

Time Series Analysis and Forecasting of In-Play Odds on a Betting Exchange

Readme Documentation

R System Requirements:

- Downloaded from: <https://cran.r-project.org/mirrors.html>
- Setup File: "R-3.2.2-win.exe" (65MB)
- *Testing is primarily on versions of Windows currently supported by Microsoft, mainly 64-bit Windows 7 and Server 2008
- Installation requires 150MB of disk space
- Memory Limit: We only test on versions of Windows currently supported by Microsoft, mainly 64-bit Windows 7 and Server 2008
- Further Requirements see: <http://ftp.heanet.ie/mirrors/cran.r-project.org/>

WEKA System Requirements:

The system requirements for WEKA are primarily based on Java, as illustrated by the screenshot below, taken from the University of Waikato website:

Requirements

The following matrix shows what Java version is necessary to run a specific Weka version. It also specifies when a changeover happened, listing date and **Subversion** revision (the number prefixed with *r*).

		Java			
		1.4	1.5	1.6	1.7
WEKA	<3.4.0	X	X	X	X
	3.4.x	X	X	X	X
	3.5.x	3.5.0-3.5.2	>3.5.2 r2892, 20/02/2006	X	X
	3.6.x		X	X	X
	3.7.x		3.7.0	>3.7.0 r5678, 25/06/2009	X

Note that Java 5.0 and later in combination with Linux/Gnome has problems with the default Look'n'Feel. A workaround for the problem was introduced with version 3.4.5/3.5.0. For more information refer to the **Troubleshooting Wiki** article. From Weka 3.6.5/3.7.4 Mac OS X users will need Java for Mac OS X 10.6 Update 3 (java 1.6.0_22) or later.

Figure 1: WEKA Java Requirements

Software Run on:

- Manufacturer: Hewlett Packard
- Model: HP Pavilion dv6 Notebook PC
- Processor: AMD Phenom™ 2 N620 Dual Core Processor (2.80GHz)
- Ram Memory: 4GB
- System Type: 64-Bit Operating System
- HDD: 320GB

Data Source:

- data.betfair.com
- www.fracsoft.com
- Email: info@fracsoft.com
- Reference: Betfair Historical Data (Payment Required)

R Code:

```
#read excel .csv file
```

```
Don1345 <- read.csv("C:/Users/Andy/Desktop/Data Analytics/Dissertation/Data/1345Doncaster.csv",  
header=T, dec=".", sep=",")
```

```
#list data variable names
```

```
names(Don1345)
```

```
#attach variable names to the code
```

```
attach(Don1345)
```

```
Don1345
```

```
#list data as a time stamped list of frequency 1sec
```

```
TSDon1345 <- ts(Don1345[-1], start = 1, frequency = 1)
```

```
#show the time stamped data
```

```
TSDon1345
```

```
#plot time stamped data, editing axis labels and plot colour
```

```
ts.plot(TSDon1345,gpars= list(xlab="Time", ylab="Odds", col=rainbow(10)))
```

```
#insert legend with variable names and font size
```

```
legend("topleft", colnames(TSDon1345), text.font=2, cex=0.64, lty=1, col=1:ncol(Don1345))
```

```
#fit a linear model for regression (time stamp compares to horses in the race)
```

```
fit <- lm(Time.Stamp~Cowards.Close+My.Dads.Horse+Victor.Hewgo)
```

```
#produce a summary of fit(Residuals, Coefficients & R-Squared)
```

```
summary(fit)
```

```
#produce an analysis of variance table
```

```
anova(fit)
```

```
#compare the winner against all other runners in the race
```

```
fit1 <- lm(Time.Stamp~Victor.Hewgo)
```

```
fit2 <- lm(Time.Stamp~Cowards.Close+My.Dads.Horse)
```

```
anova(fit1,fit2)
```

```
plot(fit)
```

Time Series Analysis in WEKA:

Open the WEKA GUI Chooser by double-clicking the WEKA 3.7 shortcut icon on the desktop

Firstly we must add the time series analysis package as it is not in use automatically

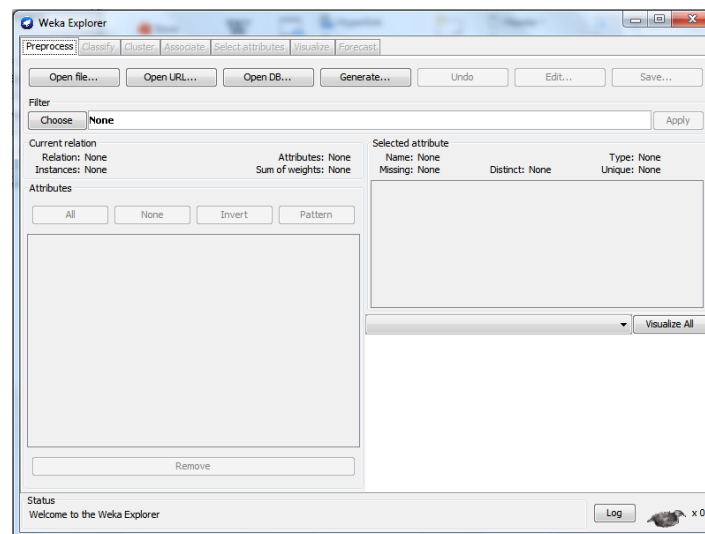
Open the package manager by clicking *tools > package manager*

Scroll to both the *timeSeriesFilters* and *timeSeriesForecasting* packages

Highlight both and click *install*

You can now close the package manager

Back at the GUI Chooser, open the *Explorer* Application (opens as per figure below)

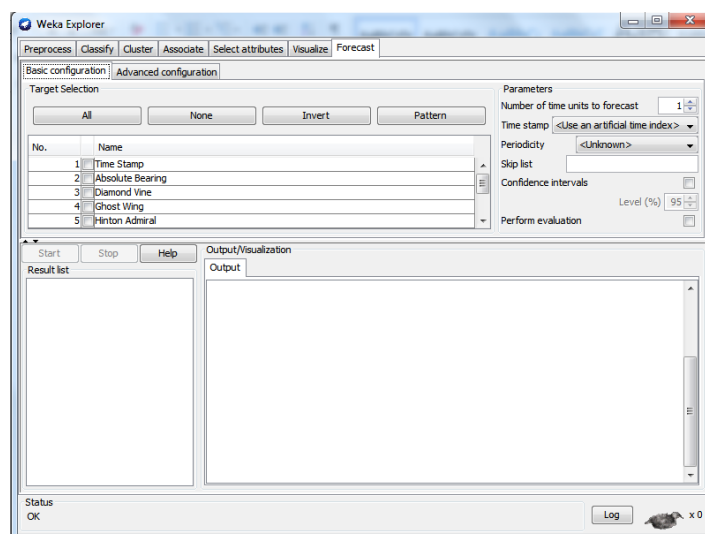


In the *preprocess* tab, click *open file*

Browse to your specific file (Note: WEKA defaults to look for .arff files)

Click on *file type* to search for .csv file formats

Open time series forecasting by clicking on the *Forecast* tab (Highlighted below)

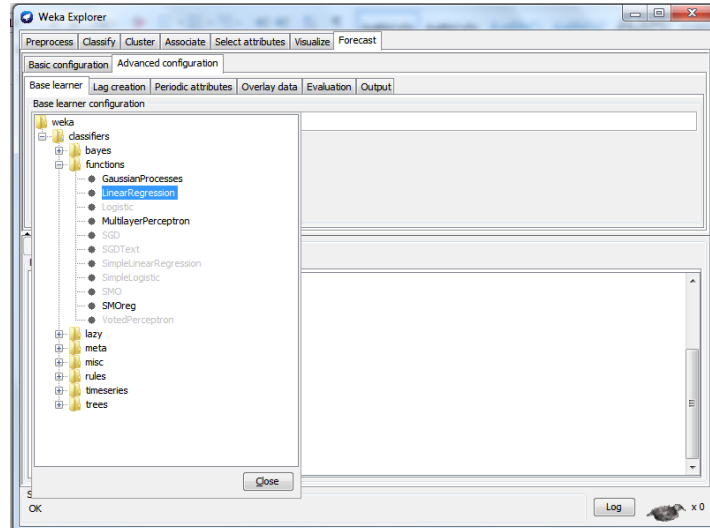


Within the *Basic Configuration* highlight the variables to forecast in the *Target Selection* pane

To the right is the *parameters* pane. Select the attribute to be time stamped (Automatic option also)

Select the *Number of time units to forecast* according to your requirements.

Select the *Advanced Configuration Tab*. In *Base Learner* tab, your forecasting function can be chosen



Click *Start* to commence analysis

