TIME SERIES FORECASTING AND ANALYSIS

PROJECT REPORT SUBMISSION

BACHELOR'S OF TECHNOLOGY

in

ELECTRONICS AND COMMUNICATION ENGINEERING

by

ARYAN (**111916012**)

PRACHHI SABITA MISHRA (**111916034**)

VIMAL GUPTA (**111916059**)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Indian Institute of Information And Technology Pune

Near Bopdev Ghat, Kondhwa Annexe, Yewalewadi, Pune, Maharashtra 411048

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **"TIME SERIES FORECAST-ING AND ANALYSIS"** submitted by **Aryan, Prachhi Sabita Mishra, Vimal Gupta** bearing the **MIS No: 111916012, 111916034, 111916059** respectively, in completion of his project work under the guidance of **Dr.Chandrakant Guled** is accepted for the project report submission in partial fulfillment of the requirements for the award of the degree of Master of Technology in Computer Science and Engineering in the Department of Computer Science and Engineering, Indian Institute of Information Technology,Pune, during the academic year 2020-21.

Dr.Chandrakant Guled   **Dr.Nagendra Khushwaha**

Project Guide   Head of the Department

Professor   Assistant Professor

Department of CSE   Department of ECE

IIIT Pune   IIIT Pune

Project Viva-voce held on   _____

Internal Examiner   **External Examiner**

ACKNOWLEDGEMENT

Abstract

First paragraph

Second Paragraph

Third Paragraphy

Keywords :   Keyword1, Keyword2

TABLE OF CONTENTS

List of Figures

Chapter 1

Introduction

1.1   What is Time Series?

A time series is simply a series of data points ordered in time. In a time series, time is often the independent variable and the goal is usually to make a forecast for the future. In particular, a time series allows one to see what factors influence certain variables from period to period.

1.2   What is Time series analysis?

Time series analysis is a broad research field covering many application domains either focusing on analysis tasks and methods like forecasting , clustering and classification , anomaly detection , changepoint analysis , pattern recognition ,or dimensionality reduction or focusing on a specific application domain.

1.3   What is Time series forecasting?

Time series forecasting is the process of analyzing time series data using many statistical parameters and modeling to make predictions and inform strategic decision-making. It's not always an exact prediction, and likelihood of forecasts can vary wildly—especially when dealing with the commonly fluctuating variables in time series data as well as factors outside our control. However, forecasting insight about which outcomes are more likely—or less likely—to occur than other potential outcomes.

Chapter 2

Motivation

Whether we wish to predict the trend in financial markets or electricity consumption, time is an important factor that must now be considered in our models. For example, it would be interesting to forecast at what hour during the day is there going to be a peak consumption in electricity, such as to adjust the price or the production of electricity.

Time series are one of the most common data types encountered in daily life. Financial prices, weather, home energy usage, and even weight are all examples of data that can be collected at regular intervals.

It can be useful to see how a given asset, security, or economic variable changes over time. It can also be used to examine how the changes associated with the chosen data point compare to shifts in other variables over the same time period.

Chapter 3

Literature Review

3.1 Univariate Forecasts

Time series analysis plays an important role in numerous applications such as healthcare, economics, finance and many more. There are several aspects of time series analysis, which includes predictions, univariate forecasts, missing value imputations, outlier detections, time series transformation, and cleaning, etc. In most of the applications, the final goal is to achieve an accurate prediction or forecasting results.

3.2 PSF Algorithm

Time series forecasting is not the only forecasting algorithm.The PSF algorithm consists of several tasks, which are broadly divided into two steps, i.e., clustering of data and the forecasting based on clustered data. The PSF is a closed-loop process, hence it forecasts values up to long-duration by appending earlier forecasted values to existing original time series data. However, it is worthy to note that PSF is particularly developed to forecast time series exhibiting some patterns in the historical data, such as weather, electricity load, or solar radiation. The application of PSF to time series without such kind of inherent patterns might lead to the generation of not, particularly competitive results.PSF consists of various tasks including the optimum cluster size, window size selection, pattern searching, and prediction processes.
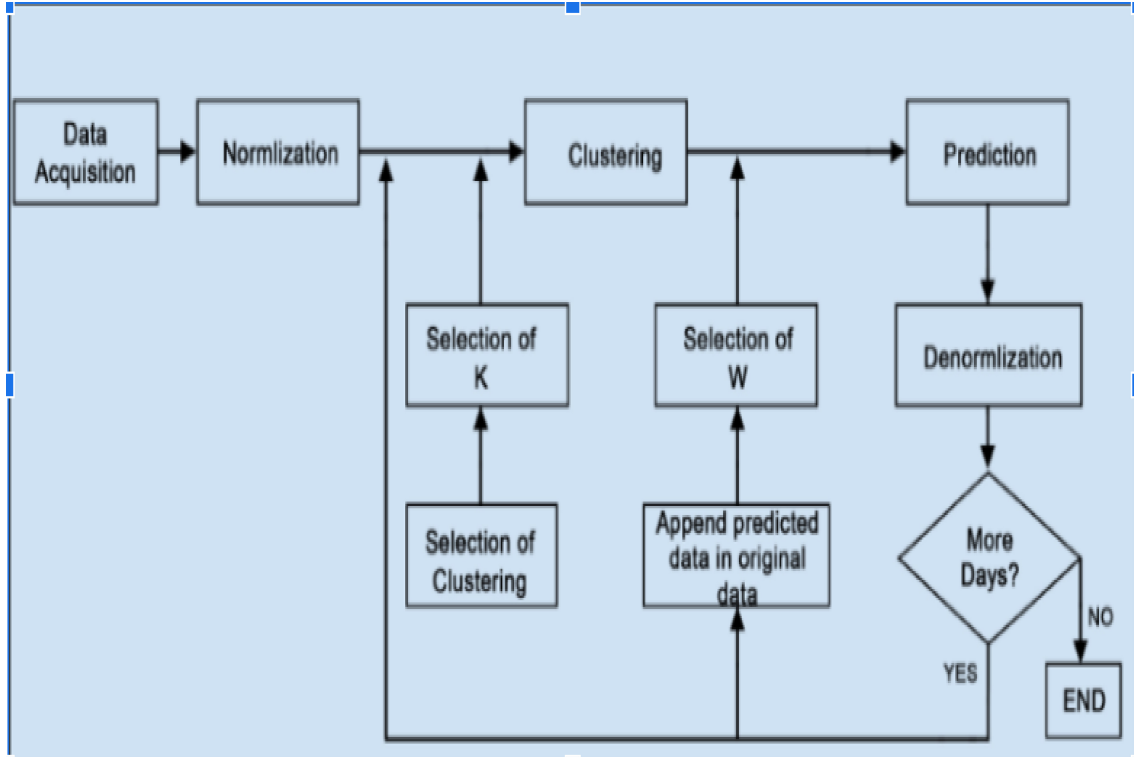
The flowchart of the algorithm:

Figure 3.1: Flowchart

A major benefit of time series analysis is that it can be the basis to forecast data. This is because time series analysis — by its very nature — uncovers patterns in data, which can then be used to predict future data points. Also time series analysis is that it can help an analyst to better understand a data set. This is because of the models used in time series analysis help to interpret the true meaning of the data, as touched on previously.

Chapter 4

Research Gaps

Over time, several formal definitions and reviews of time series analysis tasks have been published and some of those we have tried to cover and those are some limitations and some research gaps in time series analysis field:

1. Time series analysis also suffers from a number of weaknesses, including problems with generalization from a single study, difficulty in obtaining appropriate measures, and problems with accurately identifying the correct model to represent the data. While careful attempts are made to minimize these problems, each is present in the Simonton study, although sometimes in a subtle manner. Changes in how the data could be gathered are suggested that might help to solve some of these problems in future studies.

2. AR, ARCH. GARCH, ARIMA, etc. do not seem to be helpful in forecasting the coming of crisis also, Forecasting time series can be a very hard task due to the inherent uncertainty nature of these systems. It seems very difficult to tell whether a series is stochastic or deterministic chaotic or some combination of these states. More generally, the extent to which a non-linear deterministic process retains its properties when corrupted by noise is also unclear So, The noise can affect a system in different ways even though the equations of the system remain deterministic.

3. Forecasting a univariate time series is one of the challenging tasks, since such a process needs to understand, recreate, and extrapolate the patterns available in the targeted time series itself. Whereas, in the multivariate prediction process, there can be the availability of several variables with which the predicting process usually becomes easier. This is the reason that there are very few methodologies available for univariate time series forecasting than the multivariate prediction ones.

Chapter 5

Objective

The main objective of our project is to predict the upcoming forecasts by analysing the data set provided. Here we collected covid-19 data from various sources. The data includes daily cases, the total recovered and total deaths. By using the above data we can predict the further cases.

Also we will measure the how our data predictions match with the original data. Use of various mathematical tools provided we will find how much error our model generates.

Chapter 6

Methodology
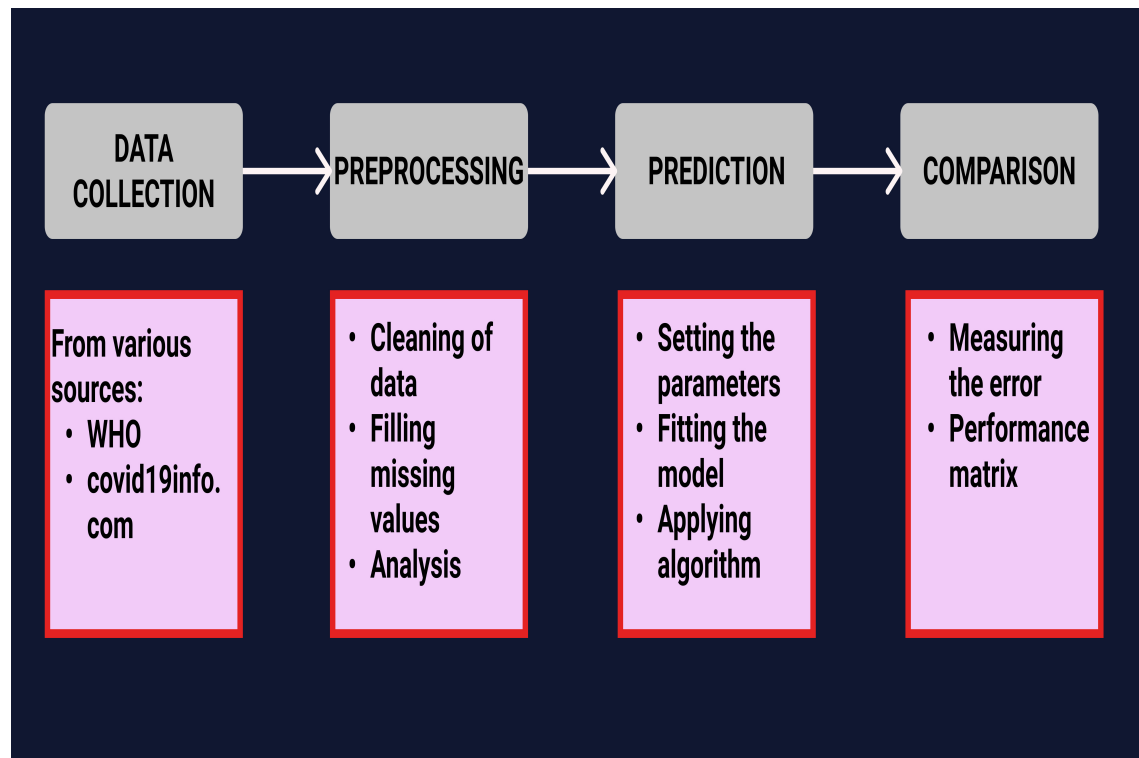
The basic working of our model:

Figure 6.1: Workflow

The very first step involves collecting data from authenticated sources and verifying it. After collection of data, it was processed into the program by using pandas library. Also by using it, we cleaned the data, sorted it according to dates and filled the missing values. The missing values were the average values of data before it.The next step involves analysis of data and finding the correlation among it.

After the analysis part we will fit it into ARIMA model using the statsmodel python library. Once the model is ready we will check for it validation by using various mathematical parameters (like squared mean error and root Mean Squared Error). Also we will visualize the real values and predicted values using matplotlib library.

Chapter 7

Results and Discussion

Time series analysis is not a new study, despite technology making it easier to access. By using the it we are able to generate a model to provide us an idea of at what time the daily cases will be high and how we can further take precaution to avoid the virus. Apart from the statistical way of forecasting there are several other parameters of how will the daily cases rise exponentially.So the final results of our project are as follows:
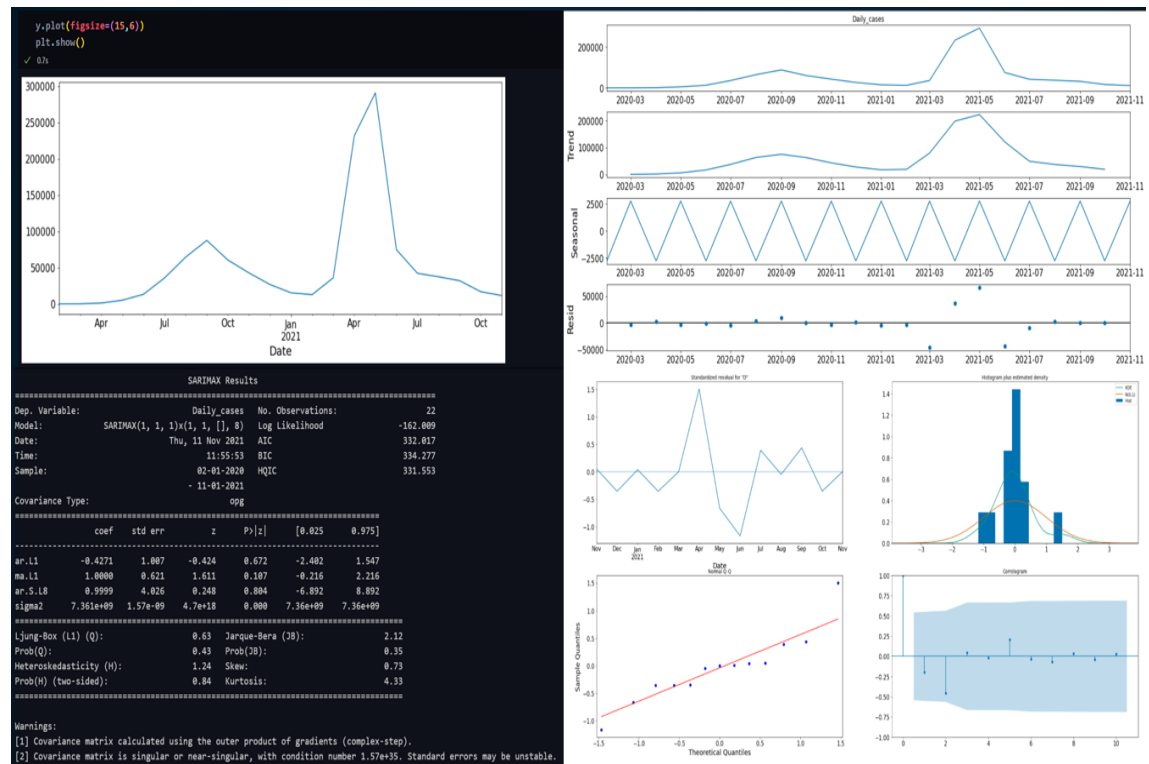

Figure 7.1: Result

Also apart from our project, time series analysis can be used for non-stationary data—things that are constantly fluctuating over time or are affected by time.

Chapter 8

Conclusion and Future Work

Modern time series forecasting methods are essentially rooted in the idea that the past tells us something about the future. Of course, the question of how exactly we are to go about interpreting the information encoded in past events, and furthermore, how we are to extrapolate future events based on this information, constitute the main subject matter of time series analysis . For our project we were able to correctly forecast the result to a good extent.


```python
mse = ((y_hat - y_truth) ** 2).mean()
rmse = np.sqrt(mse)
print('The Mean Squared Error of our forecasts is {}'.format(round(mse, 2)))
print('The Root Mean Squared Error of our forecasts is {}'.format(round(rmse, 2)))
✓ 0.6s
he Mean Squared Error of our forecasts is 146861.07
he Root Mean Squared Error of our forecasts is 383.22
```
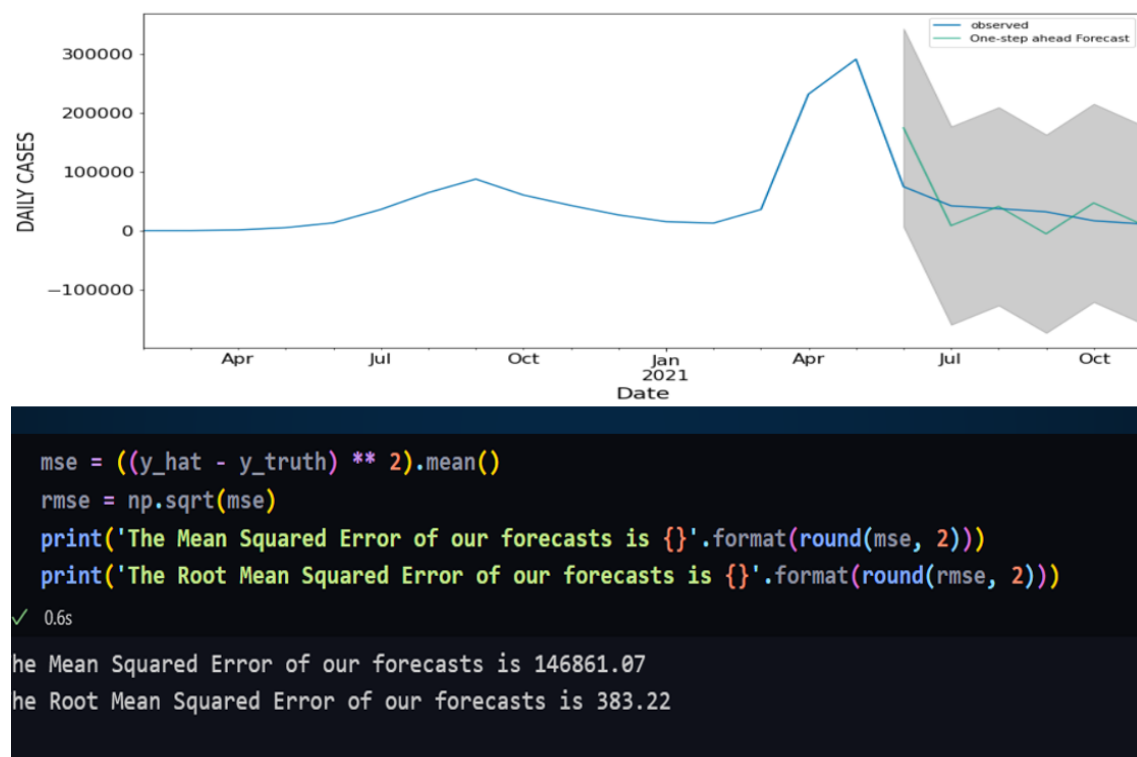
Figure 8.1: Forecast

Also, time series analysis is being researched and worked upon continuously over years So chances of these gaps not being resolved are negligible.

Chapter 9

Bibliography

1. A Comparative Study and Analysis of Time Series Forecasting Techniques.
2. A Good Alternative for ARIMA Method For Seasonal Univariate Time Series forecasting
3. Time Series Analysis
4. Weakness of Time Series
5. Big Data and TimeSeries
6. Covid-19 data