

Research Methodology in Computer Science - Assignment 2

Using Hybrid Machine Learning Models to Improve Fraud Detection in Financial Transactions

Statement of the Problem: The research falls under the purview of machine learning and financial cybersecurity, with a particular emphasis on identifying fraud in online transactions.

Financial fraud is becoming an increasing worry, with worldwide losses totaling over \$42 billion in 2023 (Association of Certified Fraud Examiners, 2024). Traditional rule-based fraud detection systems produce a lot of false positives and are unable to keep up with changing fraud trends (Ngai, 2021). Machine learning (ML) models, such as neural networks and random forests, have improved detection rates but continue to struggle with imbalanced datasets and real-time processing (Pozzolo, 2018)

Research Objectives: Banks aren't just losing money to fraud—they're hemorrhaging cash just investigating it. For every \$1 lost to fraud, they spend \$4.50 chasing ghosts (LexisNexis, 2023). Imagine that on your expense report. Better fraud detection could chop losses by 30% and maybe, finally, restore a little faith in digital banking (Security, 2024).

Existing Solutions:

- Supervised learning? Sure—XGBoost, logistic regression, you name it (Bhattacharyya, 2022)
- Deep learning models like LSTM try to catch fraud as it happens.
- The data's super skewed (fraud is 0.1% of all transactions—like finding a unicorn in a haystack), so false positives run rampant.
- Systems can't keep up with fast-changing scams. It's like fighting a shapeshifter.

Research justification:

Mixing supervised learning with anomaly detection in a hybrid ML model could be a game changer. Fewer false alarms, more actual fraudsters caught.

Build a hybrid model with XGBoost (because it's a beast) and Isolation Forest for anomaly detection.

Use real datasets (think IEEE-CIS Fraud Dataset) to actually test if this thing works, not just theory-crafting.

Judge it by accuracy, recall, F1 score—stats that actually matter. - And, of course, see if it's any good compared to the usual suspects: deep neural nets, random forests, etc.

Bonus round: see if it holds up in real-time, or totally chokes.

Literary Review

Conventional Techniques for Fraud Detection: Threshold notifications, for instance, are examples of rule-based systems, which have a high number of false positives (West & Bhattacharyya, 2021).

Nonlinear fraud patterns pose a challenge for statistical models like Logistic Regression (Dal Pozzolo et al. , 2018).

The Use of Machine Learning to Detect Fraud

Supervised Learning:

The Random Forest method achieves 85% accuracy but is prone to overfitting (Bhattacharyya et al. , 2022).

While XGBoost enhances performance, it also necessitates a lot of feature engineering

Unsupervised Learning:

The Isolation Forest identifies previously unidentified fraud trends but is not very accurate (Liu et al. , 2020).

Conclusion

To fill in the shortcomings of current fraud detection systems, this study suggests a hybrid model. The goal of the research is to lower false positives while preserving high accuracy by combining supervised learning and anomaly detection.

- Anticipated Contributions
- A unique hybrid ML model for identifying fraud.
- Comparing it to conventional techniques.
- Verification of actual real-time applicability.

References

Bhattacharyya, 2022. [Online].

Ngai, E. W. T. H. Y. W. Y. H. C. Y. & S. X., 2021. s.l.: s.n.

Pozzolo, D., 2018.