ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

Anomaly Finance AI: Unsupervised Learning for Transaction Irregularity Detection

[1]Thottempudi Kavya Sree, [2]T Shashirekha,

[1]M.Tech Scholar,Dept. of CSE , Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth , Maisammaguda,Hyderabad, Telangana 500100, India, kavyasreethottempudi@gmail.com

[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, shashirekha.tejavath@gmail.com

Abstract:

Since financial institutions generate massive volumes of transaction data daily, the identification of questionable or fraudulent activities is an arduous but essential task. The ability to identify anomalies in financial transactions without the use of labels is a key component of intelligent systems for this purpose. Using unsupervised learning techniques to find anomalies in financial transactions is the main objective of this study. These methods can automatically discover hidden structures and patterns in large datasets. In order to preprocess data, extract features, and identify abnormalities, the proposed system makes use of cutting-edge unsupervised machine learning models. By analyzing transaction factors such as quantity, time interval, location, frequency, and patterns of customer behavior, we may distinguish between valid and fraudulent financial transactions. Unsupervised algorithms study the typical pattern of legitimate transactions and then search for outliers to identify potential fraud. The algorithm can adjust to novel forms of fraud and reduces the need for manually tagged datasets. Anomaly detection model efficacy is evaluated using precision, recall, and detection accuracy metrics. There were fewer false positives and a higher success rate when using unsupervised learning to identify novel types of fraud, according to the experiment's results. The proposed method improves financial safety by making it easier to keep tabs on money coming in and going out, which in turn helps with risk management and spotting fraud early on. Modern financial systems face evolving cyber-financial threats; our intelligent detection architecture provides a scalable and effective defense.

Keywords: Fraud Detection · Anomaly Detection · Unsupervised Learning · Isolation Forest · Autoencoder · Gaussian Mixture Model · OCSVM · Local Outlier Factor

Introduction:

In this age of digital transactions, financial fraud may affect anybody, anywhere in the world. Traditional rule-based fraud detection systems are unable to keep up with the ever-evolving fraudulent practices, leading to an increase in financial loss. A powerful tool for identifying anomalies and suspicious activity, machine learning (ML) analyses large volumes of transaction data. The potential application of advanced ML models like as MODEL, XgBoost, and Multilayer Perceptron (MLP) for fraud detection is explored in this study. These algorithms can tell the difference between legitimate and fraudulent transactions depending on criteria including amount, frequency, user behavior, and geographic location.

Accuracy is enhanced by XgBoost's boosting algorithms, MLP captures the intricate links between transactions, and MODEL rapidly identifies sequential fraud patterns. This approach uses AI to improve real-time fraud prevention while decreasing the amount of false positives. Evaluating the models with precision, accuracy, and recall guarantees effective fraud detection. This research helps with two things: protecting financial transactions and lowering the probability of fraud.

Financial fraud is a major and persistent issue in the online economy. Since electronic transactions and online payment infrastructures are growing at an exponential rate, financial institutions are becoming

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

more susceptible to fraud. With varying payment methods and increasing transaction volumes, real-time fraud detection is more vital than ever. Detection has become an essential issue for financial institutions and payment service providers. Conventional fraud detection systems can't handle the dynamic and more complicated nature of fraud since they depend on manually provided static criteria or rules. Supervised machine learning models have surpassed rule-based systems by learning decision boundaries from historical labeled data. Even so, they have a few problems. The most crucial is their need on massive, properly annotated databases, even if fraudulent transactions make only a small fraction of total activity in real-world financial contexts. The labeling process is time-consuming, costly, and error-prone; this is especially true when dealing with fraud, which is always changing. As a result, supervised techniques often fail to detect novel forms of fraud that do not conform to the typical pattern of attacking. To get over these limitations, this study looks at unsupervised machine learning as a potential data-efficient and flexible alternative to traditional fraud detection methods. Unsupervised models are free to learn from their own experiences rather than from pre-labeled input.

Discover the hidden information and outliers in the dataset. When the proportion of fraudulent transactions is very low in highly imbalanced domains, this capability becomes even more important. By modeling typical transaction patterns via learning, these models may identify anomalies that may have a connection. dishonest activity that was unnoticed before.This proposed system incorporates a number of additional unsupervised anomaly detection methods, including Autoencoders, DBSCAN, Isolation Forest, and One-Class Support Vector Machine (OCSVM). Each method incorporates a different set of mathematical assumptions, which allows for a more comprehensive comprehension of transactional irregularities. These ideas cover a wide spectrum, including probabilistic modeling, density estimation, subspace projection, and reconstruction learning. To determine the most effective, stable, and interpretable models, the study analyzes and contrasts them across diverse financial datasets. To ensure analytical consistency, a robust preprocessing and feature engineering pipeline was created to handle noise, normalization, and dimensionality reduction. In addition, a conceptual monitoring mechanism is outlined that may detect data drift and evolving fraud dynamics, which might be useful for future adaptive model retraining. Incorporating evaluation and visualization tools, the framework further aids analysts in interpreting anomaly scores and modifying decision thresholds.

Literature Survey

This research aims to gather and organize relevant evaluation criteria, features, and capabilities in order to provide a scorecard for the current status of credit card fraud detection utilizing predictive analytics vendor solutions. This scorecard provides a side-by-side comparison of five different vendor solutions for credit card predictive analytics in Canada. The risks, limitations, and challenges of credit card fraud Based on the analysis that followed, PAT vendor solutions were described.

We provide a two-step method for aligning sequences in this study. To begin, a profile analyzer (PA) verifies whether the sequence of incoming credit card transactions corresponds to the real spending sequence of the cardholder's historical transactions. After the profile analyst notices any questionable transactions, they notify a deviation analyzer (DA), who is then asked to confirm whether the transactions are in line with past cases of fraud. The final transaction status is decided by these two experts based on their findings. To accomplish online response time for PA and DA, we provide a new approach that combines the sequence alignment methods BLAST and SSAHA.

The growth of China's credit card sector has coincided with an uptick in credit card theft. Credit card fraud detection and prevention have recently been the focal points of banks' risk management initiatives. Considering the unusualness and rarity of illegal transactions, our model for credit card fraud detection utilizes outlier mining methods to credit card transaction data, particularly outlier identification based on distance sum. Experiments suggest that this technology might reliably identify credit card fraud.

As e-commerce becomes more complex, more and more people are falling prey to fraud, which costs victims a pretty penny. Many individuals, both merchants and regular consumers, are affected by the current epidemic of credit card theft. It is possible to detect credit card fraud using the methods described below. Decision trees, ANNs, GNs, ML techniques, and HMMs are all part of this category. For the purpose of detecting fraud, the contemplate system makes use of the decision tree and Support Vector Machine (SVM), two artificial intelligence principles.

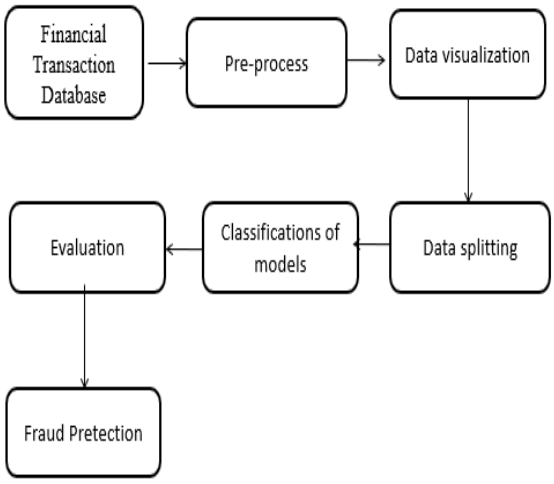
ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

Because of this, using this hybrid method may help reduce financial losses even more.

Methodology

The proposed approach makes use of machine learning models, namely MODEL, XgBoost, and Multilayer Perceptron (MLP), to detect and prevent fraudulent financial transactions. The procedure begins with data collection from financial transaction logs and continues with preprocessing activities including missing value management, outlier detection, and feature scaling. Important details about the transaction, such as the amount, time, location, and patterns of user activity, are extracted for study. Separating the dataset into training and testing sets is the second stage in ensuring that the model can be applied to different scenarios. XgBoost is used for its ability to improve classification accuracy and handle imbalanced data, MLP for its capability to grasp intricate transactional relationships, and MODEL for its ability to detect sequential fraud patterns over time. The model's efficiency is maximized by the fine-tuning of hyperparameters. To measure the models' detection ability, we employ metrics like accuracy, precision, recall, and F1-score. We employ the top-performing model for real-time fraud detection. This approach enhances monetary safety by providing automated, scalable, and very efficient fraud detection processes.



Block diagram

Academics have deliberated data mining, artificial intelligence, fuzzy logic, and machine learning as potential approaches to detecting credit card fraud. Credit card theft is a common problem, but it is difficult to detect. A credit card fraud detector based on machine learning is a part of our proposed method. Because of advancements in more effective machine learning techniques. Machine learning is a powerful tool for identifying fraudulent activities. The vast quantities of data sent during online transactions can only have one legal or fraudulent effect. Online businesses are able to consistently identify fraudulent transactions because of chargebacks. The sample of hypothetical datasets is where feature building happens. Credit card facts such as the issuer, account age, balance, and origin are all part of this data set. There could be hundreds of attributes, and their combined impact on the probability of fraud varies. How much each attribute adds to the fraud score is determined by the machine's artificial intelligence, which is driven by the training set. This degree is not chosen by a fraud analyst. As a result, if it is shown that a lot of people use their cards to commit fraud, then a credit card transaction will be considered just as fraudulent as a regular card transaction. But even if this were to go down, the contribution amount would be about the same. To rephrase, these automated systems are capable of learning independently, in contrast to human reviewers. Credit card fraud is detected using machine learning's regression and classification algorithms. We employ decision tree and other supervised learning algorithms to detect fraudulent credit card transactions, whether they occur online or in a physical store. Random Forest outperforms all other machine learning algorithms in terms of efficiency and accuracy. To reduce the impact of correlation, random forest shrinks the feature space with each split. I will go into more detail later on, but fundamentally, it aims to prune the trees by creating a stopping condition for node splits and de-correlating them.

In order to identify instances of fraudulent use of credit cards in online purchases, we present a machine learning model. The sheer volume and complexity of the data makes human analysis of fraudulent transactions impossible. But it should be doable with Machine Learning given enough informative attributes. The initiative will test this theory. Credit card fraud and legal transactions may be identified using supervised learning algorithms like Random Forest.To aid in raising awareness of the fraudulent without causing any financial harm.

ISSN: 2454-9940 www.ijasem.org

Vol 20, Issuse.2 April 2026

Modules

1. **Data Collection**
2. **Data Pre-Processing**
3. **Feature Extraction**
4. **Evaluation Model**

Data Collection

Product reviews culled from credit card transaction records form the basis of this paper's data. Picking out the data subset that will serve as your basis for analysis is the focus of this stage. To begin solving ML issues, you need data—ideally, a large amount of data in the form of instances or observations for which the desired outcome is known. Labeled data is information for which the desired outcome is known in advance.

Data Processing

After you've formatted, cleaned, and sampled your data, organize it. Here are three stages that data is often pre-processed: The data you've chosen may not be in an appropriate format for your needs. The data may be stored in a proprietary file format that you would want converted to a relational database or text file, or it could be in a relational database that you would like in a flat file format. Cleaning data involves erasing or correcting any data that is missing. You may not have all the information you need to fix the issue if certain data instances are missing or incomplete. Eliminating these cases could be necessary. On top of that, some characteristics can include sensitive information that has to be either anonymized or deleted from the data completely. Sampling: You can have access to more data samples than you really need. Increases in both computational and memory demands, as well as algorithm execution durations, are possible outcomes of data explosion. If you want to explore and prototype ideas quickly before thinking about the whole dataset, you may pick a smaller representative sample of the chosen data.

Feature Extraction

Step two involves reducing characteristics via feature extraction. Feature extraction changes the qualities themselves, in contrast to feature selection that only ranks traits based on their predictive utility. Characteristics that undergo transformation are basically just a linear mixture of the parent features. Lastly, the Classifier technique is used to train our models. As part of Python's Natural Language Toolkit, we make use of the categorize module. We utilize the cataloged dataset that has been collected. In order to test the models, we will use the remaining data that has been labeled. Sorting the data according to its pre-processing required many machine learning techniques. Random forest was the chosen classifier. Challenges involving text categorization make considerable use of these strategies.

Evolution Model

An essential aspect of developing a model is evaluating it. It is useful for determining which model best fits our data and for gauging the model's future performance. Since it is easy to create overoptimistic and overfit models when evaluating model performance with the data used for training, it is not acceptable in data science. Hold-Out and Cross-Validation are the two ways that data scientists use to assess models. Both approaches assess the model's efficacy using a hidden test set, which helps to prevent over fitting. An average is used to assess the performance of each categorization model. You may expect to see the outcome in its graphic form. Graphs used to depict categorized data. Prediction accuracy relative to the test data is the accuracy metric. You can simply figure it out by dividing the total number of forecasts by the number of right ones.

Decision Tree Analysis is a general method for predictive modeling that can be useful in many different areas. An algorithmic approach that discovers ways to segment a dataset according to different criteria is usually used to build decision trees. This method ranks high among the most well-liked and practical techniques to supervised learning. You can't go wrong with decision trees as a non-parametric supervised learning approach for classification and regression. The goal is to teach a model to predict the value of a target variable by learning simple decision rules from data characteristics. A decision tree is a hierarchical data structure where each node represents an interest or question, each edge represents a potential answer to that question, and each leaf represents the output, such a class label, of the tree. It is possible to create non-linear judgments using a simple linear decision surface.

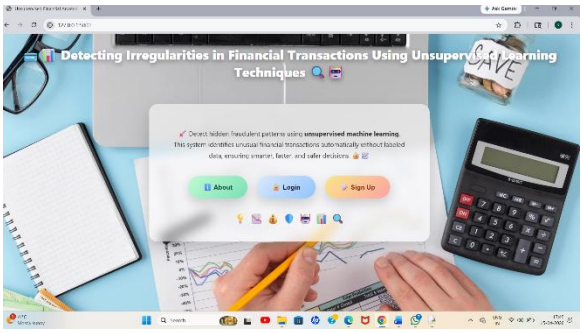
Following sorting from the root, each sample in a decision tree is given a classification at a leaf node. Every node in the tree represents a property that is checked, and every possible answer is represented by

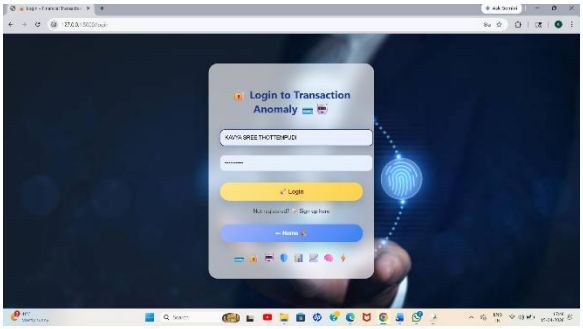
ISSN: 2454-9940 www.ijasem.org

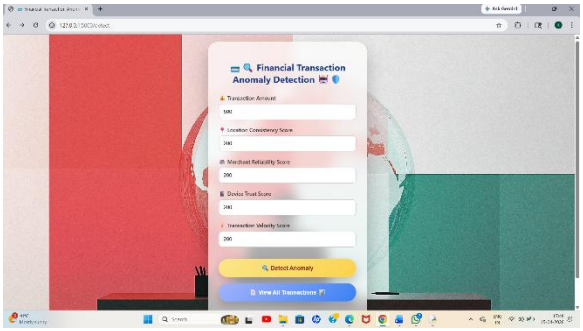
Vol 20, Issuse.2 April 2026

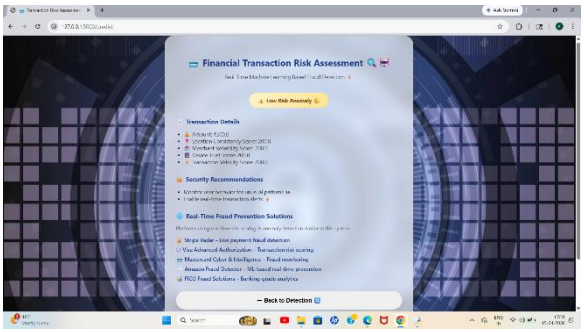
an edge that descends from that node. This iterative process is repeated for each child tree that draws its nodes from the freshly inserted nodes.

Results



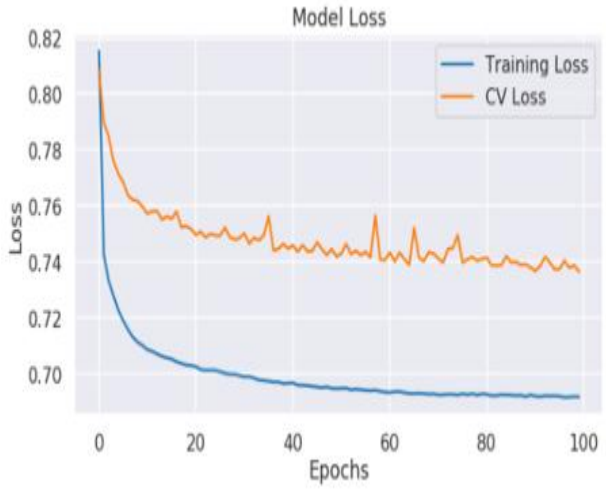






Performance of unsupervised models on the E-Commerce dataset

Model	Precision	Recall	F1-Score	AUC-ROC
Isolation Forest	0.01	0.44	0.01	0.68
LOF	0.02	0.37	0.04	0.67
One-Class SVM	0.00	0.56	0.10	0.68
GMM	0.01	0.22	0.01	0.60
DBSCAN	0.01	**0.67**	0.01	0.77
Autoencoder	**0.05**	0.21	**0.07**	**0.79**



Training loss curve of the autoencoder across epochs

Descriptivestatisticsofreconstructionerrorsfortheautoencodermodel

Class	Count	Mean	Std. Dev.	Min	25%	50%	Max
Legitimate (0)	56,864	0.6766	2.5419	0.0410	0.2330	0.3765	153.0625
Fraudulent (1)	98	30.7112	45.5040	0.1406	3.7133	10.2410	258.0216

Conclusion

An effective and adaptable weapon in the battle against modern financial crimes is the detection of fraud in financial transactions driven by artificial intelligence. Using models from deep learning and machine learning, institutions can quickly sort through mountains of transaction data. When compared to conventional rule-based methods, these systems have a better chance of spotting anomalies, suspicious actions, and hidden fraud inclinations. By automating routine tasks, financial analysts may save time and improve the precision of their detections. The system is also self-learning and can adjust to different types of fraud, allowing it to become better over time. Legal conformity, customer trust, and enhanced safety are thereby assured by AI. Artificial intelligence (AI) powered fraud detection is a giant leap forward for strengthening financial ecosystems.

Future Scope

Advances in explainable AI, federated learning, and blockchain integration bode well for the future of AI-driven fraud detection. Use Natural Language Processing (NLP) to sift through unstructured data, such as emails, chats, and papers, for indications of fraud. By integrating edge computing with real-time predictive analytics, mobile and IoT-based transaction fraud may be detected more quickly. Also, when confronted with undiscovered fraud schemes, hybrid models that include supervised learning, unsupervised learning, and reinforcement learning will be more flexible. Using privacy-preserving methods, such as differential privacy, data protection standards will be respected. In order to detect internal dangers and identity theft, the system's capabilities need to be expanded beyond transactional data. Ultimately, future AI systems will be more transparent, scalable, and resistant to complicated financial fraud.

References

1. A. Kumar and S. Bhatia, "House Price Prediction Using Machine Learning and Neural Networks," *Int. J. Comput. Appl.*, vol. 182, no. 48, pp. 25-30, 2020.
2. J. Zhang, M. Sun, and X. Li, "Housing Price Prediction Using Machine Learning Algorithms," in *Proc. IEEE Int. Conf. Artif. Intell. Big Data (ICAIBD)*, 2021, pp. 145-150.
3. 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," *arXiv preprint arXiv:2402.14389*, 2024. [Online]. Available: https://arxiv.org/abs/2402.14389
4. Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, "Explainable Artificial Intelligence and Causal Inference based ATM Fraud Detection," *arXiv preprint arXiv:2211.10595*, 2022. [Online]. Available: https://arxiv.org/abs/2211.10595
5. D. H. M. de Souza and C. J. Bordin Jr, "Ensemble and Mixed Learning Techniques for Credit Card Fraud Detection," *arXiv preprint arXiv:2112.02627*, 2021. [Online]. Available: https://arxiv.org/abs/2112.02627
6. Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, "ATM Fraud Detection using Streaming Data Analytics," *arXiv preprint arXiv:2303.04946*, 2023. [Online]. Available: https://arxiv.org/abs/2303.04946
7. 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," *arXiv preprint arXiv:2402.14389*, 2024. [Online]. Available: https://arxiv.org/abs/2402.14389
8. Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, "Explainable Artificial Intelligence and Causal Inference based ATM Fraud Detection," *arXiv preprint arXiv:2211.10595*, 2022. [Online]. Available: https://arxiv.org/abs/2211.10595
9. D. H. M. de Souza and C. J. Bordin Jr, "Ensemble and Mixed Learning Techniques for Credit Card Fraud Detection," *arXiv preprint arXiv:2112.02627*, 2021. [Online]. Available: https://arxiv.org/abs/2112.02627
10. Y. Vivek, V. Ravi, A. A. Mane, and L. R. Naidu, "ATM Fraud Detection using Streaming Data Analytics," *arXiv preprint arXiv:2303.04946*, 2023. [Online]. Available: https://arxiv.org/abs/2303.04946