms to achieve improved accuracy and reduce false positives. Experimental results demonstrate that machine learning-based techniques surpass traditional methods and offer an excellent solution to avoid fraud risk. This research emphasizes the potential of machine learning to revolutionize fraud detection in banking for enhanced security and customer satisfaction.

Keyword: Credit Card Fraud, Machine Learning, Fraud Prevention, Financial Security, Anomaly Detection, Imbalanced Data Classification.

I. INTRODUCTION

The emergence of electronic transactions at a high growth rate and extensive use of credit cards have completely transformed the financial and banking sectors. Yet, this convenience came at the price of heightened credit card fraud, which proved to be a formidable challenge for banks and consumers as well. With every innovation by fraudsters in their strategies, it became tough for the conventional fraud prevention system to remain current. Accordingly, a more sophisticated and responsive set of solutions to identify and capture fraud effectively is needed in an urgent manner. This paper discusses the use of machine learning (ML) methods to solve the increasing issue of credit card fraud within the banking sector [1].

Supervised learning [2], unsupervised learning, and deep learning algorithms are quite helpful for fraud transaction identification. Supervised learning algorithms such as logistic regression, decision trees, and random forests can be trained on labeled information to mark the transactions as fraud or real. Unsupervised learning algorithms such as anomaly detection and clustering are most suited to detect unknown fraud patterns. Deep learning techniques [3], such as neural networks, are most appropriate for highly dimensional and complicated data and therefore they function optimally in real-time fraud detection. Deep learning techniques being used together can enable banks to develop efficient systems that learn to detect changing patterns of fraud and minimize financial loss.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

The use of machine learning to identify credit card fraud is accompanied by a number of challenges. One of the significant problems is the class imbalance, where the fraudulent transactions constitute a very small percentage of all transactions. This can cause biased models towards the majority class, leading to ineffective fraud detection. Oversampling, under sampling, and creating synthetic data (e.g., SMOTE) are utilized to avoid this problem. For avoiding this issue, oversampling, under sampling, and synthetic data generation (e.g., SMOTE) are applied. Feature engineering is utilized to select and manipulate feature attributes from transactional data for optimal model performance. Since any delay in detection can lead to massive financial loss, real- time detection is required. Through data analysis of transactions and testing several ML algorithms, we aim to develop a low- false-positive, effective, and accurate real-time model. The outcome will help banks use ML-based solutions to combat fraud. Lastly, machine learning in fraud detection will enhance financial security and boost the trust of customers in the banking sector.

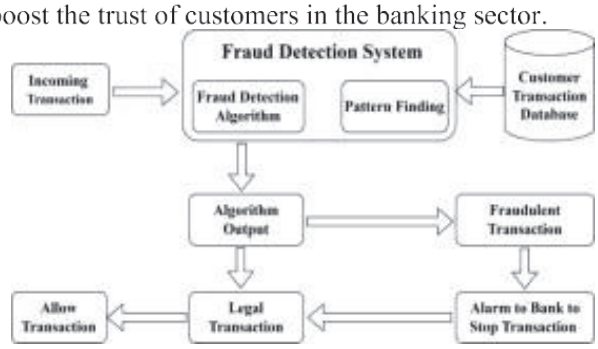

Fig. 1. Credit Card Fraud Detection Using Machine Learning.

II. RELATED WORKS

Credit card fraud detection has been a problem explored at very long lengths in the last two years with numerous techniques being employed to increase the efficiency and effectiveness of fraud detection mechanisms. Statistics and rule-based mechanisms were among early solutions that were most dependent on pre-defined thresholds and patterns in identifying suspicious transactions. Whereas these techniques had helped in the detection of well-known fraud patterns, they were unable to match the latest and refined means of operating by fraudsters. Therefore, emphasis was laid on newer techniques like machine learning, which can process enormous amounts of transaction data and detect sophisticated, non-linear patterns characteristic of fraud.

Supervised learning algorithms have been widely used for credit card fraud detection due to the ability to label the transactions as fraudulent or genuine based on labeled datasets [4]. It has been demonstrated through research that logistic regression, SVM, and decision trees are effective algorithms for detecting fraud. For instance, ensemble algorithms like random forests and gradient boosting have been shown to be very accurate in aggregating strengths of multiple models. However, one of the largest supervised learning challenges is class imbalance in fraud datasets, where fraudulent transactions significantly outnumber legitimate transactions. To address this, techniques such as oversampling, under sampling, and synthetic data creation (e.g., SMOTE) have been employed to balance the dataset as well as to improve model performance

Unsupervised learning techniques have also been identified as having the ability to detect unknown patterns of fraud without relying on labeled data. Clustering methods such as k-means and DBSCAN have been used to cluster similar transactions and separate out outliers that could be indicative of fraud. Algorithms such as isolation forests and auto encoders have proven to detect well rare and unusual transactions which do not represent usual behavior. These unsupervised algorithms are particularly beneficial for discovering new types of fraud that are not based on well-known patterns and therefore complement supervised learning algorithms.

Deep learning, which is a machine learning method [5], has been found to be a highly effective approach in detecting credit card fraud due to its ability to handle high-dimensional and complex data. Some of the neural networks utilized for handling sequences of transactions and extracting important features to detect fraud are recurrent neural networks (RNNs) and convolutional neural networks (CNNs). Research has established that deep learning models are capable of state-of-the-art performance on fraud detection problems, particularly when augmented with methods such as transfer learning and attention mechanisms. Yet, the computational complexity and resource needs of deep learning models pose a problem for There have also been recent experiments on integrating real-time fraud detection systems into banking infrastructure. Stream processing platforms such as Apache Kafka and Apache Flink were used for real-time processing of transaction data and deployment of machine learning models in favor of real-time fraud detection.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

=

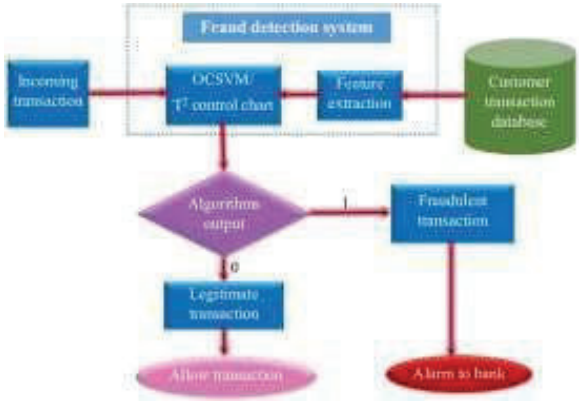

Fig. 2. Data Structure Flow

Credit card fraud detection has been a well-studied topic, with a variety of machine learning techniques [7] applied to enhance security and reduce financial loss. Traditionally, rule-based fraud detection systems have been widely used; these, however, are typified by high false positives and insensitivity to evolving patterns of fraud. In a bid to overcome these limitations, scholars have explored supervised and unsupervised learning approaches. Supervised learning models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Deep Learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have reported encouraging outcomes in identifying fraudulent transactions [8],[9]. Supervised learning models are trained on past transaction data with fraud labels, thereby enabling them to mark new transactions efficiently [10].

Recent studies have aggressively pursued real-time fraud detection via online learning models that update upon identifying evolving patterns in fraud [11]. Graph- based methods have also been investigated for detecting fraud networks based on relationships between entities and transactions. Federated learning has also been proposed as a privacy- protecting alternative to allow different banks to cooperatively detect fraud without exposing sensitive customer information [12]. Although machine learning greatly enhanced the performance of fraud detection, there remain issues like dealing with imbalanced datasets, adversarial attacks, and dynamic fraud tactics. Upcoming research focuses on explainable AI (XAI) to promote interpretability and fairness in fraud detection models.

A. Problem Statements

Credit card fraud presents a significant challenge in the banking industry, causing huge financial losses to both banks and customers. As the number of digital transactions increases, fraudsters constantly come up with advanced methods to circumvent conventional security features [6]. Traditional rule-based fraud detection systems have difficulty keeping up with changing fraudulent patterns and tend to have high false-positive rates, where genuine transactions are reported as fraud, or false negatives, where fraudulent transactions are not detected. Moreover, the extremely imbalanced nature of fraud detection datasets, where fraudulent transactions constitute a minority of all transactions, is a challenge for machine learning models. Effectively detecting fraudulent transactions without excessive false alarms is paramount to preserving customer confidence and financial integrity. Thus, there is a requirement for sophisticated machine learning methods that can identify fraud in real-time, learn dynamic fraud behavior, and enhance the overall accuracy and efficiency of fraud detection in banking systems.

III. PROPOSED METHOD AND ALGORITHM

This study proposes a machine learning-driven fraud detection system that integrates supervised and unsupervised learning approaches to identify fraudulent transactions with high accuracy. The proposed model will employ a mix of feature engineering, anomaly detection, and ensemble learning to increase fraud detection effectiveness. Raw transaction data will be preprocessed for the first time by handling missing values, encoding categorical variables, and solving data imbalance through techniques like Synthetic Minority Over- sampling Technique (SMOTE)

The framework will employ supervised learning methods like Decision Trees, Random Forest, Support Vector Machines (SVM), and Deep Learning models like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) to identify transactions as fraud or real. Besides this, other unsupervised machine learning techniques like isolation forests and auto encoders will be used in order to identify any anomaly in the pattern of transactions. For improving the overall detection rate and reducing false positives, robustness of the model will be obtained by using an ensemble approach with a list of classifiers.

A. Algorithm

Bank credit card fraud detection is based on a set of machine learning algorithms to identify fraud transactions in the right way. Supervised machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), and Gradient Boosting

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

(XGBoost, LightGBM, CatBoost) are the most popular algorithms for fraud classification. They are trained on labeled samples of known fraud historical transactions, and therefore they can detect true as well as fraud transactions.

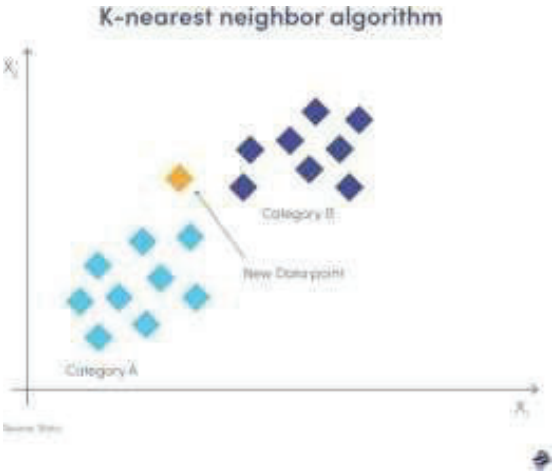

Fig. 3. K-nearest neighbor classification

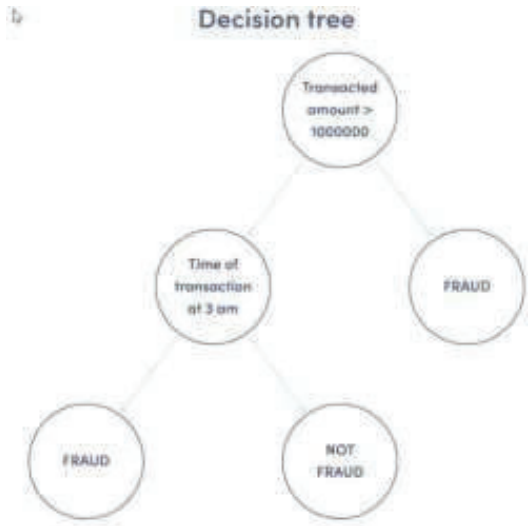

Fig. 4. Decision Tree

B. Module List and Descriptions

Data Collection and Preprocessing: This module is for gathering credit card transaction data samples from banking statements and open sources. Preprocessing data operations such as handling missing values, removing duplicates, transforming categorical variables into numerical variables, and normalizing numeric features are employed. Also, methods such as Synthetic Minority Over-sampling Technique (SMOTE) are employed to solve the imbalance in class and provide an efficient training set.

Data Set: Downloading datasets from Kaggle can be beneficial for data analysis, machine learning, and research. This dataset has 4,850 records and 11 fields, with a size of around 319KB. It seems to deal with credit card transactions, with one record per transaction. The fields have some information about the transaction and cardholder. The "Unnamed: 0" column likely is an index or auto-indexed ID of each row. The "cc_num" column has the credit card number or its masked form, and "category" has the category of the transaction, e.g., grocery shopping, gas, online shopping.The "amt" column captures the amount spent on each transaction, and the "gender" column captures the gender of the cardholder. The "is_fraud" column is a binary flag, with 1 representing a fraudulent transaction and 0 representing a valid one. The "age" column contains the age of the cardholder, and the "trans_month" and "trans_year" columns detail the date of the transaction. Lastly, the "lat_dis" and "long_dis" columns represent the geographic distance (in latitude and longitude) between the transaction location and the cardholder's known location, which may aid in spotting suspicious activity or fraud due to location irregularities.

Feature Selection and Engineering: This module emphasizes identifying and selecting essential transaction attributes for fraud detection. Significant attributes including transaction amount, location, timing, device information, and behavioral patterns of a user are used to improve the accuracy of a model. Reduction of dimension and improvement in computing efficiency are attained through techniques like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE)

Model Training and Classification: Various machine learning algorithms, including Decision Trees, Random Forest, Support Vector Machines (SVM) are implemented in this module. Supervised learning is used for labeled transaction data, while unsupervised techniques identify anomalies without predefined fraud labels.

Anomaly Detection and Fraud Identification: This module utilizes unsupervised learning algorithms like Isolation Forest, One-Class SVM, and Clustering algorithms (K-Means, DBSCAN) to identify abnormal transaction patterns that could suggest fraudulent activity. These models assist in detecting new fraud patterns which could escape supervised models.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Real-time Fraud Detection and Alert System: The framework combines stream data processing platforms to identify fraud in real-time. In case of a transaction suspected of being fraudulent, an alert is triggered and subsequent verification processes, like multi-factor authentication, are invoked to block unauthorized transactions.

Performance Evaluation and Optimization: The last module assesses model performance on measures like accuracy, precision, recall, F1-score, and ROC- AUC curves. Hyper parameter tuning methods like Grid Search and Random Search are employed to optimize the model and enhance fraud Detection.

IV. RESULTS

The machine learning approach proposed for detecting credit card fraud in banking is made up of several stages for the purpose of achieving accurate and efficient.

Data Acquisition & Preprocessing: The procedure begins with the collection of transactional data from bank statements or publicly available datasets. The data is preprocessed by handling missing values, removing duplicates, and encoding categorical features. Since fraud detection datasets are usually highly imbalanced, techniques like SMOTE and under sampling are used to balance the dataset so that the model does not lean towards non-fraudulent transactions.

Feature Selection & Engineering: Relevant transaction attributes, including transaction value, frequency, device usage, geographical location, and time-based patterns, are derived. Feature scaling and transformation methods, like Min-Max scaling or Standardization, are used to maintain consistency in data distribution.

Model Selection & Training: The data set is separated into training and testing sets, and machine learning models are trained. It is classified through supervised learning models such as Logistic Regression, Decision Trees, Random Forest, SVM. Anomaly detection is performed through the unsupervised learning algorithms such as Auto encoders and Isolation Forest.

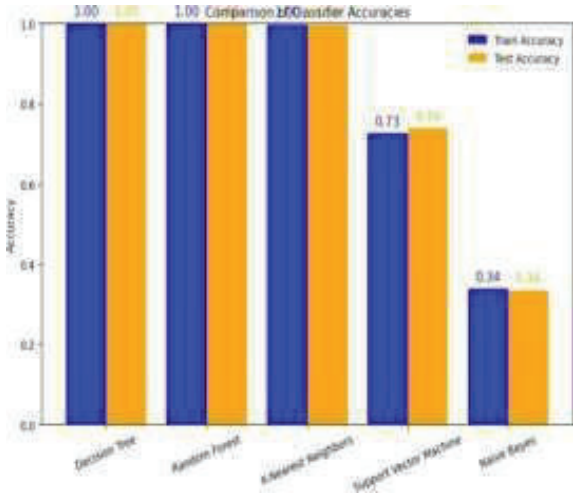

Fig. 5. Accuracy Graph

Anomaly Detection & Fraud Identification: The trained models categorize transactions into fraudulent or legitimate using learned patterns. The blended techniques that use both supervised and unsupervised techniques help accuracy and fraud discovery improvement. Other ensemble learning approaches like Boosting and Stacking help increase the performance of detection.

Real-Time Detection & Alert Generation: A real-time fraud detection system is implemented in a banking system, where streaming data is scanned in real time for dubious fraudulent transactions. In case a doubtful transaction is detected, the system initiates automatic alerts and security measures like OTP verification or locking of an account to avert fraud.

Model Evaluation & Performance Optimization: The value of fraudulent detection models is measured by Precision, Recall, F1-score, Confusion Matrix, and AUC-ROC curve. Methods of hyper parameter tuning such as Grid Search, Bayesian Optimization, and Neural Architecture Search (NAS) are applied for improving model performance.

TABLE I. ALGORITHM CLASSIFIER

SNO	Classifier	Accuracy	Precision	Recall	F1 Score:
1	decision Tree classifier :	1	1	1	1
2	Random forest classifier:	1	1	1	1
3	KNN classifier:	0.996	0.994	0.997	0.996
4	SVM classifier:	0.738	0.369	0.5	0.424
5	Naive bayes classifier:	0.335	0.641	0.549	0.31

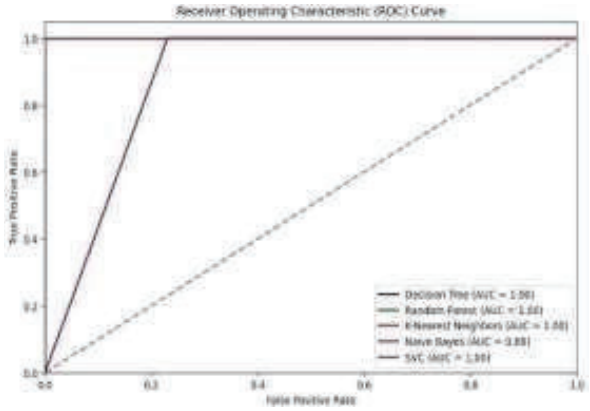

Fig. 6. ROU Curve

V. CONCLUSION

Machine learning-based credit card fraud detection in banking has been an effective tool for detecting and preventing fraudulent transactions in real time. Through the use of supervised, unsupervised, banks are able to identify sophisticated fraud patterns that rule-based systems are unable to detect. The application of techniques such as anomaly detection, ensemble learning, and feature engineering enhances the accuracy and reliability of fraud detection systems. In addition, real-time processing and automatic alerts allow financial institutions to take prompt action against suspicious activities, minimizing financial losses and boosting customer trust. As the pace of innovative fraudulent techniques accelerates, robust and elastic machine learning algorithms will become even more crucial in future years. Explainable AI, federated learning, blockchain protection, and quantum computing will steadily make fraud detection capabilities more user-friendly with assurances of privacy and transparency. As banks continue the evolution and consolidation of new technology, they are able to develop an even more secure and efficient payment process, making fraud risk obsolete and further increasing overall transaction security.

ACKNOWLEDGMENTS

The authors acknowledge the HOD, Department of Electronics and Communication Engineering, the management of Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering college, Chennai, for their continuous encouragement.

REFERENCES

[1] S. Patil, V. Nemade, and P. K. Soni, "Predictive Modelling For Credit Card Fraud Detection Using Data Analytics," *Procedia Comput. Sci.*, vol. 132, pp. 385–395, 2018, doi: 10.1016/j.procs.2018.05.199.

[2] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *J. Big Data*, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[3] I. D. Mienye and N. Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," *IEEE Access*, vol. 12, pp. 96893–96910, 2024, doi: 10.1109/ACCESS.2024.3426955.

[4] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[5] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," *Big Data Cogn. Comput.*, vol. 8, no. 1, p. 6, Jan. 2024, doi: 10.3390/bdcc8010006.

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

[7] Z. Zaffar, F. Sohrab, J. Kanniainen, and M. Gabbouj, "Credit Card Fraud Detection with Subspace Learning-based One-Class Classification," in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, Dec. 2023, pp. 407–412. doi: 10.1109/SSCI52147.2023.10372038.

[8] M. A. Talukder, R. Hossen, M. A. Uddin, M. N. Uddin, and U. K. Acharjee, "Securing transactions: a hybrid dependable ensemble machine learning model using IHT-LR and grid search," *Cybersecurity*, vol. 7, no. 1, p. 32, Nov. 2024, doi: 10.1186/s42400-024-00221-z.

[9] D. Rzayeva and S. Malekzadeh, "A Combination of Deep Neural Networks and K-Nearest Neighbors for Credit Card Fraud Detection," May 2022, [Online]. Available: http://arxiv.org/abs/2205.15300.

[10] D. H. M. de Souza and C. J. Bordin, "Ensemble and Mixed Learning Techniques for Credit Card Fraud Detection," Dec. 2021, [Online]. Available: http://arxiv.org/abs/2112.02627.

[11] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," *IEEE Access*, vol. 10, pp. 39700–39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[12] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, IEEE, Jan. 2020, pp. 680–683. doi: 10.1109/Confluence47617.2020.9057851.