

# Hausarbeit im Modul „Data Science und Machine Learning“ WS 23/24 - Teil 1

## Churn Prediction

According to the World Health Organization, strokes are the second leading cause of death worldwide, accounting for approximately 11 percent of all fatalities. This dataset can be employed to create a machine learning model for predicting strokes. The dataset includes various demographic details about patients, as well as information on whether each patient had a heart attack. The objective of this task is to develop a machine learning model for stroke prediction.

The data set contains the following attributes:

<b>Id</b>	Unique id of each observation
<b>gender</b>	Indicates the gender of the patient
<b>age</b>	Indicates the age of the patient
<b>hypertension</b>	Indicates if the patient suffered from hypertension
<b>heart_disease</b>	Indicates if the patient suffered from a heart disease
<b>ever_married</b>	Indicates the marriage status of the patient
<b>work_type</b>	Indicates the work type of the patient (“children”, “Gov”, “Never_worked”, “Private” or “Self-employed”)
<b>residence_type</b>	Residence type of the patient (rural or urban)
<b>avg_glucose_level</b>	average glucose level in blood
<b>bmi</b>	body mass index if the patient
<b>smoking_status</b>	Indicates the smoking habits of the patient (“formerly smoked”, “never smoked”, “smokes” or “unknown” if no data exist on this information)
<b>stroke (dependend variable)</b>	Indicates if the patient suffered a stroke

1. Apply three different simple machine learning models, which we discussed in lecture 02. Describe your approach, interpret the performance, and justify (based on the results) which machine learning model you would choose.

- Note:

- No hyperparameter tuning (lecture 04) is expected.
1. Select the machine learning model that you have chosen in task 1. Now try optimizing your results by applying the advanced concepts (lecture 04) and analyze if the model is overfitted or underfitted (lecture 04). Use cross-validation (lecture 04) when evaluating your results. Describe your approach, interpret the performance, and justify which machine learning model you would choose.
    - Note:
      - Use hyperparameter tuning (lecture 04).
      - Use cross-validation (lecture 04) when evaluating your results.
      - Explain if your model is over or underfitted (lecture 4).
  2. Apply one ensemble learning technique (lecture 05) to see whether those techniques lead to a better performance than in task 1 and task 2. You should also apply the advanced concepts (lecture 04) to improve the model and analyze if the model is overfitted or underfitted (lecture 04). Describe your approach, interpret the performance, and justify the final machine learning selection.
    - Note:
      - Use hyperparameter tuning (lecture 04).
      - Use cross-validation (lecture 04) when evaluating your results.
      - Explain if your model is over or underfitted (lecture 4).