DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE
(AUTONOMOUS)
(Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai)
Re-Accredited with 'A' Grade by NAAC, Accredited by TCS.
Accredited by NBA (AERO, CSE, IT & MECH)
Re-Accredited by NBA (BME, ECE & EEE)
PERAMBALUR-621 212

Online Recruitment Fraud Detection Using Deep Learning Approaches

LITERATURESURVEY

BATCH NO :14

TEAMMEMBERS:

RAUSHAN KUMAR **(810421104139)**

SANNI K. RAJAN **(810421104148)**

SANSHAY KUMAR TIWARY **(810421104149)**

SURAJ KUMAR **(810421104172)**

GUIDED BY:

Dr. R. GOPI, M.Tech., Ph.D.,(PDF).

HOD OF THE CSE,

DHANALAKSHMISRINIVASAN ENGINEERINGCOLLEGE(AUTONOMOUS), PERAMBALUR

Batch No: 14

Online Recruitment Fraud Detection Using Deep Learning Approaches

ABSTRACT

Most companies nowadays are using digital platforms for the recruitment of new employees to make the hiring process easier. The rapid increase in the use of online platforms for job posting has resulted in fraudulent advertising. The scammers are making money through fraudulent job postings. Online recruitment fraud has emerged as an important issue in cybercrime. Therefore, it is necessary to detect fake job postings to get rid of online job scams. In recent studies, traditional machine learning and deep learning algorithms have been implemented to detect fake job postings; this research aims to use two transformer-based deep learning models, i.e., Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT-Pretraining Approach (RoBERTa) to detect fake job postings precisely. In this research, a novel dataset of fake job postings is proposed, formed by the combination of job postings from three different sources. Existing benchmark datasets are outdated and limited due to knowledge of specific job postings, which limits the existing models' capability in detecting fraudulent jobs.

BATCH NO :14

TEAM MEMBERS:

RAUSHAN KUMAR (810421104139)

SANNI K. RAJAN (810421104148)

SANSHAY KUMAR TIWARY (810421104149)

SURAJ KUMAR (810421104172)

GUIDED BY:

Dr. R. GOPI, M.Tech., Ph.D.,(PDF).

HOD OF THE CSE,

DHANALAKSHMI SRINIVASAN ENGINEERINGCOLLEGE(AUTONOMOUS), PERAMBALUR

TABLEOFCONTENTS

6	An Adversary Model of Fraudsters' Behavior to Improve Oversampling in Credit Card Fraud Detection	Fraud detection, imbalance learning, machine learning, oversampling, synthetic data, time series, threat model	Limitations include adversarial adaptability, data imbalance, oversampling bias, model complexity, and real-time fraud detection challenges	
7	Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms	Fraud detection, deep learning, machine learning, online fraud, credit card frauds, transaction data analysis	Limitations include high computational cost, data imbalance, adversarial attacks, interpretability issues, and real-time detection challenges.	
8	Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Auto encoders for Real-Time Credit Card Fraud Prevention	Deep learning, credit card, fraud detection, graph neural network, auto encoders	Limitations include scalability issues, high computational costs, data privacy concerns, adversarial robustness, and real-time processing constraints.	
9	Enhancing Medicare Fraud Detection Through Machine Learning: Addressing Class Imbalance With SMOTE-ENN	Healthcare fraud, imbalanced data, machine learning (ML), noisy data	Limitations include potential overfitting, data noise sensitivity, high computational cost, and challenges in real-time fraud detection.	
10	Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection	Class imbalance, credit card fraud detection, GAN, Random Forest, SMOTE	Limitations include mode collapse in GANs, oversampling bias, high computational cost, and challenges in real-time fraud detection.	

11	Evaluating the Computational Advantages of the Variational Quantum Circuit Model in Financial Fraud Detection	anti-fraud engine, quantum computing, transaction classification, hybrid quantum-classical computing, angle encoding, quantum neural network	Limitations include hardware constraints, noise sensitivity, scalability issues, algorithm stability, and practical implementation challenges in real-world fraud detection.	
12	FedFusion: Adaptive Model Fusion for Addressing Feature Discrepancies in Federated Credit Card Fraud Detection	Credit card fraud, fraud detection system, federated learning, FedFusion, CNN, MLP, LSTM, data heterogeneity.	Limitations include communication overhead, feature inconsistency, privacy risks, model convergence issues, and scalability in federated learning environments.	
13	Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Auto encoders for Real-Time Credit Card Fraud Prevention	Detection rate, fraud detection, K-nearest neighbor, skewed instances, value-at-risk.	Limitations include high computational complexity, data privacy concerns, adversarial attacks, scalability challenges, and real-time processing constraints.	
14	Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset	Credit card fraud detection, imbalanced data, generative adversarial networks, deep learning, fraud detection	Limitations include mode collapse, data imbalance, overfitting, high computational cost, and challenges in real-world deployment.	
15	Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review	Fraud detection, nancial statement, machine learning, data mining, outlier detection, systematic literature review	Limitations include data quality issues, feature selection bias, model interpretability challenges, and computational complexity.	

16	Machine Learning Methods for Credit Card Fraud Detection: A Survey	Fraud detection, machine learning, neural networks, synthetic data	Limitations include data imbalance, adversarial attacks, model interpretability, high false positives, and real-time processing challenges	
17	Medicare Fraud Detection Using Graph Analysis: A Comparative Study of Machine Learning and Graph Neural Networks	Graph neural network, graph centrality measure, machine learning, medicare fraud detection	Limitations include scalability issues, high computational cost, data privacy concerns, class imbalance, and real-time detection challenges	
18	Online Payment Fraud Detection Model Using Machine Learning Techniques	Financial transaction fraud, deep learning, fraud defense mechanism, detection, optimization methods, classification, ResNeXt, cyber attacks.	Limitations include data imbalance, evolving fraud tactics, high false positives, real-time processing constraints, and model interpretability challenges.	
19	OptDevNet: A Optimized Deep Event-Based Network Framework for Credit Card Fraud Detection	Credit card fraud detection, OptDevNet framework, automatic fraud detection, malicious transactions, credit card security	Limitations include high computational cost, data imbalance, adversarial attacks, real-time processing challenges, and model interpretability issues.	
20	Quantum Autoencoder for Enhanced Fraud Detection in Imbalanced Credit Card Dataset	Anomalydetection, credit card fraud detection, imbalanced dataset, quantum autoencoder (QAE), quantum machine learning (QML)	Limitations include quantum hardware constraints, data encoding challenges, scalability issues, high computational cost, and real-world applicability concerns.	

1. Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries.

AUTHOURS: SYED MAHBUB, ERIC PARDEDE.

PUBLISHEDBY: IEEE.

YEAR: 2022.

ABSTRACT

The purpose of this study is to investigate the effects of contextual features on automatic detection accuracy of online recruitment frauds in Australian job market. In addition, the study aims to unearth the significance of localisation of such approaches. The study first generates a dataset based on a local and semi-structured advertising platform in Australia. The labelled dataset is then used to train a learning model on several content-based and contextual features. The existence of advertising body in relevant government and non-government registries in Australia, along with the internet presence of the advertiser, were considered as contextual features.

Methodology Used:

- Contextual feature extraction from job listings
- Deep learning models (e.g., BERT, Ro BERT)
- SMOTE for class imbalance handling
- NLP-based text classification

Advantages:

- Improved fraud detection accuracy using contextual features
- Deep learning models capture semantic nuances
- Handles class imbalance effectively

Disadvantages:

- Computationally expensive models
- Requires large labeled datasets
- Potential false positives affecting genuine job postings

2. A Credit Card Fraud Detection Algorithm Based on SDT and Federated Learning.

AUTHOURS: YUXUAN TANG, ZHANJUN LIU.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The rise of digital payment methods and the growth in financial transactions, the issue of credit card fraud has become increasingly severe. Traditional fraud detection methods are currently facing challenges such as poor model performance, difficulty in obtaining accurate results, and limitations in distributed deployment. These challenges stem from constantly evolving fraud strategies, higher volumes of transactions, and the complexity of the financial environment. This study proposes a credit card fraud detection algorithm based on Structured Data Transformer (SDT) and federated learning, which leverages the advanced capabilities of the Transformer model in deep learning. First, we organize credit card data into sequences and introduce a special, learnable token at the beginning of each sequence for classification purposes.

Methodology Used:

- Self-Adaptive Decision Tree (SDT) for dynamic fraud detection.
- Federated Learning (FL) for decentralized data processing.
- Edge computing to enhance security and efficiency.

Advantages:

- Preserves user privacy by avoiding central data storage.
- Adaptable to evolving fraud patterns.
- Reduces computational burden on central servers.

Disadvantages:

- Requires high communication bandwidth for model updates.
- Potential security risks in federated model aggregation.
- Performance depends on data distribution across clients.

3. A Novel Framework for Credit Card Fraud Detection.

AUTHOURS: AYOUB MNIAI, MOUNA TARIK, AND KHALID JEBARI.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

Credit card transactions have grown considerably in the last few years. However, this increase has led to significant financial losses around the world. More than that, processing the enormous amount of generated data becomes very challenging, making the datasets highly dimensional and unbalanced. This means the collected data is suffering from two major problems. It is characterized by a severe difference in observation frequency between fraud and non-fraud transactions, and it contains irrelevant, inappropriate, and correlated data that negatively affects their prediction performance.

Methodology Used:

- Machine Learning (e.g., Random Forest, XGBoost, SVM)
- Deep Learning (e.g., LSTM, CNN, Autoencoders)
- Anomaly Detection (Isolation Forest, One-Class SVM)

Advantages:

- High accuracy in detecting fraudulent transactions
- Real-time fraud detection capability
- Reduces false positives using advanced models

Disadvantages:

- Computationally expensive for real-time processing
- Data privacy concerns with sensitive information
- Imbalanced dataset challenges affect model performance

4. A Systematic Literature Review of Fraud Detection Metrics in Business Processes.

AUTHOURS: BADR OMAIR AND AHMAD ALTURKI.

PUBLISHEDBY: IEEE.

YEAR: 2020.

ABSTRACT

Fraud is a primary source of organization losses, amounting to up to 5% of yearly revenues. Process-based fraud (PBF) is fraud involving a deviation from the standard operating procedure (SOP) of business processes. PBF hinders the achievement of business objectives because business processes operationalize organizational strategies. A systematic content analysis of the literature was conducted on fraud detection metrics in business processes. The current state of fraud detection was surveyed by focusing on PBF metrics while including all relevant conceptual perspectives of PBF detection. The findings indicate that a large body of research has examined detection metrics for possible fraud, but less attention has been paid to PBF.

Methodology Used:

- Systematic Literature Review (SLR)
- PRISMA framework for study selection
- Keyword-based search in academic databases

Advantages:

- Comprehensive overview of fraud detection metrics
- Identifies trends and gaps in existing research
- Provides a structured comparison of techniques

Disadvantages:

- Time-consuming and resource-intensive
- Potential bias in study selection
- May not capture latest unpublished techniques

5. Advanced Credit Card Fraud Detection: An Ensemble Learning Using Random Under Sampling and Two-Stage.

AUTHOURS: IBRAHIM ALMUBARK.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The increase of Credit Card (CC) fraud in recent years requires the development of fraud detection systems that are both efficient and robust. This paper explored the utilization of machine learning models, with a particular emphasis on ensemble methods, to advance the detection of CC fraud. We present an ensemble model that incorporates different classifiers to address the dataset imbalance issue that is present in most CC datasets. We employed synthetic over-sampling and under-sampling techniques in certain machine learning algorithms to tackle the same issue.

Methodology Used :

- Ensemble learning approach
- Random Under Sampling (RUS) for handling class imbalance
- Two-stage detection method

Advantages:

- Improves fraud detection accuracy
- Reduces false positives and negatives
- Handles class imbalance effectively

Disadvantages:

- Risk of losing valuable data due to under-sampling
- Computationally expensive with multiple models
- Requires fine-tuning for optimal performance

6. An Adversary Model of Fraudsters' Behavior to Improve Oversampling in Credit Card Fraud Detection.

AUTHOURS: DANIELE LUNGHI, GIAN MARCO PALDINO, OLIVIER CAELEN.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

Imbalanced learning jeopardizes the accuracy of traditional classification models, particularly for what concerns the minority class, which is often the class of interest. This paper addresses the issue of imbalanced learning in credit card fraud detection by introducing a novel approach that models fraudulent behavior as a time-dependent process. The main contribution is the design and assessment of an oversampling strategy, called ''Adversary-based Oversampling'' (ADVO), which relies on modeling the temporal relationship among frauds. The strategy is implemented by two learning approaches: first, an innovative regression-based oversampling model that predicts subsequent fraudulent activities based on previous fraud features.

Methodology Used:

- Adversarial modeling of fraudster behavior
- Improved oversampling techniques for class imbalance
- Synthetic fraud data generation based on adversary patterns

Advantages:

- Better fraud detection accuracy
- More realistic synthetic fraud samples
- Improved model robustness against adversarial behavior

Disadvantages:

- Complexity in modeling adversarial strategies
- Potential risk of generating biased synthetic data
- Increased computational cost

7. Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms.

AUTHOURS: FAWAZ KHALED ALARFAJ, IQRA MALIK2, HIKMAT ULLAH KHAN.

PUBLISHEDBY: IEEE.

YEAR: 2022.

ABSTRACT

People can use credit cards for online transactions as it provides an efficient and easy-to-use facility. With the increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Credit card frauds cause significant financial losses for both credit card holders and financial companies. In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. However, due to low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraud losses.

Methodology Used :

- Supervised learning (SVM, Decision Trees, Random Forest, XGBoost)
- Unsupervised learning (Autoencoders, Isolation Forest, Clustering)
- Deep learning (LSTMs, CNNs, DNNs)

Advantages:

- High accuracy with deep learning models
- Automated feature extraction (CNNs, LSTMs)
- Detects complex fraud patterns

Disadvantages:

- Computationally expensive
- High false positives in some models
- Data privacy concerns

8. Enhancing Fraud Detection in Banking With Deep Learning: Graph Neural Networks and Auto encoders for Real-Time Credit Card Fraud Prevention.

AUTHOURS: FAWAZ KHALED ALARFAJ, SHABNAM SHAHZADI.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

Under the umbrella of artificial intelligence (AI), deep learning enables systems to cluster data and provide incredibly accurate results. This study explores deep learning for fraud detection, utilizing Graph Neural Networks (GNNs) and Auto encoders to enhance business practices and reduce fraudulent activities in large organizations. For real-time fraud detection, we propose Graph neural network with lambda architecture while for credit card fraud detection, we use an auto encoder, validated through case studies from two banks. The findings demonstrate that these methods effectively detect fraud with balance of precision and recall, improving the efficiency of banking systems.

Methodology Used :

- Graph Neural Networks (GNNs) for transaction relationship analysis
- Auto encoders for anomaly detection
- Real-time fraud prevention with deep learning models

Advantages:

- Detects complex fraud patterns
- Real-time transaction monitoring
- Reduces false positives

Disadvantages:

- High computational cost
- Requires large labeled datasets
- Potential for adversarial attacks

9. Enhancing Medicare Fraud Detection Through Machine Learning: Addressing Class Imbalance With SMOTE-ENN.

AUTHOURS: RAYENE BOUNAB, KARIM ZAROUR, BOUCHRA GUELIB.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The healthcare fraud detection field is constantly evolving and faces significant challenges, particularly when addressing imbalanced data issues. Previous studies mainly focused on traditional machine learning (ML) techniques, often struggling with imbalanced data. This problem arises in various aspects. It includes the risk of over fitting with Random Oversampling (ROS), noise introduction by the Synthetic Minority Oversampling Technique (SMOTE), and potential crucial information loss with Random Under sampling (RUS). Moreover, improving model performance, exploring hybrid re sampling techniques, and enhancing evaluation metrics are crucial for achieving higher accuracy with imbalanced datasets.

Methodology Used :

- Machine learning models for fraud detection
- SMOTE-ENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors) for class imbalance
- Feature selection and engineering

Advantages:

- Improves fraud detection accuracy
- Reduces class imbalance impact
- Enhances model generalization

Disadvantages:

- Computationally expensive
- Risk of over fitting
- May remove legitimate rare cases

10. Ensemble Synthesized Minority Oversampling-Based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection.

AUTHOURS: FUAD A. GHALEB, FAISAL SAEED.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

The recent increase in credit card fraud is rapidly has caused huge monetary losses for individuals and financial institutions. Most credit card frauds are conducted online by illegally obtaining payment credentials through data breaches, phishing, or scamming. Many solutions have been suggested to address the credit card fraud problem for online transactions. However, the high-class imbalance is the major challenge that faces the existing solutions to construct an effective detection model. Most of the existing techniques used for class imbalance overestimate the distribution of the minority class, resulting in highly overlapped or noisy and unrepresentative features, which cause either over fitting or imprecise learning.

Methodology Used :

- Uses GANs to generate synthetic fraudulent transactions.
- Balances class distribution by oversampling minority class.
- Combines multiple GAN models for better fraud pattern learning.

Advantages:

- Handles class imbalance effectively.
- Generates realistic fraudulent samples for training.
- Improves detection performance with ensemble learning.

Disadvantages:

- Computationally expensive.
- Requires careful tuning to avoid mode collapse.
- Potential risk of over fitting to synthetic data

11. Evaluating the Computational Advantages of the Variational Quantum Circuit Model in Financial Fraud Detection.

AUTHOURS: ANTONIO TUDISCO, DEBORAH VOLPE.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

Home banking and digital payments diffusion has greatly increased in recent years. As a result, fraud has also dramatically grown, resulting in the loss of billions of dollars worldwide every year. Therefore, banks and financial institutions are required to offer clients increasingly effective and sophisticated services for illegal transaction detection. Machine learning strategies are commonly employed for this crucial application. However, classical models are not satisfactory enough in highly unbalanced classification tasks like fraud detection.

Methodology Used :

- Variational Quantum Circuits (VQC) for fraud classification
- Hybrid quantum-classical models
- Quantum feature mapping and encoding

Advantages:

- Potential exponential speedup in complex fraud detection
- Better pattern recognition in high-dimensional data
- Enhanced security in data processing

Disadvantages:

- Current quantum hardware limitations (noise, qubit errors)
- Limited scalability for large datasets
- Requires hybrid models due to lack of full quantum advantage yet

12. Fed Fusion: Adaptive Model Fusion for Addressing Feature Discrepancies in Federated Credit Card Fraud Detection.

AUTHOURS: NAHID FERDOUS AURNA, MD DELWAR HOSSAIN.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The digitization of financial transactions has led to a rise in credit card fraud, necessitating robust measures to secure digital financial systems from fraudsters. Nevertheless, traditional centralized approaches for detecting such frauds, despite their effectiveness, often do not maintain the confidentiality of financial data. Consequently, Federated Learning (FL) has emerged as a promising solution, enabling the secure and private training of models across organizations. However, the practical implementation of FL is challenged by data heterogeneity among institutions, complicating model convergence.

Methodology Used :

- Federated Learning (FL) with adaptive model fusion
- Feature discrepancy handling across clients
- Aggregation of model updates via FedFusion strategy

Advantages:

- Preserves data privacy across institutions
- Adapts to heterogeneous client data distributions
- Improves fraud detection accuracy over standard FL

Disadvantages:

- Increased computational overhead for adaptive fusion
- Communication costs in federated settings
- Requires careful tuning of fusion weights

13. Financial Fraud Detection Using Value-at-Risk With Machine Learning in Skewed Data.

AUTHOURS: ABDULLAHI UBALE USMAN, SUNUSI BALA ABDULLAHI.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The significant losses that banks and other financial organizations suffered due to new bank account (NBA) fraud are alarming as the number of online banking service users increases. The inherent skewness and rarity of NBA fraud instances have been a major challenge to the machine learning (ML) models and happen when non-fraud instances outweigh the fraud instances, which leads the ML models to overlook and erroneously consider fraud as non-fraud instances. Such errors can erode the confidence and trust of customers.

Methodology Used :

- Value-at-Risk (VaR) for risk estimation
- Machine learning models (e.g., SVM, Random Forest, Neural Networks)
- Data preprocessing (handling skewed data with SMOTE, undersampling, or cost-sensitive learning)

Advantages:

- Effective risk estimation with VaR
- Machine learning improves fraud detection accuracy
- Handles large-scale financial data

Disadvantages :

- VaR has limitations in extreme market conditions
- Class imbalance can lead to biased predictions
- High computational cost for complex models

14. Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset.

AUTHOURS: FAHDAH A. ALMARSHAD, GHADA ABDALAZIZ GASHGARI.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

Credit card use poses a significant security issue on a global scale, with rule-based algorithms and traditional anomaly detection being two of the most often used methods. However, they are resourceintensive, time-consuming, and erroneous. Given fewer instances than legal payments, the dataset imbalance has become a serious issue. On the other hand, the generative technique is considered an effective way to rebalance the imbalanced class issue, as this technique balances both minority and majority classes before the training. In a more recent period, GAN is considered one of the most popular data generative techniques, as it is used in significant data settings.

Methodology Used :

- GANs generate synthetic fraud samples to improve detection.
- Discriminator distinguishes real vs. fake transactions.
- Model trained with adversarial learning.

Advantages:

- Handles class imbalance effectively.
- Improves fraud detection accuracy.
- Generates realistic fraudulent samples.

Disadvantages:

- Computationally expensive.
- Requires careful hyper parameter tuning.
- Risk of mode collapse in GANs.

15. Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review

AUTHOURS: MATIN N. ASHTIANI AND BIJAN RAAHEMI.

PUBLISHEDBY: IEEE.

YEAR: 2021.

ABSTRACT

Fraudulent financial statements (FFS) are the results of manipulating financial elements by overvaluing incomes, assets, sales, and profits while underrating expenses, debts, or losses. To identify such fraud ulentstatements, traditional methods, including manual auditing and inspections, are costly, imprecise, and time-consuming. Intelligent methods can significantly help auditors in analyzing a large number of financial statements. In this study, we systematically review and synthesize the existing literature on intelligent fraud detection in corporate financial statements.

Methodology Used :

- Machine learning techniques (e.g., SVM, Decision Trees, Random Forest, Neural Networks)
- Data mining approaches (e.g., anomaly detection, clustering, rule-based classification)
- Feature selection methods to identify financial fraud indicators

Advantages:

- High accuracy in fraud detection
- Automates the fraud detection process, reducing manual effort
- Can analyze large-scale financial data efficiently

Disadvantages:

- Requires high-quality labeled data for supervised learning
- Potential bias in model training due to imbalanced datasets
- Difficult to interpret complex models (e.g., deep learning)

16. Machine Learning Methods for Credit Card Fraud Detection: A Survey.

AUTHOURS: KANISHKA GHOSH DASTIDAR, OLIVIER CAELEN.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

The widespread adoption of online payments has been accompanied by a significant increase in fraudulent activities, resulting in billions of dollars in financial losses. As payment providers aim to tackle this with various preventive mechanisms, fraudsters also continuously evolve their methods to remain indistinguishable from genuine actors. This necessitates sophisticated fraud detection tools to supplement these security mechanisms. As the volume of transactions taking place per day is in the millions, relying solely on human investigation is expensive and ultimately unfeasible, leading to an emergence of research into data driven or statistical methods for fraud detection.

Methodology Used :

- **Supervised Learning** – Uses labeled fraudulent and non-fraudulent transactions.
- **Unsupervised Learning** – Detects anomalies without labeled data.
- **Hybrid Models** – Combines supervised and unsupervised approaches for improved accuracy.

Advantages:

- **High Accuracy:** Especially deep learning and ensemble models.
- **Automated Feature Extraction:** Reduces manual effort (Deep Learning).
- **Real-time Detection:** Faster response time (LSTMs, CNNs).

Disadvantages:

- **Data Imbalance:** Fraud cases are rare, leading to biased models.
- **High Computational Cost:** Deep learning requires significant resources.
- **Lack of Interpretability:** Black-box models (Neural Networks).

17. Medicare Fraud Detection Using Graph Analysis: A Comparative Study of Machine Learning and Graph Neural Networks.

AUTHOURS: YEEUN YOO, JINHO SHIN, SUNGHYON KYEONG.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

Insurance companies have focused on medicare fraud detection to reduce financial losses and reputational harm because medicare fraud causes tens of billions of dollars in damage annually. This study demonstrates that medicare fraud detection can be significantly enhanced by introducing graph analysis with considering the relationships among medical providers, beneficiaries, and physicians. We use open-source tabular datasets containing beneficiary information, inpatient claims, outpatient claims, and indications about potential fraudulent providers.

Methodology Used:

- Constructing a graph representation of Medicare claims data.
- Feature extraction from network structures (e.g., provider-patient relationships).
- Applying traditional machine learning (e.g., Random Forest, SVM) and Graph Neural Networks (GNNs) for fraud detection.

Advantages:

- Captures complex relationships in Medicare fraud cases.
- GNNs improve fraud detection by leveraging graph structures.
- More effective in detecting anomalous patterns than traditional methods

Disadvantages:

- High computational cost for large-scale graph processing.
- Requires quality graph construction for accurate results.
- Interpretability of GNNs is lower compared to traditional models.

18. Online Payment Fraud Detection Model Using Machine Learning Techniques.

AUTHOURS: ABDULWAHAB ALI ALMAZROI, NASIR AYUB.

PUBLISHEDBY: IEEE.

YEAR: 2023.

ABSTRACT

In a world where wireless communications are critical for transferring massive quantities of data while protecting against interference, the growing possibility of financial fraud has become a significant concern. The ResNeXt-embedded Gated Recurrent Unit (GRU) model (RXT) is a unique artificial intelligence approach precisely created for real-time financial transaction data processing. Motivated by the need to address the rising threat of financial fraud, which poses major risks to financial institutions and customers, our artificial intelligence technique takes a systematic approach.

Methodology Used :

- Data preprocessing (handling missing values, feature selection)
- Feature engineering (transaction patterns, user behavior analysis)
- Model selection (Logistic Regression, Decision Trees, Random Forest, SVM, XGBoost, Neural Networks)

Advantages:

- High accuracy in fraud detection
- Real-time transaction monitoring
- Reduces false positives with advanced ML techniques

Disadvantages:

- Requires high-quality labeled data
- Computationally expensive for real-time processing
- Possible model bias leading to false positives/negatives

19. Opt Dev Net: A Optimized Deep Event-Based Network Framework for Credit Card Fraud Detection.

AUTHOURS: MUHAMMAD ADIL, ZHANG YINJUN.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

In recent times, credit card fraud has emerged as a substantial financial challenge for both cardholders and the issuing authorities. To address this demanding issue, researchers have employed machine learning techniques to identify fraudulent activities within labeled transaction records. However, these techniques have primarily been evaluated on limited or specific datasets, which may not adequately represent the broader real-world scenario. These limitations motivated us to comprehensively assess the existing machine learning classifiers and propose an Optimized Deep Event-based Network (OptDevNet) framework capable of addressing these challenges.

Methodology Used :

- Deep learning-based event-driven network
- Feature extraction from transaction sequences
- Optimization techniques for performance improvement

Advantages:

- High accuracy in fraud detection
- Real-time processing capability
- Handles sequential transaction patterns effectively

Disadvantages:

- Computationally expensive
- Requires large labeled datasets
- May struggle with adaptive fraud patterns

20. Quantum Auto encoder for Enhanced Fraud Detection in Imbalanced Credit Card Dataset.

AUTHOURS: CHANSREYNICH HUOT, SOVANMONYNUTH HENG.

PUBLISHEDBY: IEEE.

YEAR: 2024.

ABSTRACT

Credit card fraud detection is crucial for financial security which entails identifying unauthorized transactions that can result in significant financial losses. Detection is inherently challenging due to the rarity and indistinguishability of fraudulent transactions from genuine ones, which makes it an anomaly detection problem. Traditional detection systems struggle with the highly imbalanced nature of transaction datasets, where genuine transactions vastly outnumber fraudulent cases. In response to these challenges, we propose a novel detection model utilizing Quantum Auto Encoders-based Fraud Detection (QAE-FD).

Methodology Used :

- Quantum autoencoder for feature extraction and dimensionality reduction.
- Hybrid quantum-classical model combining quantum encoding with classical classification.
- Variational Quantum Circuits (VQC) for efficient data representation.

Advantages:

- Efficient feature compression improves fraud detection.
- Quantum computing enhances model scalability.
- Handles high-dimensional data effectively.

Disadvantages:

- Requires quantum hardware, limiting real-world deployment.
- Noisy quantum circuits can affect accuracy.
- Training quantum models is computationally expensive.