Data scientist tasks

Portfolio assignment 102

Avans Hogeschool

March 9th, 2022

Author:

Frank Gabršek – 2171626

Lecturer:

Erco Argante

# Tasks

Data scientist is a very broad job description and includes many responsibilities, but there exist certain basic tasks a data scientist must perform that prove relevant within all business domains.

## Gathering

A data scientist must be able to gather and compile data in such a way that a useful and relevant analysis can be made for a client. In the portfolio this was done in assignment 4, when a dataset was chosen, and care was put into choosing a set that included enough data and various types of data.

## Investigating

A data scientist must perform various descriptive and predictive analyses on their data to discover correlations. This is key in understanding the data and paramount to creating meaningful predictive models. As such, this forms the core task for a data scientist. These tasks were performed in nearly all portfolio assignments. For example: finding correlations using the Pearson correlation coefficient and Chi2 test, investigating anomalies using bar plots and contingency tables and using knowledge about correlations to create clustering and predictive models.

## Drawing conclusions

While investigating data independently can be fun and interesting, data science is primarily a scientific field and as such its main purpose is to answer questions. Just like any other science, in order to find answers, conclusions must be drawn from the results that the research into the data yielded. A data scientist must interpret their results correctly, as to not reach the wrong conclusion when data is misleading. The portfolio assignments offered some practice in this regard. The meaning of confidence intervals, Pearson coefficients, Chi­2 results and the accuracy and errors of predictive trees were explained, questioned and tested.