Assignment 1: Machine Learning Methods

Due 2 Apr by 23:59   **Points** 100   **Submitting** a file upload   **File types** ipynb, py, and r
Available 28 Feb at 0:01 - 6 Apr at 23:59

This assignment was locked 6 Apr at 23:59.

⚖️ Weighting & Due Dates

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

Due: **02 April 2023, 23:59pm** (Week 5).

◎ 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.

ℹ️ Task Description

Purpose:

To practice the basic flow of machine learning and to **apply regression 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 an advanced regression model.

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

☰ Instructions

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

1. Download the [dataset (https://myuni.adelaide.edu.au/courses/82119/files/12579097?wrap=1)](https://myuni.adelaide.edu.au/courses/82119/files/12579097?wrap=1) ↓ [(https://myuni.adelaide.edu.au/courses/82119/files/12579097/download?download_frd=1)](https://myuni.adelaide.edu.au/courses/82119/files/12579097/download?download_frd=1) and [code template (https://myuni.adelaide.edu.au/courses/82119/files/12579098/download)](https://myuni.adelaide.edu.au/courses/82119/files/12579098/download) provided.

2. Read the **paper (https://myuni.adelaide.edu.au/courses/82119/files/12579096?wrap=1)** ↓ **(https://myuni.adelaide.edu.au/courses/82119/files/12579096/download?download_frd=1)** below before you start working on the task - you may find the information useful.

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/12579095?wrap=1)** ↓ **(https://myuni.adelaide.edu.au/courses/82119/files/12579095/download?download_frd=1)**

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**).

- **Identify data dependencies** that might be useful for this task and visualise those dependencies using suitable techniques and charts.
- Use this analysis to **select suitable prediction models** for experimentation and **justify** your selection.
- **Include the charts and diagrams** together with the code, e.g., in Jupyter Notebook.
- **Pre-process data**: apply suitable processing techniques such as scaling, conversion and imputation of missing values.

5. Based on your analysis:

- **Implement and train at least two prediction model(s)**.
  You can use the paper and the paper code in your assignment. If you use the code from 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)** , **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. **Do not** use any other code except the code from the seminar, workshop and the abovementioned paper.
- **Use suitable training/testing methodology**, such as data training/test split or cross-validation and justify your decision (write up to **100 words**).
- Use suitable model performance metrics and justify your selection (write up to **200 words**).

6. Test the models and print/include results for all models using machine learning methodology.

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 you selection.

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 in one file** (e.g., .ipynb), do not zip it, just submit that file.

Every submitted file name must be in the form "<your id>_<your_name>_assign1" (as in the template)

✏️ 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.

Note: each late day started is counted as one day.

📖 Assessment Criteria

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

MBD A1 Rubric

Criteria	Ratings					Pts
Technical Skills This criterion includes but are not limited 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	**40 Pts** **HD** Demonstrates a wide range of comprehensive technical skills.	**34 Pts** **D** Demonstrates a range of technical skills.	**30 Pts** **C** Demonstrates good technical skills.	**26 Pts** **P** Demonstrates appropriate technical skills.	**20 Pts** **F** Fails to demonstrates appropriate technical skills.	40 pts

assignment specification - good coding practice - making the program executable and easy to use						
Results Analysis This criterion assesses the quality of results analysis and the understanding of the results received, as well as the use of relevant libraries 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 communicated.	**20 to >17.0 Pts** **HD** Presents ideas very clearly and concisely. Ideas organised very 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.	**17 to >15.0 Pts** **D** Presents ideas clearly and concisely. Ideas organised 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	**15 to >13.0 Pts** **C** Presents some ideas clearly and mostly concisely. Presents ideas 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	**13 to >10.0 Pts** **P** Presents some ideas clearly but only sometimes concisely. Ideas 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	**10 to >0 Pts** **F** Fails to present ideas clearly and 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	20 pts

		the code.	on the code are clear and sometimes sufficient.	not always clear/sufficient.	distort the meaning. Fails to provide clear and sufficient comments on the code.	
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	