QBUS6840 Assignment 2 – Group Assignment:

Due dates: Friday 24 May 2019

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.

Description

The assignment will be done in groups of 5 (or 4 or 6 depending on the total number 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 in earlier stage. A group leader will be automatically assigned by the Canvas system.

Notes

1. The assignment will be done in groups of 5 (or 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 in earlier stage. A group leader for each group shall be automatically assigned on Canvas.
2. The assignment is due at **Friday 16:00pm 24 May 2019**. 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, 31 May 2019, 16:00pm is the last date on which an assessment will be accepted for marking.
3. Your answers shall be provided as a word-processed. Prepare one single report. Do not have separate report 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 to your Canvas folder.
4. Your report should include the Group ID and SIDs of all group members. **No names**! You may stay with the report cover sheet provided.
5. You need to provide full explanation and interpretation of any results you obtain. Output without explanation will receive **zero marks**.
6. Be warned that plagiarism between individuals is always obvious to the markers of the assignment and can be easily detected by Turnitin.
7. The data sets for this assignment can be downloaded from Canvas.
8. 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. Think about the best and most structured way to present your work, summarise the procedures implemented, support your results/findings and prove the originality of your work.
9. Numbers with decimals should be reported to the **Second** decimal point.

10. 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.

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 attached templates 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 minutes 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

    Note: Any unsatisfactory meeting minutes may attract a penalty of up to 15% of the total marks.
3. In case of a problem within a group we will request minutes of the previous meetings. We can 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. 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 mark according to their test score against the best test score.
4. Feedback will be provided on the marked submission.

Background and Dataset

The **S&P/ASX 200** index is used as the benchmark for Australian equity performance. It is a market-capitalisation weighted and adjusted stock market index, see Wikipedia. ASX 200 index is calculated based on the 200 largest ASX listed stocks. It starts from 31 March 2000 with a value of 3133.3

The ASX200 historical and current data can be downloaded from **https://au.finance.yahoo.com/quote/%5EAXJO?p=^AXJO**. The historical data can be downloaded at three different frequencies of Daily, Weekly and Monthly between any

specified duration since 31 March 2000. In this project, your group will be asked to analyze the data at both frequency of daily and monthly. To align with the due date of the project, the following strategy is recommended:

1. Use 31 March 2000 as the starting date for downloading
2. The end date can be set as a date when your group starts work on the project. You may explore this dataset and do all the possible tasks for the project.
3. The Project due date is 24 May 2019. You can download the most recent full dataset on 23 May 2019. You may use this dataset to finalise your best model and make all the predictions. Even the data on 24 May 2019 may become available before you submit your report, you shall not use it at all.
4. For the group(s) who may be granted extension due to unforeseen reasons, the competition forecasts will be the next FIVE days from the granted due date.

The dataset shall contain information of dates (daily or months), open price, highest price, lowest price, close price and adjusted close price. Your work in this project is to analyze the time series of close price. A set of Daily and Monthly datasets have been downloaded for your convenience if you don't wish to download it at early stage. You can get them directly from Canvas.

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.

Exploratory Time Series: Analyze the entire time series for both daily and monthly data. 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. Any different patterns between Daily and Monthly time series?

Benchmark Model: Based on what you have found from exploring the time series, consider using a classical model to build your benchmark model for forecasting. This could be the moving average, or decomposition method, exponential moving average etc. It is always a good idea to split the given time series data into a training subset and a validation subset. Document your findings and justification. This should be done for both Daily and Monthly time series.

Build Advanced Models: You are requested to build at least TWO advanced models such as ARIMA, State Space Model, Deep Neural Networks and Recurrent Neural Networks etc. This is your choice. In building your chosen models, you need to at least optimise models in terms of e.g. the orders in ARIMA or State Space Model, and/or other parameters as well. Simply building a model without any consideration of validation and tuning hyperparameter

does not meet the minimal requirement for this task. Document your findings and justification. This should be done for both Daily and Monthly time series.

Competent Model and Final Result(s): Finally, according to your work, decide your best model for both Daily and Monthly time series. For all of them, please make five forecasts ahead. For the monthly time series, you may report your five forecast in a table in your report, but we ask you to forecast five daily results (i.e., the forecasts for Monday 27 May to Friday 31 May 2019) and save your results into a csv file containing two columns, one for the date (named `Dates`) and the other column for the predicted values with the second decimal point. Name your file as `GroupXXX_Results_ASX200.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 results will be assessed against the actual close prices on the dates in order to decide your group performance among the entire class (competition!). It is important for you to name your csv file in the above format, otherwise our program may fail your results.

Note:

1. *The score will be based on the mean squared error.*
2. *The close price on 24 May 2019 may become available, but you shall not use that information to train you model. In the forecasting stage, for some models, you may need predict the close price on 24 May 2019, then predict close prices for the next five days based on the prediction for 24 May 2019. However, the forecast for 24 May 2019 will not be used for assessment in competition.*

Presentation

- **Please submit your project through the electronic system on Canvas**

- 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.

- The MAXIMUM page limit is 25 pages, including any computer output, graphs, tables, etc.

- 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.