**Ziel**

Erkennen von betrügerischen Kreditkartentransaktionen basierend auf vorhandenen Transaktionsdaten.

**Datenbeschaffung**

Kaggle Datasets

<https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud/data>

**Exploratory Data Analysis**

Ziel: Muster erkennen und die Struktur der Daten verstehen.

Verteilung der Klassen

* Betrug ist oft sehr selten (Imbalanced Dataset)
* Verteilung jedes Features prüfen
* Korrelation zwischen Merkmalen prüfen
* Duplikate entfernen

**Visualisierungen**

* Histogramme
* Boxplots
* PCA (Principal Component Analysis) evtl.

**Datenbereinigung**

* Fehlende Werte
* Ausreißererkennung: Insbesondere bei numerischen Merkmalen, die extrem abweichende Werte haben können.
  + Methode: Z-Score oder IQR-Methode (Interquartilsabstand).

**Daten-Transformation**

* Skalierung der Daten
* Feature Engineering: evtl. neue Features aus bestehenden ableiten, z.B. Verhältnis von Beträgen oder zeitliche Deltas

**Train-Test-Split**

beim Splitten ein gleichmäßiges Verhältnis zwischen den Klassen

Modellauswahl

* Logistische Regression
* Random Forest
* XGBoost
* Support Vector Machines

Achtung bei oft unausgewogene Klassen:

* Oversampling
* Class Weights

**Modelltraining mit Train-Daten**

**Modellbewertung**

* Vorhersagen auf Testdaten
* Accuracy
* Präzision, Recall, F1-Score
* Confusion Matrix
* ROC-AUC-Kurve

**Modelloptimierung**

* Hyperparameter-Tuning
* Cross-Validation

**Finale Evaluierung**

Ein Modell als Benchmark und danach ein CNN machen.

Zum Schluss aufzeigen was besser geht