

Mandatory Assignment 2

Classification & Clustering

Markus Kinn

Classification:

Dataset 1:

For the first dataset I chose to continue with the titanic dataset from the first mandatory assignment. The goal is the same as before; predict whether a person was going to survive the titanic disaster or not.

The has dataset has 12 columns and 891 rows.

The final dataset (after preprocessing) contains 15 columns:

Survived(the ground truth), Sex, FamilyMembersCount, Pclass_1, Pclass_2, Pclass_3, Embarked_C, Embarked_Q, Embarked_S, AgeGroup_0, AgeGroup_1, AgeGroup_2, AgeGroup_3, AgeGroup_4, AgeGroup_5.

I got the best result with: XGBoost

Dataset 2:

The second classification dataset I used was: Heart Attack Analysis & Prediction Dataset. (<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>). This dataset is used to predict if a person is likely to have a heart attack given current metrics.

The has dataset has 14 columns and 303 rows.

There are 10 columns (including ground truth):

Age: age of patient, Exng: exercised induced angina, Ca: number of major vessels, Cp: Chest Pain type, Trtbps: resting blood pressure, Chol: cholestoral, Fbs: fasting blood

sugar, Rest_ecg: resting electrocardiographic results, Thalach: max heart rate achieved, Target: 0 = lower chance of heart attack and 1 = higher chance of heart attack

I got the best result with: Support Vector Machine

Clustering:

Dataset 1:

For the clustering part of the assignment I chose to use the same clustering dataset from the first mandatory assignment: Airline Passenger Satisfaction (<https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>). This dataset is usually used as a classification dataset with the ground truth: Satisfaction, but for this task the ground truth has been removed.

The dataset has 25 columns and 103904 rows.

Due to the dimensionality, I ended up removing 90.000 rows for both the Birch algorithm and Agglomerative.

I got the best result with: Agglomerative clustering.