DSAI HW1 - Electricity Forecasting

Overview

In this HW, we will implement an algorithm to predict the **operating reserve (備轉容量)** of electrical power. Given a time series electricity data to predict the value of the operating reserve value of each day during **2021/03/23 ~ 2021/03/29**.

Evaluation

Goal
Predict the operating reserve (備轉容量) value from 2021/03/23 to 2021/03/29.

Metric
We will take the Root-Mean-Squared-Error (RMSE) between your result and real value to evaluate your model's performance.

Submission File Format
You should output the file named `submission.csv` which contain your result of prediction.
`submission.csv` format:

```
date,operating_reserve(MW)
20210323,2557
20210324,1899
20210325,1891
20210326,1811
20210327,1903
20210328,2333
20210329,1800
```

Data

Basic Data

- 台灣電力公司_過去電力供需資訊
  - Data duration: 2019/01/01 ~ 2021/01/31
  - Description:
    - Including "日期"、"尖峰負載(MW)"、**"備轉容量(MW)"**, "備轉容量率(%)" and so on.
    - "備轉容量(MW)" (operating reserve) is the prediction target. You can use this column data to train or test your model.
- 台灣電力公司_本年度每日尖峰備轉容量率
  - Data duration: 2020/01/01 ~ today
  - Description:
    - Including "日期"、**"備轉容量(萬瓩)"**, and "備轉容量率(%)".
    - 1 萬瓩 = 10 MW

Note
- The unit of the operating reserve should be "**MW**" in `submission.csv`.
- The data downloaded from "CSV" button and "檢視資料" are different. You can download both data for the model training.

資料資源下載網址	⬇ CSV	檢視資料 資料格式為CSV檔，第一列為欄位名稱。收錄前一整年度及本年度至前一月份為止之電力供需資料。

Other Data

- Side information (You **can** also use these data to train the model.)
  - 未來一週天氣預報
  - 未來三日空氣品質預報
- 台灣電力公司 (This data is for reference only. You can **not** use this data as training data or your answer.)
  - 今日預估尖峰備轉容量
  - 未來一週電力供需預測

Grade

Rank of your HW (100%)

Your grade is averaged from the RMSE result

- The best will get "100".
- The worst will get "70".
- The others will get the grade according to the distribution of RMSE.

Documentation Bonus (5%)

- If you clearly describe your method in `README.md` (e.g., data analysis, data pre-processing, feature selection, model training,...), you can get extra "5" points.

Note

- If your code cannot be legally executed or cannot output the `submission.csv` correctly, you will get some deduction.
- **You will get "0" point if one of these situations happens**:
  - Your code is the same as other classmate's.
  - Your Github repo does not contain your code.
  - You do not submit the homework before the deadline.
  - You directly use the data "未來一週電力供需預測" provided above as your answer or your trainning data.

Requirements

TA will test whether your code can run successfully and output the `submission.csv` or not.
Here are something you need to do:

1. Upload your code to Github and fill out the google form.
   - DSAI 2021 HW1 - Github Repo Submission
2. The Github repo should contain your **code**, **dataset**, **submission.csv** and **README.md**.
   a. Please put the `submission.csv` in root of your repo.
3. Please use python 3.6 or the higher version.
4. Please name your main python code `app.py`, we will test the code by running
   ```
   python app.py --training "Your Training Data" --output submission.csv
   ```
   - If you need to pass other arguments, please describe in `README.md`.
5. Please **clearly describe your idea** in `README.md` .

6. Please put all library and packages in `requirements.txt`.
    a. We will import them by running `pip install -r requirements.txt`.
    b. [Tutorial of requirements files](#)
    c. `requirements.txt` example:

```
scikit-learn==0.24.1
numpy==1.20.0
pandas==1.2.3
```

Code for TA trace

For TA traces your code, please write your main code as:

```python
# You can write code above the if-main block.
if __name__ == '__main__':
    # You should not modify this part, but additional arguments are allowed.
    import argparse

    parser = argparse.ArgumentParser()
    parser.add_argument('--training',
                        default='training_data.csv',
                        help='input training data file name')

    parser.add_argument('--output',
                        default='submission.csv',
                        help='output file name')
    args = parser.parse_args()

    # The following part is an example.
    # You can modify it at will.
    import pandas as pd
    df_training = pd.read_csv(args.training)
    model = Model()
```

```
    model.train(df_training)

    df_result = model.predict(n_step=7)

    df_result.to_csv(args.output, index=0)
```

- In your `if name == '__main__'` block, you should call your defined functions to load the data, train your model and write your file into `submission.csv`.
  Other functions should be written before the main function or in other modules. Structuring your code well is encouraged but not mandatory.
- Example Github repo

Test Environment

- Python 3.6.4
- Ubuntu 16.04.3 LTS

Deadline

- **2021/3/22 23:59 pm**.
- No Exception.
- TA will clone the code committed before the deadline. Any further change will be ignored.

Question

- If you have any question, feel free to contact the TA or use the "Comment" in the right hand side of this line.
  - TA: 陳彥儒　yjchen@netdb.csie.ncku.edu.tw
  - TA: 丁羅邦芸 lpyting@netdb.csie.ncku.edu.tw
- Every question asked by students will be listed below

[Q1] 可以使用台電的 未來一週電力供需預測 資料來訓練模型嗎？
[A1] 不行，這麼做意義上等同於直接使用台電的預測結果當你的答案，電力相關的資料只能使用上面提供的 basic data

Paper Reference

- A Survey on Data Mining Techniques Applied to Electricity-Related Time Series Forecasting [2015 Energies 8]
- A review and analysis of regression and machine learning models on commercial building electricity load forecasting [2017 *Renewable & Sustainable Energy Reviews*]
- Electricity load forecasting for Urban area using weather forecast information [2016 IEEE ICPRE]