Individual report

Aron Gaden (59023)

Project for Data Mining and Analysis
Summer Term 2020
Group 09

For the project I worked on linear regression, mostly the regression analysis together with Rebekka.
Following that I focused on task 4, fuzzy clustering with anomalous patterns. I applied both of the mentioned algorithms on our (pre-processed) chosen data set as well as the internal validation indices FPC and XBI. I also wrote the whole part of the paper concerning this exercise. We formulated the conclusions together. Rebekka worked on linear regression analysis as well as writing that part of the paper. Annemarie solved task 3 and wrote that part of the paper, as well as pre-processing the data. It was extremely important for us as a group to keep close communications and exchange our results to discuss them.
Considering the parts I focused my work on, I learned most about linear regression and (fuzzy) clustering methods. It was interesting to learn about the intricacies of linear regression, because even though the principle is quite simple, the analysis methods can become quite sophisticated to find out the details of a relationship between 2 (or more) attributes of a data set. In the past, I had a very good opinion of clustering algorithms. The project led me to curb my enthusiasm, so to speak, and really showed me the limitations of clustering algorithms, because clustering did not produce worthwhile results for our data set.
In general, I learned that data mining consists as least as much of validating the results as producing them. In that way it is quite similar to test-driven development in software engineering. Even the seemingly greatest results are of no worth if they can't be validated properly.