Data Analytics 2 - Lessons Learned

Tuomo Kareoja

Version Number	Changes	Date
1.0	Finished text and plots	18.09.2019

Brand Preference Prediction

Lesson 1: Modelling can be used to fix data problems

There are multiple ways to deal with missing values
    We can also think about missing value replacement as a completely separate project

Lesson 2: You have to rethink your metrics if multiple models are performing extremely well

If model accuracy is approaching 100 %, just comparing the accuracy numbers can be difficult. AUC curves look almost completely overlapping
    Small changes can still mean that one model makes only half as many mistakes
    Compare the metrics as percentages of each other
    Zoom to plots

Predicting future sales

Lesson 1: Data that seems only somewhat unrelated to the task at hand can still help

We were interesting in only a small group of products, but still it was better to use the whole dataset for the predictions, even though the data included products that seemed very different from the groups of interest
    Possibility of leveraging a different dataset to predict somewhat related metric

Lesson 2: Sometimes hyperparameter tuning can make the model perform radically better

Usually not but sometimes this seems to happen
    SVM

Market Basket Analysis

Lesson 1: Always keep the business question in mind

Without the business goal you would only concentrate on the learning the packages and creating strong rules

What are actually interesting rules for the business context

Does the advised approach actually best serve the business context

In real business context you should talk with the manager and ask for clarifications and be clear with the difficulties and limitation you are facing

Lesson 2: Ignore the instructions if they seem confusing

Instead of trying to follow the instructions when they seem clear but make little sense you should ignore them and keep in mind the business goals

Lesson 3: Be prepared to change your approach completely

It might be that you end spending a long time with some kind of analysis path, but then it becomes clear that this path is somehow problematic you must be ready to abandon the previous effort and try something new.