Modeling the temperature data of different areas of Bangladesh and simulate weather forecasting using Time-series analysis

By

Atanu Roy

19101267

atanu.roy@g.bracu.ac.bd

Section 1

Submitted For

CSE474: Simulation and Modelling

Final Report

Fall 2021

Submitted To

Meem Arafat Manab

Lecturer, Department of Computer Science and Engineering

Index

Introduction

Agriculture is Bangladesh's most important economic sector, with 62 percent of the country's population working in this field to earn their income. Around 87 percent of the country's population lives in rural regions and is directly or indirectly dependent on agriculture. This plays an important part in the economic development process and contributes significantly to the national GDP. It's contribution to GDP was 52 percent in 1972, according to the World Bank, but it has continuously decreased, and in FY 2016-2017, agriculture's contribution to GDP was just 14.79 percent. Agriculture production is declining as a result of climate change. Bangladesh is the most vulnerable country in the world to climate change, which causes river erosion, floods, flash floods, and salinization of the soil. Agriculture production is influenced by weather and climate. Agriculture is one of the most sensitive industries to climate change's consequences, such as temperature changes, rainfall patterns, and increased floods and droughts. But even with the changing climate of this region there is very little work done on this premise.

Keeping all this in mind, this project aims to use time-series analysis on the monthly temperature data of different districts of Bangladesh and then compare the accuracy of the result. This work will help us predict the temperature which we can use to stay ahead and take precautionary measures which can, in turn, cut down losses.

Literature review:

Here are some of the papers that I have read and repositories that I found which might help me in my project “Weather Forecasting with Bangladesh’s weather time series data”.

[1]
In the study titled “A time series analysis of weather variables and all-cause mortality in the Kasena-Nankana Districts of Northern Ghana, 1995–2010” the authors have tried to find the relationship between climate change and public health. For this they have taken the daily temperature and precipitation data and matched them against the daily mortality data grouped by age and gender from north Ghana.

They have used the time-series poisson regression method to check the involvement of daily mean temperature and precipitation data in the daily mortality data. They have also included the time factor and lagged weather predictors along with autocorrelation. Their results show strong correlation between the weather variability and mortality in northern Ghana. But they differ between age groups and genders.

In this research they concluded that temperature has high correlation with mortality rates. In order to do so they have conducted a time series analysis to model climate change. I can use this in my project as well. Although I don't agree with the conclusion. I think the conclusion here overshadows some of the more major contributors to these factors.

[2]
In the paper “A Time Series ANN Approach for Weather Forecasting” aimed towards accuracy for their weather forecast. They have used ANN and simulated modeling in MATLAB and predicted maximum and minimum temperature. They have trained their model with 60 years of data (1901-1960) then tested their model on 40 years for maximum and minimum temperature. They have concluded their model to be successful based on the MSE(mean square error) function.

Here they have only worked with temperature data, they could have tested their model on other weather parameters to compare if their model holds its effectiveness for all those other important parameters. However their integration of artificial neural networks with time series to increase efficiency is something that might be used in my project.

[3]
In the research titled “Weather Forecasting for Weather Derivatives” the researchers have worked with the daily average temperature measured in degrees obtained from four of their measurement stations: Atlanta, Chicago, Las Vegas and Philadelphia for 1/1/60
through 11/05/01, resulting in 15,285 observations at each station. Their collected data has also shown strong seasonality in the last 5 years. They have used low-ordered fourier series so model the seasonality. This produces a smooth seasonal pattern, which accords with the basic intuition that the progression through different seasons is gradual rather than discontinuous plus it promotes meanness, which enhances numerical stability in estimation. Their time series simulation finally reveals conditional mean dynamics and strong conditional variance dynamics in daily average temperature, and it reveals significant differences between the distribution of temperature and the distribution of temperature surprises.

[4]
In the paper titled “Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station”, the authors have worked with local weather station data and modern machine learning techniques for predicting the weather with time series analysis. The traditional weather forecasting models can give predictions every 6 to 18 hours with a region of every 20 km. But the main users of this kind of technology are mostly farmers who usually need much more precise location information than that, thus the need for their research. They have simulated their mode with TCN network or Temporal Convolutional Neural Network and compared the results with LSTM or Long Short-Term Memory. The results have shown that the TCNN outperforms the LSTM model in terms of prediction. They also claim their model to be deployable on a stand-alone personal computer. The local weather station is a home made raspberry pi based system.

[5]
In the study titled “Comparative Analysis of Temperature Prediction Using Regression Methods and Back Propagation Neural Network” the authors have done a comparative analysis between regression methods and a non-linear approach like neural network to predict the temperature. They have used Linear Regression, Regression Tree and Back Propagation Neural Network. They have concluded that non linear models such as neural networks perform significantly better than linear models like linear regression and regression tree due to their ability to predict with non linear patterns.

[6]
In the paper “Hybrid Model For Weather Forecasting Using Ensemble Of Neural Networks And Mutual Information”, the researchers have proposed a new hybrid model for weather forecasting based on a collection of neural network models. They have introduced a mutual information approach to tackle the challenge of combining the results of different networks and reducing the redundancy in the hybrid model. The hybridization happens in 4 levels. At the first level they have pre-processed the data The second stage included 4 different neural networks: GRNN, TDNN, RBF and MLP. In the third stage they have implemented the mutual information approach to select the best resulting model from the 4 models. The fourth stage is for combining the results from the best performing model. Their model has shown great accuracy and performed better than other similar networks like WA, Voting Scheme, SOM and other solo neural networks like RBF and GRNN.

[7]
In the study named “A Real-Time Weather Forecasting and Analysis”, the authors have used a self made weather with dht11, raspberry pi and an ethernet shield for cloud based data storing. Then they have used an ARIMA model with this time series data to predict the weather parameters. Based on their correlation results obtained from their data they have decided to go with temperature and humidity. They have used ACF (Autocorrelation Function) and PACF(Partial Autocorrelation Function) for this. In their results they have shown that their model is accurate for two consecutive years. But after that their predictions were labeled as inaccurate.

[8]
In the paper titled “Transductive LSTM for Time-Series Prediction: an Application to Weather Forecasting” the writers have used LSTM to get a weather forecasting model. They have also proposed Transductive LSTM(T-LSTM) which uses the local information of the time series prediction. In this study they have used a quadratic cost function for the regression problem. They have investigated two weighting schemes based on the cosine similarity between the training samples and the test point. In order to assess the performance of the proposed method in different weather conditions they have conducted the experiments on two different time periods of a year. The results show that T-LSTM results in better performance in the prediction task.

[9]
In this repository the owner used weather data from http://rp5.ru/. This data was collected from the weather station 2978 in Helsinki from January 2015 to September 2019. In his analysis the owner aimed to solve the following problems:
How to conduct a Time-Series Analysis project from raw data?
How to deal with a Time-Series data that has seasonality?
Implementing the Moving Average to capture long-term fluctuations
Parameter Selection for the model using GRID Search
Make Weather Forecasts using the SARIMAX model
How to diagnose the model performance using charts and other metrics (AIC score, RMSE value)?
The owner also wrote his procedures in this medium article:
https://medium.com/@llmkhoa511/time-series-analysis-and-weather-forecast-in-python-e80b664c7f71

This will be very helpful to start my project of analyzing the weather variables and forecast the weather based on the time series analysis.

[10]
This repository is titled “Deep Neural Network for Weather Time Series Forecasting”. Here the owner used the dataset from https://www.kaggle.com/selfishgene/historical-hourly-weather-data. Usually deep learning models work best with huge datasets but time series problems with that big dataset are not that common. So the owner wanted to find out what machine

learning models and neural networks might work best for a medium size dataset. In doing so he has used the following libraries: LinearRegression and RandomForest from Sci-kit Learn, ARIMA model from Statsmodel, LSTM and MLP from Tensorflow Keras Functional APIs are used.

[11]
In this repository titled “Weather forecast using recurrent neural network” the owner has used weather forecasting data with the goal to predict temperature of the next 12 or 24 hours. He has used RNN, GRU and LSTM and compared their results using mean square error and mean absolute error.

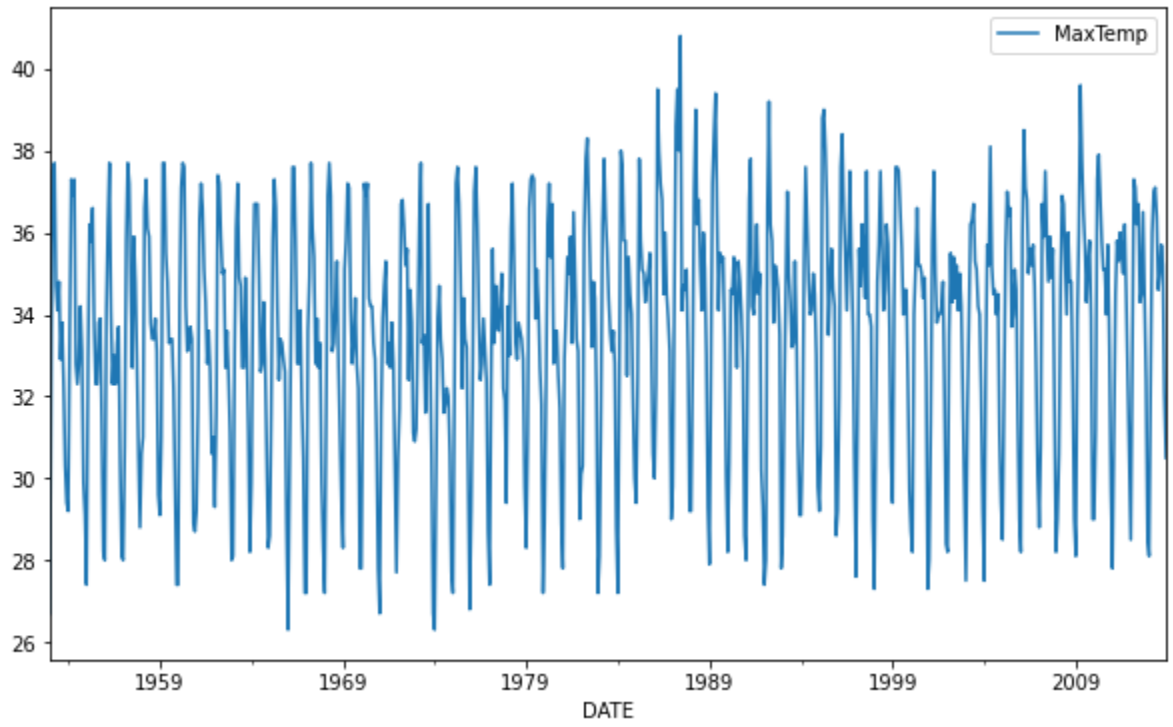
[12]
In the repository titled “Time-Series-Analysis-with-Python” the owner have used the following datasets:
Google Stocks Data
Humidity in different world cities from the Historical Hourly Weather Data 2012-2017 Dataset
Microsoft Stocks Data
Pressure in different world cities from the Historical Hourly Weather Data 2012-2017 Dataset
According to the owner this can be useful in the finance and statistics, time series decomposition and modelling using statsmodels AR, MA, ARMA, ARIMA, VAR and Dynamic Factor Models

Methodologies

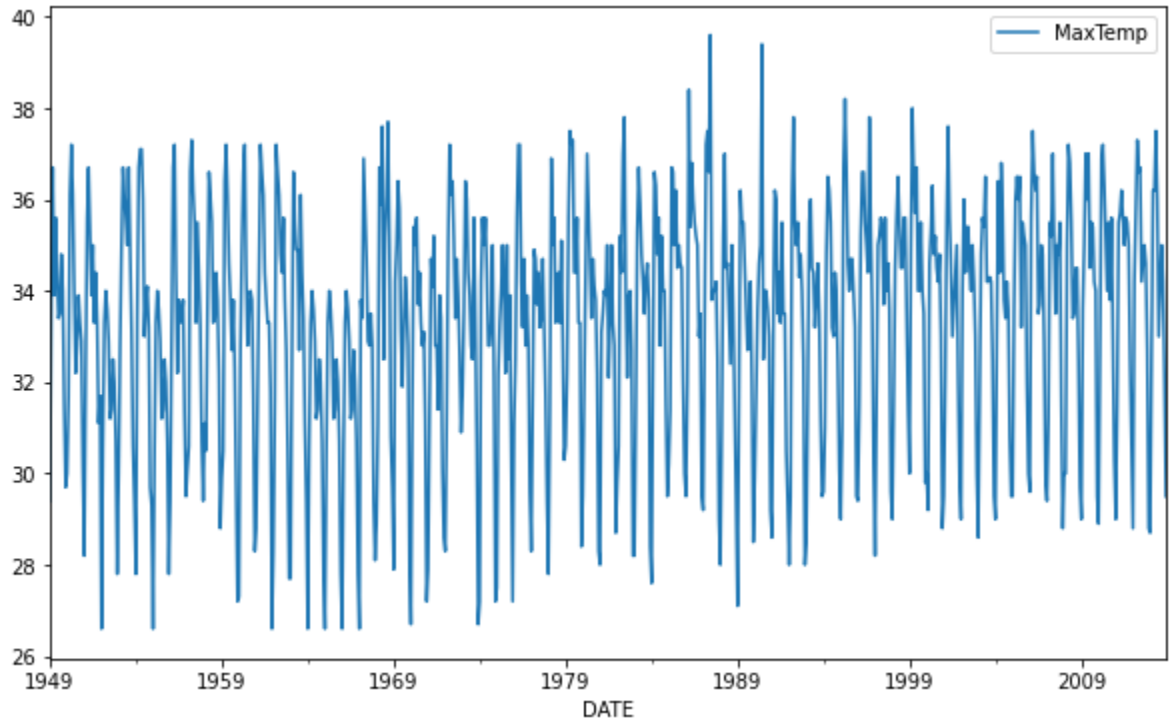
I have got the dataset from kaggle. The data was collected from Bangladesh Meteorological Department (BMD). The dataset contains Monthly average of Maximum Temperature, Minimum Temperature, Rainfall, Relative Humidity, Wind Speed, Cloud Coverage and Bright Sunshine of Bangladesh from the period 1948 to 2013 of specific areas. It also has the Weather Station Numbers, the X, Y coordinates, Latitude, Longitude and Altitude. The data was pre-processed to fill the null values. After loading the data to my notebook I have created a column

“DATE” combining the two columns “YEAR” and “Month”. For this project I have considered the MaxTemp and DATE columns. Later I have created dataframes for the following districts: Dhaka, Barisal, Chittagong(Patenga), Sylhet, Cox’s Bazar, Khulna and Rajshahi. Here are the data-frames plotted in a date vs maxTemp graph.

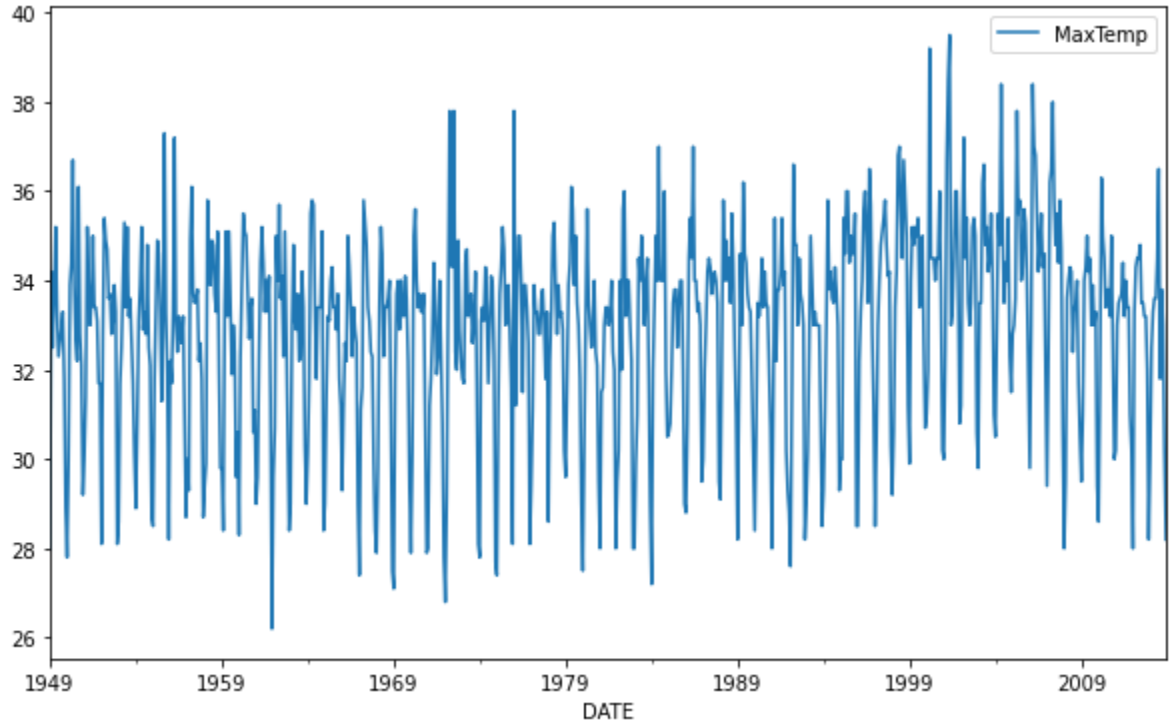
Dhaka:



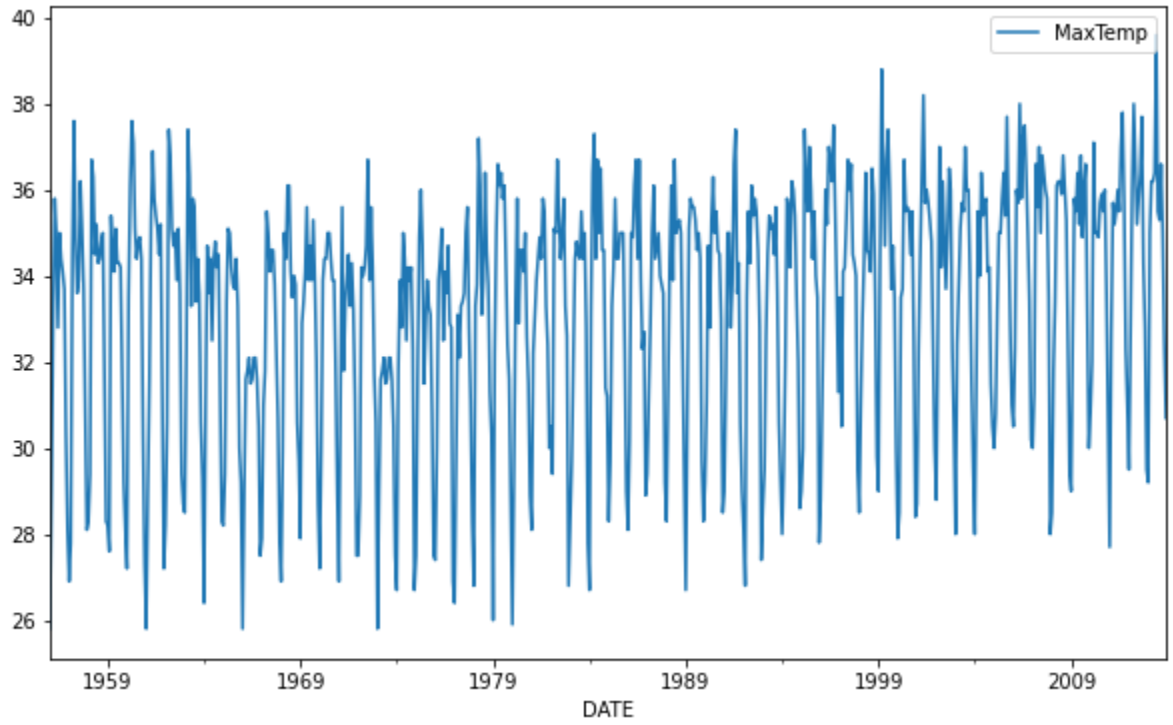
Barisal



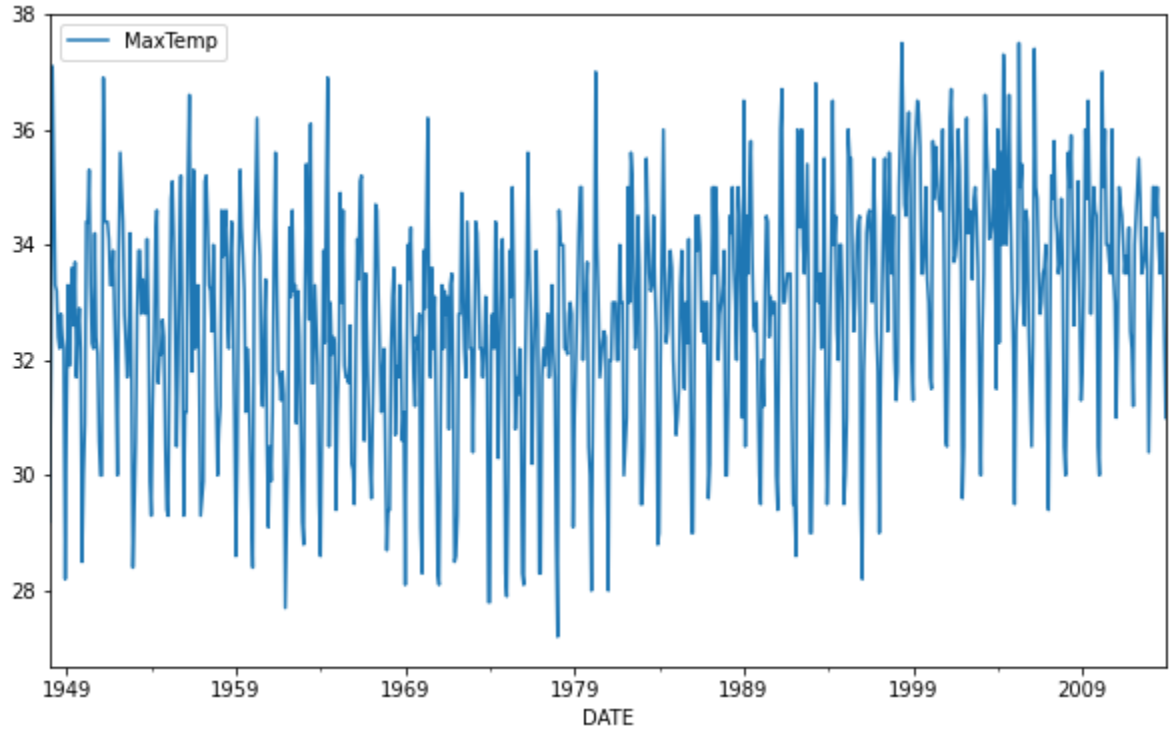
Chittagong(Patenga)



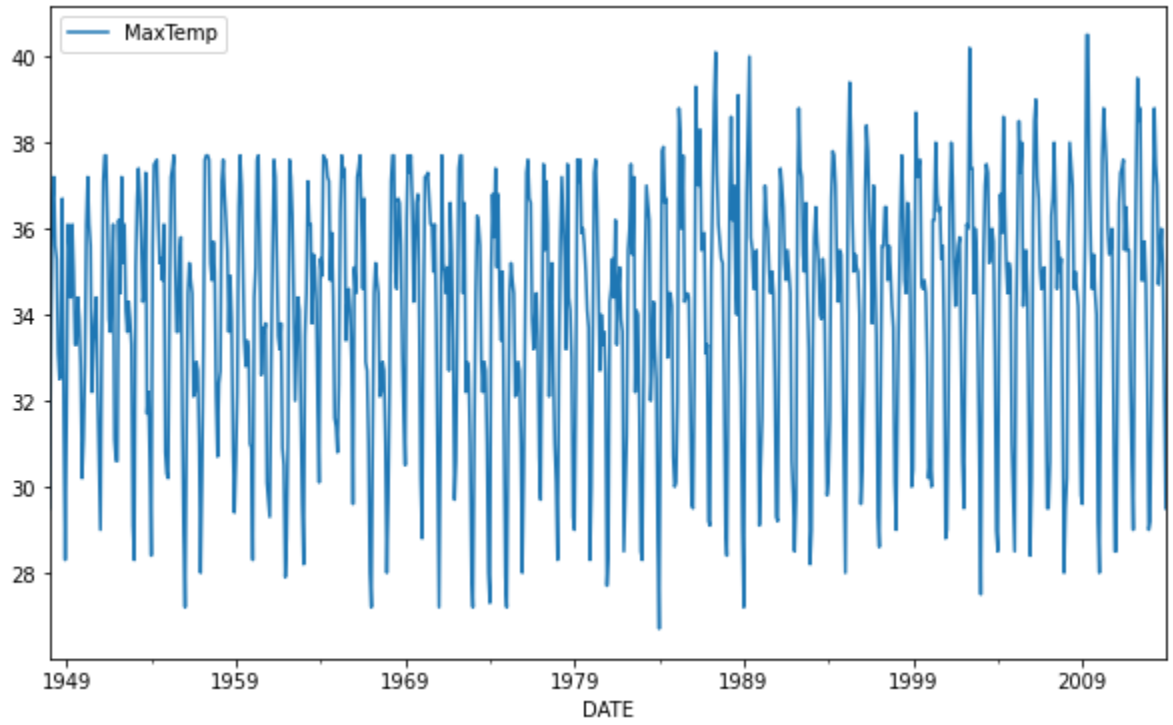
Sylhet



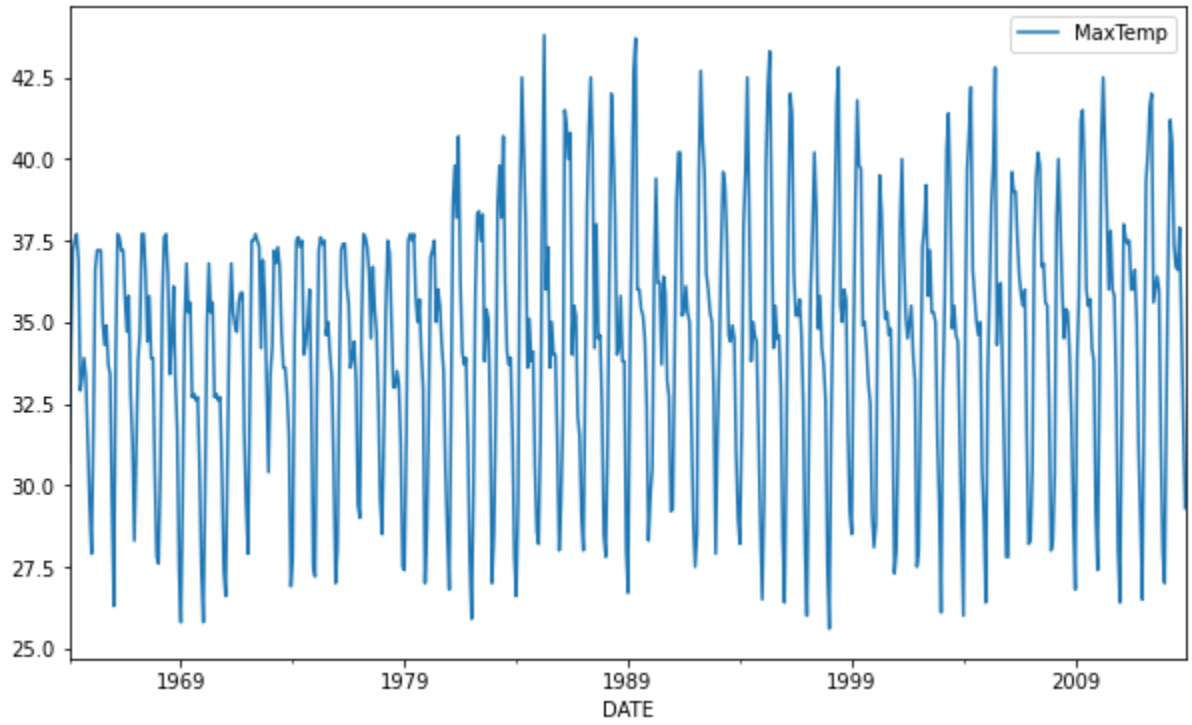
Coxs Bazar



Khulna

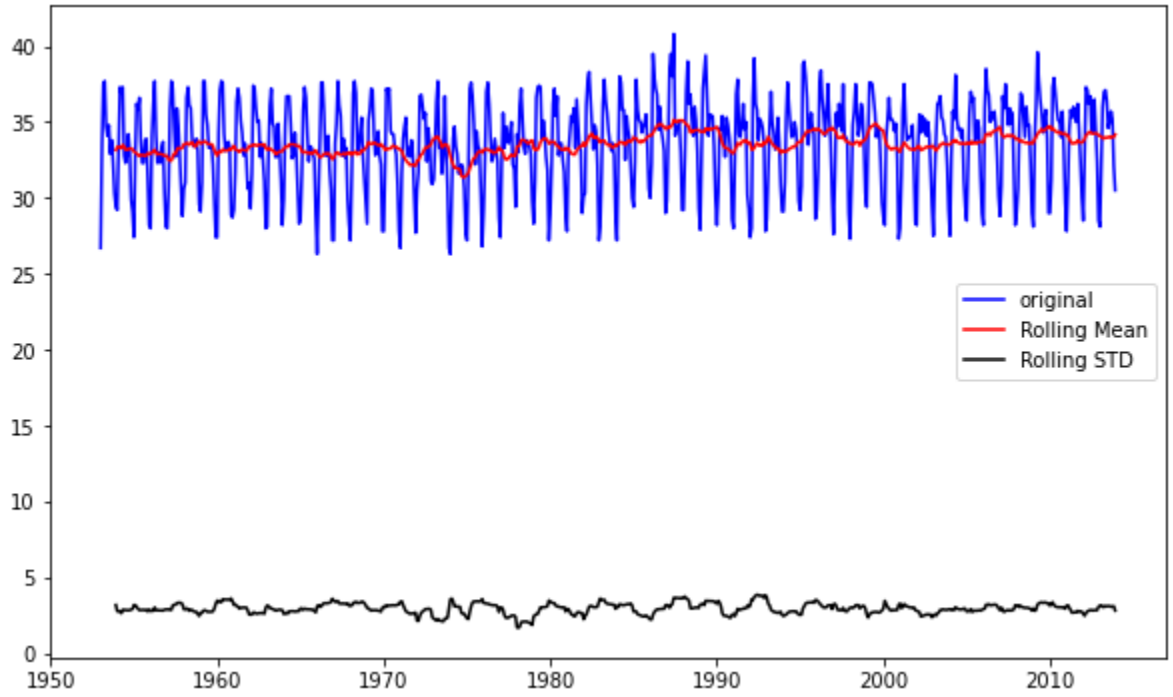


Rajshahi

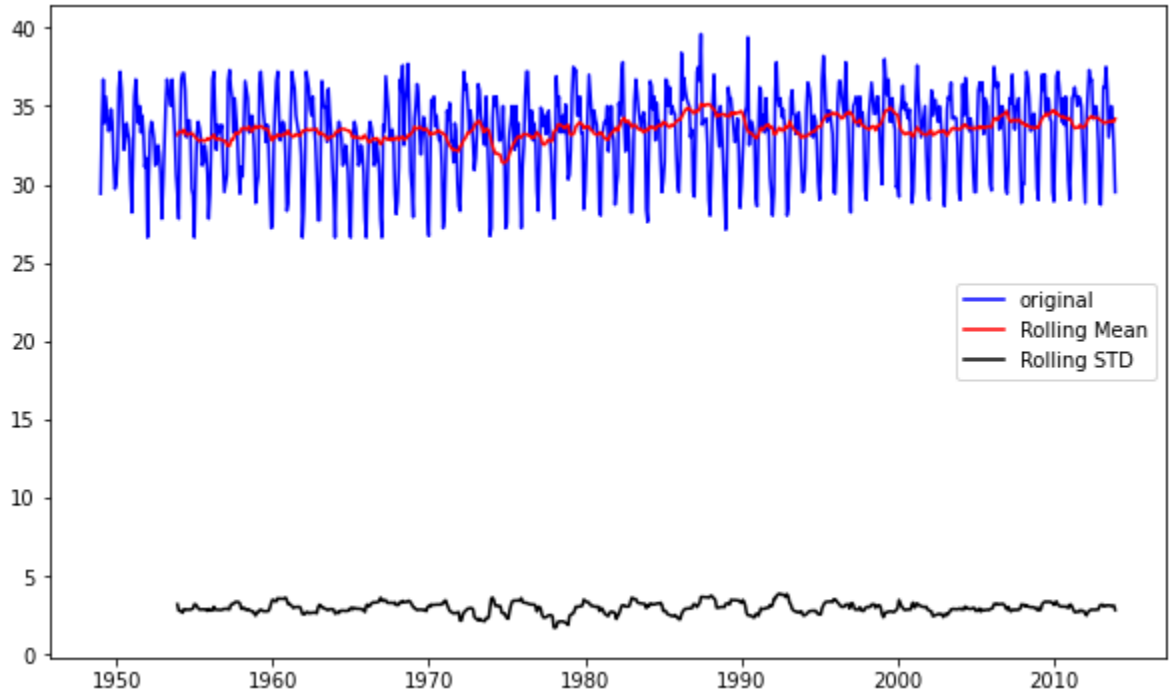


Then I have used the rolling mean and rolling standard deviation method plots to check the stationarity of the data. Here are the plots:

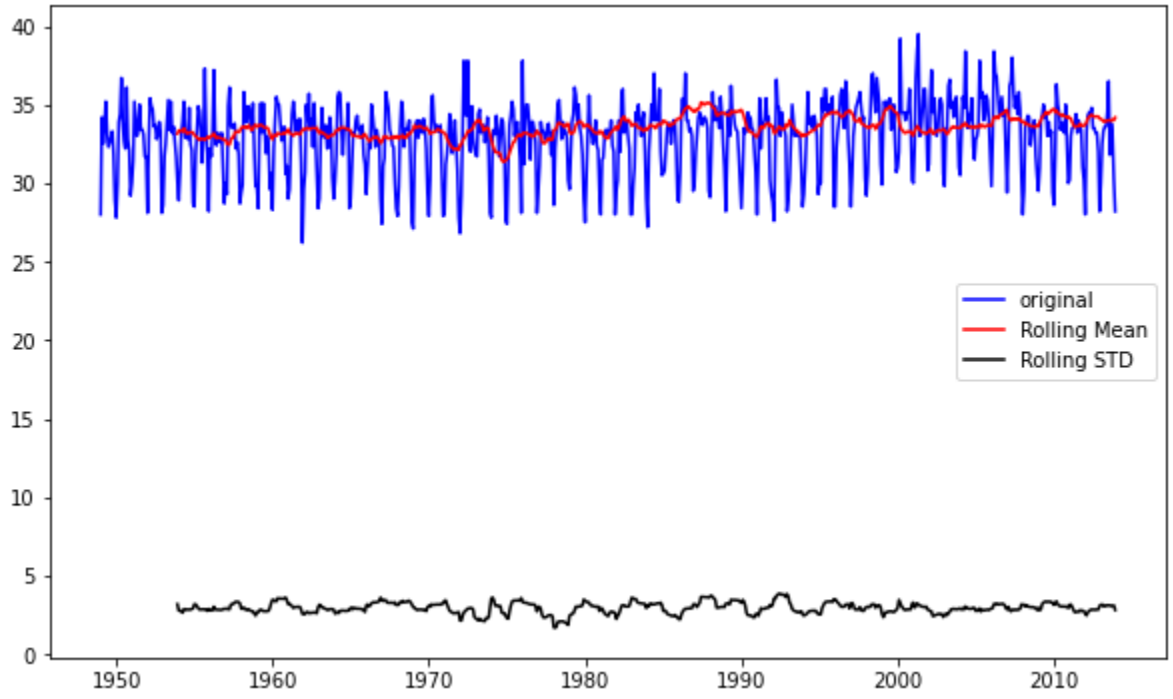
Dhaka:



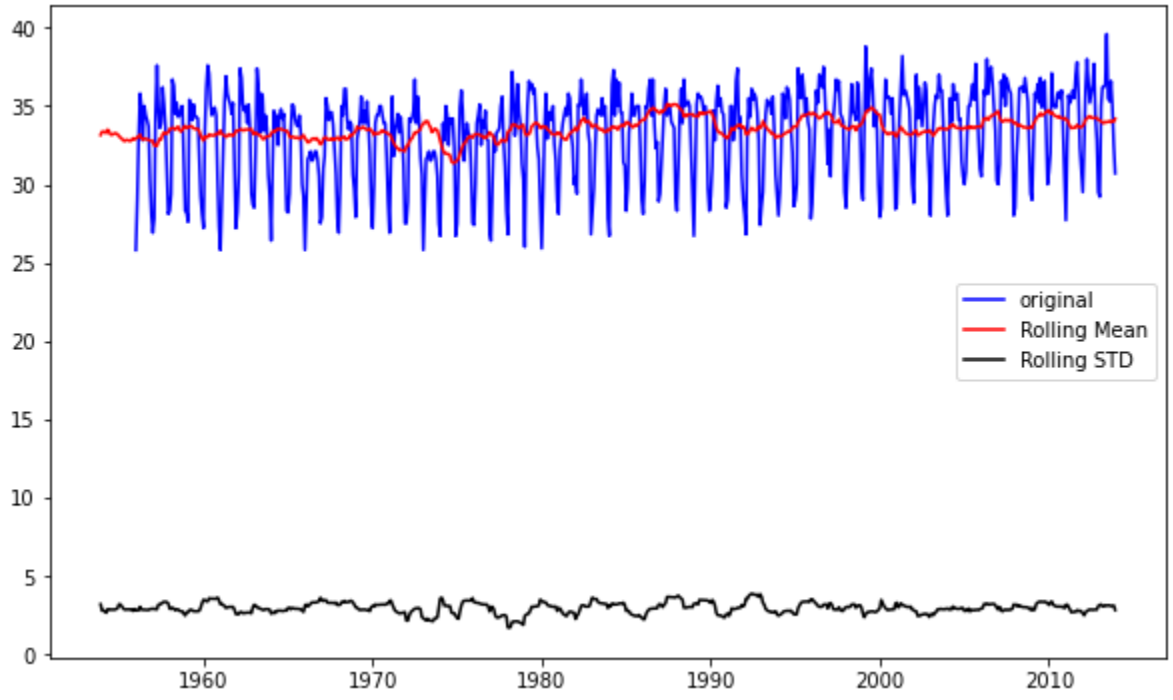
Barisal



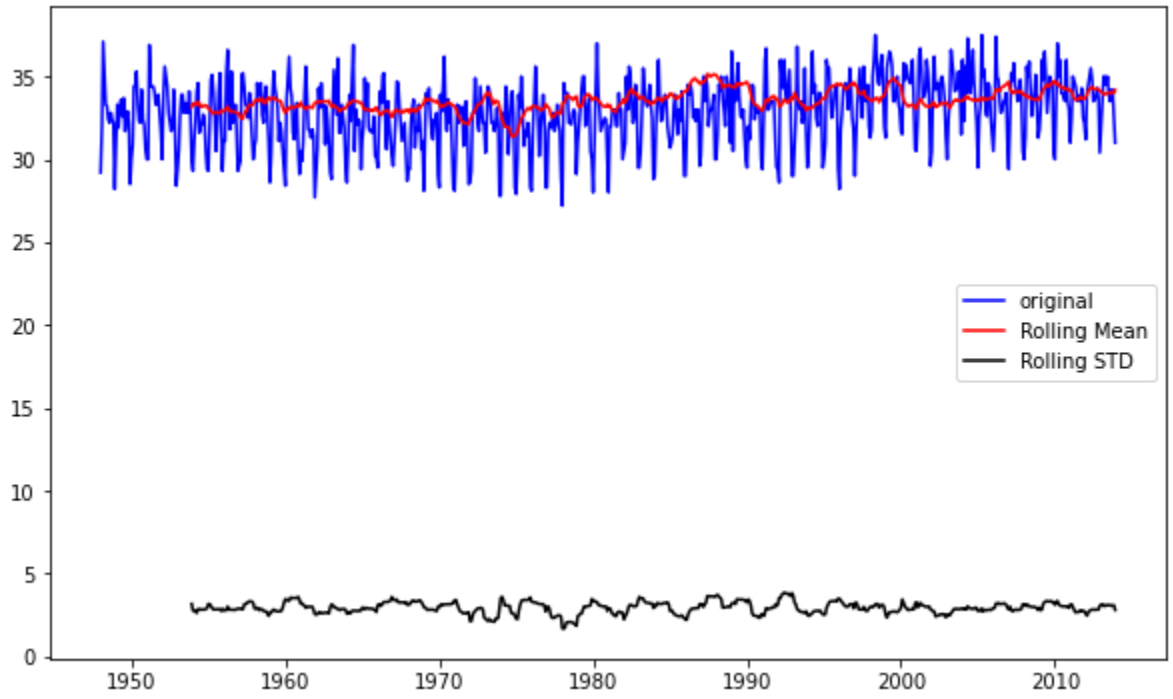
Chittagong(Patenga)



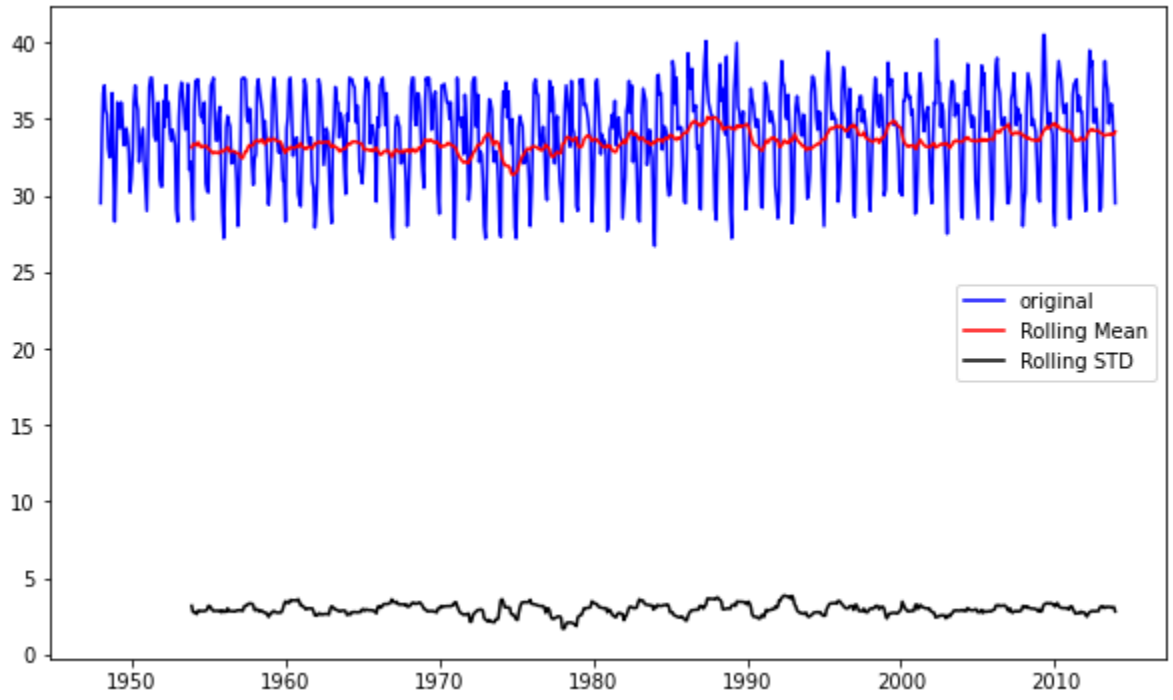
Sylhet



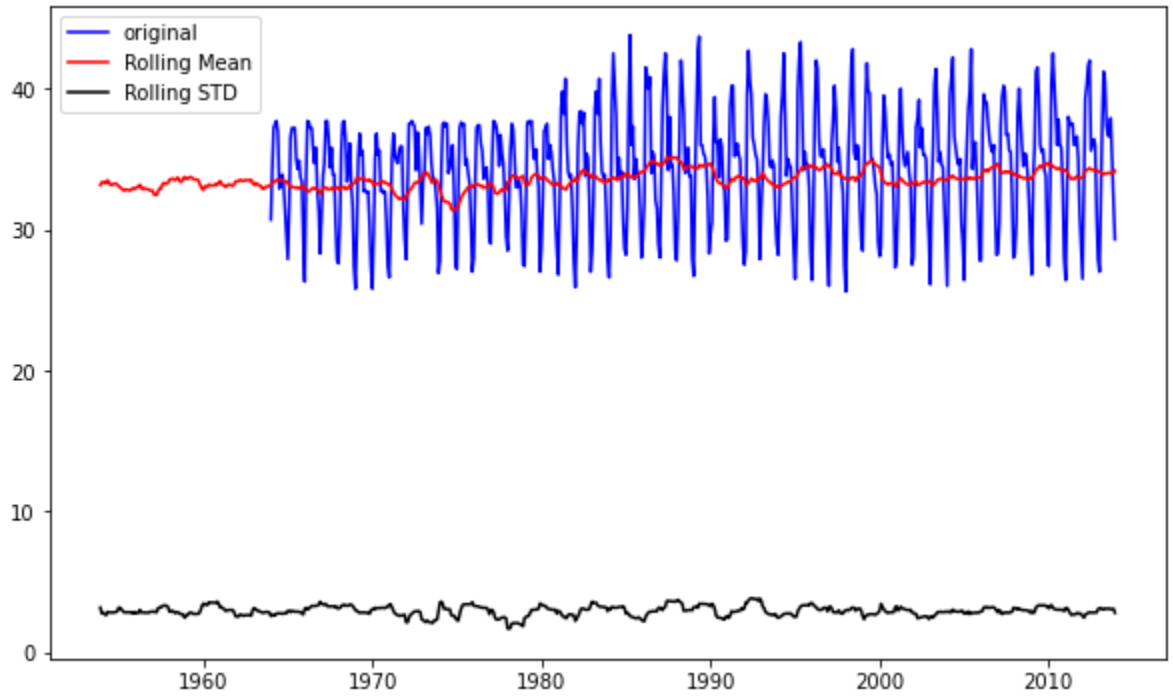
Coxs Bazar



Khulna



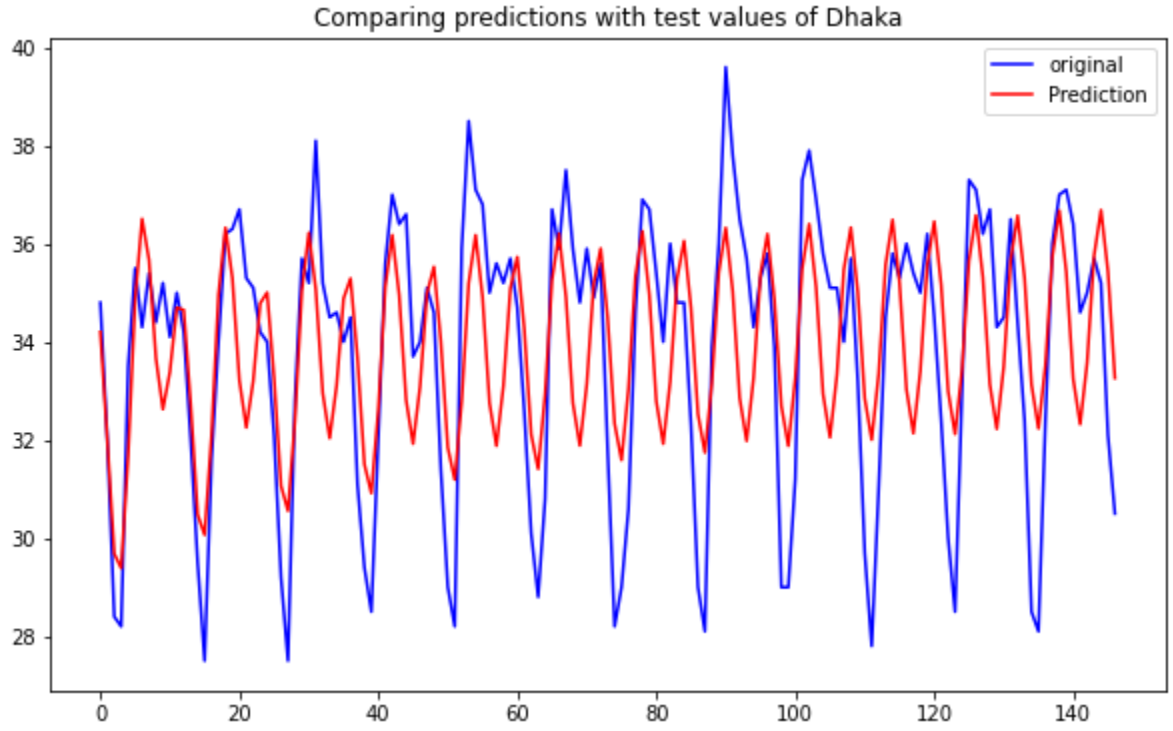
Rajshahi

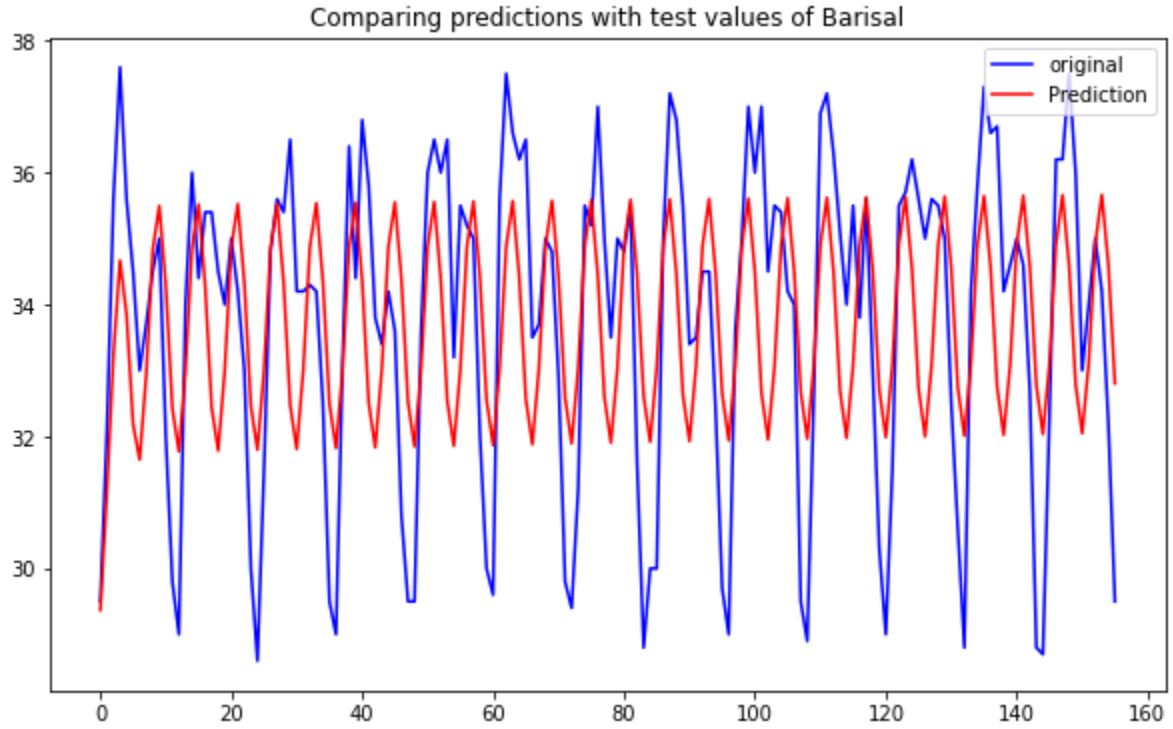


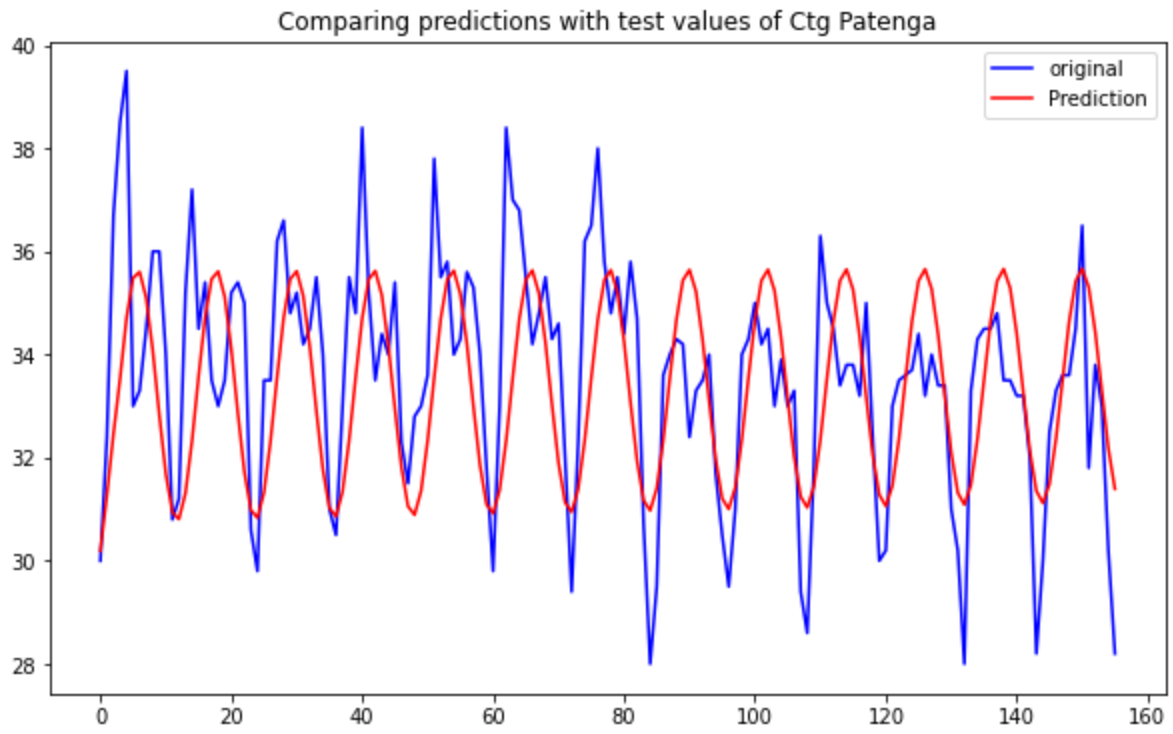
From the graphs we can see that all the datasets are fairly stationary so we move on to the next step of fitting these data to the ARIMA model. After determining the p and q value from ACF and PACF I have fitted them to the model and plotted the predicted value with the test values.

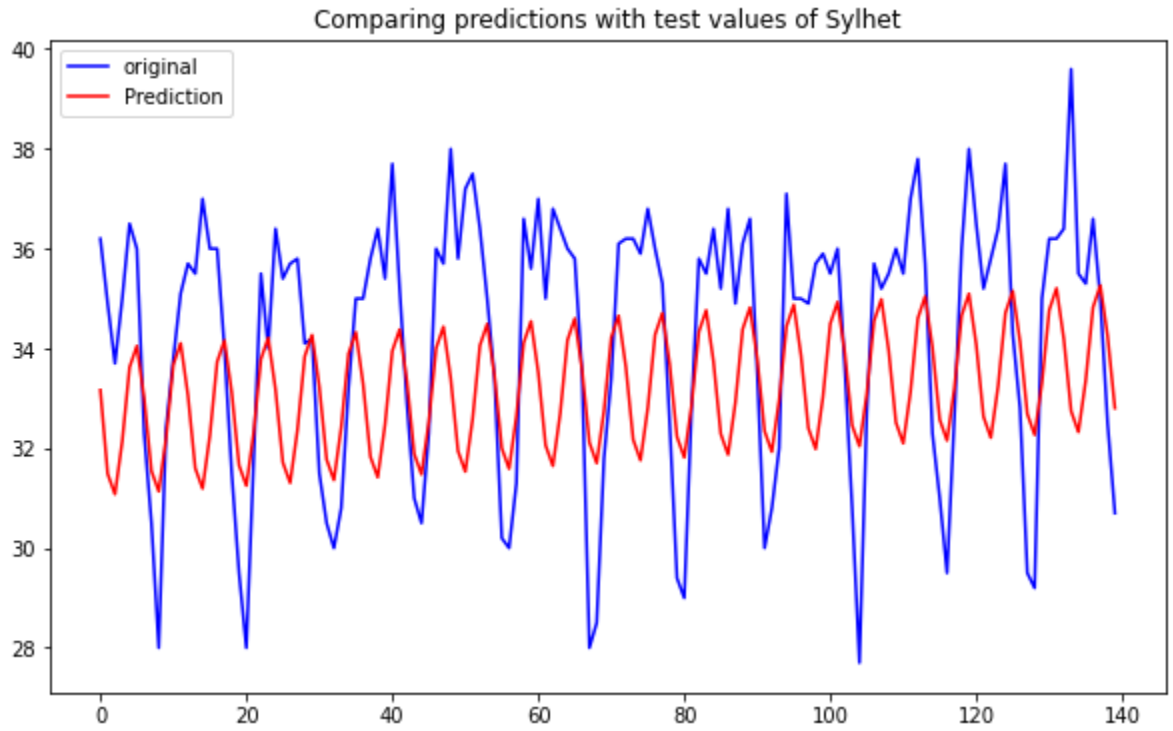
Result

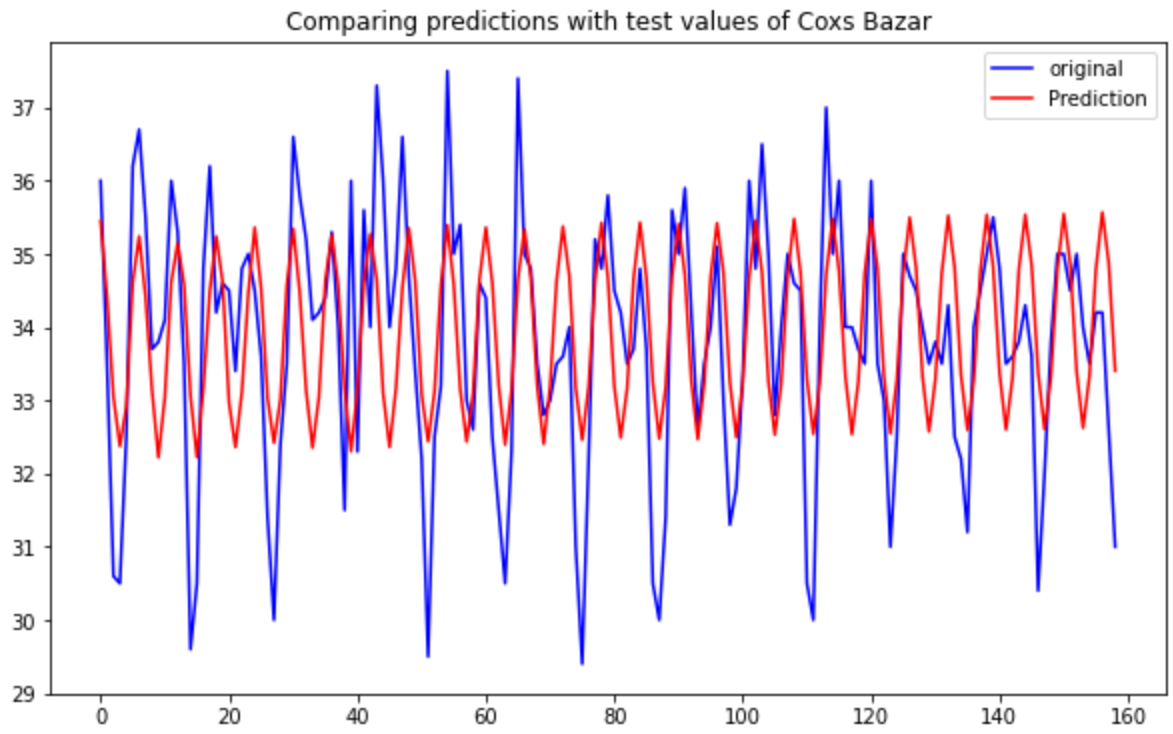
Here the results obtained from the models:

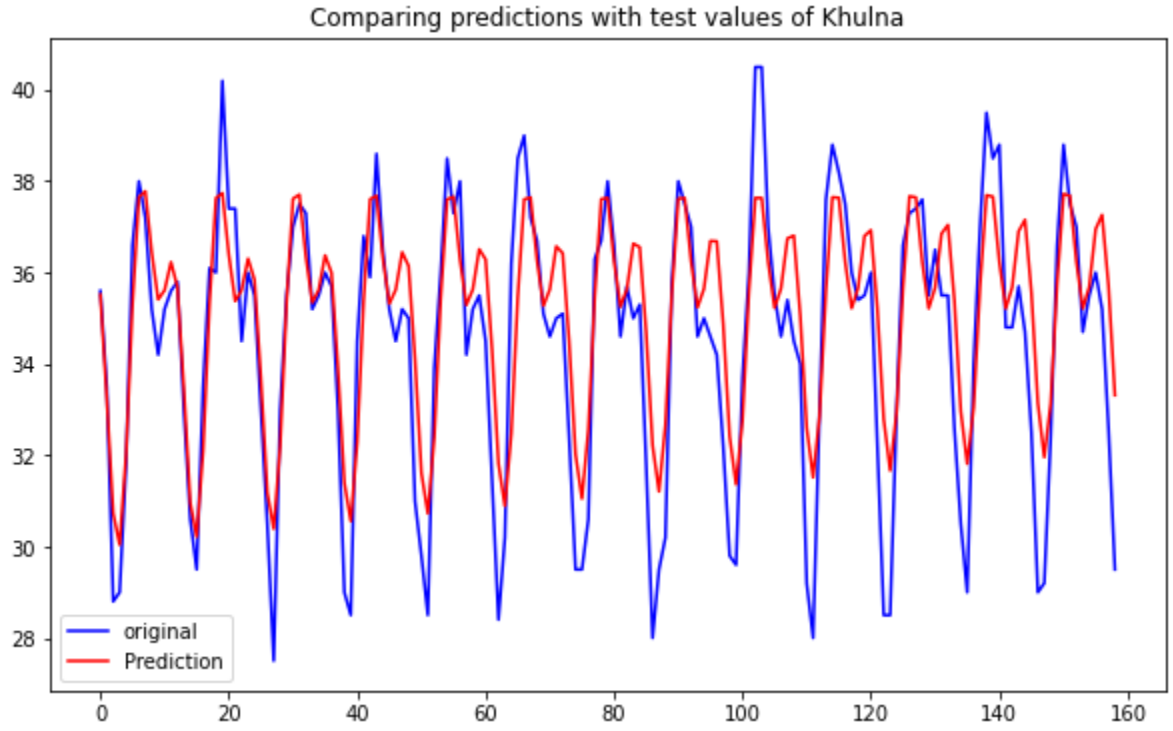


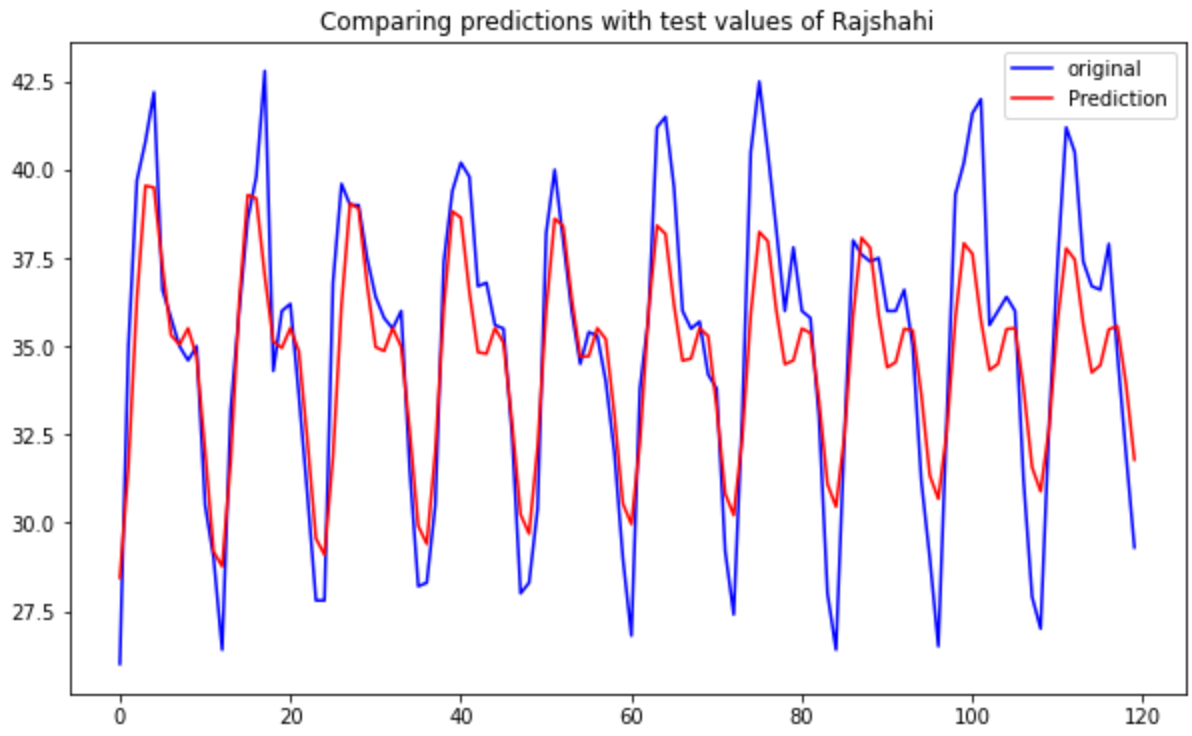



Comparing predictions with test values of Ctg Patenga
original
Prediction
40
38
36
34
32
30
28
0
20
40
60
80
100
120
140
160


Comparing predictions with test values of Sylhet
original
Prediction
40
38
36
34
32
30
28
0
20
40
60
80
100
120
140


Comparing predictions with test values of Coxs Bazar
original
Prediction
37
36
35
34
33
32
31
30
29
0
20
40
60
80
100
120
140
160


Comparing predictions with test values of Khulna
original
Prediction
40
38
36
34
32
30
28
0
20
40
60
80
100
120
140
160



AIC Value Table

District	AIC Value
Dhaka	2072.16
Barisal	2355.31
Ctg Patenga	2298.98
Sylhet	2332.67
Coxs Bazar	2235.54
Khulna	2177.50
Rajshahi	1752.19

From the results we can see that the model is working pretty well to predict the future max temperature values of these districts. Although it can also be seen that the model is not working so well with the seasonal peaks of the data.

Future Improvements

This project aimed towards modeling the maximum temperature of some districts of Bangladesh. In doing so I have used the popular ARIMA model to forecast the future max temperature of some big districts of Bangladesh due to its significance in the agricultural fields. But there are other significant factors to this field as well. Further research can be done predicting those values. Due to its relatively high mismatch with seasonality, some other models can also be implemented which are more catered towards predicting seasonality like the SARIMA model. Neural networks can also be implemented along with time series models to improve its efficiency and accuracy as per some of the existing research. I have also attached the AIC values of my model so that an improved model can be selected by comparing the AIC values of that model.

Conclusion

Bangladesh is an agricultural country. Still in 2022, one of our major sources of income is agriculture. But a lot has changed, the climate, the methods. The whole world is trying to incorporate as much technology as possible into this field so ensure food for all mankind. Due to a lot of infrastructural problems and lack of advancement in technology we are lagging behind. So as students of science we should focus on this significantly important field.

This project models the maximum temperature readings of different districts of 65 years then forecasts the future maximum temperature values using time series analysis and simulates the result. This could help farmers to predict the weather ahead of time and prepare their measures accordingly.

Reference:

[1] A time series analysis of weather variables and all-cause mortality in the Kasena-Nankana Districts of Northern Ghana, 1995â “2010. (2012). Taylor & Francis. https://www.tandfonline.com/doi/full/10.3402/gha.v5i0.19073%40zgha20.2012.5.issue-s1

[2] Campbell, S. D., & Diebold, F. X. (2005). Weather Forecasting for Weather Derivatives. Journal of the American Statistical Association, 100(469), 6–16. https://doi.org/10.1198/016214504000001051

[3] Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Computing, 24(21), 16453–16482. https://doi.org/10.1007/s00500-020-04954-0

[4] Sharaff, A., & Roy, S. R. (2018). Comparative Analysis of Temperature Prediction Using Regression Methods and Back Propagation Neural Network. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). https://doi.org/10.1109/icoei.2018.8553803

[5] Ahmadi, A., Zargaran, Z., Mohebi, A., & Taghavi, F. (2014). Hybrid model for weather forecasting using ensemble of neural networks and mutual information. 2014 IEEE Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/igarss.2014.6947305

[6]Kothapalli, S., & Totad, S. G. (2017). A real-time weather forecasting and analysis. 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). https://doi.org/10.1109/icpcsi.2017.8391974

[7]Karevan, Z., & Suykens, J. A. (2020). Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, *125*, 1–9. https://doi.org/10.1016/j.neunet.2019.12.030

[8]Lai, K. (2021, December 13). Time Series Analysis and Weather Forecast in Python. Medium. https://medium.com/@llmkhoa511/time-series-analysis-and-weather-forecast-in-python-e80b664c7f71

[9] Nury, Ahmad. (2013). Time series analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh.

[10] Gupta, N. (2019). GitHub - ngupta66/weather_forecast: Deep Neural Network for Weather Time Series Forecasting. GitHub. https://github.com/ngupta66/weather_forecast

[11]GitHub - exchhattu/TimeseriesWeatherForecast-RNN-GRU-LSTM: Weather forecasting using recurrent neural network. (2020). GitHub. https://github.com/exchhattu/TimeseriesWeatherForecast-RNN-GRU-LSTM

[12]GitHub - kennedykwangari/Time-Series-Analysis-and-Forecasting-with-Python: The aim of the project is to provide a reference container guide of the numerous broad topics in the field of Time Series Analysis. (2020). GitHub. https://github.com/kennedykwangari/Time-Series-Analysis-and-Forecasting-with-Python