**Soccer Match Classification**

A soccer match results in one of three outcomes - a home win, away win, or draw.

Using publicly available game data, how good can we get at classifying outcomes?

**Sample Data Source:**

http://www.football-data.co.uk/englandm.php

**Pros & Cons:**

+

Plenty of free, compiled data sources available online

Data is broad enough to enable multiple machine learning techniques to be tried

Naturally leads into a more complex soccer results prediction project

High personal interest

-

Not original

**Aircraft Accident Causes**

The US National Transportation Safety Board makes their final accident investigation reports available online (1965 to present). There have been significant maintenance technology developments over the past 40 years which, in theory, should have reduced mechanical failure accidents.

How have accident causes changed over time?

**Sample Data Source:**

https://www.ntsb.gov/investigations/AccidentReports/Pages/AccidentReports.aspx

**Pros & Cons:**

+

Data available in single location

Career related

-

Lots of web scraping required to get data organized

Analysis probably requires NLP Techniques which are a bit beyond me right now

**Electrical Power Demand Prediction for Ontario**

If you can accurately predict the 5 annual occurrences of peak power demand in Ontario, there are opportunities to reduce annual electricity costs by over 50% per year.

Using publicly available data, how good can we get at predicting peaks?

**Data Sample:**

http://www.ieso.ca/Power-Data/Data-Directory

**Pros & Cons:**

+

Data available online - provincial forecasting, historical data, weather

Commercial applications

Career related

-

Need to build a fair bit of domain knowledge to make sense of historical data

Lots of web scraping and merging