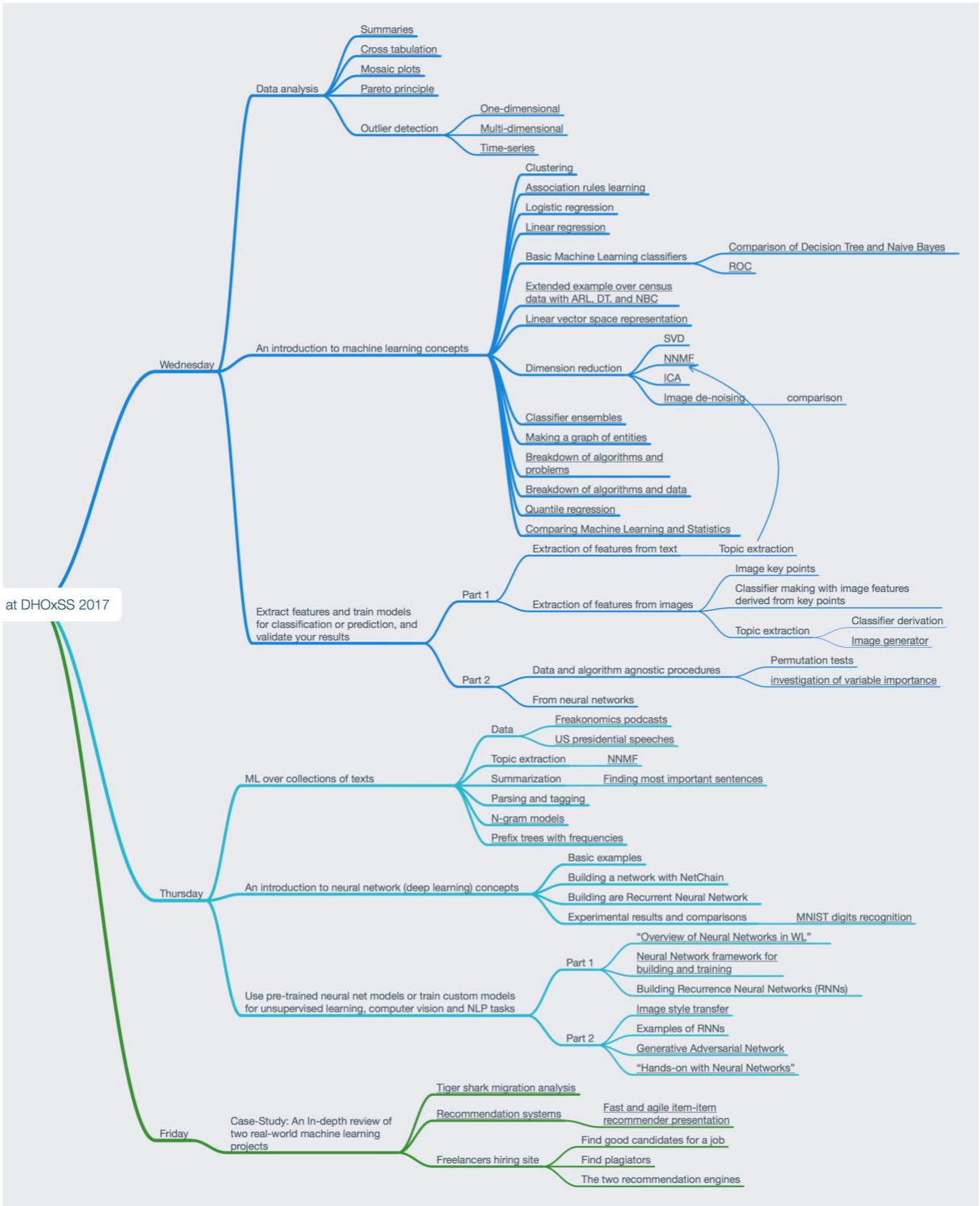
Oxford Summer School 2017 -- Data Science track



The data science workshop was very intensive. First two days, we trained on Wolfram Mathematica software. I did not have prior experience with Wolfram Mathematica before and I found the program interesting. Mathematica is a computation program well known for technical computing such as machine learning and it uses Wolfram language, that is a multi-paradigm, functional language. Below is a sample of Mathematica code, where a dataset is divided into training and testing data and printing the summary.

```
mlData = MapThread[Rule, {Training, Tlabels[[All, 1]]}];
```

Separating the data at 0.75:

```
SeedRandom[123]
{trainingData, testData} =
  TakeDrop[RandomSample[mlData], Floor[0.75 * Length[mlData]]];

meanRow = Mean[trainingData[[All, 1]]];
RecordsSummary[meanRow]
```

```
1 column 1
Min      -1131.99
1st Qu   8.22631
Median   14.6834
3rd Qu   20.9983
Mean     25.0205
Max      83377.9
```

Especially the sessions on understanding Supervised learning vs. Unsupervised learning vs. Reinforcement learning was very interesting for me. Furthermore, a learnt about an unsupervised machine learning technique called Associate rules learning, also known as "market basket analysis". The technique can be summarized as follows:
Given a set of baskets (or transactions) T:

1. Find μ-frequent sets -- sets of items that appear together more than $\mu \mid T \mid$ times

2. Extract rules antecedent $\rightarrow$ consequent from each μ-frequent set

2.1. apply the measures confidence, lift, leverage, conviction, and others;
2.2. order and filter the rules.

For instance, after examining 5K shopping charts within a day in K-market, we can conclude that people who buy beer and chips together, 80% likely to buy salmiakki candy, as well. Thanks to this method and the summer school, I utilized this technique during my PhD studies and submitted publications, that are part my PhD dissertation.