Practical work 4

Analyse & compare classification models

Dataset

Dataset: Credit Card Fraud Detection
(https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud)

Goal

Develop a machine learning model for detecting fraudulent transactions using Logistic Regression, K-Nearest Neighbours (KNN), SVM and Decision Tree classifier, analysing transaction data to improve fraud detection accuracy while minimizing false positives.

Implementation and Results

Model Selection & Training:

- Implement the selected classifier for fraud detection.
- Optimize hyperparameters to improve model performance.

Performance Evaluation & Visualization

- Evaluate the model using performance metrics such as accuracy, precision, recall, specificity, and F1-score.
- Visualize the model's effectiveness using a confusion matrix.
- Apply t-SNE (t-Distributed Stochastic Neighbour Embedding) to visualize high-dimensional data in a 2D space. Interpret the results to identify key patterns in fraudulent transactions.

Conclusions

- Compare the model performances with other classifiers.
- Provide recommendations for improving fraud detection in real-world applications.