

Dataset bereitgestellt von Kaggle:

<https://www.kaggle.com/ealtman2019/credit-card-transactions>

Transformiert und bereinigt durch mich

Tabelle ccfraud in Schema uccfraud

Spalten		
userid	integer	Nutzerkennung
card	integer	Kreditkartennummer
ts	timestamp	Zeitpunkt der Transaktion
amount	float	Betrag der Transaktion
use_chip	varchar	
mname	integer	Verkäufer, als Zahl kodiert
mcity	varchar	
mstate	varchar	
zip	varchar	
mcc	integer	
err1	varchar	
err2	varchar	
err3	varchar	
is_fraud	varchar	'Yes' oder 'No'
ccfid	integer	Anonymer Schlüssel

Aufgabenstellung

- Analyse der Daten mit SQL
- Visualisierung mit Pandas Dataframes
https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html
- Präsentation der Analyse im LV-Termin
- Entwicklung eines Klassifikators mittels SQL-Abfragen
 - Materialisieren Sie ggf. Teilmengen der Daten im eigenen Schema
<https://www.postgresql.org/docs/14/sql-createtables.html>
 - Unterteilung in Trainings- und Testdatenmenge
 - Definition einer komplexen Abfrage, die als Ergebnis 'Yes' oder 'No' abgeliefert
- Analyse der Qualität des Klassifikators
 - Anwendung des Klassifikators (d.h. der Abfrage) auf der Testdatenmenge und Erstellung Ergebnistabelle mit Spalten "true" und "pred"
 - Berechnung Confusion Matrix
 - Berechnung Accuracy, Recall, Precision, F1