

F21DL: Data Mining and Machine Learning

Coursework A

Andrew Beveridge
ab441@hw.ac.uk

11/10/2013

1 Sports Data Repository

This data repository is probably not a very high quality data set, and probably doesn't really fit the bill as a "repository", since it is really just the last 8 years' worth of match scores for a (surprisingly popular!) pub game down in southern england called "Aunt Sally". However, I personally find it rather interesting as I have a vested in it, in that I recently redeveloped the website <http://abingdonauntsally.com/> and have been working with this data (parsing it into the database, running very basic statistics on parts of it) for weeks. I couldn't resist passing it to you to see what you thought!

<http://abingdonauntsally.com/oldresults/>

2 Time Series Data Repository

This is a data repository which you have probably seen before, but I found it very interesting because of the direct relevance and interesting trends which may be able to be pulled out of it. It's the World Health Organisation's set of data from 1961 to present day, on "Alcohol and Health: Levels of Consumption". I reckon there are a whole load of correlations worthy of a Daily Mail article which can be found in here, and since scotland is such an alcohol-plagued country, perhaps we can find some useful insight into which countries to pick as role models! <http://apps.who.int/gho/data/node.main.A1022>

3 Anything Interesting Data Repository

Given that I'm a web geek, I wanted to find something specific to the internet. This website, titled "Internet-Wide Scan Data Repository" seems to be a public repository of data collected by scanning publicly accessible hosts on the internet. I like the idea of this as datasets such as the "HTTPS Ecosystem Scans" could provide interesting insights into adoption of secure websites and widespread use of SSL encryption across the internet. Either way, it's a big, fairly high quality data set with hopefully interesting data. I figured it was a wild card, I'm interested to hear what you think of it. <https://scans.io/>