

## UPI Fraud Detection System Using CNN Machine Learning Model

A.B. Koli, Prerana Patil, Bhakti Patil, Nikita Patil

Department of Computer Engineering,

P. S. G. V. P. Mandal's D. N. Patel College of Engineering, Shahada, Dist- Nandurbar, Maharashtra, India

### ABSTRACT:

*Unified Payments Interface (UPI) has revolutionized digital transactions in India offering convenience and real-time processing. However, this rapid adoption has also led to a surge in fraudulent activities, challenging the efficacy of traditional rule-based fraud detection methods. These conventional systems often struggle to adapt to evolving fraud patterns, necessitating more robust and adaptive solutions. In response to these challenges, researchers have explored machine learning techniques to detect fraudulent activities within UPI transactions. While existing strategies have shown promise, they are frequently validated on limited or synthetic datasets, which may not fully capture the complexities of real-world scenarios. To address these limitations, a comprehensive evaluation of prevalent machine learning classifiers—including Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Decision Trees, Random Forests, and Convolutional Neural Networks (CNNs)—was conducted. Building upon this analysis, we propose the Convolutional Neural Networks (CNNs) framework, designed to enhance fraud detection capabilities in UPI transactions.*

### I. INTRODUCTION

Humans have the innate ability to learn from experiences, adapting their knowledge over time. Similarly, machines can be designed to emulate this learning process through machine learning (ML), a branch of artificial intelligence focused on developing algorithms that enable computers to learn from data autonomously. With the proliferation of online transactions, digital payments have become common place, leading to thousands of transactions occurring every second on various platforms. However, this convenience also opens avenues for cybercriminals to exploit vulnerabilities, resulting in financial fraud that threatens both consumer interests and the stability of the digital economy. A significant challenge in fraud detection is the inherent class imbalance in transaction data—fraudulent transactions are rare compared to legitimate ones. Additionally,

the characteristics of fraudulent activities evolve over time, complicating detection efforts. To combat these issues, a range of ML and deep learning models, such as neural networks, logistic regression, random forests, and support vector machines, are utilized to identify and flag suspicious transactions. These models analyze historical user data to detect anomalies in payment behaviors, enabling real-time classification of transactions as either "fraudulent" or "genuine." ML-driven solutions are instrumental across various sectors, including banking, e-commerce, and fintech, enhancing their capabilities to manage extensive datasets and improve the accuracy of UPI fraud detection mechanisms.

### II. RELATED WORK

[A] The article from the *International Journal of Innovative Science and Research Technology* (Vol. 8, Issue 10, Oct 2023) discusses the use of machine learning, particularly the Random Forest (RF) algorithm, for detecting online payment fraud. RF is highlighted for its high accuracy, resistance to overfitting, and ability to handle large, imbalanced datasets. It also provides useful insights through feature importance analysis. Supported by past studies (Kumar et al., 2018; Chawla et al., 2019), RF is shown to outperform other methods like SVM and Logistic Regression. In this project, the RF model achieved 99% accuracy in Python and 93% in R, confirming its effectiveness. The authors suggest future enhancements could involve training on larger datasets for even better performance.[1]

[B] The study by Vedant Mayekar et al. (IRJET, May 2021) highlights the effectiveness of XGBoost, a Gradient Boosting algorithm, for online fraud transaction detection. XGBoost excels due to its ability to handle imbalanced data, avoid overfitting, and deliver high accuracy. Compared to models like Random Forest and Logistic Regression, it shows superior performance in terms of accuracy and AUC. The project achieved strong results with XGBoost, and the authors note its potential for further improvement with larger and dynamic datasets.[2]

[C] The paper by Vedant Mayekar et al. (IRJET, May 2021) explores the use of Convolutional Neural Networks (CNNs) for fraud detection, highlighting their ability to capture spatial and sequential patterns in transaction data. The proposed CNN model improves efficiency by using feature rearrangement, reducing computational cost and making it suitable for real-time detection. The study shows that CNNs can effectively learn complex fraud patterns, and suggests future enhancements by integrating LSTM models to further improve accuracy through better sequence analysis.[3]

[D] The study by Zhaohui Zhang et al. (2018) presents a novel CNN-based fraud detection model (RXT-J) for real-time online transactions. Designed to handle large and complex financial datasets, the model outperforms traditional machine learning and deep learning approaches in accuracy and speed, effectively identifying sophisticated fraud patterns. It addresses key limitations of earlier methods and shows strong potential for future enhancements, such as integrating fraud location and timing data. This work marks a significant advancement in securing financial transactions.[4]

[E] The paper by Mr. Ch Mahesh Babu et al. (INT-JECSE, 2023) demonstrates the effectiveness of XGBoost in detecting online payment fraud. Using historical transaction data and a structured methodology—covering preprocessing, feature selection, training, and evaluation—the XGBoost model achieved high accuracy and outperformed other algorithms. Its ensemble-based design captures complex, non-linear relationships and resists overfitting. Continuous monitoring and updates enhance adaptability to new fraud patterns, making XGBoost a powerful tool for securing online payment systems.[5]

[F] The paper by Yash Patil et al. (September 2024) presents an SVM-based model for detecting UPI fraud. The model achieves high precision in identifying fraudulent transactions but struggles with data imbalance, scalability, and interpretability. While Random Forest may outperform SVM in some areas, SVM remains promising with proper optimization. The study emphasizes the need for continuous updates, feature engineering, and possibly hybrid models to adapt to evolving fraud tactics, ensuring both security and a good user experience in financial systems.[6]

### III. PROBLEM STATEMENT

- UPI, a mobile-based payment system, has become the backbone of digital transactions in India. While offering convenience and accessibility, the surge in UPI transactions has exposed vulnerabilities to fraudsters. The problem lies in identifying and preventing fraudulent UPI transactions before they impact users and the financial system.
- The rapid adoption of UPI has led to a parallel increase in fraudulent activities, posing a significant threat to user trust and financial stability.
- Fraudulent transactions often involve sophisticated techniques like account takeovers, phishing scams, and unauthorized payments, making detection challenging.
- The need for real-time fraud detection is crucial to prevent financial losses and ensure the integrity of UPI transactions.

### IV. PROPOSED SYSTEM

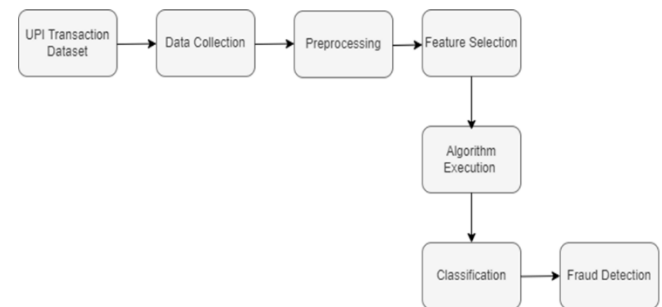


Fig. System architecture

The architecture diagram represents the workflow of a UPI (Unified Payments Interface) fraud detection system. It begins with the **UPI Transaction Dataset**, which contains historical transaction records, including attributes such as transaction amount, sender ID, timestamps, etc. and whether a transaction was fraudulent. The process then moves to **Data Collection**, where this raw data is gathered from relevant sources such as banking servers, UPI gateways, or APIs. Once collected, the data undergoes **Preprocessing**, which involves cleaning the dataset by handling missing values, removing duplicates, encoding categorical variables, and normalizing numerical features. This step ensures the data is in a suitable format for further analysis.

Following preprocessing, the system performs **Feature Selection** to identify the most relevant attributes that influence fraud detection. This helps in reducing dimensionality and improving model performance. The selected features are then used in the **Algorithm Execution** phase, where a machine learning or deep learning model is applied. Convolutional Neural Network (CNN) is used here, the transaction data transformed into a structured format that CNNs can process—such as a time-series matrix or reshaped feature grid. CNNs are particularly useful in capturing spatial or sequential patterns in transaction behavior, which can be indicative of fraudulent activity.

The output of the model is passed to the **Classification** step, where each transaction is classified as either fraudulent or legitimate. Finally, the **Fraud Detection** module uses this classification to flag suspicious transactions, enabling the system to take further action such as alerting the user or blocking the transaction. This end-to-end pipeline ensures that fraudulent activities can be detected effectively and in real-time, leveraging CNN's ability to recognize complex patterns in structured financial data.

## V. METHODOLOGY AND ALGORITHM

### A. Data Preprocessing

We began by collecting publicly available datasets containing UPI transaction records, each labeled as either fraudulent or genuine. These records included various attributes such as transaction amount, timestamp, location, user behavior, and device information. To ensure data quality, we performed preprocessing steps like handling missing values, removing duplicates, and normalizing numerical features. This process ensured that the data was clean and suitable for training machine learning models.

### B. Feature Engineering

From the raw transaction data, we extracted meaningful features that could help in distinguishing between fraudulent and genuine transactions. These features included transaction frequency, average transaction amount, time since last transaction, and patterns in user behavior. By transforming and encoding these features appropriately, we aimed to provide the model with informative inputs that capture the underlying patterns associated with fraudulent activities.

### C. Model Architecture (Convolutional Neural Network - CNN)

For the fraud detection task, we employed a Convolutional Neural Network (CNN) architecture tailored to handle structured transaction data. The CNN model was designed to capture complex patterns and interactions among features that might indicate fraudulent behavior. The architecture consisted of multiple convolutional layers with ReLU activation functions, followed by pooling layers to reduce dimensionality and capture essential features. After the convolutional and pooling layers, the output was flattened and passed through fully connected dense layers, culminating in a sigmoid activation function for binary classification (fraudulent or legitimate). This architecture allowed the model to learn hierarchical representations of the data, enhancing its ability to detect subtle anomalies.

**Input Layer:** Receives preprocessed transaction data, such as user behavior, transaction time, amount, and device details.

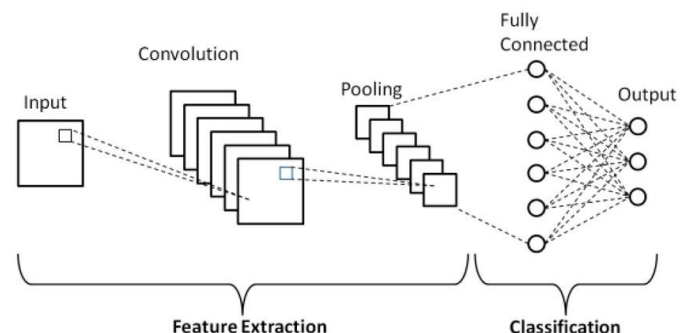
**Convolutional Layers:** Apply filters to detect patterns and anomalies indicative of fraudulent activities.

**Activation Function:** ReLU (Rectified Linear Unit) introduces non-linearity, enabling the network to learn complex patterns.

**Pooling Layers:** Reduce the spatial dimensions of the data, retaining essential features and reducing computational load.

**Fully Connected Layers:** Integrate features from previous layers to make final predictions.

**Output Layer:** Produces a binary classification: 0 for legitimate transactions and 1 for fraudulent ones.



#### D. Training & Evaluation Strategy

The dataset was split into training and testing sets to evaluate the model's performance on unseen data. We trained the CNN model for 200 epochs using the Adam optimizer, which adapts the learning rate during training for efficient convergence. The binary cross-entropy loss function was used to measure the discrepancy between predicted and actual labels. To assess the model's effectiveness, we employed evaluation metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics provided a comprehensive view of the model's performance, particularly its ability to correctly identify fraudulent transactions while minimizing false positives.

**Epochs:** The model processes the entire dataset 200 times. Ensures the model learns from the data thoroughly, adjusting weights to minimize error. Track metrics like accuracy and loss to evaluate performance and prevent overfitting.

#### E. Handling Class Imbalance (SMOTE)

In fraud detection datasets, genuine transactions typically outnumber fraudulent ones, leading to class imbalance issues that can bias the model. To address this, we applied the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic examples of the minority class (fraudulent transactions) by interpolating between existing minority instances. This technique helped balance the dataset, enabling the model to learn from a more representative distribution of classes and improving its ability to detect fraud. Through this comprehensive methodology, we developed a robust CNN-based model capable of detecting fraudulent UPI transactions with high accuracy, while addressing challenges such as data quality, feature representation, model architecture, training strategies, and class imbalance.

### VI. DATASET

The efficacy of any machine learning model heavily relies on the quality and characteristics of the dataset used for training and evaluation. For this UPI (Unified Payments Interface) fraud detection project, a dedicated dataset, referred to as the UPI Fraud Detection Dataset, was utilized. This section provides a detailed overview of its structure, contents, and key properties.

**1. Dataset Origin and Size:** The dataset was provided for the purpose of this study. It is a tabular dataset comprising a total of 2667 records and 11 distinct features. Each record represents

a unique UPI transaction, capturing various attributes associated with it.

1. The dataset is structured with several columns, each representing a specific characteristic of a UPI transaction. Based on the provided dataset sample, the key features include:
2. **trans\_hour:** Represents the hour of the day when the transaction occurred, providing a temporal aspect at a granular level.
3. **trans\_day:** Indicates the day of the month the transaction took place.
4. **trans\_month:** Represents the month of the year for the transaction.
5. **trans\_year:** Denotes the year in which the transaction was processed.
6. **category:** A categorical feature that likely describes the type or nature of the transaction, or a specific category associated with it.
7. **upi\_number:** A unique identifier or reference number associated with the UPI transaction.
8. **age:** The age of the user involved in the transaction.
9. **trans\_amount:** The monetary value of the transaction.
10. **state:** The state associated with the transaction, possibly the user's state or the transaction's origin/destination state.
11. **zip:** The zip code associated with the transaction, providing more specific geographical context.

**3. Data Types:** The features within the dataset consist of a mix of data types, including:

- Numerical: trans\_hour, trans\_day, trans\_month, trans\_year, age, trans\_amount, state, zip.
- Categorical/Object: category, upi\_number.
- Boolean/Integer: fraud\_risk.

**4. Target Variable Distribution:** The "Is fraud" column serves as the target variable for classification. It is common in fraud detection datasets for there to be a significant class imbalance, where the number of legitimate transactions (0) vastly outnumbers fraudulent transactions (1). This imbalance presents a challenge for model training and necessitates specific handling techniques (e.g., oversampling, undersampling, or synthetic data generation) during the preprocessing phase to prevent models from being biased towards the majority class.



## VII. IMPLEMENTATION

### • Training

Prior to model training, the raw dataset typically undergoes several preprocessing steps to ensure data quality and suitability for machine learning algorithms. While training a model it has been feed with a large dataset of labeled UPI transactions. Each sample is processed through the network, and the model's predictions are compared to the actual labels using a loss function (such as binary cross-entropy). An optimization adjusts the weights of the network to minimize this loss over time.

During training:

- The model learns to detect intricate patterns within the transaction data that are often associated with fraud.
- The weights of the filters are adjusted in such a way that they respond strongly to suspicious or unusual transaction characteristics.
- Techniques such as dropout and early stopping are employed to prevent overfitting and improve generalization to unseen data.

After training the model with various algorithms like CNN, SVM, Random Forest, Naïve Bias, Decision Tree, K-Nearest Neighbors, Logistic Regression the model is tested.

### • System Testing

The primary objective of testing in the UPI Fraud Detection Project is to ensure the accuracy, reliability, and robustness of the fraud detection mechanisms implemented. Testing focuses on validating all functionalities against specified requirements and simulating real-world fraudulent and legitimate transaction scenarios to evaluate system performance. By comparing new data with an existing, trained dataset, CNN are a powerful tool for UPI fraud detection. In order to arrive at a final binary prediction of 0 or 1.

	Algorithm Name	Accuracy Score
1	CNN	96.23
2	Random Forest	95.50
3	Decision Tree	94.75
4	K-Nearest Neighbors	83.00
5	Support Vector Machine	81.50
6	Naïve Bias	81.25
7	Logistic Regression	80.25

Table: Algorithm and Accuracy

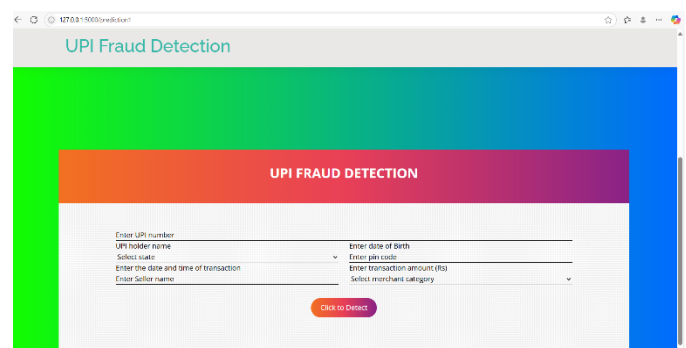
Among the evaluated models, the Convolutional Neural Network (CNN) demonstrated the highest performance with an accuracy 96.23%, making it the most effective algorithm for detecting fraudulent transactions.

Random Forest and Decision Tree classifiers also performed strongly, each achieving accuracy scores above 90%. The K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, and Logistic Regression models followed with slightly lower accuracies, all hovering around the 80% mark.

These results highlight the superior capability of deep learning techniques like CNNs in identifying complex fraud patterns in UPI transaction data, while traditional machine learning models still offer reasonably good performance.

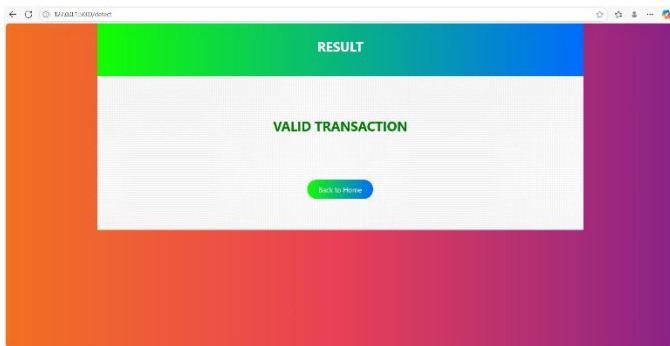
## VIII. DESIGN

### 1. Data Entry Page



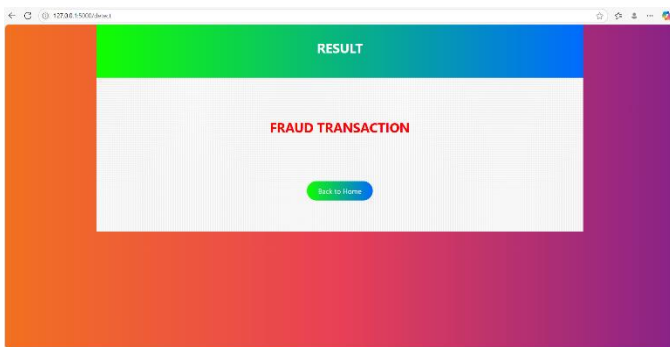
The screenshot shows a web application titled "UPI Fraud Detection". The page has a green header bar with the title. Below the header, there is a form with a pink background. The form contains several input fields and a "Click to Detect" button. The input fields are labeled: "Enter UPI number", "UPI holder name", "Select state", "Enter the date and time of transaction", "Enter Seller name", "Enter date of birth", "Enter pin code", "Enter transaction amount (Rs)", and "Select merchant category".

## 2. Result Page (1)



Our algorithm predicts and displays the outcome based on the training phase's lessons when the user gives the necessary transaction inputs. It presents a result based on the model's prediction that **VALID TRANSACTION**.

## 3. Result Page (2)



Our algorithm predicts and displays the outcome based on the training phase's lessons when the user gives the necessary transaction inputs. It presents a result based on the model's prediction that **FRAUD TRANSACTION**.

## CONCLUSION

UPI fraud is becoming an increasingly serious issue in India's digital payment landscape. As more people use UPI for quick and easy transactions, fraudsters are finding new ways to exploit the system. To combat this, advanced detection methods are needed. Machine learning offers promising solutions, but challenges like data privacy and real-time processing still persist. Convolutional Neural Networks (CNNs) have emerged as effective tools in detecting fraudulent UPI transactions. They excel at identifying complex patterns in transaction data, leading to higher accuracy compared to traditional models. For instance, a CNN-based model achieved an impressive accuracy of 96.23% by analyzing user behavior and transaction details in real-time, thus preserving user privacy. CNNs

can lead to scalable and secure UPI fraud detection systems that protect both users and financial institutions. Future research should focus on balancing accuracy, privacy, scalability, and real-time processing capabilities to develop effective solutions in this domain

## REFERENCES

[1] Mallipudi Devi Siva Sai, Palaparthi Prudhvi, Gollapudi M Naga Venkata Sai Gopi, Indla Ganeswara Naga Sai Ram, Mandadi Ram Sandeep, "Online Payment Fraud Detection"-International Journal of Innovative Science and Research Technology ISSN No:-2456-2165 Volume 8, Issue 10, October – 2023

<https://ijisrt.com/assets/upload/files/IJISRT23OCT1251.pdf>

[2] Vedant Mayekar, Siddharth Mattha, Sohan Choudhary, Prof Amruta Sankhe "ONLINE FRAUD TRANSACTION DETECTION USING MACHINE LEARNING"-International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 08 Issue: 05 May 2021.

<https://www.irjet.net/archives/V8/i5/IRJET-V8I5133.pdf>

[3] Zhaohui zhang, xinxin Zhou, Xiaobo zhang, lizhi wang, pengwei wang "A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection"-Security and Communication Networks, Volume 2018, Issue 1.

<https://onlinelibrary.wiley.com/doi/10.1155/2018/5680264>

[4] Mr. CH MAHESH BABU, BOJJA SWEEHONEY, PADAM PRATHYUSHA, BOMMEPALLI DEVENDRA REDDY, MAROJU SATHVIKA "ONLINE PAYMENT FRAUD DETECTION"- International Journal of Early Childhood Special Education (INTJECSE) DOI:10.48047/INTJECSE/V15I4.86 ISSN: 1308-5581 Vol 15, Issue 04 2023.

[https://www.intjecse.net/media/article\\_pdfs/1ONLINE\\_PAYMENT\\_FRAUD\\_DETECTION.pdf](https://www.intjecse.net/media/article_pdfs/1ONLINE_PAYMENT_FRAUD_DETECTION.pdf)

[5] Yash Patil, Amar Shinde, Yash Parthe, Sameer Sayyad "UPI FRAUD DETECTION USING MACHINE LEARNING"-International Research Journal of Modernization in Engineering Technology and Science, e-ISSN: 2582-5208, Volume:06, Issue:09, September-2024

[https://www.irjmets.com/uploadedfiles/paper//issue\\_9\\_september\\_2024/61840/final/fin\\_irjmets1727797829.pdf](https://www.irjmets.com/uploadedfiles/paper//issue_9_september_2024/61840/final/fin_irjmets1727797829.pdf)

[6] Shabreshwari R M, Shafiya Mehrooz, Sidra Fatima, Tanmai R B, Prof. Ganesh Manasali "UPI Fraud Detection Using Machine Learning"-International Journal of Advances in Engineering and Management (IJAEM), ISSN: 2395-5252 Volume 6, Issue 06 June 2024.  
[https://ijaem.net/issue\\_dcp/Upi%20Fraud%20Detection%20Using%20Machine%20Learning.pdf](https://ijaem.net/issue_dcp/Upi%20Fraud%20Detection%20Using%20Machine%20Learning.pdf)

[7] Viswanatha V, Ramachandra A.C, Deeksha V, Ranjitha R "Online Fraud Detection Using Machine Learning Approach"-International Journal of Engineering and Management Research, Volume-13, Issue-4(August2023).  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4533856](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4533856)

[8] Prof. Varsha Pande, Mainak Saha, Shaiban Mulla, Sumedh Gamre, Fraud Detection In UPI Transaction Using AI Volume 12, Issue 4 April 2024 | ISSN: 2320-2882.  
<https://www.ijcrt.org/papers/IJCRT24A4311.pdf>

[9] Jain R., Gour B., Dubey S., A hybrid approach for credit card fraud detection using rough set and decision tree technique, International Journal of Computer Applications 139(10) (2016).  
[https://www.researchgate.net/publication/301335558\\_A\\_Hybrid\\_Approach\\_for\\_Credit\\_Card\\_Fraud\\_Detection\\_using\\_Rough\\_Set\\_and\\_Decision\\_Tree\\_Technique](https://www.researchgate.net/publication/301335558_A_Hybrid_Approach_for_Credit_Card_Fraud_Detection_using_Rough_Set_and_Decision_Tree_Technique)

[10] Dermala N., Agrawal A.N., Credit card fraud detection using SVM and Reduction of false alarms, International Journal of Innovations in Engineering and Technology (IJJET) 7(2) (2016).  
<https://www.sciencedirect.com/science/article/pii/S187705092030065X>

[11] Phua C., Lee V., Smith, Gayler K.R., A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119 (2010).  
<https://arxiv.org/pdf/1009.6119>

[12] Bahnsen A.C., Stojanovic A., Aouada D., Ottersten B., Cost sensitive credit card fraud detection using Bayes minimum risk. 12th International Conference on Machine Learning and Applications (ICMLA) (2013), 333-338.  
<https://ieeexplore.ieee.org/document/6784638>

[13] Carneiro E.M., Dias L.A.V., Da Cunha A.M., Mialaret L.F.S., Cluster analysis and artificial neural networks: A case study in credit card fraud detection, 12th International Conference on Information Technology-New Generations (2015),22\_126.  
[https://www.academia.edu/47749764/Cluster\\_Analysis\\_and\\_Artificial\\_Neural\\_Networks\\_A\\_Case\\_Study\\_in\\_Credit\\_Card\\_Fraud\\_Detection](https://www.academia.edu/47749764/Cluster_Analysis_and_Artificial_Neural_Networks_A_Case_Study_in_Credit_Card_Fraud_Detection)

[14] Hafiz K.T., Aghili S., Zavarsky P., The use of predictive analytics technology to detect credit card fraud in Canada, 11th Iberian Conference on Information Systems and Technologies (CISTI) (2016), 1-6.  
<https://ijrsrset.com/paper/9487.pdf>

[15] Sonapat H.C.E., Bansal M., Survey Paper on Credit Card Fraud Detection, International Journal of Advanced Research in Computer Engineering & Technology 3(3) (2014).VarrePerantalu K., BhargavKiran, Credit card Fraud Detection using Predictive Modeling (2014).  
<https://ijrsrset.com/paper/9487.pdf>

[16] Stolfo S., Fan D.W., Lee W., Prodromidis A., Chan P., Credit card fraud detection using meta-learning: Issues and initial results, AAAI-97 Workshop on Fraud Detection and RiskManagement(1997).  
[https://www.researchgate.net/publication/2282588\\_Credit\\_Card\\_Fraud\\_Detection\\_Using\\_Meta-Learning\\_Issues\\_and\\_Initial\\_Results](https://www.researchgate.net/publication/2282588_Credit_Card_Fraud_Detection_Using_Meta-Learning_Issues_and_Initial_Results)

[17] Maes S., Tuyls K., Vanschoenwinkel B., Manderick, B., Credit card fraud detection using Bayesian and neural networks, International Journal of Pure and Applied Mathematics Special Issue 836Proceedings of the 1st international naiso congress on neuro fuzzy technologies (2002),261-270.  
[https://www.researchgate.net/publication/2524707\\_Credit\\_Card\\_Fraud\\_Detection\\_Using\\_Bayesian\\_and\\_Neural\\_Networks](https://www.researchgate.net/publication/2524707_Credit_Card_Fraud_Detection_Using_Bayesian_and_Neural_Networks)

[18] Chan P.K., Stolfo S.J., Toward Scalable Learning with Non- Uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection, In KDD (1998), 164-168  
[https://www.researchgate.net/publication/2256438\\_Toward\\_Scalable\\_Learning\\_with\\_Non-uniform\\_Class\\_and\\_Cost\\_Distributions\\_A\\_Case\\_Study\\_in\\_Credit\\_Card\\_Fraud\\_Detection](https://www.researchgate.net/publication/2256438_Toward_Scalable_Learning_with_Non-uniform_Class_and_Cost_Distributions_A_Case_Study_in_Credit_Card_Fraud_Detection)