Comparing Decision Tree and K-Nearest Neighbor for Cerebral Stroke Prediction

by Stefan Diener

1) Description and Motivation of the Problem

In 2020, England recorded cerebrovascular diseases are the third leading cause of death [1]. A cerebral stroke is a subtype of these diseases in which the blood supply to part of the brain is interrupted, depriving brain tissue of oxygen and nutrients and causing the death of brain cells within minutes. Therefore, predicting whether a person will have a cerebral stroke can be of great importance to public health, as it can help ensure adequate preparation and quick action in an emergency. The goal of this project is to train and compare the ability of a Decision Tree and a K-Nearest Neighbor classifier to predict whether a person will suffer a heart stroke.

2) Initial analysis of the data set including basic statistics

Target Variable

Correlation Matrix

Density of Continous Variables

%-True Values of Binary Variables

Countplot of Categorical Variables