

BIRLA INSTITUTE OF TECHNOLOGY

MESRA



MASTER OF COMPUTER APPLICATIONS (SEM - IV)

2020-2021

Temperature Predictions of Different Cities

MOOC COURSE (MACHINE LEARNING)

SUBMITTED TO –

Rashmi Rathi Upadhyay

(Assistant Professor)

SUBMITTED BY –

RAHUL GUPTA

ROLL – MCA/10030/19

ABSTRACT

In today's modern world, most of the products that we produce have a very crucial factor affecting it i.e. temperature. Starting from the traditional crops to the artificial ones in food industries, from drugs to chemicals manufactured in the pharmaceutical industries, all of them need the right amount of temperature to be maintained for manufacture which is why the monitoring of temperature constantly is an indispensable part of these sectors.

Our homes too have thermostat installed which monitor and regulate the temperature. Maintaining the right temperature is required for having a healthy growth of plants. If the right temperature is not maintained, the plants will die.

Effects that scientists had predicted in the past that would result from global climate change are now occurring: loss of sea ice, accelerated sea level rise and longer, more intense heat waves and so on.

So far, global warming has been limited to a rise of around 0.75°C since the end of the 19th century. This sounds like a small change, but scientific evidence suggests it is already leading to a range of impacts around the world.

More importantly, however, is that the temperature rise observed so far is not the chief cause of concern. More worrying is the significantly larger temperature rise (and associated changes in rainfall, snowfall, sea level and other phenomenal) that scientists expect the world to experience in the coming decades and centuries.

Temperature play an important role in shaping weather patterns, guiding the life cycle of various organisms and maintaining ocean levels. Shifting the temperature a couple of degree can throw an entire ecosystem into chaos.

I, as computer science students can contribute to climate change and global warming by predicting future temperature changes and making others aware of the consequences they will need to face. I have taken up this topic to dwell upon the rising concern of global warming and to gain insights from data that is available to us in the real world.

Using this project, you will be able to build such a monitoring system where you will be able to monitor the temperature of the environment in the form of visual graphs. This project can predict the future sensor values too.

PROBLEM STATEMENT

To predict global temperatures by analyzing historical monthly city temperatures between 1750 and 2014.

GOALS AND TARGET VARIABLES

1. Average Temperature

Predicted temperature for a given city over a specified time period.

2. Affected Cities

The top 3 cities in the United States that will experience the most temperature change from 2014-2024

3. Analysis of Global Warming

Predicting future Global Warming and will have most correlation with temperature change.

DATA SOURCES

Temperature Data

Climate Change: Earth Surface Temperature Data from Kaggle

MODEL FOR FORECASTING

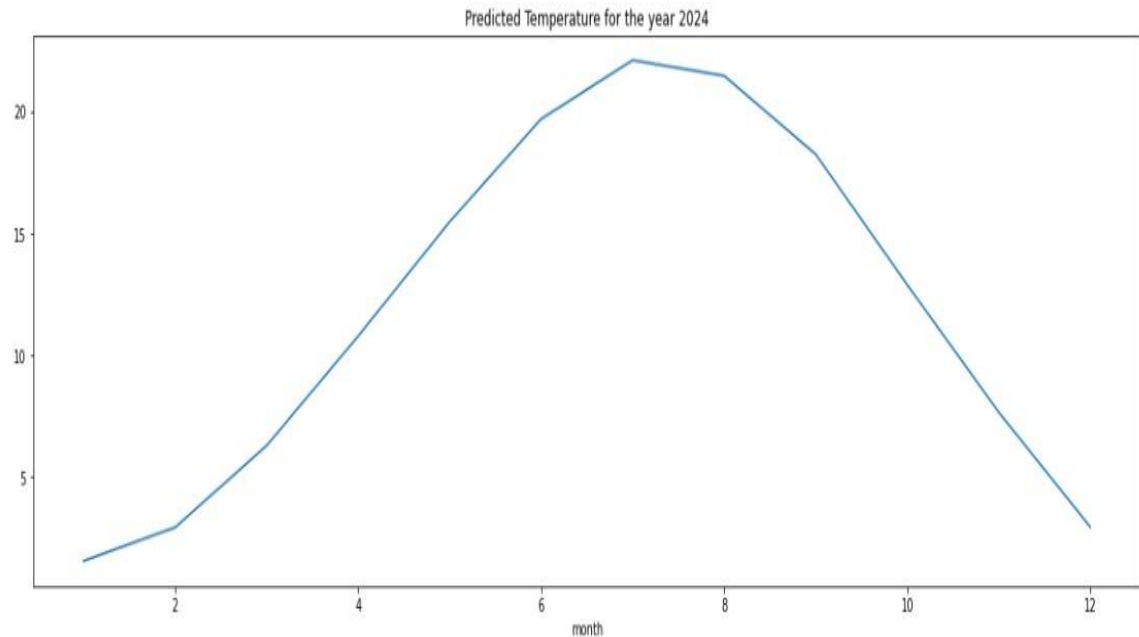
Since I have time series data, I have chosen to use an **ARIMA (Auto Regressive Integrated Moving Average) Model** for time series forecasting.

I have automated the process of estimating the best p, q parameters for the ARIMA model, for any given time series data. The p,q values were chosen based on the returned the lowest AIC (Akaike Information Criterion) Values.

The model was used to successfully forecast the temperature for a given city across a specified time period.

```
In [115]: monthly_seasonality.plot(figsize=(20,6),title='Predicted Temperature for the year 2024')
```

```
Out[115]: <AxesSubplot:title={'center':'Predicted Temperature for the year 2024'}, xlabel='month'>
```



EVAULATION

We have used Mean Squared Error (MSE) and Mean Absolute Error (MAE) to evaluate the performance of the ARIMA model.

For New York, the ARIMA model achieved:

- Mean Squared Error: 2.3934235122562058
- Mean Absolute Error: 1.512023989745657

ASSUMPTIONS /LIMITATIONS

The city-wise monthly temperature data had several missing values for certain cities, instead of replacing these NaN values with the mean temperature, I chose to remove rows having missing values.

Also, I were not able to fully analyze the Global Warming effect on temperature change, due to the lack of data for missing values.

TOOLS

For coding Implementation, Python language is used and Platform used is “Jupyter Notebook”.

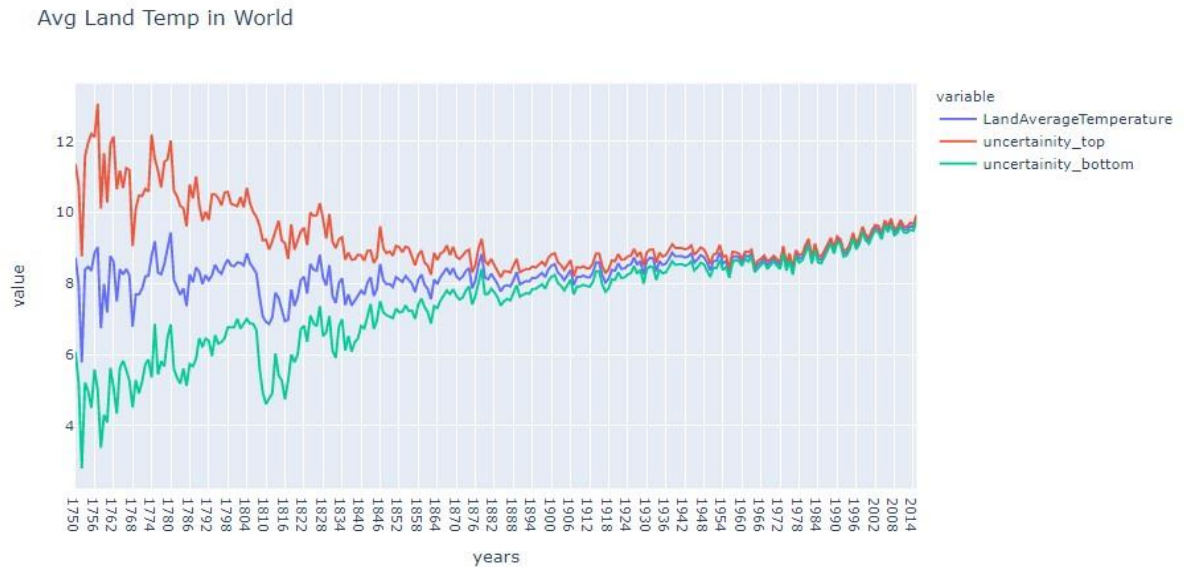
I have used libraries including pandas, numpy, matplotlib, sklearn, seaborn, plotly and statsmodels.

- **NUMPY**: is the fundamental package needed for scientific computing with python. This contains: a powerful N-dimensional array object.
- **Pandas**: is a software library written for the python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating tables and time-series.
- **MATPLOTLIB**: Matplotlib is a cross-platform ,data visualization and graphical plotting library for python and its numerical extension NumPy.
- **SKLEARN**: the most useful library for machine learning in python. The skLearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

The crux of my problem lies in Time Series Forecasting.I have made use of ACF and PACF plots, in combinations with AIC Values, to estimate the best p.q value for the ARIMA model and have forecasted the temperatures for cities over a given period of time.

RESULTS

I predicted the temperature of any given city across a specific time period. I also identified the top-3 cities in the USA that are likely to experience the most temperature change. Additionally, I predict the Average Temperature of each season.



This figure tell us, global warming exist in last 3 decades.



Average Seasonal Temperature

Objective: Predicting the temperature change of a given city across a specified Time period

To accomplish this, I used data from the [Climate Change: Earth Surface Temperature Data](#) dataset on Kaggle.

Pre-Processing:

1. Converting the 'dt' (date) column to DateTime format
2. Dropping irrelevant columns and removing rows with NaN values

Processing:

ARIMA models need the data to be stationary i.e. the data must not exhibit trend and/or seasonality. To identify and remove trend and seasonality, I used the following methods:

- A. Plotting the time series to visually check for trend and seasonality
- B. Checking if histogram of the data fits a Gaussian Curve, and then splitting data into two parts, calculating means and variances and seeing if they vary
- C. Calculating the Augmented Dickey-Fuller Test Statistic and using the p-value to determine stationary

If the data was not stationary, I performed differencing to make it stationary.

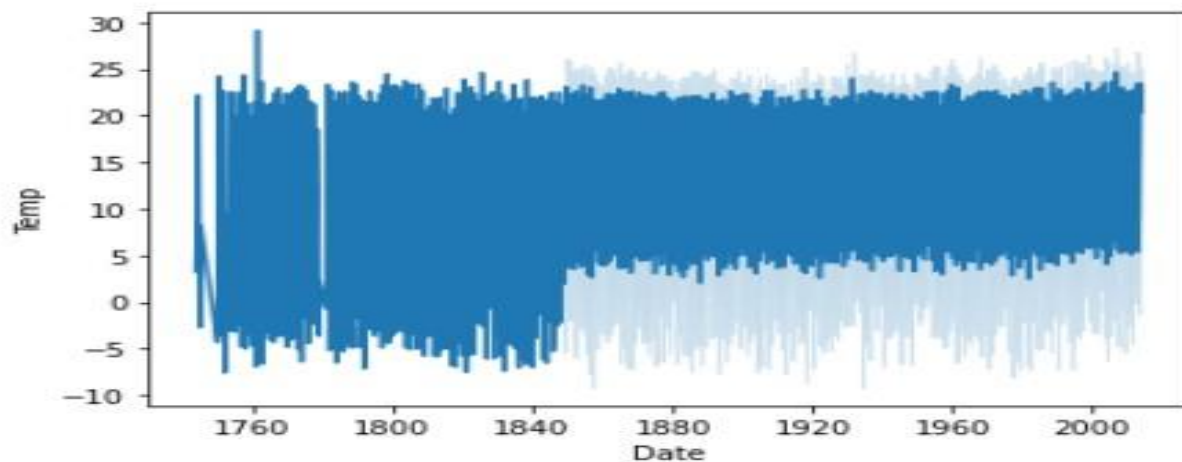
Fitting the ARIMA model:

I performed a Grid-search to estimate the best p,q values for the model, for the given data.

I then fit the ARIMA model using the calculated p,q values.

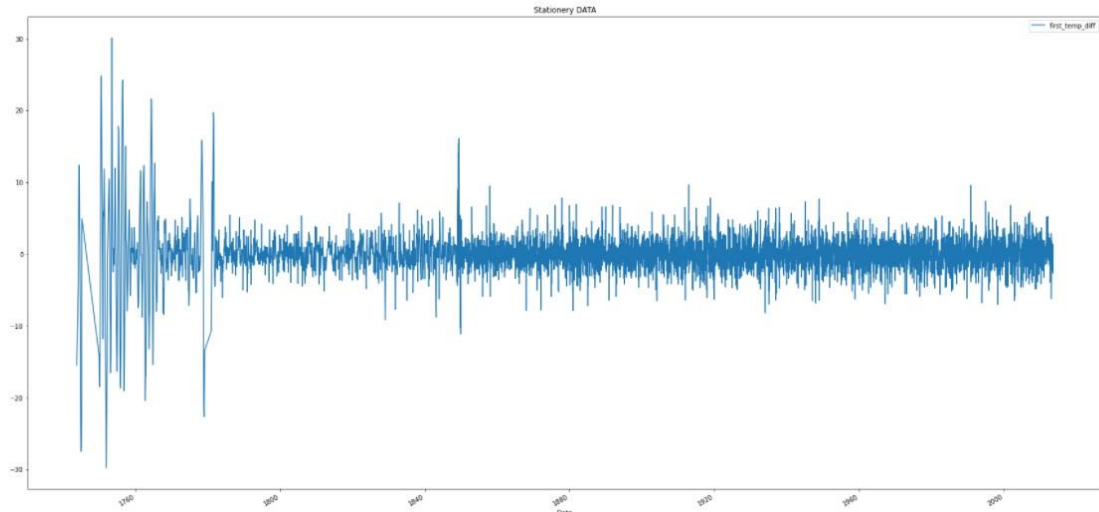
Evaluation:

I calculated the **Mean Squared Error (MSE)** to estimate the performance of the model.



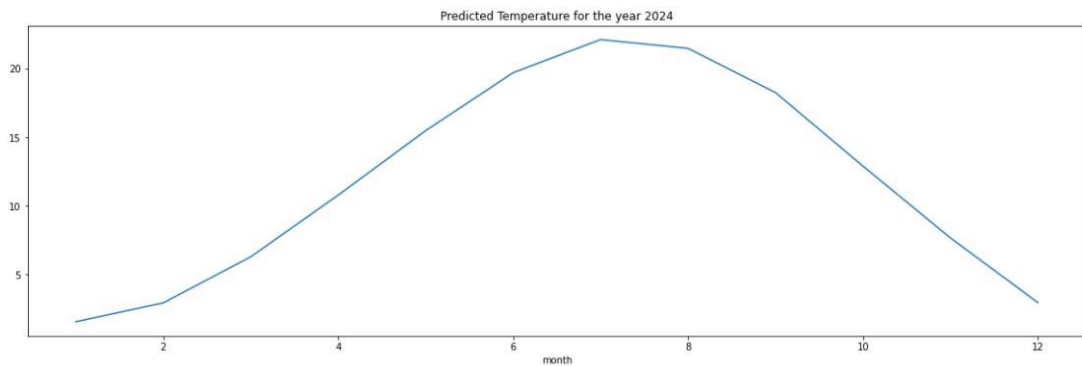
Above Histogram shows, It is not Stationary so making its stationary I need to perform some operations.


```
In [117]: df[['first_temp_diff']].plot(figsize=(30,15), title='Stationery DATA')
Out[117]: <AxesSubplot:title={'center':'Stationery DATA'}, xlabel='Date'>
```



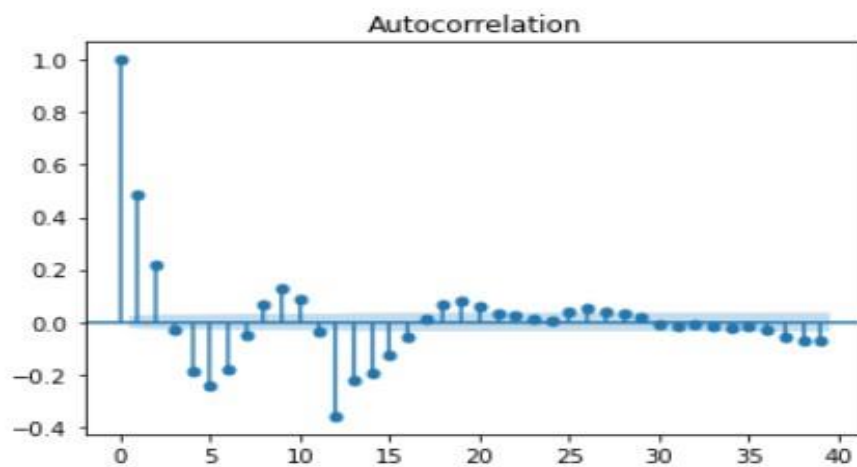
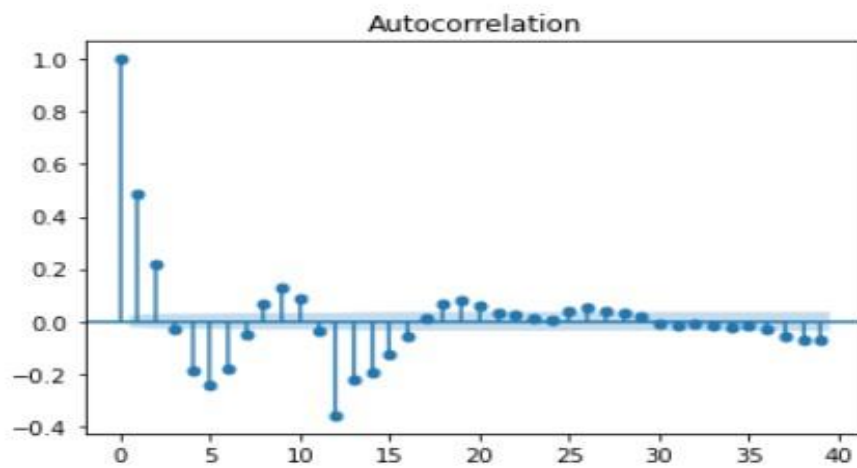
After performing some operations, now it is changed into Stationary data.

```
In [115]: monthly_seasonality.plot(figsize=(20,6),title='Predicted Temperature for the year 2024')
Out[115]: <AxesSubplot:title={'center':'Predicted Temperature for the year 2024'}, xlabel='month'>
```



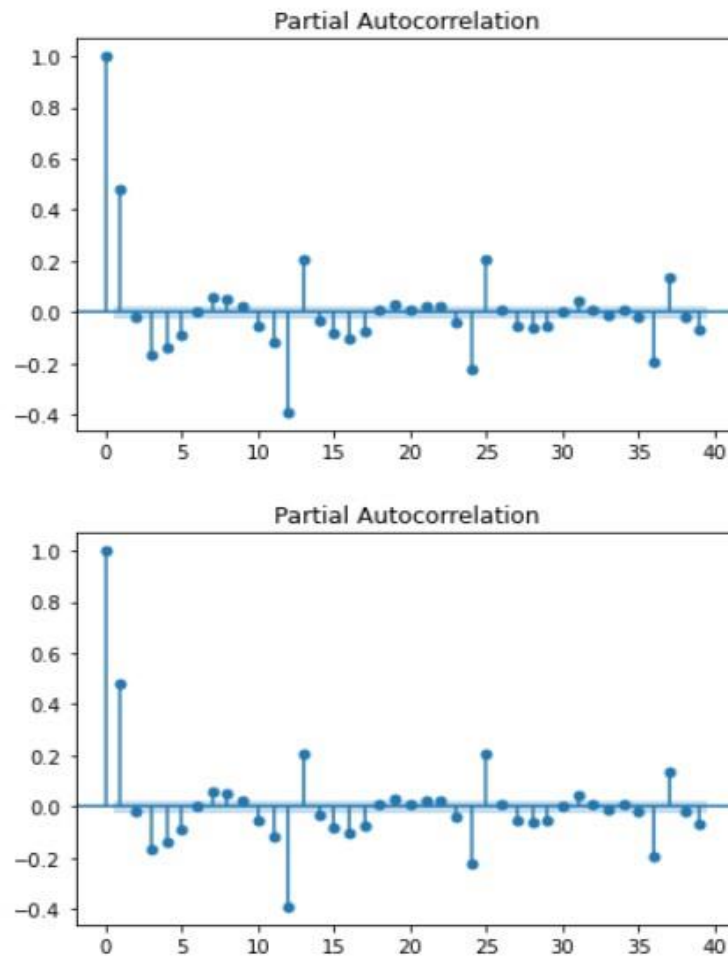
This figure is prediction of 10 years of data.

```
plot_acf(df['first_temp_diff'].dropna())
```



```
In [114]: plot_pacf(df['first_temp_diff'].dropna())
```

```
Out[114]:
```



Conclusion:

In this project, I:

- Forecasted the temperature of a given city over a given period of time.
- Predicted the top-3 cities in the USA which will experience the most temperature change from 2014-2024.
- Analyzed the Global Warming exist or not, as well as the analyzed the temperature of each season.

Code: <https://github.com/bitmesra-rru/mooc-project-rahul22nrp/upload/master>