Weather Forecasting with Machine Learning

BY

Md. Ramim Hossain
ID: 183-15-2246
AND

Mohammad Abdul Karim
ID: 183-15-2271

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Taslima Ferdaus Shuva
Sr. Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Aliza Ahmed Khan
Lecturer
Department of CSE
Daffodil International University

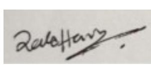
DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
SEPTEMBER 14, 2022

APPROVAL

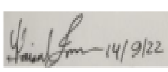
This Project/internship titled "Weather Forecasting with Machine Learning"", submitted by **Ramim Hossain** , ID No: 183-15-2246 and **Mohammad Abdul Karim** , ID No: 183-15-2271 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 14.09.22.

BOARD OF EXAMINERS

Dr. S M Aminul Haque **Chairman**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Md. Zahid Hasan **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Mr. Faisal Imran (FI) **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md Sazzadur Rahman **External Examiner**
Associate Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Taslima Ferdaus Shuva, Sr. Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

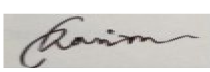
Taslima Ferdaus Shuva
Sr. Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Aliza Ahmed Khan
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Md. Ramim Hossain
ID: 183-15-2246
Department of CSE
Daffodil International University



Mohammad Abdul Karim
ID: 183-15-2271
Department of CSE
Daffodil International University

ACKNOWLEDGMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Taslima Ferdaus Shuva**, **Sr. Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Weather forecasting with Machine Learning*" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Taslima Ferdaus Shuva, Aliza Ahmed Khan, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Due to its numerous applications in industries including agriculture, utilities, and daily life, weather forecasting has been a significant factor. In the past ten years, the world has faced real-time difficulties with weather forecasting. Because of the constantly shifting weather, the prediction is getting more difficult. The goal of weather forecasting is to foresee future changes in the atmosphere. Understanding the numerous contributing elements that lead to weather changes is essential for effective weather analysis. The process of recording meteorological variables, such as wind direction, wind speed, humidity, rainfall, temperature, etc., is known as weather forecasting. Since machine learning techniques are more robust to perturbations, in this project we applied Neural Network with DNN regressor models and LSTM to predict the weather such as temperature, humidity etc. and compare both approaches and analyzed it. We used two different datasets for the same. Coming to result that we got from each approach was quite amazing. In the Neural Network with DNN regressor approach, we got mean absolute error about mean absolute 1.49 mm and median absolute error 0.94 Celsius and explained variant 0.90 when performing rainfall and temperature prediction respectively whereas in the deep learning approach, the mean absolute error was 0.002268 degree Celsius, when performing temperature, wind speed and pressure prediction respectively. We could clearly see the difference between the outcomes.

Keywords: Data Mining, Machine Learning, UCI Dataset, Weather Forecasting, Deep Learning.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Weather conditions change swiftly, so weather forecasting is vital. The practice of gathering information about atmospheric conditions, such as temperature, humidity, rainfall, wind speed and direction, etc., is known as weather forecasting. High-speed computers, wired and wireless sensors, meteorological satellites, and weather radars are the instruments used to collect weather data for weather forecasting.

The economy, particularly agriculture but also other sectors, can be significantly impacted by the weather. Energy prices significantly increased in 2005 after the hurricanes Katrina and Rita shut down several oil and gas rigs in the Gulf of Mexico, and in 2006 the Californian freeze cost citrus growers hundreds of millions of dollars in damages and resulted in massive unemployment. There are many applications that this system uses. Pattern recognition forecasting methods, this relied detecting patterns of events. The sunset is exceptionally red, the following day's weather has been proven to be pleasant. All of the forecasts, however, this turns out incorrect. We used characteristics such as average temperature and cloud cover to forecast rainfall in this system. Linear regression and artificial neural networks are two machine and deep learning algorithms that were used. These algorithms are built to use a corpus of historical weather data spanning Denmark and India.

By iterating and adding a portion of the mistake to the input data, the Linear Regression method is changed to get the lowest possible error percentage. Using various atmospheric data, such as the average temperature and cloud cover, this method offers a rainfall estimate. The collection of data is subjected to linear regression, and the coefficients are utilized to forecast the amount of precipitation based on the corresponding values of the parameters. The primary benefit of this model is that it predicts rainfall based on historical correlations between several atmospheric parameters. As a result, it is simple to find an estimated number for the amount of rainfall that might occur at a specific location and time. Globally, the weather is constantly and quickly changing. Today's world depends on accurate forecasts to get through. We significantly depend on weather forecasts for everything from agriculture to industry, transport, and daily commuting. In order to maintain simple and seamless movement as well as safe day-to-day operations, it is crucial to predict the weather accurately given that the entire planet is suffering from ongoing climate change and its repercussions. Applying science and technology to forecast the weather is known as weather forecasting. It involves determining the atmosphere's state at a specific location and moment in time. Using meteorology to predict how the atmosphere will change, weather forecasts are created by gathering quantitative data about the atmosphere's current condition at a specific location.

The usage of weather forecasting in the twenty-first century spans a wide range of applications, from monitoring industrial conditions to keeping track of weather conditions in agricultural fields. Temperature, wind, humidity, and other climatic behaviors can all be recorded with the use of weather forecasting.

In this project, our goal is to forecast the weather in Jaipur, India. The project's dataset includes several weather-related features from 1996 to 2017. Numerous features, including the date, time, wind, temperature, humidity, climate, and weather, are included. In this project, time series forecasting is the major methodology used.

Motivation

1.2 On many facets of human life, the weather has a significant influence. In order to ensure a higher quality of life, efforts have been done for years to increase weather prediction accuracy. We're going to learn about the major significance of weather forecasting in this post.

Rationale of the Study

1.3 The accuracy of weather forecasts has never been perfect. This does not imply, however, that people shouldn't rely on it. The likelihood that the scientists' forecasts will come true or not is typically expressed as a probability. Most of the time, the probabilities are less than 0.6. It is questioned if weather forecasters gather all the information necessary to make weather predictions. To anticipate the weather, these experts can gather almost 10,000 elements. They make use of advanced equipment like computers.They make use of balloons and satellites. This is typically done every six hours. These components come from different parts of the world. Despite the fact that some of these pieces are still uncollected, it seems like there are too many of them.

Research Question

1.4

A. What issues and theories are raised in the area of global warming?
B. What number of weather observatories does your country's hydrological or weather service have? Do you believe they will be entirely automated in two years?
C. Exist any financial institutions or businesses that factor predicted climatic changes into their business decisions?

Project Management and Finance

1.5 The definition of a forecast is the first step in the chapter's description of the fundamentals of strategic decision forecasting for project financing. It dispels myths about what makes a successful forecast. The chapter provides examples of what needs to be projected for project financing negotiations as well as sources of forecasts from organizations other than the sponsor. It covers the methods for creating forecast assumptions and carrying out sanity tests. The chapter provides an overview of the complex project forecasting procedure. It discusses the meaning, function, and application of project demand analysis as well as the choice of a forecasting strategy.

CHAPTER 2

Background

Preliminaries

2.1 The state of the atmosphere, or weather, can be described in terms of things like how hot or cold, wet or dry, quiet or stormy, clear or foggy, etc. The troposphere, the lowest part of the atmosphere that lies just below the stratosphere on Earth, is where the majority of meteorological events take place. While the term "climate" refers to the average of atmospheric conditions over longer periods of time, "weather" refers to the day-to-day variations in temperature, precipitation, and other atmospheric factors. "Weather" is typically taken to refer to Earth's weather when used without more explanation.

Related Works

2.2 In similar studies, attempts to forecast the weather used a wide range of fascinating techniques. Applications of modern forecasting technology frequently involve simulations based on physics and differential equations, but many approaches for artificial intelligence that have recently been developed Machine learning methods are dominated by neural networks. Some people employed probabilistic models, while others used Bayesian models networks. Three papers on machine learning combined into one Two of the three weather prediction techniques we examined used neural networks, while one used vector machines. Neural networks tend to be the most popular machine learning models for weather forecasting because they knowledge of historical non-linear relationships in contrast to past weather patterns and predictions Among these are models of linear regression. The benefit of not assuming that all features in our models have simple

linear dependence is provided by this. Of among the two neural network methods, one [3] utilized a hybrid a physics-based model that made use of neural networks forecasting the weather, while the other [4] applied learning how to forecast weather more directly. Similarly, the support vector machine strategy [6] used the classifier directly to predict the weather but had a smaller range than the neural network approaches.

Scope of The Problem

2.3 Weather a sophisticated and frequently difficult talent, weather forecasting requires intensive observation and data processing. Weather systems can range in size from small, brief thunderstorms with a few miles in diameter that last a few hours to massive showers and snowstorms with a diameter of up to a thousand miles that continue for days. In the end, forecasting is a three-step procedure. These comprise

a. Observing
b. Forecasting
c. Communicating

The first thing a forecaster does when they start working is familiarize themselves with the current weather conditions. This include studying satellite images, surface information, precipitation reports, and getting a briefing from other forecasters who are currently on duty. The following task is to predict weather changes in the future in order to create a prediction. Looking just a few hours into the future, short-range forecasting often relies on closely monitoring the development and tracking of existing weather systems and extrapolating their future motion using knowledge of the physics of the atmosphere

Challenge

2.4 A lot of data has to be collected to get weather forecast. Then it becomes very difficult to tend the data to different algorithms. Because we have to work with many algorithms for best results.

CHAPTER 3

Research Methodology

Research Subject and Instrumentation

3.1 Our research topic is how to easily predict the weather. As a benefit of this, we can easily know about the weather, even though it is not possible to get 100% weather results, still about 90% can be known with it.

Data Collection Procedure

3.2 The creation of weather forecasts involves gathering information about the atmosphere's current condition and predicting its future evolution using knowledge of atmospheric processes. What makes forecasting challenging is the chaotic nature of the atmosphere together with the lack of a thorough understanding of atmospheric processes. A meteorologist must first be aware of the current weather conditions and the factors that are causing them before they can make a forecast. This is accomplished by checking over a sizable amount of observation data, including surface observations, satellite images, radar, radiosonde, upper-air, wind profilers, airplane observations, river gauges, and just plain gazing outdoors. We collect data from Kaggle and uci website.

3.3 Data Description:

mean_temp: mean air temperature

max_temp: mean daily maximum air temperature

min_temp: mean daily minimum air temperature

meanhum: mean relative humidity

meandew: mean dew point temperature

pressure: mean daily air pressure

heat: true when mean air temperature is over or equal to 30

wet: true when mean relative humidity is over or equal to 80

Mean_cloud: mean cloud

population: population density

Sunshine_hour: mean number of hour of sunshine

Wind_direction: mean wind direction

Wind_speed: mean wind speed

Air_health_quality: mean daily air health quality

System Requirement:

Python3.6, BeautifulSoup, Pandas, Numpy, Matplotlib, Seaborn, Openpyxl, Sklearn, wxPython

Developer Tools:

Programming Language: Python

IDE: PyCharm

GUI: wxPython, wxFormBuilder

Web Scraping: BeautifulSoup, ParseHub

Debugging & Testing: Jupyter Notebook

Data Format: Microsoft Excel

Deferent Approaches

3.4 For the weather forecasting, we try to solve the problem using two methods and those were-

A. MACHINE LEARNING
B. DEEP LEARNING

We divide the problem using two methods and we have done so because of the following reasons-

- Tackle the problem with different approaches.
- Will know what are the strengths and weaknesses of each method
- Will know what are the difficulties one could face while approaching the problem with these two methods.
- Will know what is the efficiency of each process.

Coming to the first approach is Machine Learning, here we have solved the problem of weather forecasting with linear regression and Neural Network with DNN regressor models. In this approach, weather forecasting like rainfall prediction and temperature prediction is done separately. Also, the linear model which was trained for temperature predictions and rainfall predictions were trained using ordinary least square and regularization respectively.Now coming to the second approach is Deep learning, here we have solved the problem of

weather forecasting in a single stroke. Mean, only in one solution, we have done the forecasting of the temperature, wind speed, and pressure. We have trained our model on LSTM (advanced version of the RNN). Forecasting the weather using machine learning is a straightforward yet effective application. Climate is regarded as a complex system by physicists. Although there are many ways to understand it, in this particular instance we can say that complexity is the inability to be solved analytically. This may be discouraging, but it really makes a wide range of numerical techniques that are meant to address the issues brought on by climate change possible. The most recent developments in computers unquestionably include machine learning algorithms. I would want to discuss an issue that occurs when attempting to estimate the average temperature using traditional machine learning techniques, such as Neural Network with DNN regressor. Despite the fact that this post's main focus is not on the theoretical underpinnings.

CHAPTER 4

Experimental Result and Discussion

FIRST APPROACH USING MACHINE LEARNING

4.1 The first set of rules that we used turned into auto-regressive integrated moving average models, which seeks to expect excessive and coffee temperatures as a linear mixture of the features. Since linear regression can't be used with category data, this set of rules no longer uses the climate category of every day.

The libraries:

For data processing, plotting, and mathematical operations, the most well-known libraries have been employed (pandas, matplotlib, numpy etc). There are also some of them for sophisticated data visualization (such as folium), and some of them are particular libraries for Neural Network with DNN regressor models (like statsmodels). The import code can be found below.

Collecting Data Set, and Filter Process:

We started to collect the dataset and we look out to various uci websites and found a dataset on () which consists of various attributes.

Data Visualization

```
city_data = data.drop_duplicates(['City'])

city_data.head()
```

	dt	AverageTemperature	AverageTemperatureUncertainty	City	Country	Latitude	Longitude
0	1849-01-01	26.704	1.435	Abidjan	Côte D'Ivoire	5.63N	3.23W
1977	1850-01-01	15.986	1.537	Addis Abeba	Ethiopia	8.84N	38.11E
3942	1796-01-01	19.649	2.286	Ahmadabad	India	23.31N	72.52E
6555	1791-05-01	20.836	1.993	Aleppo	Syria	36.17N	37.79E
9224	1791-05-01	20.772	1.848	Alexandria	Egypt	31.35N	30.16E

Figure 1.1. Average Temperature Uncertainy

These cities latitude and longitude must be slightly modified if we wish to place them on a world map. Let's use a few lines of sudo code to achieve that:

```
LAT = []
LONG = []
for city in city_data.City.tolist():
    locator = Nominatim(user_agent="myGeocoder")
    location = locator.geocode(city)
    LAT.append(location.latitude)
    LONG.append(location.longitude)
```

Figure 1.2

2 number point is Jaipur location [REFERENCE 2]

The Data Preprocessing and Visualization, Stationarity Model:

Isolating Chicago allowed me to focus on its data, which I am using as my dataset. There is no unique justification for doing it. Simply said, I enjoy Chicago. Naturally, you are free to select your own city and carry out the subsequent procedures using your own dataset.

Figure 1.3 This picture is NonNaN Elements and NaN Elements Deference

I've chosen to fill the missing values with the previous ones because they are not a consistent component of the dataset. In the same way, I calculated the average temperature uncertainty

The year and month details are in the column labeled "dt." It is more practical to convert this column into a datetime object for use in the next operations and to clearly state the year and the month in two separate columns. You can achieve it using the following lines.

	dt	AverageTemperature	AverageTemperatureUncertainty	City	Country	Latitude	Longitude	Year	Month	Day	Week
2	1744-01-01	8.766	2.357	Chicago	United States	42.59N	87.27W	1744	1	1	
14	1745-01-01	-0.901	2.649	Chicago	United States	42.59N	87.27W	1745	1	1	
26	1746-01-01	-0.966	2.577	Chicago	United States	42.59N	87.27W	1746	1	1	
38	1747-01-01	-0.966	2.577	Chicago	United States	42.59N	87.27W	1747	1	1	
50	1748-01-01	-0.966	2.577	Chicago	United States	42.59N	87.27W	1748	1	1	

Figure 1.4

	meantempm	meandewptm
date		
2018-03-12	27	3
2018-03-13	28	1
2018-03-14	27	2
2018-03-15	26	5
2018-03-16	26	4
2018-03-17	24	2
2018-03-18	25	1
2018-03-19	26	3
2018-03-20	27	7
2018-03-21	26	2

Figure 1.5

This is the dataset meantempm and meandewptm

	meantempm	meandewptm	meantempm_1
date			
2016-05-01	34	-1	NaN
2016-05-02	36	4	34.0
2016-05-03	35	6	36.0
2016-05-04	34	7	35.0
2016-05-05	31	11	34.0
2016-05-06	28	13	31.0
2016-05-07	30	10	28.0
2016-05-08	34	8	30.0
2016-05-09	34	11	34.0
2016-05-10	34	16	34.0

Figure 1.6

The target measurement of mean temperature and total number of rows

a list of the feature's Nth previous measurements. Observe that N None values must be added to the head of the list to preserve the same number of rows for each N. Then add a new column to the data frame with the name "feature."

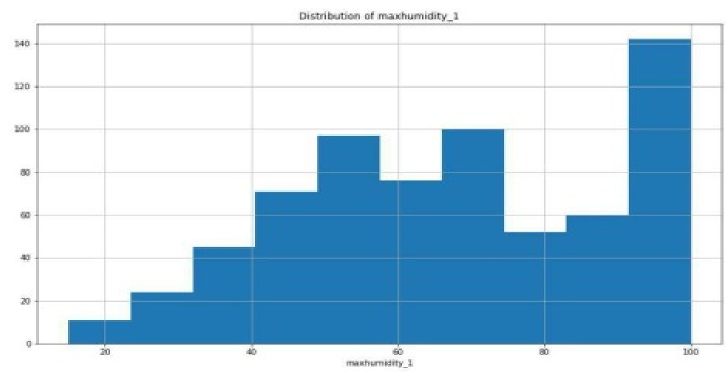


Figure 1.7 This Figure 1.7 and Figure 1.8 Distribution of max humidity graph

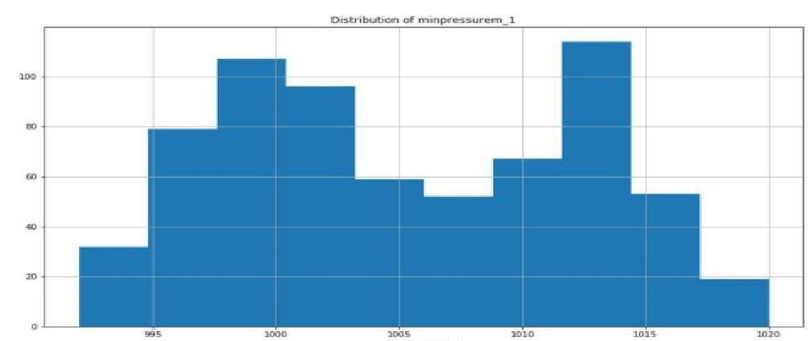


Figure 1.8 Distribution of max humidity graph

This dataset can be used to create a scatter plot like this one:

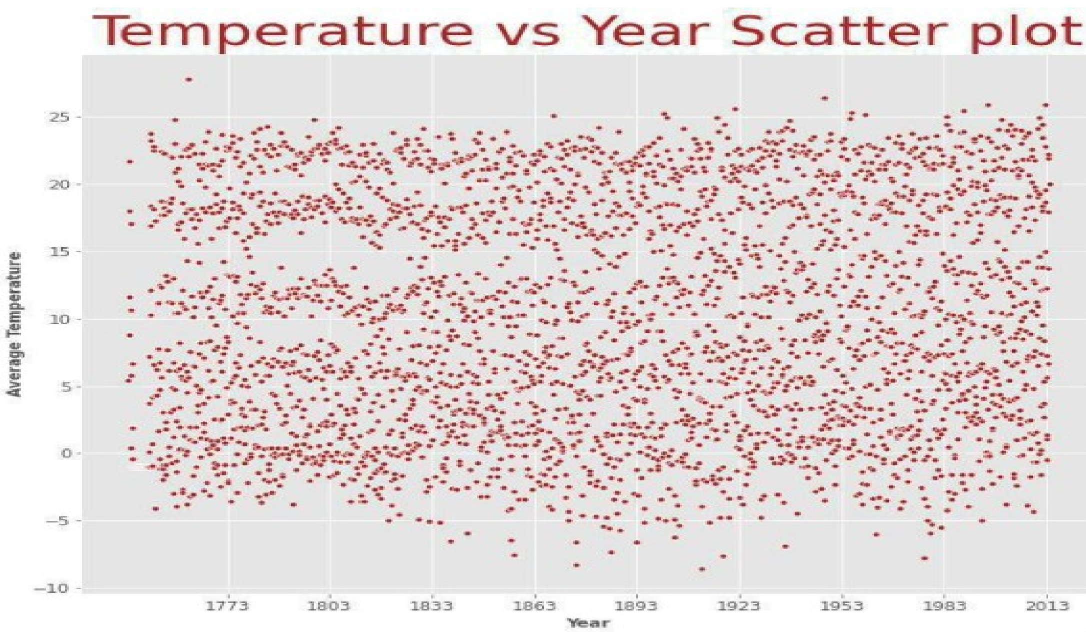


Figure 1.9 Temperature vs Year Scatter Plot

But because reading it is difficult, we need to improve. Let's now discuss three incredibly fundamental functions:

•The get timeseries (start year, end year) function extracts the data between the two years from the dataset.

•The readable timeseries acquired by get timeseries are plotted using plot timeseries (start year, end year).

•The readable data (Average Temperature) is plotted wrt the time (dt) using the plot from data (data, time, display options) function.

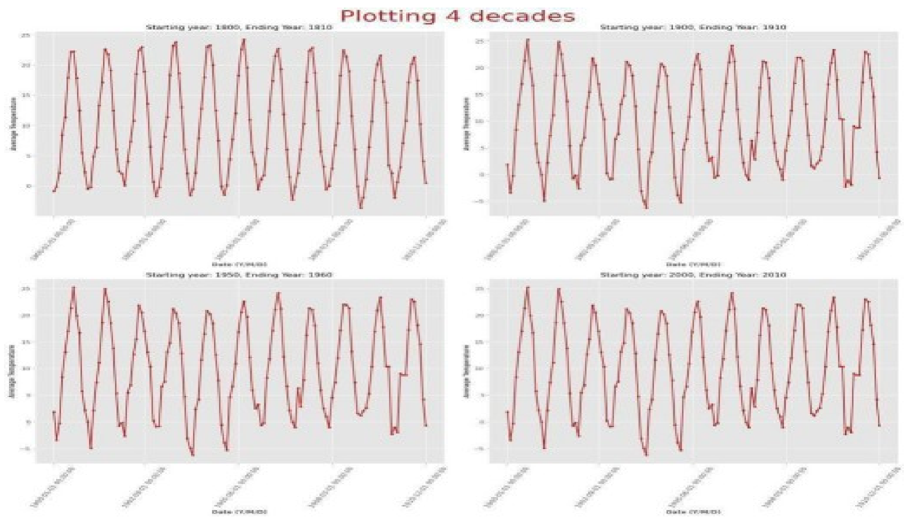


Figure 1.10 Plotting 4 decades

When applying Neural Network with DNN Regresor models, stagnant time series should be taken into account. We can look at the correlation and autocorrelation plots to see if the timeseries under consideration is stationary

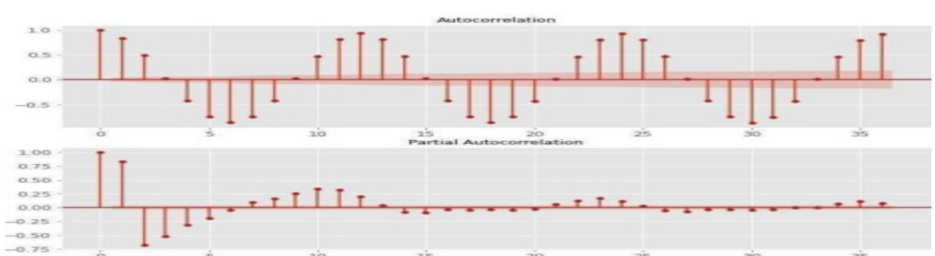


Figure 1.11 Checking on Stationary

It hints to the non-stationarity of the timeseries. However, the AD Fuller Test reveals that the dataset is stationary when applied to the complete collection.

```
ADF Statistic on the entire dataset: -6.029493829973582
p-value: 1.429253097986233e-07
Critical Values:
    1%: -3.4323875260668344
    5%: -2.862440255934873
    10%: -2.5672492261933377
```

But simply because we are examining the complete dataset, it is accurate. In reality, it is obvious that the dataset is not stable over a decade if we only look at one decade of data

```
ADF Statistic on the first decade: -2.096122487386037
p-value: 0.2460766247103494
Critical Values:
    1%: -3.4901313156261384
    5%: -2.8877122815688776
    10%: -2.5807296460459184
```

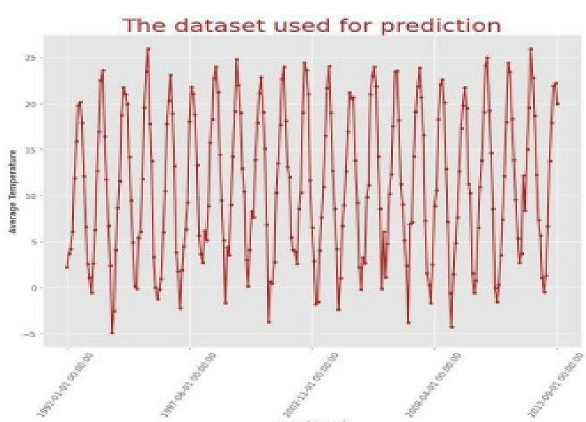


Figure 1.12 Dataset Used for Prediction

The Neural Network with DNN Regresor models are the machine learning algorithms. These have an optimization process that uses the Maximum Likelihood function as their foundation.

```
result_d_0.head()
```

	(p, d, q)	AIC
0	(5, 0, 7)	1095.997844
1	(4, 0, 6)	1097.294458
2	(3, 0, 5)	1097.693577
3	(4, 0, 5)	1097.920777
4	(5, 0, 6)	1098.095676

First differentiated model:

The (2,1,5) and (2,1,6) models are the best ones, according to the overall summary, which is highlighted by this function. As can be seen, the statistical summary results are nearly identical.

SARIMAX RESULT:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:     AverageTemperature   No. Observations:                  248
Model:               SARIMAX(2, 1, 5)   Log Likelihood                -534.958
Date:                Sat, 17 Apr 2021   AIC                           1085.917
Time:                        11:42:44   BIC                           1113.992
Sample:                             0   HQIC                          1097.220
                                - 248
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.7320      0.001   3440.332      0.000       1.731       1.733
ar.L2         -1.0000      0.000  -6492.021      0.000      -1.000      -1.000
ma.L1         -2.4616      0.449     -5.478      0.000      -3.342      -1.581
ma.L2          2.0863      0.543      3.841      0.000       1.022       3.151
ma.L3         -0.5374      0.302     -1.779      0.075      -1.130       0.055
ma.L4          0.0227      0.232      0.098      0.922      -0.432       0.477
ma.L5         -0.1099      0.078     -1.403      0.161      -0.263       0.044
```

```
sigma2          4.2305      0.114     37.023      0.000       4.007       4.454
===================================================================================
Ljung-Box (Q):                       68.29   Jarque-Bera (JB):                 8.90
Prob(Q):                              0.00   Prob(JB):                         0.01
Heteroskedasticity (H):               1.36   Skew:                             0.12
Prob(H) (two-sided):                  0.17   Kurtosis:                         3.90
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.48e+19. Standard
errors may be unstable.
                               SARIMAX Results
==============================================================================
Dep. Variable:     AverageTemperature   No. Observations:                  248
Model:               SARIMAX(2, 1, 6)   Log Likelihood                -535.530
Date:                Sat, 17 Apr 2021   AIC                           1089.060
Time:                        11:42:45   BIC                           1120.644
Sample:                             0   HQIC                          1101.776
                                - 248
```

```
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.7320      0.001   3339.099      0.000       1.731       1.733
ar.L2         -0.9999      0.000  -3122.916      0.000      -1.001      -0.999
ma.L1         -2.4409      0.783     -3.116      0.002      -3.976      -0.906
ma.L2          1.9998      1.250      1.599      0.110      -0.451       4.451
ma.L3         -0.4058      0.523     -0.777      0.437      -1.430       0.618
ma.L4         -0.0739      0.271     -0.273      0.785      -0.605       0.457
ma.L5         -0.0742      0.192     -0.387      0.698      -0.450       0.301
ma.L6         -0.0049      0.077     -0.064      0.949      -0.155       0.145
sigma2         4.1416      3.348      1.237      0.216      -2.420      10.703
===================================================================================
Ljung-Box (Q):                       65.25   Jarque-Bera (JB):                 9.13
Prob(Q):                              0.01   Prob(JB):                         0.01
Heteroskedasticity (H):               1.34   Skew:                             0.13
Prob(H) (two-sided):                  0.18   Kurtosis:                         3.90
===================================================================================
```

```
Dep. Variable:     AverageTemperature   No. Observations:                  248
Model:               SARIMAX(2, 1, 5)   Log Likelihood                -534.958
Date:                Sat, 17 Apr 2021   AIC                           1085.917
Time:                        11:42:44   BIC                           1113.992
Sample:                             0   HQIC                          1097.220
                                - 248
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.7320      0.001   3440.332      0.000       1.731       1.733
ar.L2         -1.0000      0.000  -6492.021      0.000      -1.000      -1.000
ma.L1         -2.4616      0.449     -5.478      0.000      -3.342      -1.581
ma.L2          2.0863      0.543      3.841      0.000       1.022       3.151
ma.L3         -0.5374      0.302     -1.779      0.075      -1.130       0.055
ma.L4          0.0227      0.232      0.098      0.922      -0.432       0.477
ma.L5         -0.1099      0.078     -1.403      0.161      -0.263       0.044
```

So, this is SARIMAX result and the statistical graphs also exhibit the same behavior

RESULT MODEL (2,1,5)

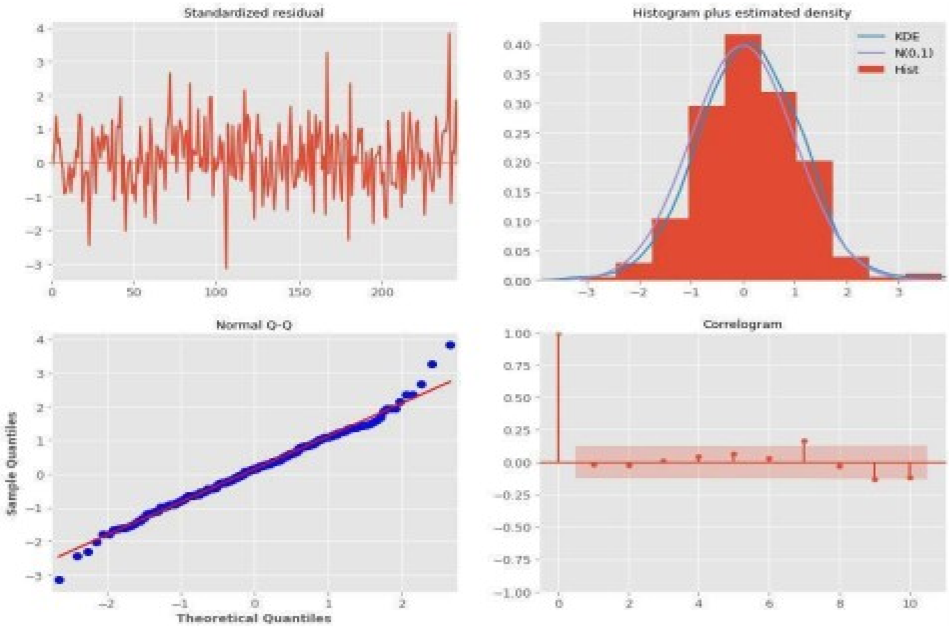


Figure 1.13 Result Model (2,1,6)

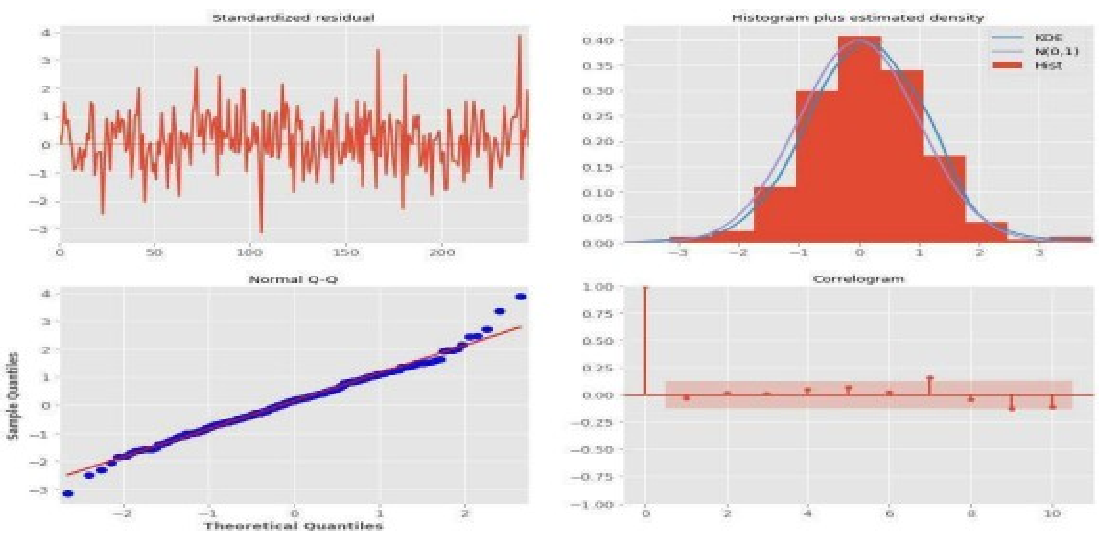


Figure 1.14

However, using low index models is preferred because it prevents overfitting and puts less strain on your computer's processing power. The (2, 1, 5) model has been taken into consideration because of this. Let's plot the results of the forecasting operation:

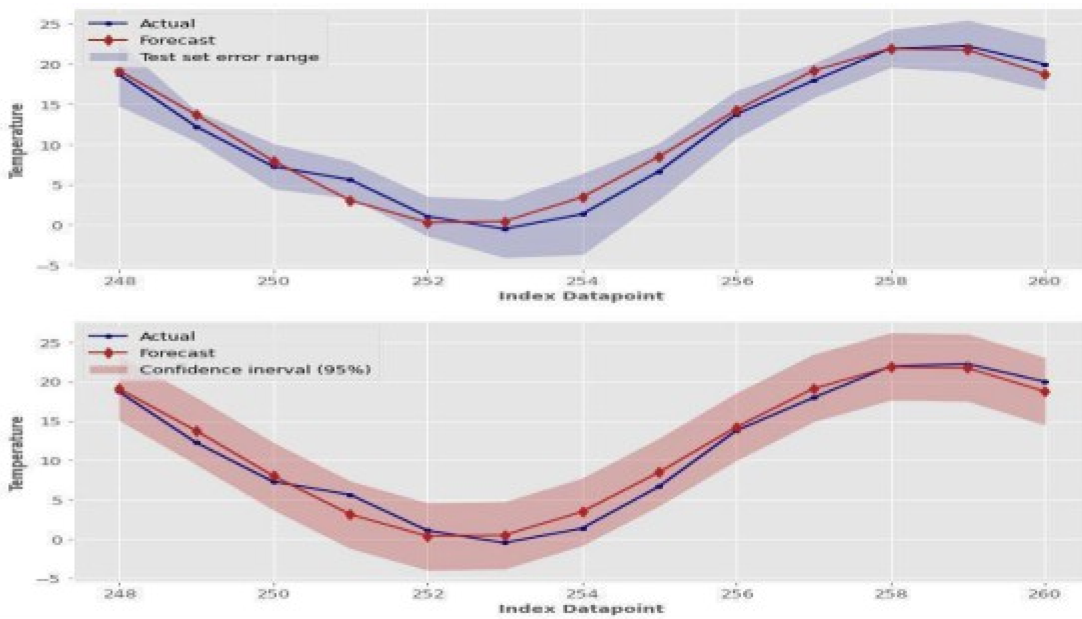
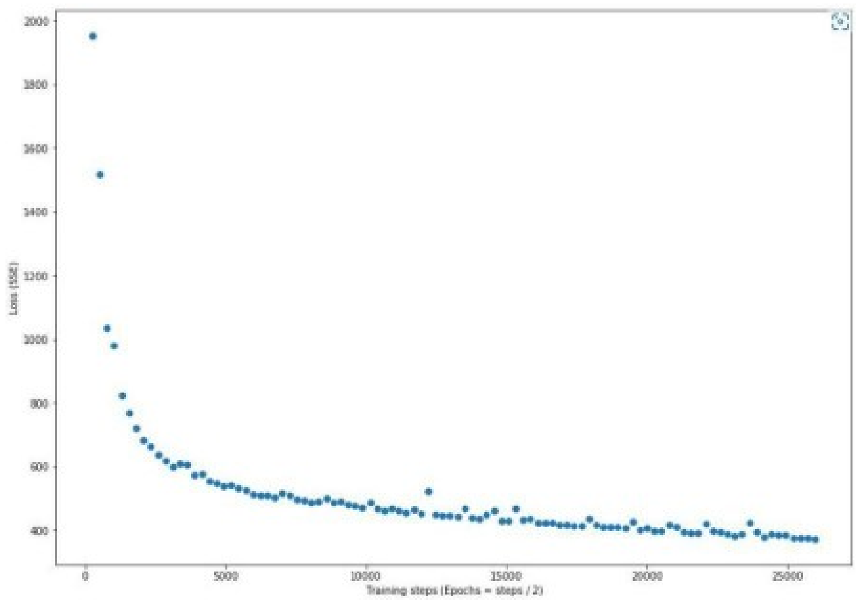


Figure 1.15

Neural Network with DNN Regressor Model Accuracy Result:

```
Training instances   540, Training features   36
Validation instances 68, Validation features 36
Testing instances    68, Testing features    36
```




```
INFO:tensorflow:Restoring parameters from tf_wx_model\model.ckpt-26000
The Explained Variance: 0.90
The Mean Absolute Error: 1.49 degrees Celcius
The Median Absolute Error: 0.94 degrees Celcius
```

Figure 1.16

Time Series Equation:

```
Accuracy for class Rain  is: 90.0 %
---------------------------------------
Network Accuracy: 90.0 %
---------------------------------------
              precision    recall  f1-score   support

           0       0.97      0.99      0.90        68
           1       1.00      1.00      1.00        36
           2       1.00      0.96      0.90        48
           3       0.97      0.99      0.90        74

    accuracy                           0.90       226
   macro avg       0.90      0.98      0.89       226
weighted avg       0.90      0.98      0.88       226
```

SECOND APPROACH USING DEEP LEARNING

4.2 Dataset

For five Danish cities, we'll use weather information from 1980 to 2018.The National Climatic Data Center (NCDC), in the USA, provided the original raw weather data. We are attempting to forecast the weather for the Danish city of "Odense" over the next 24 hours using weather data from 5 cities (although the flowchart below only shows 2 cities). In the dataset, there are four features basically temp, pressure, wind speed and wind direction. A small snippet is given below of the dataset. Time varies for every 1 hour.

	Aalborg				Aarhus			
	Temp	Pressure	WindSpeed	WindDir	Temp	Pressure	WindSpeed	WindDir
DateTime								
1980-03-01 11:00:00	5.000000	1007.766667	10.2	280.000000	5.0	1008.300000	15.4	290.0
1980-03-01 12:00:00	5.000000	1008.000000	10.3	290.000000	5.0	1008.600000	13.4	280.0

Figure 2.1 Temp, pressure,windspeed [Ref: kaggle website]

LSTM

Long Short-Term Memory (LSTM) recurrent neural networks (RNNs) were developed to more accurately replicate temporal sequences and their long-range relationships. As information spreads ahead, it analyses data and passes it along. The actions taking on within the LSTM's cells are what differ. The cell state and its many gates form the basis of LSTMs. Recurrent neural networks are given long-term memory thanks to LSTM. It reduces the vanishing gradient problem, which occurs when a neural network stop learning because the updates to its various weights get progressively smaller. It accomplishes this by using a number of "gates."

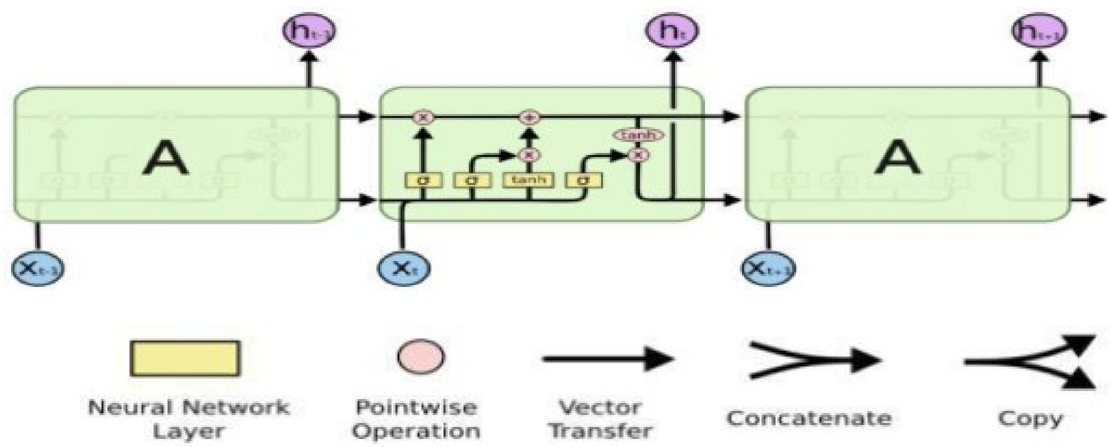


Figure 2.2 LSTM Cell

We have three different gates that regulate information flow in an LSTM cell. A forget gate, input gate, and output gate.

Forget gate

First is the lost gate. What data should be destroyed or kept is decided by this gate. Both data from the current input and data from the prior hidden state are processed by the sigmoid function. The range of values is 0 to 1. To maintain implies moving farther from 1, and to forget means moving farther from 0.

Input Gate

We have the input gate to change the state of the cell. First, a sigmoid function is used to process the current input and the prior concealed state. By converting the data to a range of 0 to 1, where 0 denotes importance and 1 signifies importance, that determines which values will be updated. Additionally, you supply the tanh function with the hidden state and current input to squish values between -1 and 1, which aids in network regulation. The output of the sigmoid is then multiplied by the output of the tanh. The information that should be retained from the tanh output will be determined by the sigmoid output.

Output Gate

The output gate comes last. What should be the next concealed state is determined by the output gate. Keeping this in mind, the concealed state comprises data about prior inputs. For forecasts, the concealed state is also utilized. We start by feeding a sigmoid function the current input as well as the prior concealed state. Then, we provide the tanh function the newly updated cell state. The information the hidden state should contain is determined by dividing the tanh output by the sigmoid output. The concealed state is what is produced. Following that, the new cell state and new concealed are carried over to the following time step.

Cell State

Now, we ought to have enough knowledge to determine the cell state. The forget vector is first pointwise multiplied by the cell state. If this is multiplied by values close to 0, there is a chance that the values in the cell state will drop.

Result LSTM Algorithm

Here is the Temperature prediction, with temp (in F) & days on y-axis and x-axis respectively.

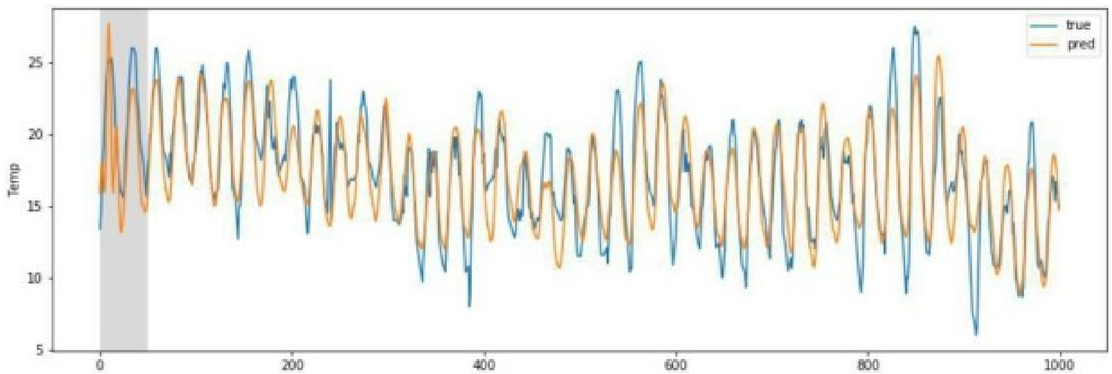


Figure 2.3 Temperature Prediction

In the above graph, we could observe that there is not much variance in the predicted result and observed the result except for some points.

Here is the wind speed prediction, with wind speed (km/hr) & days on y-axis & x-axis respectively.

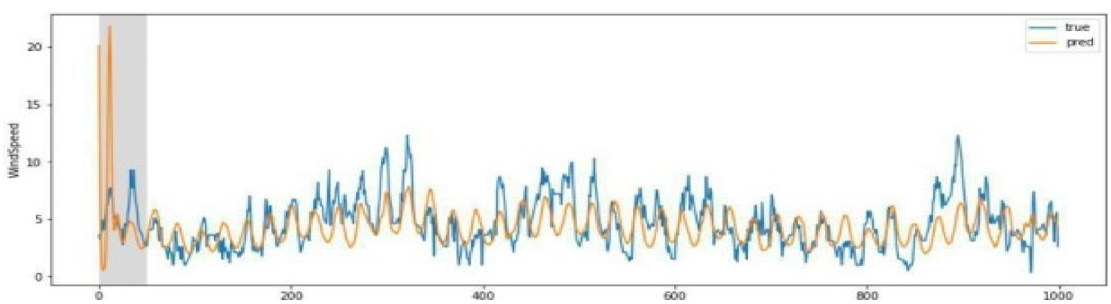


Figure 2.4 Wind Speed Prediction

In the above graph, we could observe that there are is not much variance in the predicted result and observed the result except for some points.

Here is the pressure prediction, with pressure (in pascal) & days on y-axis and x-axis respectively.

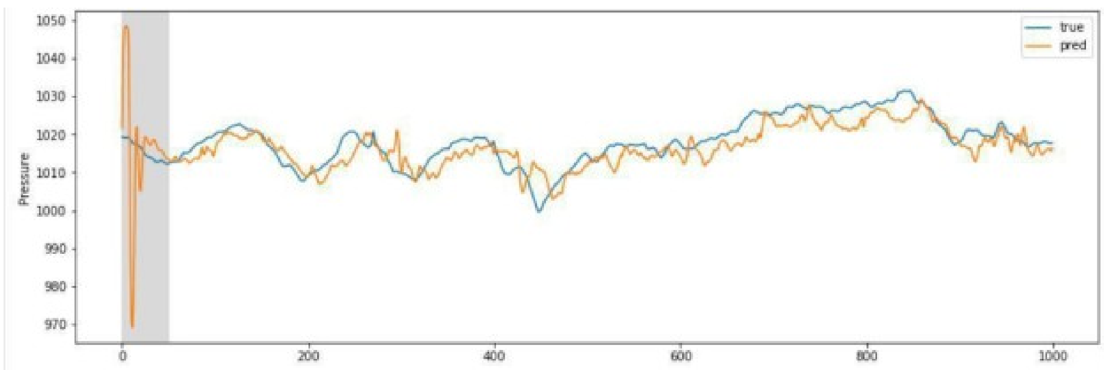


Figure 2.5

In the above graph, in starting you could see that variation is large but later on it is good. Also, we could observe that there are is not much variance in the predicted result and observed the result.

```
print("The temperature loss is : %f degree celsius"%temp_loss)
print("The wind speed loss is : %f kmh"%windspeed_loss)
print("The pressure loss is : %f pascal"%pressure_loss)
```

```
The temperature loss is : 0.002268 degree celsius
The wind speed loss is : 0.003266 kmh
The pressure loss is : 0.003068 pascal
```

Figure 2.6

CHAPTER 5

Impact on Society, Environment and Sustainability

Impact on Society

5.1 Now a days society has become more sensitive about weather and it's forecasts. Climates are changing and many storms, disasters, floods and other kinds of various climate disaster are affecting our normal life . it also affects our economy systems. Predicting weather has become a key part of everyone's daily routine . Improving weather forecast has now been an important issue So that, We can predict the weather more accurately and decrease the risk of our health and economic system.

Impact on Environment

5.2 Climates are changing along with day to day. Natural disasters like floods, Storms are affecting our environment. It has become a common issue for some of the areas every year. If we couldn't follow the weather forecast it will be a larger problem in future . Our economic systems are affecting also. To reduce the risk of these disaster it has now been an issue to predict the weather quite easily so that we can find a solution to save our environment.

Ethical Aspects Sustainabiliy Plan

5.3 Climate change is totally a highly responsible and critical task. Depending on weather and climate ethics and to understand the concept of it's sustainability.To define the reason of th ethics of sustainability and also to predict the weather more accurately .

CHAPTER 6

Summary Conclusion Recommendation and Implication for Future Research

Summary of the Study

6.1 On the integration of meteorological and physical principles, weather forecasting is built. Weather forecasting allows for the prediction of surface changes caused by atmospheric conditions (snow and ice cover, storm tides, floods, etc.). The basis for scientific weather forecasting is a combination of computer-controlled mathematical models and empirical and statistical techniques, such as measurements of temperature, humidity, air pressure, wind speed and direction, and precipitation.

Conclusions

6.2 In this project, machine learning and deep learning are used to predict the weather forecasting considered Date, Minimum Temperature, Humidity, and Wind Direction as predictors for rainfall, and they have adopted Supplied test set as a test option. The Correlation coefficient of all base classifiers is greater than 0.8. In this study we considered only seven predictors for rainfall prediction, if we use some more climate factors such as atmospheric pressure, sea surface temperature, so we may obtain more accurate predictions. Also, if Ensemble methods have been applied the results may be improved. Weather forecasts are increasingly accurate and useful, and their benefits extend widely across the economy. While much has been accomplished in improving weather forecasts, there remains much room for improvement.

Implication for Further Study

6.3 For future improvements, following step we thought to took-

- Replacing model with a latest/different model
- Using other robust datasets
- Predicting result on more attributes
- Training model on higher-end GPU
- Weather application

Also, while performing weather forecasting, there was a lot of complexities involved. There are a lot of variables/attributes to consider for forecasting weather and if all or most of them are used, then we need a lot of computation power to get weather information. And, Real time weather forecasting is very difficult to forecast correctly.

REFERENCE

[1] Parmar, Aakash & Mistree, Kinjal & Sompura, Mithila. (2017). Machine Learning Techniques for Rainfall Prediction: A Review.

[2] Mateo, Fernando & Carrasco, Juan José & Sellami, Abderrahim & Millán-Giraldo, Mónica & Domínguez, Manuel & Olivas, Emilio. (2013). Machine learning methods to forecast temperature in buildings. Expert Systems with Applications. 40. 1061–1068. 10.1016/j.eswa.2012.08.030.

[3] Pooja Malik, Prof. Saranjeet Singh, Binni Arora. An Effective Weather Forecasting Using Neural Network

[4] Abhishek, Kumar & Singh, M. & Ghosh, Saswata & Anand, Abhishek. (2012). Weather Forecasting Model using Artificial Neural Network. Procedia Technology. 4. 311–318. 10.1016/j.protcy.2012.05.047.

[5] Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. Has¸im Sak, Andrew Senior, Franc¸oise Beaufays.

[6] Xiong, L., O'Connor, K.M.: An empirical method to improve the prediction limits of the glue methodology in rainfall–runoff modeling. J. Hydrol. 349(1), 115–124 (2008)

[7] Wu, J., Huang, L., Pan, X.: A novel Bayesian additive regression trees ensemble model based on linear regression and nonlinear regression for torrential rain forecasting. In: Third International Joint Conference on Computational Science and Optimization (CSO), vol. 2, pp. 466–470 (2010)

[8] Lin, G.F., Chen, L.H.: Application of an artificial neural network to typhoon rainfall forecasting. Hydrol. Process. 19(9), 1825–1837 (2005)

[9] Hong, W.C.: Rainfall forecasting by technological machine learning models. Appl. Math. Comput. 200(1), 41–57 (2008)

[10] Mellit, A., Pavan, A.M., Benghanem, M.: Least squares support vector machine for short-term prediction of meteorological time series. Theor. Appl. Climatol. 111(1–2), 297– 307 (2013)

[11] Rasel, R.I., Sultana, N., Meesad, P.: An efficient modeling approach for forecasting financial time series data using support vector regression and windowing operators. Int. J. Comput. Intell. Stud. 4(2), 134–150 (2015)

Ramim

ORIGINALITY REPORT

24%	21%	6%	22%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	15%
2	Submitted to University of Bolton Student Paper	2%
3	www.trp.org.in Internet Source	1%
4	towardsdatascience.com Internet Source	1%
5	www.ametsoc.org Internet Source	1%
6	Submitted to Middle East Technical University Student Paper	1%
7	Submitted to University of Hertfordshire Student Paper	1%
8	Submitted to Dr. S. P. Mukherjee International Institute of Information Technology (IIIT-NR) Student Paper	<1%
9	eur-lex.europa.eu Internet Source	<1%
10	Submitted to BITS, Pilani-Dubai Student Paper	<1%
11	sigport.org Internet Source	<1%
12	www.upgrad.com Internet Source	<1%
13	Submitted to Daffodil International University Student Paper	<1%
14	Submitted to National Institute Of Technology, Tiruchirappalli Student Paper	<1%
15	www.kdnuggets.com Internet Source	<1%