Fraud Detection

Zahra Khalafi
UIN: 655096362
Zkhala3@uic.edu

Github link for this project:
https://github.com/zahrakhalafi/Project_CS_418.git

Problem statement:

Developing a fraud detection system using machine learning techniques that can effectively identify fraudulent transactions

Fraudulent activities in financial transactions cause significant financial losses to individuals, businesses, and financial institutions. Moreover, fraud can damage the reputation and trustworthiness of financial institutions.

fraud is a pervasive problem that affects everyone, from individuals to large corporations. A robust fraud detection system can significantly reduce the risk of financial losses

I chose this topic because I can use several machine learning techniques in predicting the fraud transactions.

My hypothesis is that by using supervised machine learning algorithms, we can develop a highly accurate fraud detection system that can detect fraudulent transactions in real-time with high precision and recall.

Data

The data that I will be using is Credit card fraud detection - Date 25th of June 2015

I was able to access the data through this link:

https://www.openml.org/search?type=data&sort=runs&id=1597&status=active

I believe I can do several machine learning algorithms on this dataset on a reasonable amount of time.

Quality Name	Value
number of instances	284807
number of features	31
number of classes	2
number of missing values	0
number of instances with missing values	0
number of numeric features	30
number of symbolic features	1

Project Plan

Next steps: We plan to perform exploratory data analysis (EDA) to gain insights into the data and identify any data quality issues or imbalances. We will then preprocess the data by scaling the features and addressing any missing or outlier values. Next, we will split the data into training, validation, and testing sets, and experiment with different machine learning algorithms, such as logistic regression, decision trees, and random forests, to determine the best approach for this problem. We will evaluate the model's performance using metrics such as precision, recall, F1-score, and AUC-ROC. Finally, we will use the best-performing model to make predictions on new, unseen data.

End result: We envision a machine learning model that can accurately predict whether a given credit card transaction is fraudulent or not, based on the transaction's features. We hope to achieve a high level of accuracy and precision in our predictions, while minimizing false positives and false negatives.

Techniques: We plan to use various techniques, such as feature engineering, feature scaling, model selection, and cross-validation, to preprocess the data and develop an accurate machine learning model.

System: We envision our system to be a static model that takes in a transaction's features as input and outputs a prediction of whether the transaction is fraudulent or not.

Progress report: We hope to have an initial understanding of the strengths and weaknesses of our approach, and any challenges that we may encounter in the next stages of the project.