Assignment 2: Time Series Forecasting

Due 7 May by 23:59   **Points** 100   **Submitting** a file upload   **File types** ipynb, py, and r
Available 3 Apr at 0:01 - 11 May at 23:59

This assignment was locked 11 May at 23:59.

⚖ Weighting & Due Dates

This assessment is worth **20%** of your overall grade.

Due: **07 May 2023, 23:59pm** (Week 8).

ⓘ Task Description

Purpose:

To **apply time series forecasting techniques** to solve a practical problem.

Task description:

The task is to **predict future energy use in a household** based on weather conditions by building time series forecasting models.

Note: The ultimate prediction objective is the **same** as in Assignment 1, however, the prediction methods you are expected to use are **different**.

You need to write **Python or R code** to **predict** energy use and **analyse** the impact of different factors based on your model.

◎ Course Learning Outcomes

- CLO 2: Apply suitable algorithms for particular data mining problems.
- CLO 3: Design and develop processes and products to solve business problems related to data mining.
- CLO 5: Communicate effectively in a variety of forms using appropriate terminology.

☰ Instructions

Please **read and follow the instructions** below to complete the task.

1. Download the [dataset (https://myuni.adelaide.edu.au/courses/82119/files/12758046?wrap=1)](https://myuni.adelaide.edu.au/courses/82119/files/12758046?wrap=1) ↓

(https://myuni.adelaide.edu.au/courses/82119/files/12758046/download?download_frd=1) and **code template (https://myuni.adelaide.edu.au/courses/82119/files/12758047/download)** provided. Use this template for your assignment.

2. Read the **data description** ⤷ **(https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction)** .

3. Construct a code in Python Jupyter notebook or R Notebook/Markdown. Python is preferred.

4. Analyse and visualise the data (word limit: **200 words**).

- Use suitable time series analysis techniques learned in **Module 6 (https://myuni.adelaide.edu.au/courses/82119/pages/module-6-online-learning)** .
- **Discuss** your findings and how can they be used to select a time series forecasting model.
- **Compare** this analysis with the one from your analysis in Assignment 1 and comment on any similarities. Consider whether the time series analysis techniques have given you more insights, and comment on what are those insights were.
- Include the **charts and diagrams together with the code**, e.g. in Jupyter Notebook.

Note: You can use the pre-processed data from your Assignment 1, if suitable.
If for some reason you haven't done it in Assignment 1, then consider applying the following processing techniques: scaling, feature selection and imputation of missing values.
If any of these techniques are not suitable, explain your decision in less than 100 words. You may use other pre-processing techniques if needed.

5. Based on your analysis:

- **Use at least two time series forecasting methods** learned in **Module 6 (https://myuni.adelaide.edu.au/courses/82119/pages/module-6-online-learning)** (naïve method, seasonal naïve method, ARIMA, LSTM) and **justify your methods selection**.
- You can use the **paper (https://myuni.adelaide.edu.au/courses/82119/files/12579095?wrap=1)** ↓ **(https://myuni.adelaide.edu.au/courses/82119/files/12579095/download?download_frd=1)** mentioned below and the paper code in your assignment.
  Candanedo, LM, Feldheim, V & Deramaix, D 2017, 'Data driven prediction models of energy use of appliances in a low-energy house', *Energy and buildings***, vol. 140, pp. 81-97. (https://myuni.adelaide.edu.au/courses/82119/files/12758045?wrap=1)** ↓ **(https://myuni.adelaide.edu.au/courses/82119/files/12758045/download?download_frd=1)**
  If you use the code from the paper, **clearly identify which part of the code is used** and where, and how it has been adapted to your task. You can also use common Python and R libraries.
- **If you use any other code** except the code from the seminar, workshop and the abovementioned paper, clearly identify the source (r.g. URL), which part of the code is used, what is the task that the code is used for, and how it has been adapted to your task.
- **Use training/testing methodology suitable** for time series and suitable model performance metrics.

6. Test the models using the same performance metrics as you selected in Assignment 1, and show the results for all models.

7. Compare the results from all candidate models, **choose the best** model, **justify your choice** and **discuss the results** (word limit: **200 words**).

- Show the results of all models in the form of suitable charts and tables.
- Select the best-performing model, show the final results for this model and justify your selection.

- Compare the forecasting results obtained in this assignment with the results you have obtained in Assignment 1. Make sure that the results are comparable in terms of metrics. Use conversion or adjustments if needed.

8. Reflect on what you have learned by completing this assignment (word limit: **200 words**).

Submission requirements:

You are required to **submit all the runnable code(s), analysis and results**. **Do not** include the dataset in the submission.

Submit one file (e.g., .ipynb), and do not zip it.

Submitted file name must be in the form "<your id>_<your_name>_assign2".

✎ Late Submission Rules

If you hand in your work late, your mark will be capped, based on the number of late days.

A part of the late day is counted as full day.

- 1 day late – mark capped at 75%
- 2 days late – mark capped at 50%
- 3 days late – mark capped at 25%
- more than 3 days late – no marks available.

📖 Assessment Criteria

Please familiarise yourself with the **assessment rubric** below.

MBD A2 Rubric

Criteria	Ratings					Pts
Technical Skills This criterion includes but are not limited	**40 to >34.0 Pts** **HD** Demonstrates a wide range of	**34 to >30.0 Pts** **D** Demonstrates a range of	**30 to >26.0 Pts** **C** Demonstrates good technical	**26 to >20.0 Pts** **P** Demonstrates appropriate	**20 to >0 Pts** **F** Fails to demonstrates	

	comprehensive technical skills.	technical skills.	skills.	technical skills.	appropriate technical skills.	
to: - choice of the appropriate visualising tools/charts, - application of suitable pre-processing techniques - choice of training/testing methodology, - choice of measuring metrics, - adherence to assignment specification - good coding practice - making the program executable and easy to use						40 pts
Results Analysis This criterion assesses the quality of results analysis and the understanding of the results received, as well as the use of relevant codes and algorithms.	**30 to >28.0 Pts HD** Communicates comprehensive analysis and demonstrates excellent understanding of the results received. Uses highly relevant code and algorithms .	**28 to >24.0 Pts D** Communicates very good analysis and demonstrates very good understanding of the results received. Uses relevant code and algorithms.	**24 to >20.0 Pts C** Communicates good analysis and demonstrates very good understanding of the results received with minor exceptions. Mostly uses relevant code and algorithms.	**20 to >15.0 Pts P** Communicates appropriate analysis and demonstrates understanding of the results received with some exceptions. Uses some appropriate code and algorithms.	**15 to >0 Pts F** Fails to provide appropriate analysis and demonstrates little/no understanding of the results received with some exceptions. Fails to use appropriate code and algorithms.	30 pts
Organisation of ideas This criterion assesses how ideas are organised and	**20 to >17.0 Pts HD** Presents ideas very clearly and concisely. Ideas organised very	**17 to >15.0 Pts D** Presents ideas clearly and concisely. Ideas organised	**15 to >13.0 Pts C** Presents some ideas clearly and mostly concisely. Presents ideas	**13 to >10.0 Pts P** Presents some ideas clearly but only sometimes concisely. Ideas	**10 to >0 Pts F** Fails to present ideas clearly and	

communicated.	effectively in logical order. Ideas skilfully linked. Excellent range of lexical and grammatical devices. Rare minor errors/no errors. Sufficient and explicitly clear comments on the code.	effectively in logical order. Ideas well-linked. Very good range of lexical and grammatical devices. Only occasional errors that do not hinder the comprehension. Mostly sufficient and clear comments on the code.	clearly and mostly concisely. Ideas organised appropriately in logical order and linked appropriately. Good range of lexical and grammatical devices. Some errors that do not cause difficulty for the audience. Most comments on the code are clear and sometimes sufficient.	partially organised in logical order. Some linking between ideas. Satisfactory range of lexical and grammatical devices. Some errors that cause some difficulty for the audience. Comments are not always clear/sufficient.	concisely. Fails to organise ideas in logical order. Very limited use of linking between ideas. Very limited range of lexical and grammatical devices. Frequent errors that distort the meaning. Fails to provide clear and sufficient comments on the code.	20 pts
Formatting This criterion assesses how your document is presented.	**10 to >8.5 Pts HD** Document professionally presented. Clear and consistent formatting overall.	**8.5 to >7.5 Pts D** Document professionally presented with minor exceptions.	**7.5 to >6.5 Pts C** Document professionally presented with several exceptions.	**6.5 to >5.0 Pts P** Document professionally presented with significant exceptions.	**5 to >0 Pts F** Document poorly presented. Poor formatting overall.	10 pts
					Total points: 100	