**Literature review Credit card fraud Detection**

**Introduction**

Credit card fraud detection has been a significant area of research due to the increasing number of fraudulent transactions in the financial industry. Various machine learning and deep learning approaches have been proposed to detect fraudulent activities efficiently. This literature review explores the traditional and contemporary methodologies employed in credit card fraud detection, emphasizing supervised and unsupervised learning techniques, hybrid models, and feature engineering strategies.

**Traditional Methods for Fraud Detection**

Traditional methods for fraud detection rely on rule-based systems, statistical models, and expert-driven heuristics. These methods have been foundational in fraud detection but suffer from various limitations in terms of adaptability and scalability.

* **Rule-Based Systems:** Rule-based approaches involve predefined sets of conditions and thresholds that determine whether a transaction is fraudulent. These rules are crafted by domain experts based on known fraud patterns, such as unusually large transactions, frequent transactions in a short period, or purchases from high-risk locations (Bhattacharyya et al., 2011). While effective in detecting known fraud types, rule-based systems struggle to adapt to new and evolving fraud techniques and require constant updates.
* **Statistical Models:** Statistical approaches such as logistic regression (LR), decision trees, and Bayesian networks have been widely applied in fraud detection (Bolton and Hand, 2002). Logistic regression helps in classifying transactions based on probability scores, while decision trees create a hierarchical structure to classify fraud and non-fraud instances. Bayesian networks utilize probabilistic reasoning to detect anomalies. However, these methods often require extensive feature engineering and domain knowledge to yield optimal results.
* **Expert-Driven Heuristics:** Many financial institutions traditionally relied on expert-driven heuristics where fraud analysts manually inspect suspicious transactions. This approach incorporates human expertise and intuition, but it is highly time-consuming, labour-intensive, and difficult to scale for high-volume transactions.
* **Outlier Detection:** Traditional anomaly detection techniques such as Z-score analysis and clustering methods (e.g., k-means) have been employed to identify outliers in transaction data (Bolton and Hand, 2002). These techniques assume that fraudulent transactions deviate significantly from normal spending behaviour. However, they may generate high false positives if legitimate transactions also exhibit unusual patterns.

Despite their contributions, traditional methods suffer from several drawbacks, including high false positive rates, lack of adaptability to emerging fraud patterns, and the need for extensive manual oversight. These limitations have led to the adoption of more sophisticated machine learning and deep learning approaches.

**Machine learning and Deep learning Methods**

With the rise of e-commerce and digital payments, credit card fraud detection has become crucial. Machine learning models often struggle due to redundant and irrelevant features in real-world data. A hybrid feature-selection technique is introduced, combining filter and wrapper methods to enhance model performance. Information gain ranks features, while a genetic algorithm optimized with the geometric mean refines selection. Using extreme learning machines, the approach achieves sensitivity of 0.997 and specificity of 0.994, surpassing baseline methods (Mienye and Sun, 2023).

Credit card fraud detection is crucial due to increasing fraudulent activities. The study applies a Random Forest (RF) classifier to detect fraud, addressing the challenge of imbalanced datasets where most transactions are legitimate. The Synthetic Minority Over-Sampling Technique (SMOTE) is used to balance the data, and hyperparameter tuning enhances model performance. Results show an accuracy of 98% and an F1-score of 98%, demonstrating effectiveness (AlsharifHasan Mohamad Aburbeian and Ashqar, 2023). The model is practical and improves fraud detection in imbalanced datasets.

The paper by (Awoyemi, Adetunmbi and Oluwadare, 2017) explores credit card fraud detection using data mining techniques, addressing the challenges of evolving fraud patterns and highly skewed datasets. It evaluates Naïve Bayes, k-Nearest Neighbour (k-NN), and Logistic Regression on a dataset of 284,807 European credit card transactions. A hybrid sampling method is applied to balance the data. The study, implemented in Python, assesses models based on multiple performance metrics. Results indicate that k-NN outperforms Naïve Bayes and Logistic Regression, achieving 97.69% accuracy. While the paper effectively compares these techniques, it could benefit from exploring additional methods like deep learning for improved fraud detection.

The paper explores credit card fraud detection using both traditional machine learning and deep learning techniques. It highlights the challenges of fraud detection, such as data imbalance and high false alarm rates. Various models, including Decision Trees, Random Forest, and Support Vector Machines, are evaluated before introducing convolutional neural networks (CNNs) to enhance performance. Through empirical analysis on a European benchmark dataset, the study demonstrates that increasing hidden layers and epochs improves detection accuracy. The proposed deep learning model achieves a remarkable 99.9% accuracy, surpassing state-of-the-art methods. This research effectively showcases CNNs as a superior approach for fraud detection (Alarfaj et al., 2022). The research explores credit card fraud detection amid increased online transactions due to COVID-19. A total of 66 machine learning models are evaluated using a European credit card fraud dataset and stratified K-fold cross-validation. In a two-stage process, nine algorithms are tested, with the top three further assessed using 19 resampling techniques. The KNN-CatBoost model achieves the best performance, with an AUC of 97.94%, Recall of 95.91%, and F1-Score of 87.40%. Results show that this approach outperforms previous models, providing an effective solution for fraud detection (Alfaiz and Fati, 2022).

**Comparison of all studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study**  **(authors)** | **Methods Used** | **Dataset** | **Challenges Addressed** | **Best Model & Performance** |
| **Mienye & Sun (2023)** | Hybrid feature selection (Information Gain + Genetic Algorithm), Extreme Learning Machines (ELM) | Real-world credit card transactions | Redundant/irrelevant features, Imbalanced data | Sensitivity: 0.997, Specificity: 0.994 |
| **Aburbeian & Ashqar (2023)** | Random Forest (RF) with SMOTE, Hyperparameter tuning | Credit card transactions dataset | Imbalanced dataset | Accuracy: 98%, F1-Score: 98% |
| **Awoyemi, Adetunmbi & Oluwadare (2017)** | Naïve Bayes, k-Nearest Neighbors (k-NN), Logistic Regression | European credit card transactions (284,807) | Fraud pattern evolution, Data skewness | k-NN: Accuracy 97.69% |
| **Alarfaj et al. (2022)** | Decision Trees, Random Forest, SVM, Convolutional Neural Networks (CNNs) | European benchmark dataset | Data imbalance, High false alarm rates | CNN: Accuracy 99.9% |
| **Alfaiz & Fati (2022)** | 66 ML models, Two-stage evaluation, KNN-CatBoost | European credit card dataset with stratified K-fold validation | Selecting best ML model, Performance optimization | KNN-CatBoost: AUC 97.94%, Recall 95.91%, F1-Score 87.40% |

**Reference list**

Alarfaj, F.K., Malik, I., Khan, H.U., Almusallam, N., Ramzan, M. and Ahmed, M. (2022). Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. *IEEE Access*, 10, pp.39700–39715.

Alfaiz, N.S. and Fati, S.M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics*, 11(4), p.662.

AlsharifHasan Mohamad Aburbeian and Ashqar, H.I. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. *Lecture notes in networks and systems*, pp.605–616. doi:https://doi.org/10.1007/978-3-031-33743-7\_48.

Awoyemi, J.O., Adetunmbi, A.O. and Oluwadare, S.A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI)*, pp.1–9.

Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), pp.602–613. doi:https://doi.org/10.1016/j.dss.2010.08.008.

Bolton, R.J. and Hand, D. (2002). *Statistical fraud detection: A review*. [online] Quality Engineering. Available at: https://www.semanticscholar.org/paper/Statistical-fraud-detection%3A-A-review-Bolton-Hand/d9665992ee36699b8ae4a2e2294552cd4be9003a.

Mienye, I.D. and Sun, Y. (2023). A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection. *Applied Sciences*, [online] 13(12), p.7254. doi:https://doi.org/10.3390/app13127254.