Coursework
Data Science Development (CMM535)

Andrew Tait, *1504693@rgu.ac.uk*

March 7, 2018

---

1 Research

[your text goes here] The paper that was chosen for this work is [1]. The authors provided a full review on different streaming algorithms and methos that handle concept drfit?. Below is a review of this paper that includes problem statement, related work and methods applied.

Notice how I cited the paper, and how does it appear in the document, to do so, you need to have a file in your working director (references.bib), this file simply contains the bibtext items for the papers you chose. These BibTex items are often avaiable to download from publishers website, see Figure 1

1.1 Problem Statement

What is this paper about? your text goes here, your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here

1.2 Relevant Work

[your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here]

1.3 Methods

[your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here]

1.4 Results

[your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here]

1.5 Conclusion

[your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here your text goes here]

2 Data Streams

The dataset that has been chosen for this part of the course work is IRIS. This is available on the UCI repository. The set was chosen because of .... It proves to be a good set for evaluating 'x' methods ....

2.1 Data Exploration

Explore the dataset as required by the course work. Discuss, visualise, summarise ... Make sure utilise what you have learnt in the class .... Presenting results is important. Here are some expampoles....

It is a good practice to separate the code from the results, this makes the document more readable and easier to reproduce. Here is an example, where I want to show the class distribution in my dataset.

```
table(iris$Species)
```

```
    setosa versicolor  virginica
        50         50         50
```

Tables: If you want to show results in a table format, the package xtable is very useful. Here is a an example: suppose you want to show the head of iris data, a simple way is to do the following:

```
head(iris)
```

```
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

This is fine, but sometimes you want to produce the results in a table-format at a better standard. See the code below. Often, I won't show the code below, nor make it executable (i.e. set eval=FALSE). Instead, run the commands below in the console, then copy the resulting latex code from consle and paste it below, and add caption and label as shown below

```
# load package
library(xtable)
# run the following command in the console
print(xtable(head(iris))
# copy the resulting latex code from consle and paste it below
# Add caption and label as shown below
```

The good thing about the above approach is you can always refer to the table in your document. For example 'As can be seen in Table 1,...'
Figures: are important part of your analysis, and also good way to give insight about the dataset you are working with. You will use packages like *ggplot2* to produce some visuals. Lets start with a simple example to show how to present your visuals in the report with proper labels and captions. Remember, I need to see the code the produced the figure as well.

Suppose, I just want to create a plot that shows the relation between Petal width and length and map the colour and shape to the Species (class label in my dataset).

Table 1: Iris Data

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.10	3.50	1.40	0.20	setosa
2	4.90	3.00	1.40	0.20	setosa
3	4.70	3.20	1.30	0.20	setosa
4	4.60	3.10	1.50	0.20	setosa
5	5.00	3.60	1.40	0.20	setosa
6	5.40	3.90	1.70	0.40	setosa

First, I will write my code, but notice that my chunk definition is set as follows *(«warning=FALSE,message=FALSE,eval=FALSE»=)*, I set warning and message to FALSE, because I don't want these warnings/ messages to appear in my final output.

```
library(ggplot2)

p <- ggplot(iris, aes(x=Petal.Length,
                      y = Petal.Width,color=Species) )
p <- p+ geom_point(aes(shape=Species))
p <- p + xlab('Petal Length')
p <- p + ylab('Petal Width')
p <- p + theme_bw()
p
```

Make sure that your code is running, and once everything is OK, then you need to insert the above code within a Latex code used to insert images (check the .rnw file to see how we achieved this). Again, remember the caption and the label which allows you to refer to this figure from anywhere in your document (Figure 1). Notice the header of the chunk code in the *.rnw* file.

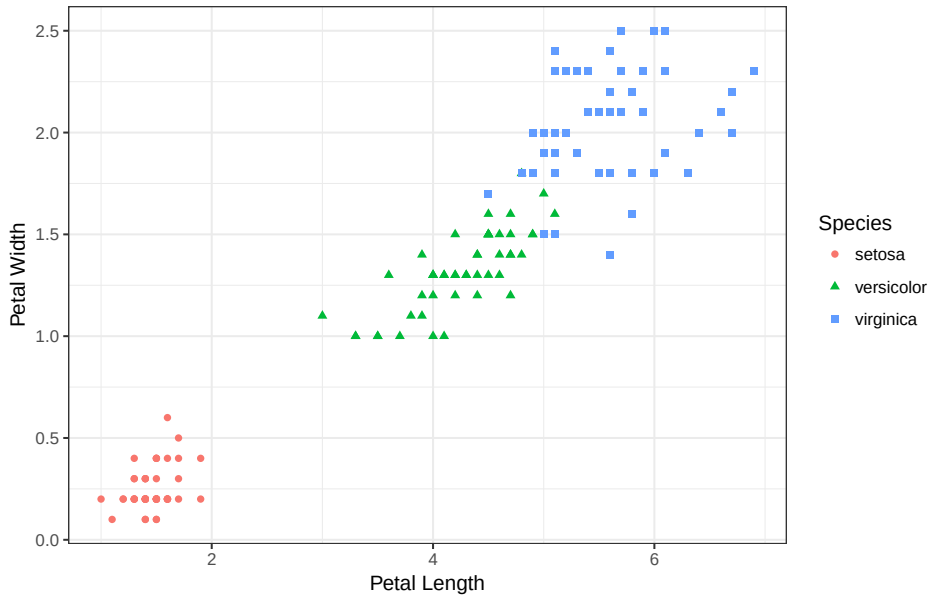

Figure 1: Petal Length /Width per species in IRIS set

2.2 Build Classifier

Complete this part as required by the coursework sheet. Again, be clear, visuals always helps in communicating results. Justify your choices and explain your methods.

2.3 Build Stream Classifier

Same as above. Complete this part as required by the coursework sheet. Again, be clear, visuals always helps in communicating results. Justify your choices and explain your methods.

3 Text Classification

Same as above. Complete this part as required by the coursework sheet. Again, be clear, visuals always helps in communicating results. Justify your choices and explain your methods.

3.1 Preprocessing

Record, explain and justify your appraoch. Record, explain and justify your appraoch. Record, explain and justify your appraoch. Record, explain and justify your appraoch.Record, explain and justify your appraoch.Record, explain and justify your appraoch.Record, explain and justify your appraoch.Record, explain and justify your appraoch. Record, explain and justify your appraoch.

3.2 Text Analysis

Record, explain and justify your appraoch.Complete this part as required by the coursework sheet. Again, be clear, visuals always helps in communicating results. Justify your choices and explain your methods.

3.3 Text Classification

Use one of the 'R' packages to build a classifier that classifies the tweets as leave tweet or remain tweets. Complete this part as required by the coursework sheet. Again, be clear, visuals always helps in communicating results. Justify your choices and explain your methods.

4 Reproducing Results

You don't need to have a section called **Reproducing Results**, your report itself will be the answer for this section.

5 References

[1] P. B. Dongre and L. G. Malik. "A review on real time data stream classification and adapting to various concept drift scenarios". In: *2014 IEEE International Advance Computing Conference (IACC)*. 2014, pp. 533–537. DOI: 10.1109/IAdCC.2014.6779381.