National College of Ireland

Configuration Manual

MSc Research Project
MSc Data Analytics

Aiswarya Kallumpuram Prabhakaran
Student ID: x22155961

School of Computing
National College of Ireland

Supervisor: Abdul Shahid

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Aiswarya Kallumpuram Prabhakaran
Student ID:	X22155961
Programme:	MSc in Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Abdul Shahid
Submission Due Date:	31/01/2024
Project Title:	Forecasting of climatic influence on energy generation from renewable resources in Spain using Neural Network Models
Word Count:	414
Page Count:	3

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Aiswarya Kallumpuram Prabhakaran
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Aiswarya Kallumpuram Prabhakaran
Student ID: x22155961

1 System Configuration

1.1 Hardware

Ram: 8GB
System OS: Windows 11

1.2 Software

Software Computing Tools Used: Python Jupyter Notebook, Microsoft Excel
Browse Engine: Google Chrome

2 Project Development

The project consists of four stages such as Data Collection, Data Preprocessing, Implementation of Neural Network Models and finally evaluation of results.

2.1 Data Collection

In this project two datasets is being used. One is Energy Dataset and the second one is Weather dataset. Both of the dataset are taken from an open source platform named Kaggle.
Name : Hourly energy demand generation and Weather
The dataset contains the energy consumption, generation, pricing and the weather data for Spain from the year 2015 to 2018. The energy consumption and generation data is obtained from the ENTSOE, a public portal for Transmission Service operator data and the weather data for 5 cities in Spain is retrieved from the Open Weather API, which is helpful in identifying the influence of weather on energy generation and it's consumption. The pricing data is retrieved from the Red Electric Espana, which gives the cost of electricity during specific periods.
Weather dataset contains 17 columns with 178396 observations and energy dataset contains 29 columns with 35064 observations.

1. URL of the location of the dataset: https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather

2. License of usage : https://creativecommons.org/publicdomain/zero/1.0/

3. Data Size: Energy dataset - 6126 KB, Weather dataset – 2120 KB

2.2 Dataset Loading and Implementation of Codes

The finalised two dataset is downloaded into the system. For doing this project a system with 8GB RAM is used. Python Jupyter Notebook (anaconda 3) is used for the execution of the code. A folder named 'RIC FINAL PROJECT' is created and the two datasets energy and weather is uploaded into this folder. The dataset is in csv format (energy.csv and weather.csv).

3 Libraries Used

For the execution of these codes several libraries have been imported and they are numpy, pandas, matplotlib, seaborn, tensorflow, sklearn.

```
#importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import zscore
import seaborn as sns
from statsmodels.stats.outliers_influence import variance_inflation_factor
import statsmodels.api as sm
import tensorflow as tf
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
from tensorflow.keras.layers import LSTM, Dense, Conv1D, MaxPooling1D, Flatten
import warnings
warnings.filterwarnings('ignore')
```

Figure 1: List of imported libraries

Data Cleaning (of energy data) is done by removing the unwanted columns and by filling the null values using interpolation method.

```
energy=energy.drop(['generation fossil coal-derived gas','generation fossil oil shale',
                    'generation fossil peat', 'generation geothermal',
                    'generation hydro pumped storage aggregated', 'generation marine',
                    'generation wind offshore', 'forecast wind offshore eday ahead',
                    'total load forecast', 'forecast solar day ahead',
                    'forecast wind onshore day ahead'],
                    axis=1)
```

```
#Drop columns which are not renewable resources
energy=energy.drop(['generation fossil brown coal/lignite','generation fossil gas','generation fossil hard coal',
                    'generation fossil oil','generation nuclear','generation other',
                   'forecast wind offshore eday ahead',
                    'total load forecast', 'forecast solar day ahead',
                    'forecast wind onshore day ahead'],
                    axis=1)
```

```
# Fill null values using interpolation
energy.interpolate(method='linear', limit_direction='forward', inplace=True, axis=0)
```

Figure 2: Data Cleaning

Outliers also removed using z score from both energy and weather data.

The datasets are then merged using the time index as follows:

Merging of two datasets

```
# Merge the two DataFrames using 'time' column as index
final_df = pd.merge(energy_data, weather_data, on='time', how='outer')
print(final_df.columns)
```

```
Index(['generation biomass', 'generation hydro pumped storage consumption',
       'generation hydro run-of-river and poundage',
       'generation hydro water reservoir', 'generation other renewable',
       'generation solar', 'generation waste', 'generation wind onshore',
       'total load actual', 'price day ahead', 'price actual', 'city_name',
       'temp', 'temp_min', 'temp_max', 'pressure', 'humidity', 'wind_speed',
       'wind_deg', 'rain_1h', 'rain_3h', 'snow_3h', 'clouds_all'],
      dtype='object')
```

Figure 3: Merging of datasets

Then all the neural network models is performed using the desired libraries and the outputs are evaluated using the evaluation metrics by plotting bar graph of all results.