DATA201 Project

June 2020

The purpose of this project is to enable you to demonstrate the various skills you have learned during this course.

Deliverable: *A single iPython notebook containing code and graphs giving clear answers to the questions below, in order.*

Submission: *The deadline is Friday 26th June (the last possible date, so no extensions) before midnight. Upload your solution to the ECS assignment system at* `https: // apps. ecs. vuw. ac. nz/ submit/ DATA201 .`

The El Niño-Southern Oscillation (ENSO) is a cycle of warm and cold surface sea water temperatures in the southern Pacific Ocean. As the area of warm sea water moves the air pressure gets higher on either the east or west of the Pacific. This has an effect on rainfall and temperature globally, but given our geographic location, it has a very strong effect on New Zealand weather. The cycle varies between two and seven years, with an average of four years.

In this project you are going to investigate exactly what kind of effect it has on New Zealand, and see if you can predict the strength of the El Niño into the future.

There are three datasets provided as `projectdata.zip` in the dataset section of the course webpage. They are all from the Ministry for the Environment (`https://data.mfe.govt.nz/`. The data is in the `csv` files, which there is various information in the other files in the zip package.

1. Load all three files into Python and take a look at them. Ensure that you understand what is recorded in each. Compare the geographic locations of data collected, and the years. Create new dataframes that contain only the data for matched years and locations where appropriate. Deal appropriately with NaN datapoints. (3 marks)

2. Make plots of the various timeseries and check if you can see the expected temporal pattern in the ENSO index. (3 marks)

3. Test whether or not there is a correlation between the rainfall and temperature and the index. (4 marks)

4. Test whether or not this correlation is stronger or weaker in different parts of the country. (5 marks)

5. Using the Auckland data, predict the strength of the ENSO index for each year using the rainfall and temperature data. You will need to:

   - Choose the appropriate data and normalise it
   - Split it into training and testing sets
   - Choose an appropriate algorithm
   - Perform the training
   - Report appropriate metrics

   (15 marks)

6. Using just the ENSO index data, see if you can predict the strength of the ENSO index for date $t + 1$ using the index values for dates $t - 1, t - 2, t - 3$. The easiest way to do this is to assemble the data into an array with 3 columns. The first few rows would be (where $I_{m-y}t$ is the index value for date month-year):

   I_{01-86} I_{02-86} I_{03-86}
   I_{02-86} I_{03-86} I_{04-86}
   I_{03-86} I_{04-86} I_{05-86}

   Then use the first line to predict I_{04-86} and so on.

   You need to follow the same data preparation steps as in the previous question.

   Compare the model with two others:

   (a) Using the last 6 monthly values instead of the last 3
   (b) Using the last 12 monthly values instead of the last 3

   Explain why you get the results you do. (15 marks)

7. Test whether or not the ENSO is getting stronger or weaker over time. (5 marks)