Coursework Brief

Task To Be Done

You are given the ML Coursework Dataset - A3 (access it via the link) to build a multilayer perceptron model that can be used for forecasting relative humidity (in %) for any geographical region one month into the future (e.g. given historical data up until or at January 2030, your model can predict the relative humidity for February 2030). For more information about the dataset, see the METADATA file included in the dataset folder.

What To Submit

You are to submit THREE separate documents – a report as a **pdf** file; your code as **ipynb** file (Jupyter notebook file); and your test output as a **csv** file.

A. (80 marks) **Report (submit as a pdf file)** - Prepare a report that answers ALL 4 questions below. Use the EXACT question themes highlighted in bold as section titles for your report, so that your report has 4 (titled) sections exactly. Submit the document as a single pdf file.

1. (20 marks) **Performance**
    - (10 marks max) Report the performance of your model for generalization to unseen years. Using both distance and correlation metrics, describe each metric used (using maths formulae if appropriate).
    - (10 marks max) State clearly the total number of instances used and how the data was split to train and evaluate the model.

2. (20 marks) **Model**
    - (8 marks max) Describe your model (using maths formulae when appropriate).
    - (6 marks max) List and describe the steps that you took to prevent overfitting for your model (using maths formulae when appropriate).
    - (6 marks max) List and describe the steps that you took in your model to address imbalance in your labels (using maths formulae when appropriate).

3. (20 marks) **Features & Labels**
    - (20 marks max) Report the features and label(s) used and describe (as a list) how they were derived and/or why they were selected from the given data (using maths formulae when appropriate).

4. (20 marks) **Preprocessing**

- (20 marks max) List and describe the preprocessing that you did on your data for building your model (using maths formulae when appropriate).

B. (10 marks) **Code (submit as a ipynb file)** - Prepare all the code used to complete all the tasks above and submit as a single ipynb file that can be run.

C. (10 marks) **Model outputs for the test set/sets (submit as a csv file)** - Prepare the prediction outputs for your model in a single csv file and using an appropriate header row. The file should have at least 3 columns: 1 column for the data instance id, 1 column for the corresponding true label, at least 1 column for the corresponding prediction(s) for your model. Submit the document as a single csv file.

Other Important Notice

1. You are not required to code machine learning algorithms from scratch. You can use standard machine learning libraries including Scikit-learn, PyTorch, TensorFlow. However, you must only use standard libraries.

2. You are ONLY allowed to use standard Python libraries for data extraction, preparation, and the rest of the machine learning pipeline.

3. You are NOT permitted to use or submit someone else's code, output, or report as yours (you are permitted to use code snippets from the lab materials given by the teacher, from machine learning software library documentations as stated in #1 above, or from recommended textbooks).

4. You must NOT use generative artificial intelligence (AI) to generate any materials or content for your assessment submissions.

   Note that:

   - The baseline position at the University is Sussex is that the use of generative AI material in assessment submissions is prohibited, unless explicitly permitted by the module convenor.
   - Students registered with the Disability Advice team and in receipt of reasonable adjustments are still permitted to use other assistive technology as required.
   - If in any doubt about what is permissible, students should check with the module convenor.

Link to learning outcomes

This assessment is designed to evaluate how well you have achieved the module's learning outcomes:

- To know the fundamentals of machine learning and understand the theory (maths) behind standard algorithms;
- To be able to build, optimize, and evaluate machine learning models appropriately with data;
- To be aware of ethical issues relevant to machine learning;

by how you are able to systematically and creatively apply these knowledge and experience to a new problem.