Mandatory Assignment 2

Classification & Clustering

Markus Kinn

**Classification:**

**The 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 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.

**The 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 columns are:

Age,

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, 1 = higher chance of heart attack