

# KLASSIFIZIERUNG MITTELS LOGISTISCHER REGRESSION

---

**Kurzbeschreibung:** Hauptziel des Projektes ist es, anhand verschiedener Variablen wie Alter, Geschlecht, Passagierklasse, etc. die Überlebenswahrscheinlichkeit einer Person der versuchten Titanic-Überfahrt zu schätzen. Dabei wird hauptsächlich auf die Logistische Regression zurückgegriffen.

**Datensätze:** `titanic.csv`, `diabetes.csv`

---

**Projektziel:** Das Hauptziel des Projektes ist es, mittels logistischer Regression und anhand verschiedener Prädiktoren zu klassifizieren, ob ein Passagier die Unglücksfahrt der Titanic überlebte oder nicht. Mittels des `Pipeline`-Moduls können dabei die Inputvariablen wie die Wahl der Loss-Funktion optimiert werden. Außerdem soll die Qualität der Klassifikation anhand verschiedener Gütekriterien wie des AUC-Scores evaluiert und grafisch dargestellt werden (ROC-Kurve).

Zusätzlich kann diese Klassifikationsuntersuchung noch mit dem Datensatz `diabetes.csv` durchgeführt werden.

**Informationen zum Datensatz:** Der Datensatz `titanic.csv` enthält neben der Zielvariablen, die angibt, ob ein Passagier das Titanic-Unglück überlebt hat oder nicht, eine Reihe von Informationen über den jeweiligen Passagier, wie Alter, Geschlecht, Reiseklasse, Anzahl Angehöriger etc.

Der Datensatz `diabetes.csv` enthält verschiedene medizinische Prädiktoren, um über das Vorliegen von Diabetes zu entscheiden.

**Verwendete Methoden:** Logistische Regression, ROC-Kurven, Parameteroptimierung mit `GridSearchCV`

**Eigene Implementierung:** Implementieren der logistischen Regression und eines Moduls, welches eine ROC-Kurve ausgibt.

**Hilfreiche Python-Pakete:**

- `sklearn.linear_model` (LogisticRegression),
- `sklearn.pipeline`,
- `sklearn.model_selection` (`train_test_split`, `GridSearchCV`),
- `sklearn.metrics` (`accuracy_score`, `roc_curve`, `auc`),
- `scipy.optimize` (`minimize`)