Transaction Fraud Detection (Combination of Nature Inspired Computing and Machine Learning)

Rostislav Kirsanov
Innopolis University
r.kirsanov@innopolis.university

Dzhamilia Fatkullina
Innopolis University
d.fatkullina@innopolis.university

Sofia Pushkareva
Innopolis University
s.pushkareva@innopolis.university

I. Goal

Online transaction fraud poses a significant threat to financial security, resulting in substantial consequences for both financial institutions and consumers. Banks face reputational damage and direct financial losses, while affected individuals endure not only monetary harm but also psychological distress. As the financial sector evolves to adopt innovative consumer engagement strategies, fraud detection has emerged as a critical area of research. The project objective is to develop an adaptive, accurate, and efficient system capable of addressing the dynamic challenges posed by sophisticated fraudulent activities in digital transactions.

II. Method

Classical machine learning techniques methods will be used in combination with nature-inspired algorithms to create a hybrid model:

- Genetic Algorithms (GA) based on the principles of natural selection and evolution. It is used to optimize hyperparameters of machine learning models.
- Swarm Intelligence (SI) includes methods such as Particle Swarm Optimization (PSO) algorithm and Ant Colony Optimization (ACO). It is used to find optimal solutions in classification problems.

III. Data

A. Dataset explanation

The project employs the publicly available IEEE-CIS Fraud Detection Dataset from Kaggle as its primary data source for model development and evaluation.

Key stages of data utilization:

1) Exploratory data analysis (EDA):
   - Analyzing feature distributions, correlations, and anomalies.
   - Processing outliers and missing values to ensure data integrity.
   - Structural analysis to guide preprocessing and modeling.
2) Machine Learning Model development:
   - Training classifiers (Random Forest, Gradient Boosting, SVM) for fraud detection.
   - Optimizing hyperparameters to improve accuracy and robustness.
3) Nature-Inspired optimization:
   - Applying Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Artificial Immune Systems (AIS) to refine model parameters and enhance performance.
4) Performance evaluation:
   - Metrics: accuracy, recall, F1-score.
   - Benchmarking against existing methods to assess competitiveness.

In addition to the primary dataset, supplementary resources, including pre-implemented fraud detection methodologies, will be leveraged. This will enable:

- Exploration of existing approaches and their effectiveness in fraud detection.
- Comparative analysis of the proposed model's performance against established solutions.

B. Accessible data link

https://www.kaggle.com/c/ieee-fraud-detection/overview

IV. Timeline

Week 1: Data analysis and preprocessing.
Responsible: Rostislav Kirsanov, Dzhamilia Fatkullina.
Week 2: Selecting and tuning algorithms.
Responsible: Sofia Pushkareva.
Week 3: Model Training.
Responsible: Sofia Pushkareva.
Week 4: Testing and Assessment
Responsible: Rostislav Kirsanov, Dzhamilia Fatkullina.
Week 5: Report and presentation preparation
Responsible: All participants of the group

V. References

Lecture materials are used as references, as well as articles and resources on nature-inspired computing and fraud detection techniques.