National College of Ireland

Configuration Manual

M.Sc. Research Project
Data Analytics

Punit Lohani
Student ID: x18127339

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Punit Lohani
Student ID:	x18127339
Programme:	M.Sc.-Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	12/12/2019
Project Title:	Daily Precipitation Forecasting using Neural Network- A case study of Punjab, India
Word Count:	XXX
Page Count:	8

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:	
Date:	12th December 2019

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):	

Contents

Configuration Manual

Punit Lohani
x18127339

1 Introduction

Configuration Manual basically provides a detailed description about the environment that has been set for the Research. This involves, the configuration of the system, describing the coding language used along with the various packages and libraries.

The document includes the important information about the procedures that have been followed while conducting the research and also contains the results that have been generated after performing the tests and running our model.

This document is important so that all the relevant and useful information about the research can be captured for better understanding.

2 Specification of the Environment

2.1 System Specifications

Below are the details about the local machine on which the research has been conducted:

- Operating System: Windows 10, 64 bit
- Processor: Intel® Core™ i5, 8th Gen, 1.80 GHz
- Installed Memory (RAM): 12 GB

2.2 Technical Specifications

Python programming language has been used for conducting the research and the details are mentioned below:

Python [1]

- Version 3.7.3
- Anaconda 1.9.7 is used as a platform for Python.
- jupyter notebook 5.7.8 has been used.
- The following packages in Python are used in research project:
  - Pandas 0.24.2 [2]
  - Numpy 1.16.2 [3]
  - Scikit-learn 0.20.3 [4]
  - Keras 2.2.4 [5]
  - Matplotlib 3.0.3 [6]
  - Tensorflow 1.9.0 [7]

[1] https://www.python.org
[2] https://pandas.pydata.org/
[3] https://numpy.org/
[4] https://scikit-learn.org/stable/index.html
[5] https://keras.io/
[6] https://matplotlib.org/
[7] https://www.tensorflow.org/install/

3 Collection of Dataset

- The Dataset has been collected from the spatial science website of Texas A&M University and the data is publicly available and we need to place the request for obtaining it.

- For this the request has been raised and the screenshot of the request is shown below:

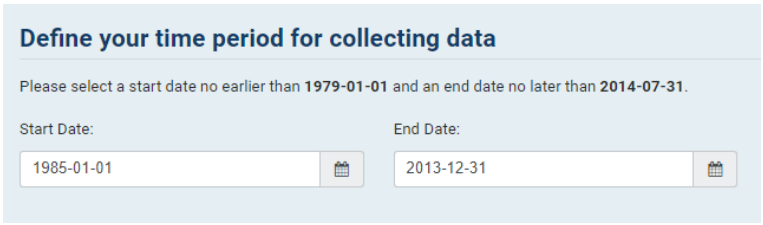

Figure 1: Specifying the date range

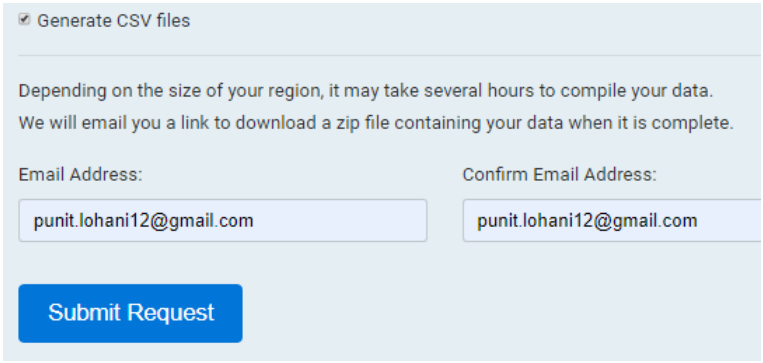

Figure 2: Providing email id and file format

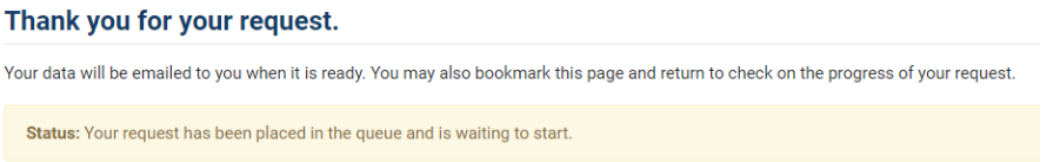

Figure 3: Confirmation on the submission of request

- After placing the request, the mail has been received stating the link of zip file from where it can be downloaded. Below is the screenshot of the email received:

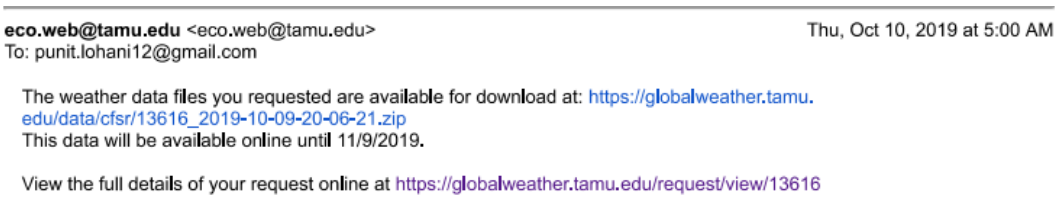

Figure 4: Mail received containing dataset information

4 Outputs

4.1 Augmented Dickey Fuller Test

ADF test is basicall used to find out if the time series is stationary or non stationary. The sig level we have taken is 0.05. The time series in our case is stationary and as we can see that the null hypothesis is rejected.

```
   Augmented Dickey-Fuller Test on "Precipitation"
   -----------------------------------------------
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level    = 0.05
Test Statistic        = -86.2745
No. Lags Chosen       = 0
Critical value 1%     = -3.431
Critical value 5%     = -2.862
Critical value 10%    = -2.567
=> P-Value = 0.0. Rejecting Null Hypothesis.
=> Series is Stationary.
```

Figure 5: Output from ADF Test

4.2 Seasonal ARIMA Model

- **Result Summary of Seasonal ARIMA**

The result summary of the SARIMA model is shown in the below screenshot.

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.3133	0.004	74.447	0.000	0.305	0.322
ma.L1	-0.9111	0.003	-359.292	0.000	-0.916	-0.906
ma.S.L30	-1.0000	1.702	-0.587	0.557	-4.336	2.336
sigma2	133.0820	226.643	0.587	0.557	-311.130	577.294

Figure 6: Result summary

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2008	0.012	17.046	0.000	0.178	0.224
ma.L1	-0.8321	0.006	-133.981	0.000	-0.844	-0.820
ar.S.L7	-0.7454	0.077	-9.689	0.000	-0.896	-0.595
ma.S.L7	0.6634	0.084	7.869	0.000	0.498	0.829
sigma2	87.6044	0.645	135.762	0.000	86.340	88.869

Figure 7: Result summary

- **Seasonal ARIMA Forecast**

For forecasting the daily precipitation amount first we used SARIMA model and the RMSE values were obtained. The forecast graph representing the precipitation derived from the model is shown in the below screenshot:

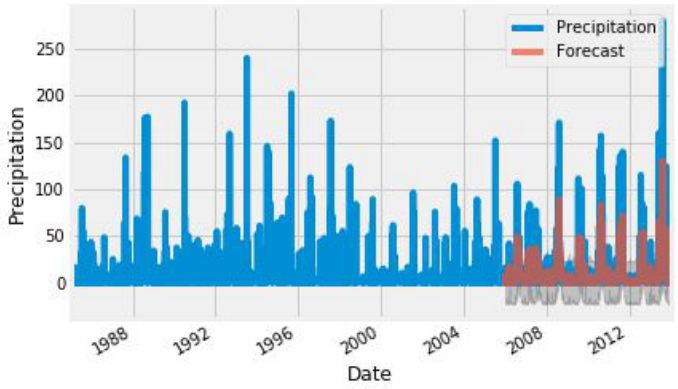

Figure 8: Seasonal ARIMA Forecast

4.3 Long Short Term Memory

After tuning the parameters and getting the values of epoch, batch size and neurons, the execution of the LSTM has done. In total, 100 epochs and 30 batch size has been taken into account. Below is the screenshot for the same:

- **LSTM epochs run**

```
Epoch 1/100
 - 1s - loss: 0.0023
Epoch 2/100
 - 1s - loss: 0.0021
Epoch 3/100
 - 1s - loss: 0.0021
Epoch 4/100
 - 1s - loss: 0.0021
Epoch 5/100
 - 1s - loss: 0.0021
Epoch 6/100
 - 0s - loss: 0.0021
Epoch 7/100
 - 0s - loss: 0.0021
```

Figure 9: LSTM Model epochs

- **Summary of LSTM Model**

The summary of the LSTM model after the execution of all the epoch is shown below:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_10 (LSTM)               (None, 128)               66560
_________________________________________________________________
dropout_10 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_10 (Dense)             (None, 1)                 129
=================================================================
Total params: 66,689
Trainable params: 66,689
Non-trainable params: 0
_________________________________________________________________
None
```

Figure 10: LSTM Model summary

- **Forecast through LSTM Model**

The forecast is then finally done by the LSTM model and it is then plotted. Below is the graph representing the forecast:

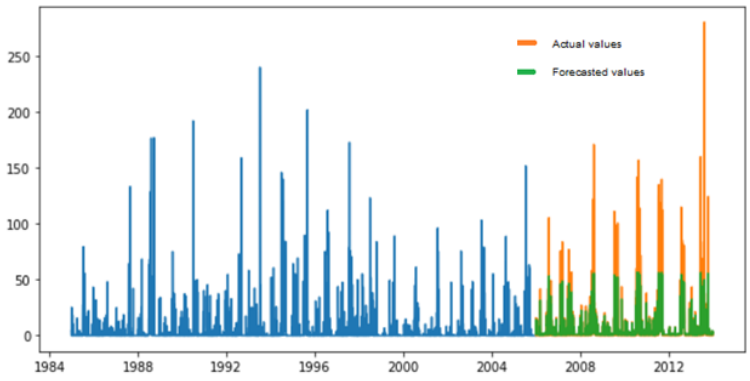

Figure 11: LSTM Model Forecast