A magnifying glass over people

AI-generated content may be incorrect.

Fraud Detection Analysis Using   
Machine Learning and  
 Deep Learning Techniques

**Start** : December 2024 | **End** : February 2025

developed by [Nardeen Gerges](http://www.linkedin.com/in/narden-gerges-716452220) & [Yousef Emad](https://www.linkedin.com/in/youssifemad)

**Abstract** :

This project develops a fraud detection system using machine learning and deep learning to classify transactions as fraudulent or non-fraudulent. Addressing imbalanced data, we experiment with different models and Neural Networks, supported by extensive visualizations for performance analysis. The final model is deployed into production, complemented by interactive dashboards for real-time fraud monitoring and insights .

Contents

[**Introduction :** 3](#_Toc192203572)

[**Problem statement :** 3](#_Toc192203573)

[**Goals of the Project :** 3](#_Toc192203574)

[**Project overview :** 4](#_Toc192203575)

[**Methodology :** 4](#_Toc192203576)

[Data Collection & Preprocessing : 4](#_Toc192203577)

[EDA and Visualizations : 5](#_Toc192203578)

[Machine Learning Models : 6](#_Toc192203579)

[Deep Learning Models : 9](#_Toc192203580)

[visualize to compare Models : 12](#_Toc192203581)

[**Conclusion :** 12](#_Toc192203582)

# **Introduction :**

## **Problem statement :**

Fraudulent activities, such as unauthorized transactions, identity theft, and financial scams, pose significant challenges to businesses worldwide. Detecting fraud in real-time is critical, as it directly impacts revenue, customer trust, and operational efficiency. Traditional rule-based systems often fail to adapt to evolving fraud patterns, leading to high false positives and missed detections. Additionally, fraud detection is complicated by imbalanced datasets, where fraudulent cases are rare compared to legitimate transactions. This project aims to address these challenges by leveraging machine learning and deep learning techniques to build a robust, scalable, and accurate fraud detection system that minimizes financial losses and enhances business security.

**Impact on Business :**

1. **Financial Losses**: Fraud results in billions of dollars in losses annually, affecting profitability.
2. **Customer Trust**: Frequent fraud incidents erode customer confidence and damage brand reputation.
3. **Operational Costs**: Manual fraud detection processes are time-consuming and resource-intensive.
4. **Regulatory Compliance**: Failure to detect fraud can lead to legal penalties and non-compliance with industry regulations.

By implementing an advanced fraud detection system, businesses can mitigate these risks, improve decision-making, and ensure a secure environment for their customers.

## **Goals of the Project :**

1. **Accurate Fraud Detection**: Build a model that achieves high precision and recall to minimize false positives and missed fraud cases.
2. **Handling Imbalanced Data**: Implement techniques like SMOTE, undersampling, or ensemble methods to address dataset imbalance.
3. **Real-Time Detection**: Develop a system capable of detecting fraud in real-time to prevent losses.
4. **Interpretability**: Provide clear visualizations and insights into fraud patterns for better decision-making.
5. **Scalability**: Ensure the solution is scalable and adaptable to different business environments.
6. **User-Friendly Dashboards**: Create interactive dashboards for stakeholders to monitor and analyze fraud trends.

By achieving these goals, the project aims to deliver a comprehensive and practical solution for fraud detection, empowering businesses to combat fraud effectively.

# **Project overview :**

Fraud detection is a critical aspect of modern financial and business operations, as fraudulent activities continue to evolve in complexity. This project aims to develop an advanced **fraud detection system** using **machine learning and deep learning techniques** to enhance security, minimize financial losses, and improve operational efficiency.

The system will be designed to handle **highly imbalanced datasets**, where fraudulent transactions are significantly fewer than legitimate ones. To address this, techniques such as **Synthetic Minority Over-sampling Technique (SMOTE), undersampling, and ensemble methods** will be employed. The model will be optimized for **high precision and recall**, ensuring both fraud detection accuracy and reduced false positive rates.

A key feature of this project is **real-time fraud detection**, enabling businesses to take immediate action against suspicious activities. Additionally, the system will provide **interpretability and insights** through clear visualizations, helping stakeholders understand fraud patterns and improve risk management strategies.

To make the solution **scalable and user-friendly**, a web-based dashboard will be developed, allowing businesses to monitor transactions and analyze fraud trends interactively. The final deployment will ensure seamless integration into different business environments, offering a robust and efficient approach to fraud prevention.

# **Methodology :**

### **Data Collection & Preprocessing :**

I utilized a simulated credit card transaction dataset obtained from Kaggle. This dataset encompasses transactions conducted between January 1, 2019, and December 31, 2020, including both legitimate and fraudulent activities. It comprises transaction records from 1,000 customers interacting with a network of 800 merchants, offering a comprehensive basis for analyzing credit card fraud patterns. [[Dataset](https://www.kaggle.com/datasets/kartik2112/fraud-detection)]

### **EDA and Visualizations :**

**A graph of a graph

AI-generated content may be incorrect.A diagram of a violin plot

AI-generated content may be incorrect.class Imbalance:**

### 

**A diagram of a diagram

AI-generated content may be incorrect.A blue line graph with white text

AI-generated content may be incorrect.Geospatial Analysis: Temporal Analysis :**

**Features:**

A graph of a person with a number of bars

AI-generated content may be incorrect.A blue circle with black text

AI-generated content may be incorrect.A graph with red and blue squares

AI-generated content may be incorrect.

A graph of a city population

AI-generated content may be incorrect.

A graph of a distribution of population

AI-generated content may be incorrect.A graph of a distribution of a number

AI-generated content may be incorrect.

### **Machine Learning Models :**

1. **Isolation Forest :**

Isolation Forest (IF) is an unsupervised anomaly detection algorithm designed to identify fraudulent transactions by isolating anomalies rather than profiling normal instances. It constructs multiple isolation trees by randomly selecting features and splitting values between their observed ranges. Since anomalies have unique or extreme values, they require fewer splits to be isolated, resulting in a lower path length in the tree structure. This makes Isolation Forest an efficient method for detecting fraud without needing labeled data, making it particularly useful in highly imbalanced datasets where fraudulent cases are rare.

A diagram of a tree

AI-generated content may be incorrect.

A screenshot of a computer

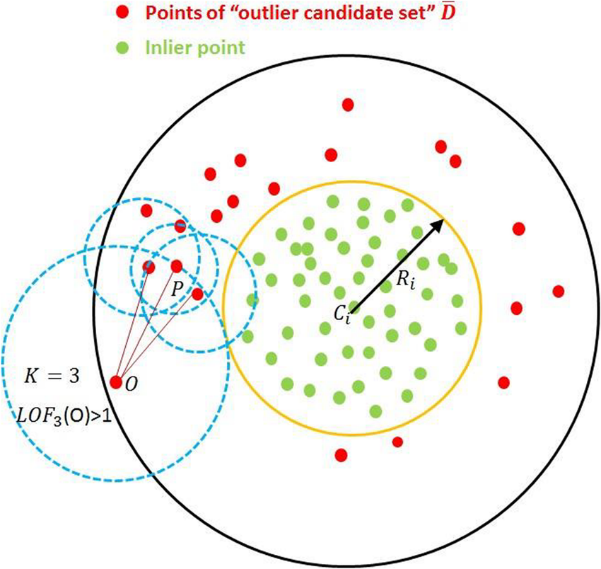
AI-generated content may be incorrect.**Evaluation :**

**A graph with blue squares and white text

AI-generated content may be incorrect.**

**A graph of a curve

AI-generated content may be incorrect.**



1. **Local Outlier Factor (LOF) :**

Local Outlier Factor (LOF) is an unsupervised anomaly  
 detection algorithm that detects fraud by measuring the   
local density deviation of a data point relative to its neighbors.  
 A lower density suggests a higher likelihood of fraud, as  
 anomalies tend to have significantly different densities  
 compared to surrounding data points. LOF is particularly  
 useful for identifying subtle fraud patterns in imbalanced  
 datasets where fraudulent transactions blend with  
 normal ones.

**Evaluation :**

A screenshot of a test results

AI-generated content may be incorrect.

A blue square with white text

AI-generated content may be incorrect.

A line graph with a blue line

AI-generated content may be incorrect.

1. A diagram of a molecule

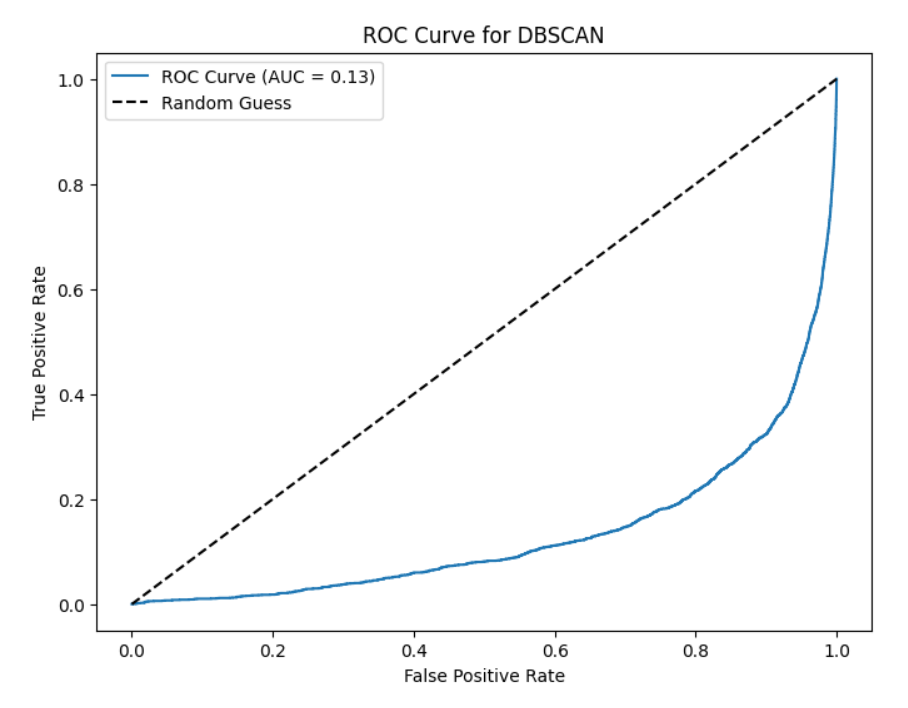
   AI-generated content may be incorrect.**DBSCAN :**Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an unsupervised clustering algorithm that identifies fraud by detecting dense regions of normal transactions while treating anomalies as noise. It groups closely packed points together and marks points in low-density areas as potential fraud. DBSCAN is effective in handling complex fraud patterns and does not require prior knowledge of the number of fraud cases, making it suitable for highly imbalanced datasets.

A screenshot of a computer

AI-generated content may be incorrect.**Evaluation :**

**A blue squares with white text

AI-generated content may be incorrect.**

****

### Deep Learning Models :

1. **Autoencoders**[**¶**](https://www.kaggleusercontent.com/kf/226192431/eyJhbGciOiJkaXIiLCJlbmMiOiJBMTI4Q0JDLUhTMjU2In0..BdfpSp-xck0jeH1eiF-vYQ.pS7VI1AQ7Tt1Fmxtub2nae8rRK7wuCqu3tjVeT80Lo139ljTo8tGpgtln60TD5xvHv5BAG1jhC53vA3Y_Wb65xwcEv9JrSLOVrRFBB9flQ6AogL605dpSIe99VLkPB_2RiDd1LtjDyB28cFDqCk62FsfhucwWR1Aw9K3ptHcsn4gouZ4e_E1SGNmVElXAFnMpPg7i8ZfvFbjPB6IjjYWmxybXz_SU5D2FpFgjy0DdXPkwv1sgzk6LcWcAQR3HqcfYJqJ1k7KMHQmrufUCYsBJceNCCAn8NVcsyidOqffPOdUutW7DhSzlAcWy7rcIHXZ9UeydFhgulMz3_IlnVTJvkCu3a1WTJcIYrR7r-pLwsG-9_j8sQM1WXL87UJ1GBU8E918JU10_hBTf52Jj-4B4poStdDF1iXZiHh7ZoHRMwi1kIxpl9kX950lZV6x0g-0FV3Pg-HXQuRTdgKDYAY3m1ogeflu52b11uqjX3wpSGa3VFNRvCDdqEfIm77kzloo3BHUNr-8Bx_MQCCT7jAQFBh2rkZnaDmvkgpu4b3WtI76_uluBqfg2o9NevfH-P8EFLiIBDK5LamtGPKGcYPGQ22raurkLNzXAC7mffBbxWgUbOPQ7uDKdpsFYNEl0djD-NKoPh29-NmpA3lBd967gA.2RooDbbGaYNg_sw39lS16Q/__results__.html#Autoencoders)

An autoencoder is a type of neural network used for unsupervised learning, and it's highly effective for anomaly detection. It works by learning to compress (encode) and then reconstruct (decode) data. If a sample cannot be reconstructed well (i.e., has a high reconstruction error), it is likely an anomaly.

**Evaluation :**

A blue square with white text

AI-generated content may be incorrect.

**A screenshot of a computer

AI-generated content may be incorrect.**

1. **One-Class Neural Networks (OC-NN) :**

One-Class Neural Networks (OC-NN) is the best model for anomaly detection, as it excels at learning the underlying patterns of normal data and identifying deviations with high precision. Unlike traditional approaches, OC-NN leverages deep learning to capture complex feature representations, making it highly effective in detecting outliers in high-dimensional datasets. Its adaptability and robustness make it superior for applications such as fraud detection, cybersecurity, and medical diagnostics, where identifying rare anomalies is crucial.

A blue square with white text

AI-generated content may be incorrect.**Evaluation :**

A screenshot of a computer

AI-generated content may be incorrect.

1. **Deep Support Vector Data Description (Deep SVDD) :**

Deep Support Vector Data Description (Deep SVDD) is a powerful anomaly detection model that learns a compact representation of normal data by mapping it into a hypersphere. It minimizes the distance between normal points and a fixed center, identifying anomalies as those that fall outside this boundary. Its ability to handle high-dimensional data makes it highly effective for applications like fraud detection, medical diagnostics, and industrial fault detection

**Evaluation :**

**A blue square with white text

AI-generated content may be incorrect.**

A screenshot of a computer

AI-generated content may be incorrect.

### A graph of different colored bars AI-generated content may be incorrect.visualize to compare Models :

A graph of a line graph

AI-generated content may be incorrect.A graph of a graph

AI-generated content may be incorrect.

# 

# **Conclusion**

This fraud detection project applied various machine learning and deep learning models to identify fraudulent transactions. The primary goal was to minimize false negatives, reducing financial risk. After extensive evaluation, **One-Class Neural Networks (OC-NN)** proved to be the most effective model. This highlights the importance of AI-driven fraud detection in securing business transactions and reducing losses. Implementing such systems enhances financial security and builds customer trust. Future improvements may focus on real-time detection and adaptive learning for even better fraud prevention.