

Coursework A

Joseph Davidson - 071729468

1 Introduction

The purpose of this coursework was to acquaint myself with the UCI Machine Learning repository as well as other ones scattered around the Internet. With this in mind, I have found these four other ones. (One is a 'proper' financial repository, whilst the other is just interesting).

- *Economagic*
- The Executive PayWatch Database
- The Time Series Data Library
- The Cooperative Association for Internet Data Analysis (CADIA)

2 *Economagic*

URL: <http://www.economagic.com/>

Economagic.com is a financial time series data repository that mostly features data from the US government, although Australia, Japan and Europe have a few entries. The data is diverse and, due to the nature of the world economic markets, changing regularly. A quick exploration of the daily exchange rates (updated at noon) offers links to the history of a given currency against the U.S. Dollar. It also appears that you can purchase a membership to the site. This membership has 'levels' that presumably allows you to deeper archives depending on your level. The site has many sources including: The Federal Reserve, the Census Bureau and the Department of Energy.

3 PayWatch

URL: <http://www.aflcio.org/corporatewatch/paywatch/ceou/database.cfm>

The American Federation of Labor and Congress of Industrial Organizations is a union organisation that recognises the plight of the common working man by keeping a record of the the salaries of executives of many major public companies. It also has some interesting comparisons between the executive, a common worker and the president of the USA (who makes \$3.20 per minute compared to the Boeing CEO who raked in \$155 per minute last year). The site offers up more fun stats which demonstrate its clear political slant, this site could more be classified as a use for data mining rather than an actual data repository. It cites www.salary.com as its source, but investigation offers up no data sets that can be found.

4 The Time Series Data Library

URL: <http://robjhyndman.com/TSDL/>

This extensive repository has time series data on topics as diverse as finance to tree rings (which has a surprising number of files in it). All of the data is free to access and each file (presented as .DAT) has a quick line of description as well as a source citation. There is also ancillary information relating to forecasting and time series, such as useful books and links to organisations who love time series forecasting as well as links to other time series data repositories.

5 CADIA

URL: <http://www.caida.org/data/>

CADIA is a data repository for the sharing of data to do with networking topics such as: routing, topology, traffic and more. The dataset list is not massive, but is diverse with security, worm summaries and topological statistics in its list. A few of the sets have public access, but many of them need access to be granted. This screening is probably to stop large corporations making use of the data for nefarious marketing purposes instead of helping do actual research. The public topological data sets are particularly in depth (one going monthly from may 2005 to the present).