

DSC Exam Project: Chronic Kidney Disease detection

Simon Schønberg Bojesen - cph-sb339@cphbusiness.dk
Frederik Blem - cph-fb114@cphbusiness.dk
Martin Høigaard Cupello - cph-mr221@cphbusiness.dk
Kenneth Leo Hansen - cph-kh415@cphbusiness.dk

June 2021

1 About

The problem that we want to solve is that many cases of Chronic Kidney Disease go undetected around the globe. We will solve this problem by getting information from a blood test, then use machine learning to train a model that can correctly classify if a person has chronic kidney disease or not.

Our goal is to create a model that has an error rate of under 5%, or we won't be able to use the model for actual medical application. We have developed a model that helps predict chronic kidney disease. It was developed using a dataset found on: <https://www.kaggle.com/mansoordaku/ckdisease/>.

The dataset consists of data gathered in India, with 25 different features that may determine if a person has chronic kidney disease or not. First we clean the data, convert text values to numeric values and replace outliers with NaN values, and then in the end replace all NaN values with means. Then we take a sample of 25 entries and put it aside for later validation. Then we clean data, we train our models (KNN, Gaussian Naive Bayes, Multinomial Naive Bayes and Decision Tree), using a split of 0.3/0.7. From these results, we determine that the Decision Tree is most accurate with a score of 99%.

To further improve upon our model, we looked at a heatmap of our data so we could implement feature engineering. By looking at the heatmap we decided to cut away the least significant data columns, but our scores got a little worse: 98% for Decision Tree, where it was 99% before, still well within our acceptance criteria. To test the models against real data we took our validation sample from the original dataset, where the Decision Tree model was able to predict all 25 correctly.

We then exported our best performing model which was the original decision tree model to a file, ready for future use.

2 Project files

Project files, dashboard and readme can be found here: https://github.com/MHR1985/DSC_ExamProject