Student:
se21m024

Software version:
Python 3.8.5

Operating System:
Windows 10

Algorithm used for prediction:
The best performing setting was the k-NN algorithm with k=2, which accomplished an accuracy of 0.954 and an F1 measure of 0.936 on the learning set and was therefore chosen for the actual prediction.

Processing applied to the data before training/predicting:
The non-numeric features in the data sets have been mapped to numeric values before further processing. The learning data was split into the input features (everything except of the 'ID' and 'stroke' column) and the target feature (the 'stroke' column).

Then the learning data was split into 2/3 trainig set and 1/3 test set. With this split 4 different algorithms with different parameters were trained:

- k-NN (1-NN)
- k-NN (2-NN)
- k-NN (3-NN)
- Decision Tree (max features: None)
- Decision Tree (max features: sqrt)
- Decision Tree (max features: log2)
- SVM (SVC classifier)
- Random Forests (num trees: 10, max features: sqrt)
- Random Forests (num trees: 10, max features: log2)
- Random Forests (num trees: 100, max features: sqrt)
- Random Forests (num trees: 100, max features: log2)

The best performing algorithm on this test split was than retrained with the whole learning data and used to create the prediction on the test set.