QBUS6840 Assignment 2 – Group Assignment:

Value: 30%

Rationale

This assignment has been designed to help students develop valuable communication and collaboration skills and to allow students to apply their predictive analytics skills on a real world datasets.

Background and Dataset

The unemployment rate is the number of people unemployed divided by the number of people in the civilian labor force. The unemployment rate is reported by the relevant government agency in most countries. The **unemployment rate** can be used as an **important** indicator by the central bank of the country to determine the health of the economy when setting monetary policy. Also investors use **unemployment statistics** to plan future operating activities, see **https://www.thebalance.com/unemployment-rate-3305744**.

For this project, we have obtained the historical and current unemployment rates in a country from January 1986 to December 2018. The data have been deliberately added noises to de-identify the country. The dataset `UnemploymentRateJan1986-Dec2018.csv` can be downloaded from the Canvas.

The dataset shall contain information of month-year and unemployment rates. Similarly a test dataset file of 2019 unemployment rate, named `UmemploymentRate2019.csv` in the same format, has been hidden from you. They will be used to assess the prediction errors of your produced model(s). Your work in this project is to analyze the time series of unemployment rates.

Tasks

Please note most tasks are deliberately designed open. This gives more freedom for you to explore a better solution.

Data Pre-processing: Conduct initial analysis over the entire data. Write python program to clean up the data, e.g., checking/deleting incomplete information if any, to make sure data is complete, or normalising the data, etc. It is up to you how to normalise or transform the data so that the resulting dataset can be well incorporated in training your chosen model(s). You **MUST** retain your python program (or code section) used for all the pre-processing work.

Exploring Time Series: Analyze the entire time series: you may plot them or do what you can to reveal any patterns. Summarise what you have revealed or observed. In your report, carefully present your analysis and findings.

Model(s) and Competition: Based on what you have found from exploring the time series, consider fitting an appropriate model to the data for predictions. This could be the moving average, or decomposition method, exponential moving average, advanced models such as

ARIMA, Deep Neural Networks and Recurrent Neural Networks etc. This is your choice. Document your findings and justify your choice from a couple of different models. We suggest you shall try at least two different types of models.

Once you decide your best model for prediction, your final task is to use the trained model to predict all the 12 unemployment rates of 2019. Please note that this is a multiple–step-ahead forecast. As we assume you were in December 2018 in time and have no knowledge about 2019.

The 12 predicted values of the unemployment rates shall be written into a csv file named as `GroupXXX_Project_Results.csv`, where XXX is your group number in form of e.g., 008 (for Group 8), 085 (for Group 85) and 123 (for Group 123). The format of this file shall be the same as the given file `UnemploymentRateJan1986-Dec2018.csv`, except for the title of the second column to be `Predicted_Unemployment_Rates`.

Prediction/Test Error: For the measure of forecast accuracy, please use mean squared error (MSE). The MSE, computed on the prediction/test data, is defined as follows. Let $\hat{Y}_{T+h|1:T}$ be the h-step-ahead forecast of $\hat{Y}_{T+h}$, based on the training data $Y_{1:T} = \{Y_1, Y_2, ..., Y_T\}$, while Y_{T+h} is the h-th true unemployment value in the test data `UmemploymentRate2019.csv`. The test error is computed as follows

$$\text{test\_error} = \frac{1}{12} \sum_{h=1}^{12} \left(\hat{Y}_{T+h|1:T} - Y_{T+h} \right)^2,$$

where 12 is the number of observations in the test data.

Grouping Rules, Deadline, Late Penalty and Plagiarism

1. The assignment will be done in groups of 5 (a very few groups with 4 or 6 depending on the total of students in the class) without exception. The group can be formed freely or assigned by the Coordinator. Please get close contact with your members at the earlier stage. A group leader for each group shall be automatically assigned on Canvas.
2. The assignment is due at **Friday 16:00pm 29 May 2020**. The late penalty for the assignment is 5% of the assigned mark per day, starting after 16:00 pm on the due date. The closing date, 6 June 2020, 16:00pm is the last date on which an assessment will be accepted for marking.
3. The University of Sydney takes plagiarism very seriously. Please be warned that plagiarism between individuals is always obvious to the markers and can be easily detected by Turnitin

Meeting Minutes

1. Your group is required to submit meeting minutes, which are to be attached to the report as the second appendix. Your group may use the templates on Canvas for preparing agendas and meetings minutes.
2. You should document at least 3 meeting minutes for this group assignment, using the template provided/or a template you choose. Each minute should at least record the following information:
   a. Meeting dates/time/venue/duration;
   b. Key points of the process of discussion such as who said what (key points);

  c. Action list and responsible members, task due time etc

  d. Review/group judgement on the quality of individually completed/responsible tasks; The purpose of this is to infer whether a member is doing their share of jobs

3. In case of a problem within a group we will request minutes of the previous meetings. We will make an individual adjustment to the group mark if there is sufficient evidence that a student has done very little. If the student has truly done little, we will award a mark of 0.

Peer Assessment, Marks and Feedback

1. We may ask for peer assessment from each student. The instruction how to do this will be released later on.

2. Each group will be awarded a group mark per the marking criteria. In some cases, individual marks may be applied if there is dispute in a group and the quality or quantity of contributions made by individuals are significantly different, in which cases the unit coordinator will seek peer assessments reports from individuals in a group and meeting minutes.

3. Feedback will be provided on the marked submission.

Marking Criteria

1. The content in your report **Group_xxx_document.pdf** contributes **55% marks** of total project marks:

- Focus on the appropriateness of the chosen forecasting method and provide full explanation and interpretation of any results you obtain in your report. Output without explanation will receive **zero marks.**

- Describe your data analysis procedure in detail: how the Exploratory Data Analysis (EDA) step is done, what and why models/methods are used, how the models are trained, etc. with sufficient justifications. The description should be detailed enough so that other data scientists, who are supposed to have background in your field, understand and are able to implement the task.

- Clearly and appropriately present any relevant graphs and tables.

- Presentation of the assignment is part of the assignment. Markers will assign up to **10%** of the mark for clarity of writing and presentation. It is recommended that you should include your Python code as appendix to your report, however you may insert small section of your code into the report for better interpretation when necessary.

2. The Python implementation contributes **30% marks**. The main program file should be named as **Group_xxx_Prediction.py** (Jupyter Notebook .ipynb). If the marker cannot get your program run, some partial marks (maximum 10%) will be allocated.

3. We will allocate **15% marks** for competition among the groups. The group with the highest test score will secure full **15% marks** while other groups will secure a scaled mark according to the following rules, based on your test error:

- If the accuracy (or test error) of your forecast is within 5 percent of the best test error in the Class, then a full 15% marks will be award;

- If the accuracy (or test error) of your forecast is between 5.1percent and 20 percent of the best test error, then 12% marks will be award;

- If the accuracy (or test error) of your forecast is between 20.1percent and 50 percent of the best test error, then 8% marks will be award;
- Otherwise 4% marks will be awarded if the program is runnable.

Presentation

- Your answers shall be provided as a word-processed. Prepare one single report. Do not have separated reports for each question/task. Add your Python code as appendix to the report. At the same time, we will ask you to upload your python code and prediction results to your Canvas folder.

- The report should be NOT more than **25 pages**, with font size no smaller than 11pt, including everything like text, figure, tables, small sections of inserted codes etc but excluding the cover pages, appendix containing Python code and the meeting minutes. A violation of this rule will attract a penalty of 5% marks.

- Numbers with decimals should be reported to the **Third** decimal point in the report.

- The assignment material to be handed in will consist of a final report that:

   i) Takes a research article form in which you shall have a number of sections such as introduction, methodology, experiment results, findings/interpretation, and conclusion. All references should be properly cited and take a full bibliographical format. Here are couple of examples
   http://cs229.stanford.edu/proj2015/007_report.pdf
   http://cs229.stanford.edu/proj2015/188_report.pdf
   http://cs229.stanford.edu/proj2015/031_report.pdf

   ii) Details ALL steps and decisions taken by the group regarding requirements above.

   iii) Demonstrates an understanding of the relevant principles of predictive analytics approaches used.

   iv) Clearly and appropriately presents any relevant graphs and tables.

- Your group is required to submit meetings minutes. Your group may use the attached templates for preparing agendas and meetings minutes. You should document at least 3 meetings during the semester. Documentation should be in terms of attendance, discussion points, actions decided, review etc. You may use your own form or find something online.

- You, as a member of a group, may be also required to submit your peer assessment. Please use the attached criteria sheet and assessment form for this purpose. You will be informed of how to use online form when it becomes available.