Stock Market Prediction Time Series Analysis – ARIMA model

B. Tech-IT Semester- VI

Prepared at

ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

Bhaskaracharya National Institute for Space Applications & Geo-informatics
Ministry of Electronics and Information Technology, Govt. of India.

Gandhinagar

Prepared By

Ishan Pansuriya	**Ketul Patel**	**Smit Borasaniya**
ID No. M1	**ID No. M1**	**ID No. M1**

Guided By:

Anitha Suji
Project Scientist
BISAG- N, Gandhinagar.

SUBMITTED TO

Indus University

Bhaskaracharya National Institute for Space Applications and Geo-informatics

MeitY, Government of India

Phone: 079 - 23213081 Fax: 079 - 23213091

E-mail: info@bisag.gujarat.gov.in, website: https://bisag-n.in/

ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

CERTIFICATE

This is to certify that the project report compiled ***Mr.Ishan P. Pansuriya, Mr.Ketul V. Patel ,Mr.Smit G. Borasaniya*** *students of 6th Semester* ***B. Tech - IT from Indus University, Ahmedabad*** *have completed their summer internship project satisfactorily. To the best of our knowledge this is an original and bonafide work done by them. They have worked on Data Science application for "****Stock Market Prediction Time Series Analysis – ARIMA model****", starting from June 15th, 2021 to September 15th, 2021.*

During their tenure at this Institute, they were found to be sincere and meticulous in their work. We appreciate their enthusiasm & dedication towards the work assigned to them.

We wish them every success.

Anitha Suji

Project Scientist,

BISAG- N, Gandhinagar

T. P. Singh

Director General,

BISAG- N, Gandhinagar

Certificate of College

CERTIFICATE

This is to certify that the 6th Semester Internship Project entitled " **Stock Market Prediction Time Series Analysis – ARIMA model**" has been carried out by **Ishan Pansuriya (IU1841220028), Ketul Patel (IU1841220032) and Smit Borasaniya (IU1841220053)** under my guidance in fulfilment of the degree of Bachelor of Engineering in INFORMATION TECHNOLOGY (6th Semester) of Indus University, Ahmedabad during the academic year 2021.

Head of the Department

Dr. Seema Mahajan

About BISAG-N

ABOUT THE INSTITUTE

Modern day planning for inclusive development and growth calls for transparent, efficient, effective, responsive and low-cost decision-making systems involving multi-disciplinary information such that it not only encourages people's participation, ensuring equitable development but also takes into account the sustainability of natural resources. The applications of space technology and Geo–informatics have contributed significantly towards the socio-economic development. Taking cognizance of the need of geo-spatial information for developmental planning and management of resources, the department of Ministry of Electronics and Information Technology, Government of India, established "Bhaskaracharya National Institute for Space Applications and Geo-informatics" (BISAG- N). BISAG- N is an ISO 9001:2008, ISO 27001:2005 and CMMI: 5 certified institute. BISAG- N which was initially set up to carryout space technology applications, has evolved into a centre of excellence, where research and innovations are combined with the requirements of users and thus acts as a value-added service provider, a technology developer and as a facilitator for providing direct benefits of space technologies to the grass root level functions/functionaries.

BISAG- N's Enduring Growth

Since its foundation, the Institute has experienced extensive growth in the sphere of Space technology and Geo-informatics. The objective with which BISAG- N was established is manifested in the extent of services it renders to almost all departments of the State. Year after year the institute has been endeavouring to increase its outreach to disseminate the use of geo-informatics up to grassroots level. In this span of nine years, BISAG- N has assumed multi-dimensional roles and achieved several milestones to become an integral part of the development process of the Gujarat State.

BISAG-N Journey

2003-04

Gujarat SATCOM Network

2007-08

Centre for Geo-informatics Applications

2010-11

Academy of Geo-informatics for Sustainable Development

2012-13

A full-fledged Campus

Activities

Satellite Communication..

for promotion and facilitation of the use of broadcast and teleconferencing networks for distant interactive training, education and extension.

Remote Sensing..

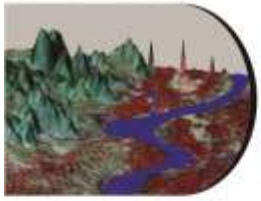
for Inventory, Mapping, Developmental planning and Monitoring of natural & man-made resources.

Geographic Information System..

for conceptualization, creation and organization of multi purpose common digital database for sectoral/integrated decision support systems.

Global Navigation Satellite System..

for Location based Services, Geo-referencing, Engineering Applications and Research.



Photogrammetry..

for Creation of Digital Elevation Model, Terrain Characteristic, Resource planning.

Cartography..

for thematic mapping, value added maps.

Software Development..

for wider usage of Geo-spatial applications, Decision Support Systems (desktop as well as web based), ERP solutions.

Education, Research and Training..

for providing Education, Research, Training & Technology Transfer to large number of students, end users & collaborators.

Applications of Geospatial Technology for Good Governance: Institutionalization

Through the geospatial technology, the actual situation on the ground can be accessed. The real-life data collected through the technology forms the strong foundation for development of effective social welfare programs benefiting directly the grass root level people. The geospatial data collected by the space borne sensors along with powerful software support through Geographic Information System (GIS), the vital spatio-temporal maps, tables, and various statistics are being generated which feed into Decision Support System (DSS).

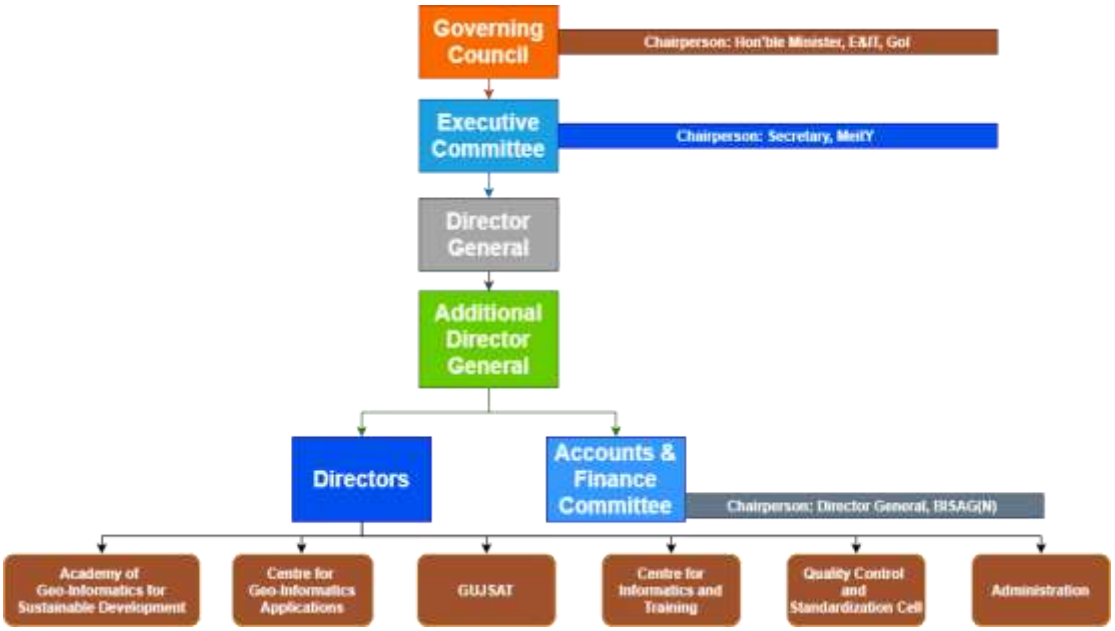
A multi-threaded approach is followed in the process of institutionalization of development of such applications. The 5 common threads which run through all the processes are: *Acceptability, Adaptability, Affordability, Availability and Assimilability*.
These are the "Watch Words" which any application developer has to meet. The "acceptability" addresses the issue that the application developed has met the wide acceptability among the users' departments and the ultimate end beneficiary by way of providing all necessary data and statistics required. The "affordability" addresses the issue of the application product being cost effective. The "availability" aspect looks into aspect of easily accessible across any platform, anywhere and anytime. The applications should have inbuilt capability of easy adaptability to the changing spatio- and temporal resolutions of data, new aspects of requirements arising from time to time from users. The assimilability aspect ensures that the data from various sources / resolutions and technologies can be seamlessly integrated.

ACCEPTABILITY	Problem definition by users Proof of Concept development without financial liability on users Execution through collaboration under user's ownership
ADOPTABILITY	Applications as per present systems & database Maximum Automation Minimum capacity building requirement at the user end
AFFORDABILITY :	Multipurpose geo-spatial database, common, compatible, standardized (100s of layers) In house developed/open source software Full Utilization of available assets
AVAILABILITY:	Departmental /Integrated DSS Desired Product delivery anytime, anywhere in the country
ASSIMILABILITY	Integration of Various technologies like RS, GIS, GPS, Web MIS, Mobile etc.

Organizational Setup

The Institute is responsible for providing information and technical support to different Departments and Organizations. The Governing Body and the Empowered Executive Committee govern the functioning of BISAG- N. The Institute is registered under the Societies Registration Act 1860. Considering the scope and extent of activities of BISAG- N, its organizational structure has been charted out with defined functions.

Organizational Setup of BISAG- N



Governing Body

For smoother, easier and faster institutionalization of Remote Sensing and GIS technology, decision makers of the state were brought together to form the Governing Body. It is the supreme executive authority of the Institute. The Governing Body comprises of ex-officio members from various Government departments and Institutes.

Secretary means : Additional Chief Secretary / Principal Secretary / Secretary of Respective Department of Government of Gujarat

- Chief Secretary, Government of Gujarat .....Chairperson
- Secretary, Science and Technology.....Member
- Secretary, Finance Department .....Member
- Secretary, Education Department.....Member
- Secretary, Revenue Department .....Member
- Secretary, Panchayats, Rural Housing and Rural Development Department....Member
- Director, Space Applications Centre, ISRO, Ahmedabad .....Member
- Dr. George Joseph, Former Director, SAC, ISRO, Ahmedabad.....Member
- Vice Chancellor, Gujarat University, Ahmedabad.....Member
- Chief Executive Officer, Gujarat Infrastructure Development Board.....Member
- Jt. Secretary ..... Invitee
- Director, BISAG .....Member Secretary

Centre for Geo-informatics Applications

Introduction

The objective of this technology group is to provide decision support to the sectoral stake holders through scientifically organized, comprehensive, multi-purpose, compatible and large scale (village level) geo-spatial databases and supporting analytical tools. These activities of this unit are executed by a well-trained team of multi-disciplinary scientists. The government has provided a modern infrastructure along with the state-of-the-art hardware and software. To study the land transformation and development over the years, a satellite digital data library of multiple sensors of last twenty years has been established and conventional data sets of departments have been co-registered with satellite data. The geo-spatial databases have been created using conventional maps, high resolution satellite 2D and 3D imagery and official datasets (attributes).The geo-spatial databases include terrain characteristics, natural and administrative systems, agriculture, water resources, city survey maps, village maps with survey numbers, water harvesting structures, water supply, irrigation, power, communications, ports, land utilization pattern, infrastructure, urbanization, environment data, forests, sanctuaries, mining areas, industries. They also include social infrastructure like the locations of schools, health centres, institutions, aganwadies, local government infrastructure etc. The geospatial database of nagar-palikas includes properties and amenities captured on city and town planning maps with 1000 GIS layers. Similar work for villages has been initiated as a pilot project.

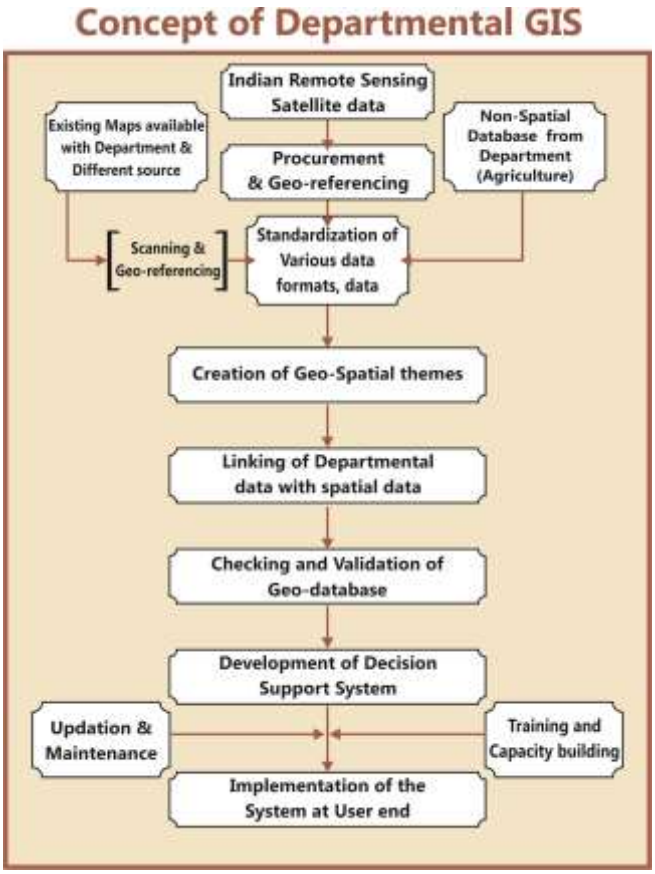
The applications of space technology and geo-informatics have been operational in almost all the development sectors of the state. Remote sensing and GIS applications have provided impetus to planning and developmental activities at grass root level as well as monitoring and management in various disciplines.

The GIS based Applications Development

The GIS software is a powerful tool to handle, manipulate and integrate both the spatial and non-spatial data. The GIS system operates on the powerful backend data base and Sequential Query Language (SQL) to inquiry the data bases. It has the capability to handle large volume of data and process to yield values of parameters which can be input to very important government activity as Decision Support System (DSS). Its mapping capabilities help the users and specialists in generating single and multi-theme wise maps.

The GIS based applications development has been institutionalized in BISAG- N. This process can be listed as (Refer Figure for Details)

- Making the users aware of the GIS capabilities through introductory training programme and by exposing to already developed projects as success stories.
- Helping the users in defining the GIS based projects.
- Digitizing the data available with the users and encouraging them to collect any additional data as may be required.
- Generating the appropriate data bases with the full involvement of the users following the data bases standards

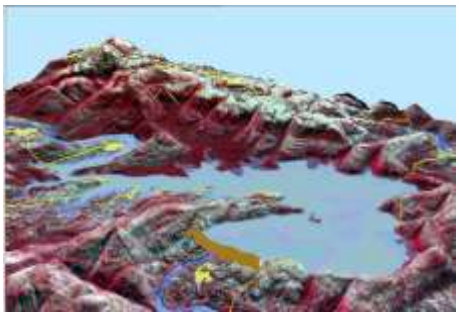


Remote Sensing and GIS Sectoral Applications:

Geo-informatics-based Irrigation Management and Monitoring System

- The Geo-spatial information system for Irrigation water Management and Monitoring system for command areas in Sardar Sarovar Narmada Nigam Limited (SSNL) has been developed. Satellite image-based Irrigation monitoring system has been developed in GIS. From the multi-spectral Satellite images of every month, the irrigated areas were extracted.
- The irrigated area was overlaid on the geo-referenced cadastral maps and the statistics of area irrigated has been estimated.
- The user-friendly Customized Decision Support System (DSS) has been developed.

Preparation of DPR of Par–Tapi-Narmada Link using Geo-informatics for National Water development Agency (NWDA)



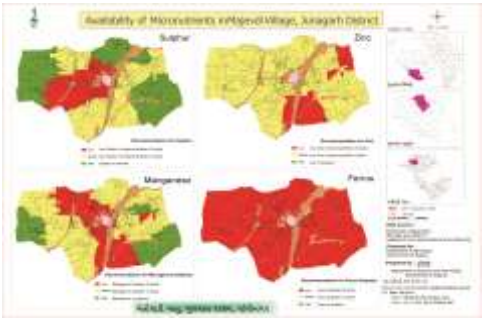
- The main objective of Par–Tapi-Narmada Link project is to divert surplus water available in west flowing rivers of south Gujarat and Maharashtra for utilization in the drought prone Saurashtra and Kachcha. On the request from NDWA, preparation of various maps for proposed DPR work was undertaken by the BISAG- N. Land use and submergence maps of proposed dams along with its statistics have been prepared by the BISAG- N. The detailed work consisted of generation of Digital Elevation Model (DEM), contour generation, Land use mapping, forest area generation of submergence extent at different levels etc.

Agriculture

District and Village-level Crop Inventory

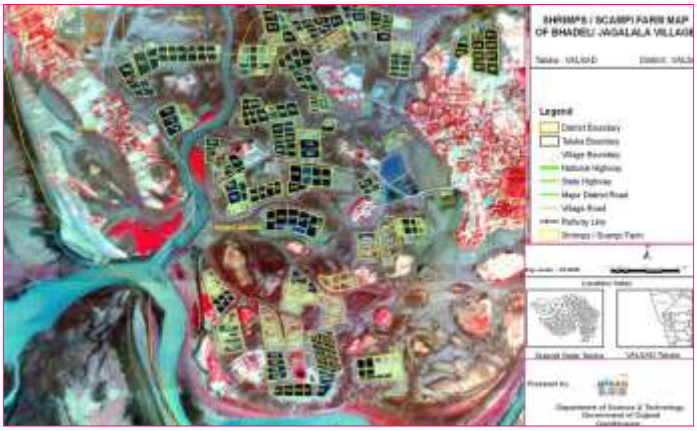
- Remote Sensing (RS) based Village-level Crop Acreage Estimation at was taken up in two villages of Anand and Mehsana districts of Gujarat state. The major objective of this study was to attempt village-level crop inventory during two crop seasons of Kharif (monsoon season) and Rabi (winter season) using single-date Indian Remote Sensing (IRS) LISS-III and LISS-IV digital data of maximum vegetative growth stage of major crops during each season.
- District-level crop acreage estimation during three cropping seasons namely Kharif, Rabi and Zaid (summer) seasons was also carried out in all the 26-districts of Gujarat State. Summer crop acreage estimation Gujarat State was carried out during 2012.

Spatial Variability Mapping of Soil Micro-Nutrients



- The spatial variability of soil micro-nutrients like Fe, Mn, Zn and Cu in various villages of different districts, Gujarat state was mapped using geo-informatics technology. The major objectives of this study were i) to quantify the variability of Mn, Fe, Cu and Zn concentration in soil; ii) to map the pattern of micro-nutrient variability in cadastral maps, iii) suggest proper application of micro-nutrients based on status of deficiency for proper crop management and iv) preparation of village-level atlases showing spatial variability of micro-nutrients.

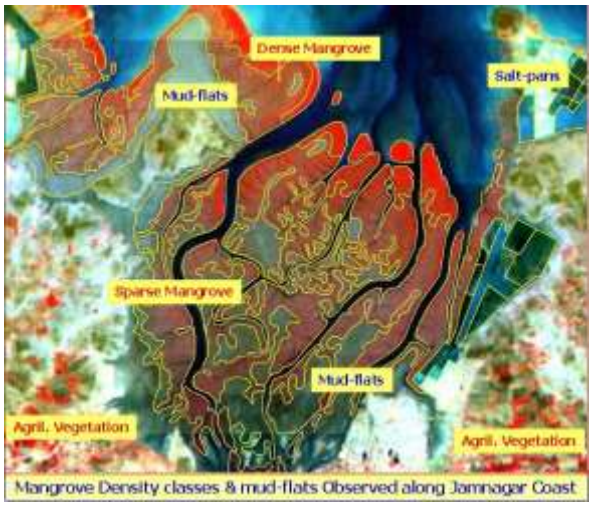
Geo-spatial Information System for Coastal Districts of Gujarat



- The project on development of Village-level Geo-spatial Information System for Shrimp Farms in Coastal Districts of Gujarat, was taken with major objective of development of Village-level Geo-spatial Information System for Shrimp/Scampi areas using Remote Sensing (RS) and GIS. This project was sponsored by the Marine Products Export Development Authority (MPEDA), Ministry of Commerce & Industry, Government of India for scientific management of Scampi farms in the coastal districts which can help fishermen to better their livelihood and increase the economic condition on sustainable basis. The customized query shell was developed using the open-source software for sharing the information amongst the officers from MPEDA and potential users. This has helped the farmers to plan their processing and marketing operations so as to achieve better remunerations.

Environment and Forest

Mapping and Monitoring of Mangroves in the Coastal Districts of Gujarat State



Mangrove Density classes & mud-flats Observed along Jamnagar Coast

- Gujarat Ecology Commission, with technical inputs from the Bhaskaracharya National Institute for Space Applications and Geo-informatics - N (BISAG- N) made an attempt to publish Mangrove Atlas of the Gujarat state. Mangrove atlas for 13-coastal districts with 35-coastal talukas in Gujarat, have been prepared using Indian Remote sensing satellite images. The comparison of mangrove area estimates carried out by BISAG- N and Forest Survey of India (FSI) indicates a net increase in the area under mangrove cover. The present assessment by BISAG- N, has recorded 996.3 sq. km under mangrove cover, showing a steep rise to the tune of 88.03 sq. km. In addition to the existing Mangrove cover, the present assessment also gives the availability of potential area of 1153 sq. km, where mangrove regeneration program can be taken up.

Academy of Geo-informatics for Sustainable Development

Introduction

- Considering the requirement of high end research and development in the areas having relevance of geo-informatics technology for sustainable development, a separate infrastructure has been established. In collaboration with different institutes in the state as well as in the country, R&D activities are being carried out in the areas of climate change, environment, disaster management, natural resources management, infrastructure development, resources planning, coastal hazard and coastal zone management studies, etc. under the guidance of eminent scientists.

- Various innovative methodologies/models developed in this academy through the research process have helped in development of various applications. There are plans to enhance R&D activities manifold during coming years.
- This unit also provides training to more than 600 students every year in the field of Geo-informatics to the students from various backgrounds like water resources, urban planning, computer Engineering, IT, Agriculture in the areas of Remote sensing, GIS and their applications.

- This Academy has been established as a separate infrastructure for advanced research and development through following schools:
  - School of Geo-informatics
  - School of Climate & Environment
  - School of Integrated Coastal Zone Management
  - School of Sustainable Development Studies
  - School of Natural Resources and Bio-diversity
  - School of Information Management of Disasters
  - School of Communication and Society

During XIIth Five-year Plan advance applied research through above schools shall be the main thrust area. Already M. Tech and Ph.D. students of other Universities/ Institutes are doing research in this academy in applied sciences under various collaborative programmes.

M. Tech. Students' Research Programme

The academy started M. Tech. students' research programme in a systematic way. It admitted 11 students from various colleges and universities in Gujarat, Rajasthan and Madhya Pradesh for period of 10 months from August 2011 to May 2012. All the students were paid stipend of Rs. 6000 per month during the tenure. The research covered the following areas:

- Cloud computing techniques
- Mobile communication
- Design of embedded systems
- Aquifer modelling
- Agricultural and Soils Remote Sensing
- Digital Image processing Techniques (Data Fusion and Image Classification).

The research resulted in various dissertations and publications in national and international journals.

• Now nine students, one from IIT, Kharagpur, three from GTU, one from M. S University, Vadodara and four from GU, are undergoing their Ph. D programme. Out of nine, two thesis have been submitted. Two students are from abroad. One each from Vietnam and Yemen. Since then (after approval of research programme from the Governing Body), 200+ papers have been published by the Academy

CANDIDATE'S DECLARATION

We declare that 6th Semester Summer internship project report entitled "**Stock Market Prediction Time Series Analysis – ARIMA model**" is our own work conducted under the supervision of the external guide **Anitha Suji** from BISAG-N (Bhaskaracharya National Institute for Space Applications & Geo-informatics). We further declare that to the best of our knowledge the report for this project does not contain any part of the work which has been submitted previously for such project either in this or any other institutions without proper citation.

Candidate 1's Signature

Ishan Pansuriya

Student ID: M1

Candidate 2's Signature

Ketul Patel

Student ID: M1

Candidate 3's Signature

Smit Borasaniya

Student ID: M1

Submitted To:

BISAG- N, Gandhinagar

ACKNOWLEDGMENT

We are grateful to **T.P. Singh,** Director General (BISAG-N) for giving us this opportunity to work the guidance of renowned people of the field of MIS Based Portal also providing us with the required resources in the company.

We would like to express our endless thanks to our external guide **Anitha Suji**, And Admin Department **Mr. Sidhdharth Patel** at Bhaskaracharya National Institute of Space Application and Geo-informatics for their sincere and dedicated guidance throughout the project development.

Ishan Pansuriya.
Student ID: M1

Ketul Patel.
Student ID: M1

Smit Borasaniya.
Student ID: M1

ABSTRACT

The purpose and essence of Time Series Analysis is to provide the right information in the right place, in the right order, at the right time for the right person for better insights. This is achieved by a complex algorithm used to find the trends in the market values. There are some other forecasting algorithms in existence, but they are not optimised for such predictions including delicate balancing on the model and finding the perfect fit for the predictions. Now a days people are readily investing in stock; the need and the market for algorithms like ARIMA model is proliferating. The use of the algorithm not only mollifies the mistakes of making risky decisions but also help in insinuating the effects it will have on one's finance.

CONTENTS

INTRODUCTION

Time Series Analysis: Definition, Types, Techniques, and When It's Used

For as long as we have been recording data, time has been a crucial factor. In time series analysis, time is a significant variable of the data. Times series analysis helps us study our world and learn how we progress within it.

What is time series analysis?

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.

What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data. Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data.

It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

Why organizations use time series data analysis?

Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. With modern analytics platforms, these visualizations can go far beyond line graphs. When organizations analyze data over consistent intervals, they can

also use time series forecasting to predict the likelihood of future events. Time series forecasting is part of predictive analytics.

It can show likely changes in the data, like seasonality or cyclic behavior, which provides a better understanding of data variables and helps forecast better. For example, Des Moines Public Schools analyzed five years of student achievement data to identify at-risk students and track progress over time.

Today's technology allows us to collect massive amounts of data every day and it's easier than ever to gather enough consistent data for comprehensive analysis.

When time series analysis is used and when it isn't?

Time series analysis is not a new study, despite technology making it easier to access. Many of the recommended texts teaching the subject's fundamental theories and practices have been around for several decades. And the method itself is even older than that. We have been using time series analysis for thousands of years, all the way back to the ancient studies of planetary movement and navigation.

Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms.

Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow's weather report to future years of climate change.

Examples of time series analysis in action include:

- Weather data
- Rainfall measurements

- Temperature readings
- Quarterly sales
- Stock prices
- Automated stock trading
- Industry forecasts
- Interest rates

Because time series analysis includes many categories or variations of data, analysts sometimes must make complex models. However, analysts can't account for all variances, and they can't generalize a specific model to every sample.

Models that are too complex or that try to do too many things can lead to lack of fit. Lack of fit or overfitting models lead to those models not distinguishing between random error and true relationships, leaving analysis skewed and forecasts incorrect.

Theory:

Auto Regressive (AR only) model

A pure **Auto Regressive (AR only) model** is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

where, $Y*{t-1}$ is the lag1 of the series, $\beta*1$ is the coefficient of lag1 that the model estimates and α is the intercept term, also estimated by the model.

This video covers the intuition and workings Auto Regressive model. And how PACF can be leveraged for building AR models.

Moving Average (MA only) model

Likewise, a pure **Moving Average (MA only) model** is one where Yt depends only on the lagged forecast errors.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + .. + \beta_0 Y_0 + \epsilon_{t-1}$$

[Error: The beta coefficients in the second equation above is incorrect.]

That was AR and MA models respectively.

So, what does the equation of an ARIMA model look like?

An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

Predicted Yt = Constant + Linear combination Lags of Y (up to p lags) + Linear Combination of Lagged forecast errors (up to q lags)

ARIMA (Autoregressive Integrated Moving Average Model) Model in words:

An ARIMA model is a class of statistical models for analyzing and forecasting time series data.
It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

ARIMA is an acronym that stands for Autoregressive Integrated Moving Average. It is a generalization of the simpler Autoregressive Moving Average and adds the notion of integration.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

- AR: *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.

- I: *Integrated*. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.

- MA: *Moving Average*. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA (p, d, q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

- p: The number of lag observations included in the model, also called the lag order.

- d: The number of times that the raw observations are differenced, also called the degree of differencing.

- q: The size of the moving average window, also called the order of moving average.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e., to remove trend and seasonal structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

Applications of Machine learning in the stock market and artificial intelligence trading

The development of Machine learning and artificial intelligence technologies have started playing a significant role in trading. Since these technologies began to provide quick and accurate results, they have been a part of trading.

Machine learning is the development of Artificial Intelligence. It has also shown it’s great to work in the field of trading and make trading more comfortable and profitable. Machine learning has many applications in the domain of trading.

The forms of Machine learning in the stock market are listed below:

- **Prediction of stock prices by using historical data:**

By now, every one of us realized that Machine Learning is nothing but storing the outcomes and the parameters that lead to those outcomes. ML stores the data in the database; it gives us results using the historical data present.

The stock prices that have to be predicted are called target variables. The historical data used to predict these target variables are called predictor variables. To make these predictions, ML uses the algorithm which uses the predictor variables to predict the result for target variables.

- **Trading at high frequency:**

ML uses algorithms to give results. High-frequency algorithms came into the picture that analyses several thousand trades in a day. They analyze the trades and provide traders with the information that is required to invest in a market. In case the investment bankers or traders need to analyze multiple financial markets to execute large orders.

Pension funds, investment banks, mutual funds, and hedge funds currently use these algorithms. In 2019-20, United States trading gained almost 60 to 70 percent of profit using high-frequency trading algorithms.

- **Detection of frauds:**

Every one of us usually has a fear of being subjected to frauds when we are into trading. We always want to make sure whether the amount we are investing in is safe or not. ML helps us to detect the frauds in the market.

Machine learning usually stores large amounts of data, and it can scan through that data. It can also say if any data is out of box or unusual since it can browse through the data. Using this information, we can easily crack the frauds in the markets.

Purpose

The purpose if the project is to make meaningful insights and help make a meaningful investment for the user.

Scope

With increases in new technologies and new trading apps that provide an ease to the millennials to invest in stock and thus more people indulge creating a huge market gap for beginners who does not know the ins and outs of trading. Thus, the program helps to provide a more insightful and learned decision among the market.

Objective

This program/algorithm can be used to make predictions while including all the hidden patterns which are computed in a model to make it a best option for people to make an erudite investment.

Details of Tools Used

Below is the list of the tools and technologies which we have used in this project:

Tools:

- Visual Studio Code
- Google Collaboratory
- GitHub
- Quandl - API

Main technology/programming language:

- Python
- HTML
- CSS
- JS
- Bootstrap

Framework:

- Django

Database connection:

- MySQL – SQL Lite3

Other Libraries Used:

- Asgiref
- autopep8
- Django
- Pycodestyle
- Pytz
- Sqlparse
- Unipath
- dj-database-url python-decouple
- whitenoise
- Quandl
- Pmdarima
- Sklearn
- Statsmodels
- Matplotlib
- Plotly

Program Analysis

Program Planning:

We defined 3 main pipeline which were essential for the working of the web side application

- Search module
- Fetching data using API call
- ARIMA model fitting and prediction

1. Search module:
   - Objective of the pipeline is to get a query set in return of the searched string by the user
   - The string needed to be passed from HTML page to JavaScript to Django backend
   - Which will then be cross searched in the metadata module and return a query set
   - Which will then be parsed to give us the company name and BSE code of best fit to the string searched by the user

2. Fetching data using API call
   - Objective of this pipeline is to get the data from the API call
   - The BSE code received from the search module is then passed to keep for reference
   - Which will then in return provide us with the historical data of the name company

3. ARIMA model fitting and prediction
   - the objective of this pipeline is to get the best fit model and the prediction using the Arima model
   - the historical data received from the API call is passed into auto Arima function which yields the best fitting model for Arima which we can use for prediction

(function tries the different combination and permutation of the P Q and D values)

- The best fitting value are then passed to predictor function of the Arima model to get the final output

Problem Identification

Though there are other variants of Arima model which profoundly with seasonality / trend in data set. The bases Arima model which is used here is not very accurate with the seasonality in the data.

If a company shows a seasonal trend in sales due to its certain product, if we take that product out of the equation than despite of the given indications Arima model is going to predict the same trend which it has trained on.

Hence every time a data is received a seasonality checked must be performed on the data in order to ensure that the Arima model is not going to pick up on the seasonality and predict the wrong output.

The other major issue can be e when the user search for particular company but the query set returns a different set of company name and code.

Which requires a robot so search module that you can search precisely on the string provided by the user.

Stable connection with your module, API, and API's server as the data fetching is dynamic.

System Requirement

Minimum Hardware Requirements

- **CPU:** Quad Core CPU (>= 2ghz)
- **Memory:** 4096 MB
- **Free disk space:** 10 GB

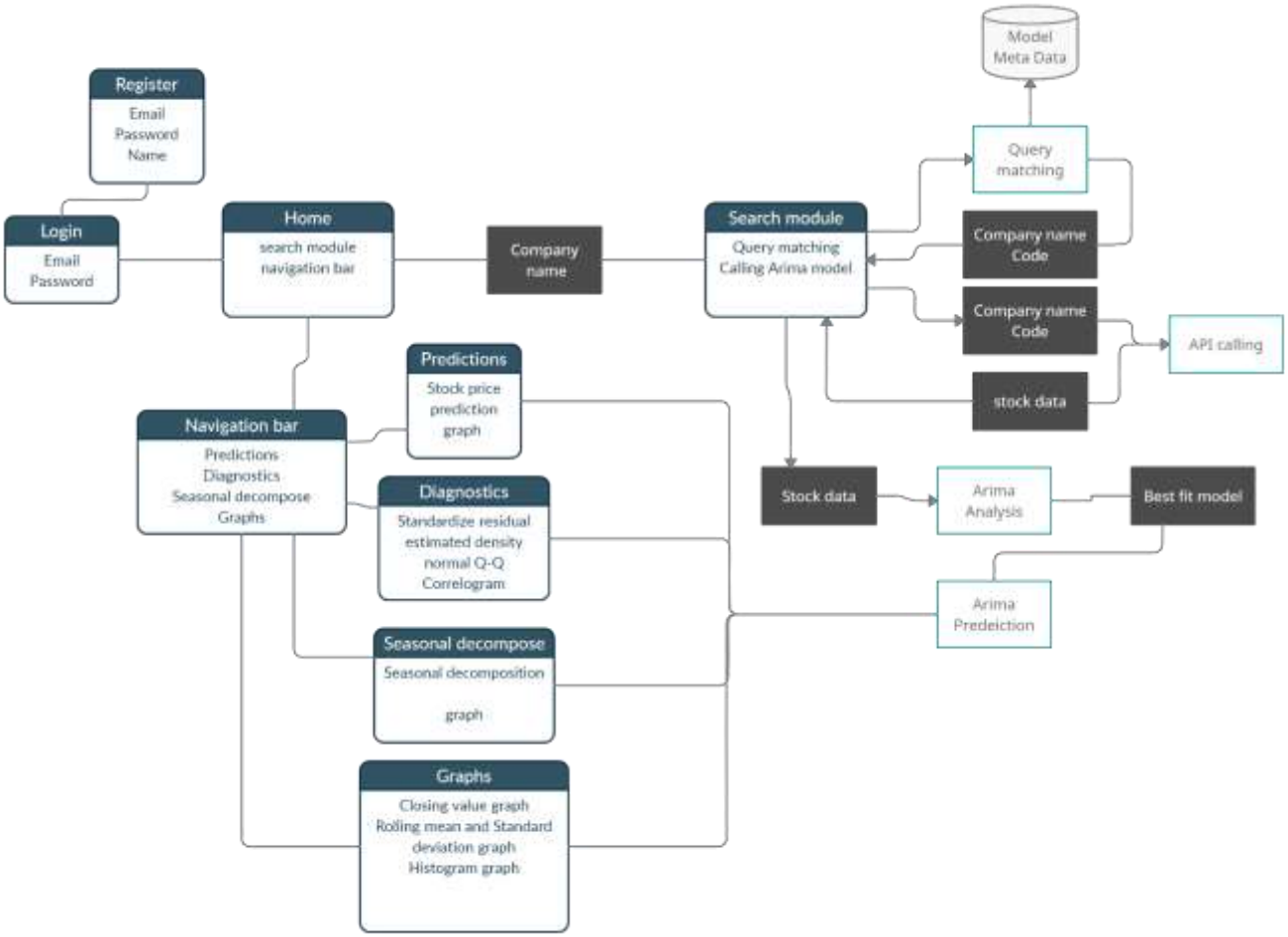
Minimum Software Requirements

- **Ubuntu 18 or above**
- **Windows 8 or above**
- **Mac OS 12 or above**

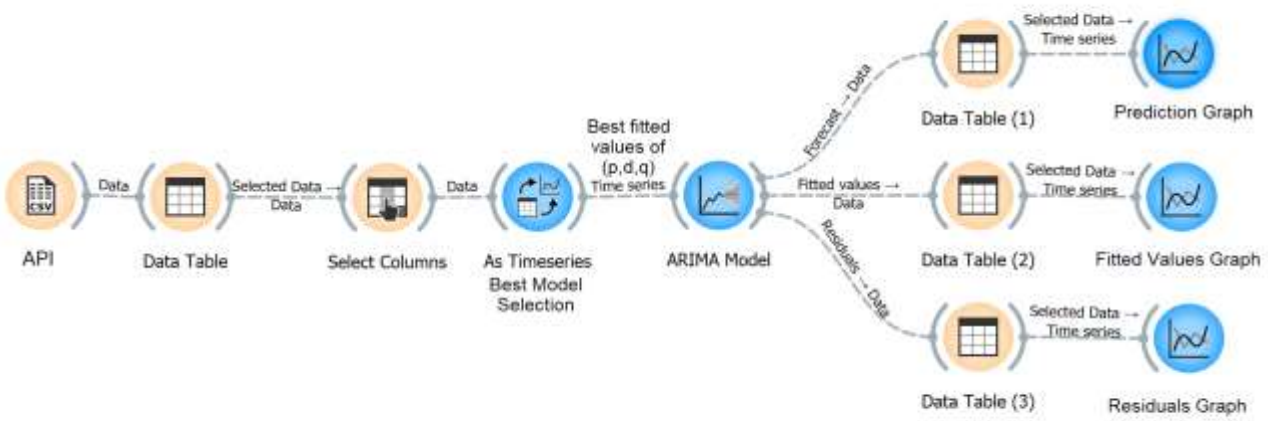
Note: Should support above mentioned tools and technologies.

System Design

- Site Map with Data Flow Diagram

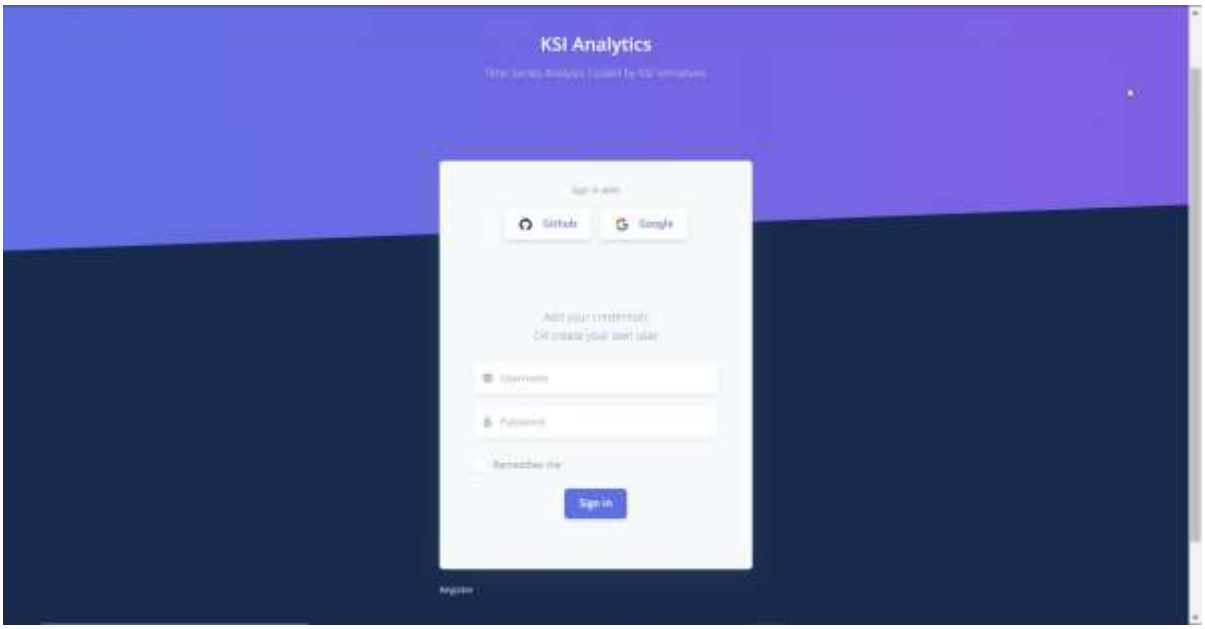


- System Implementation: Pipeline for Arima model

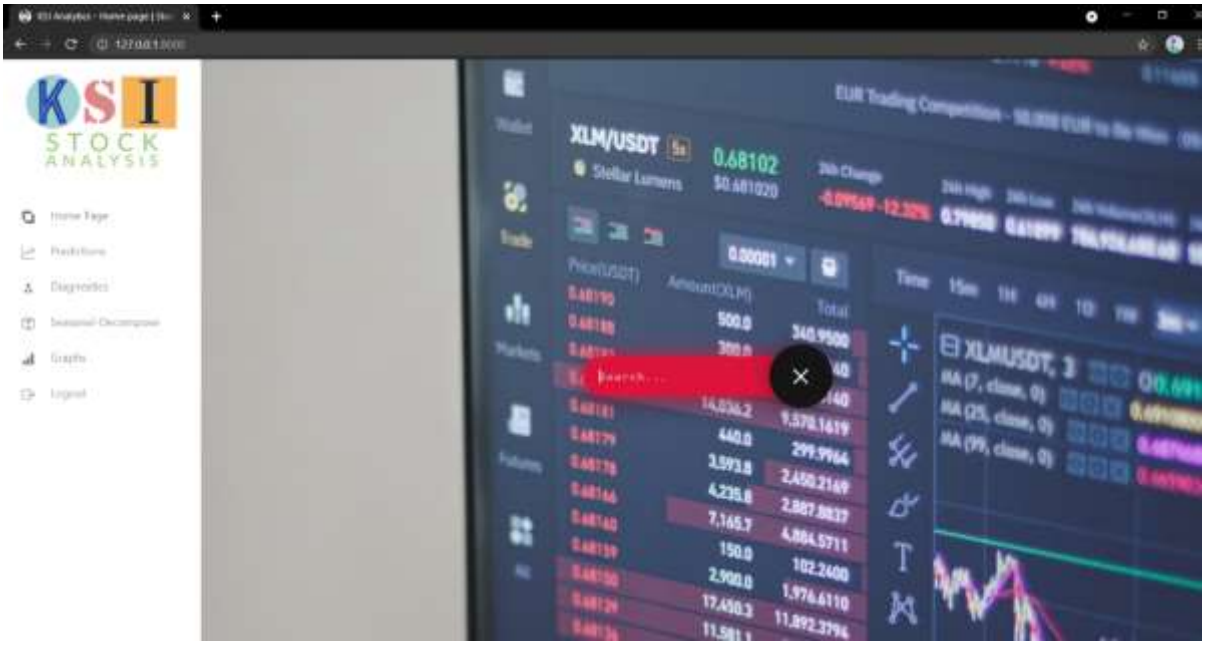


Screen Shots

Login Page



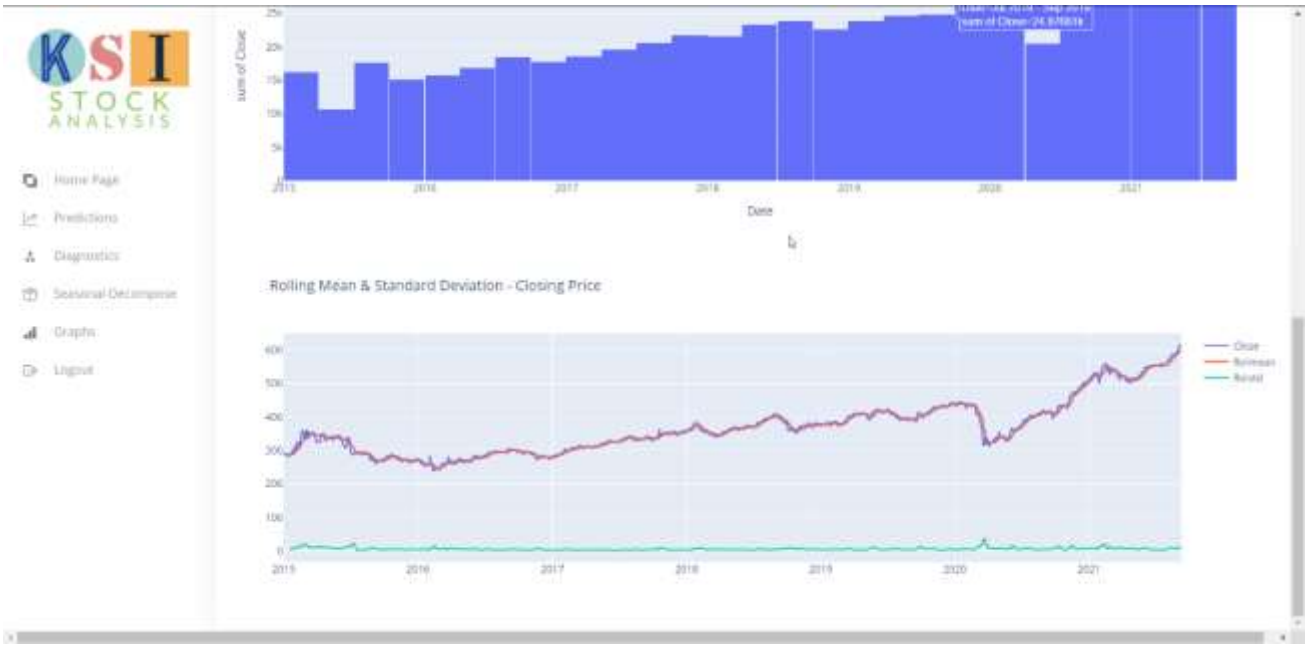
Home Page



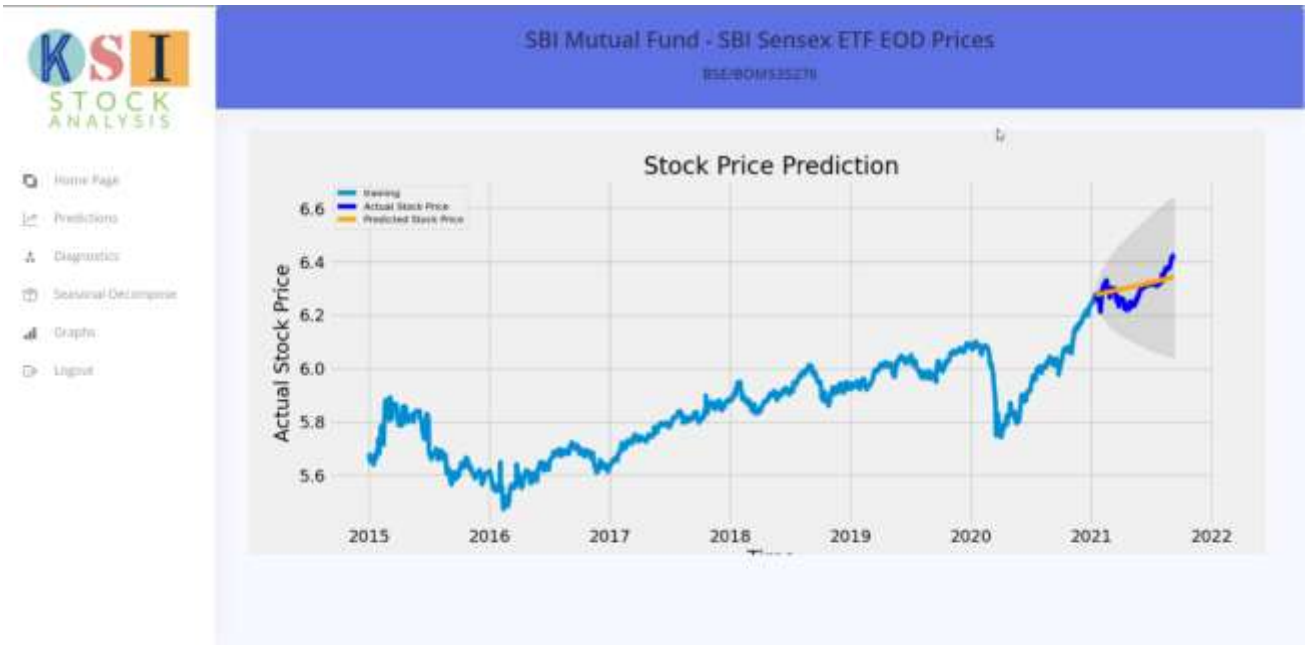
Line Chart

Histogram

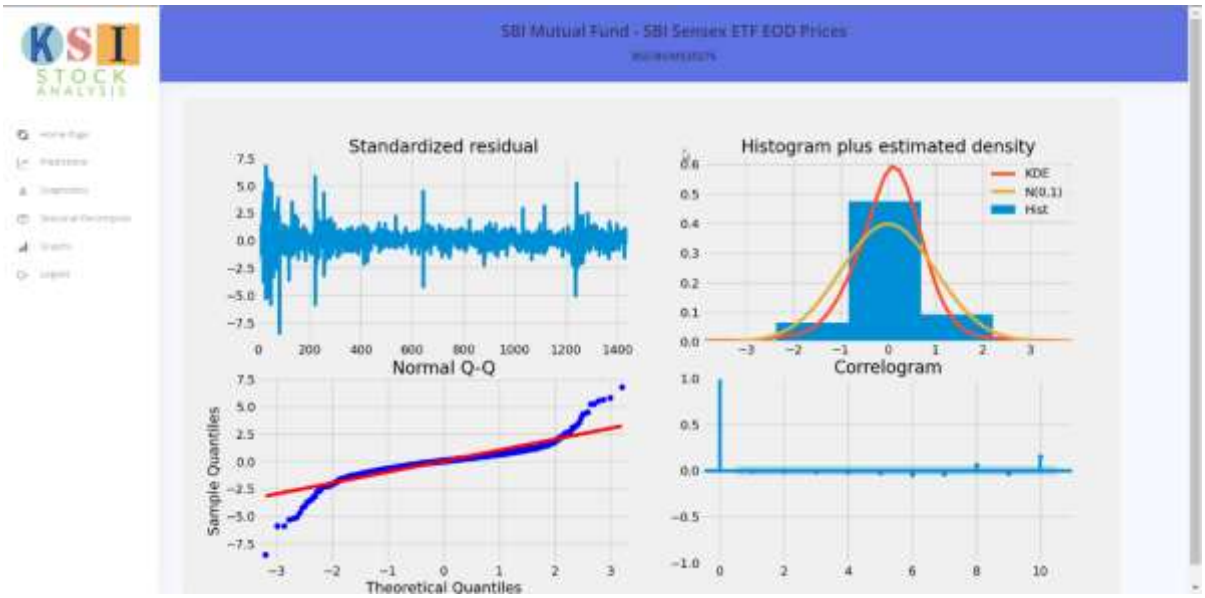
Rolling Mean and Standard Deviation



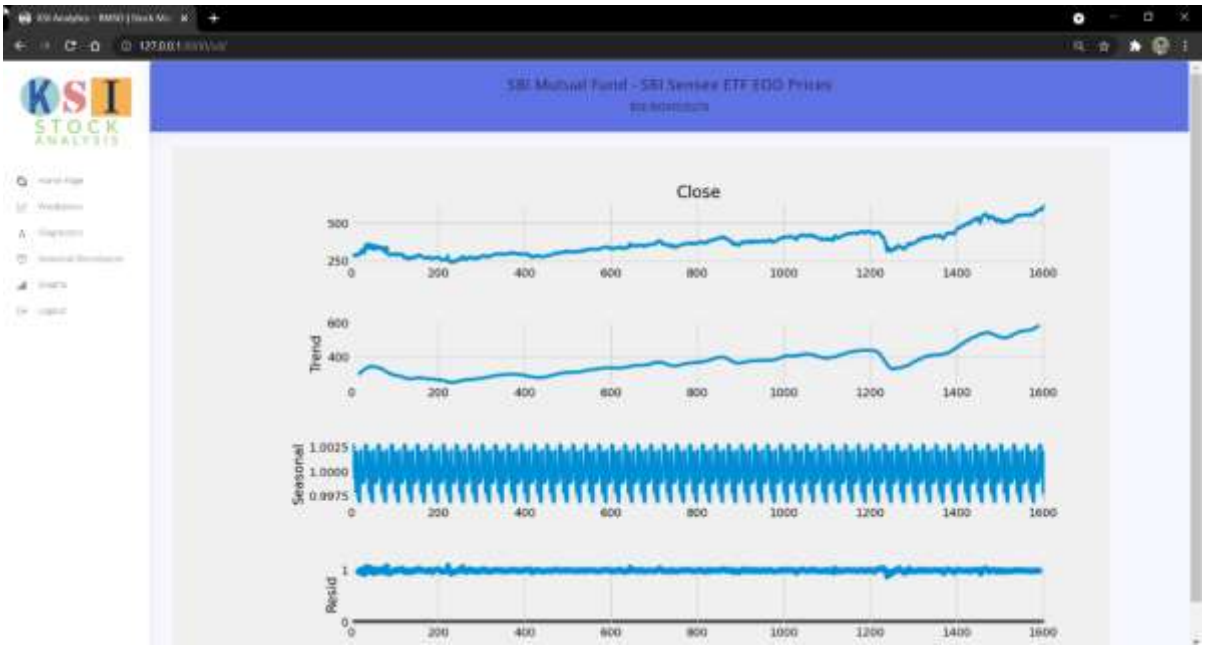
Prediction: SBI Mutual Fund



Seasonal Decomposition

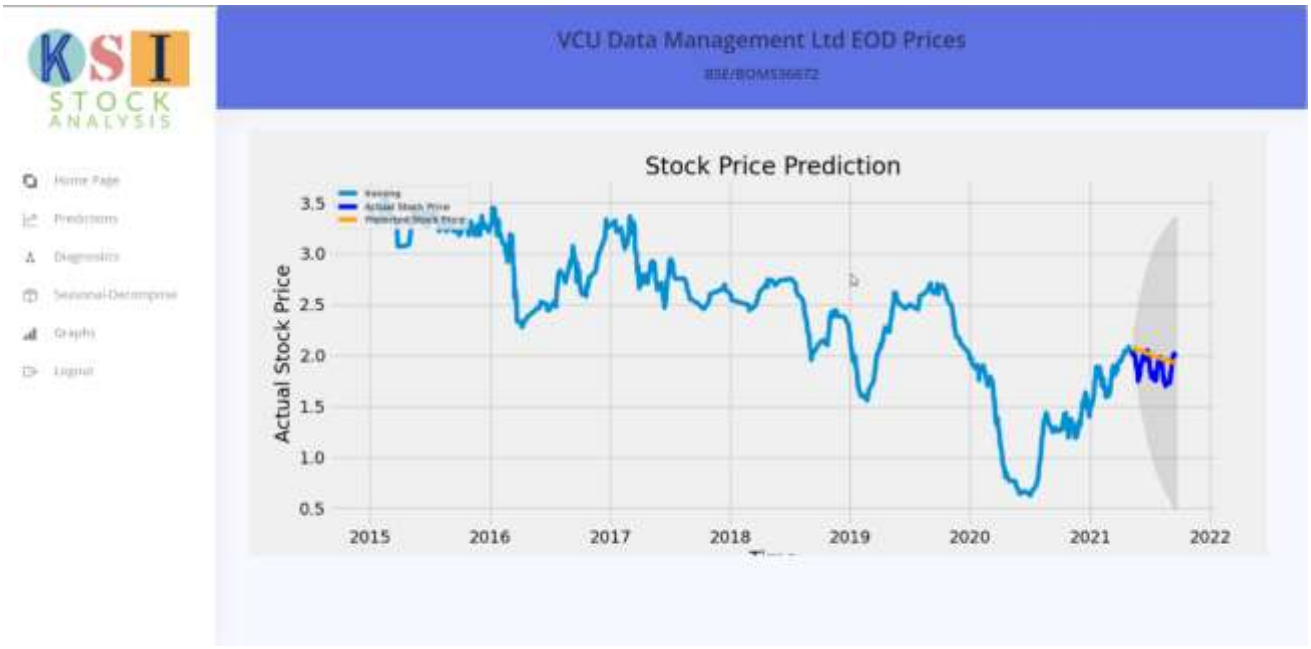


Model Diagnostics



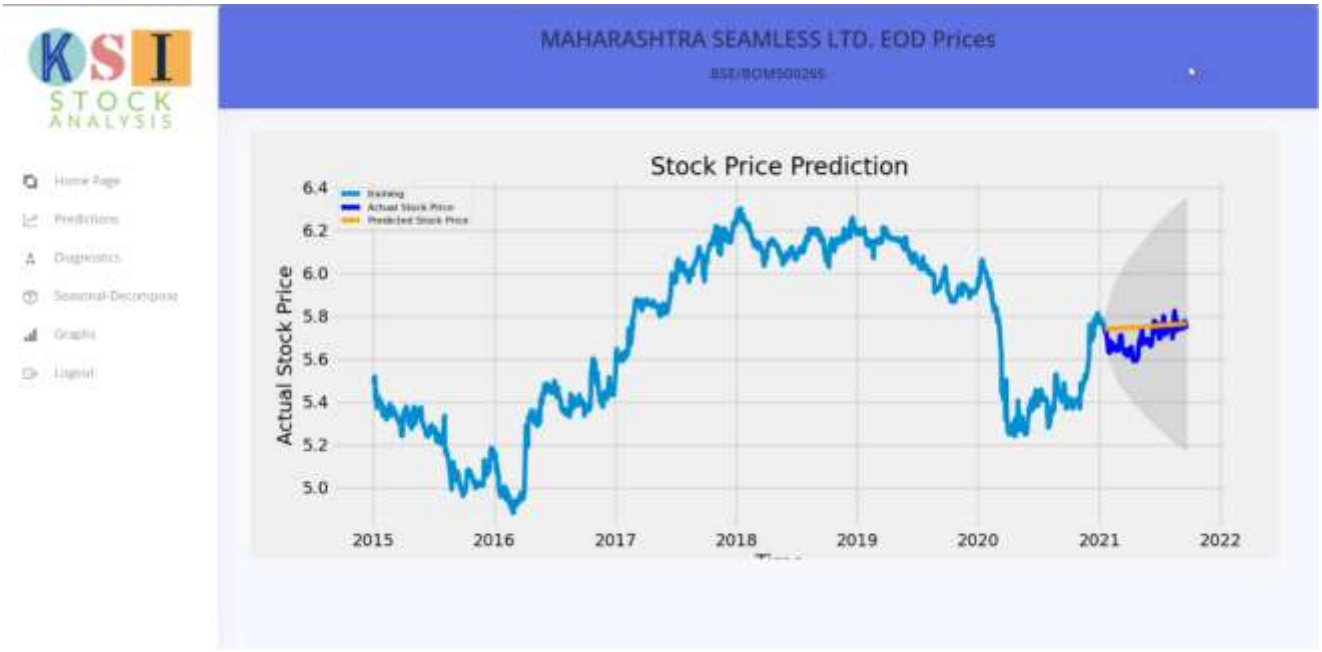
Some predictions for other Companies:

Prediction: VCU Data Management Ltd.



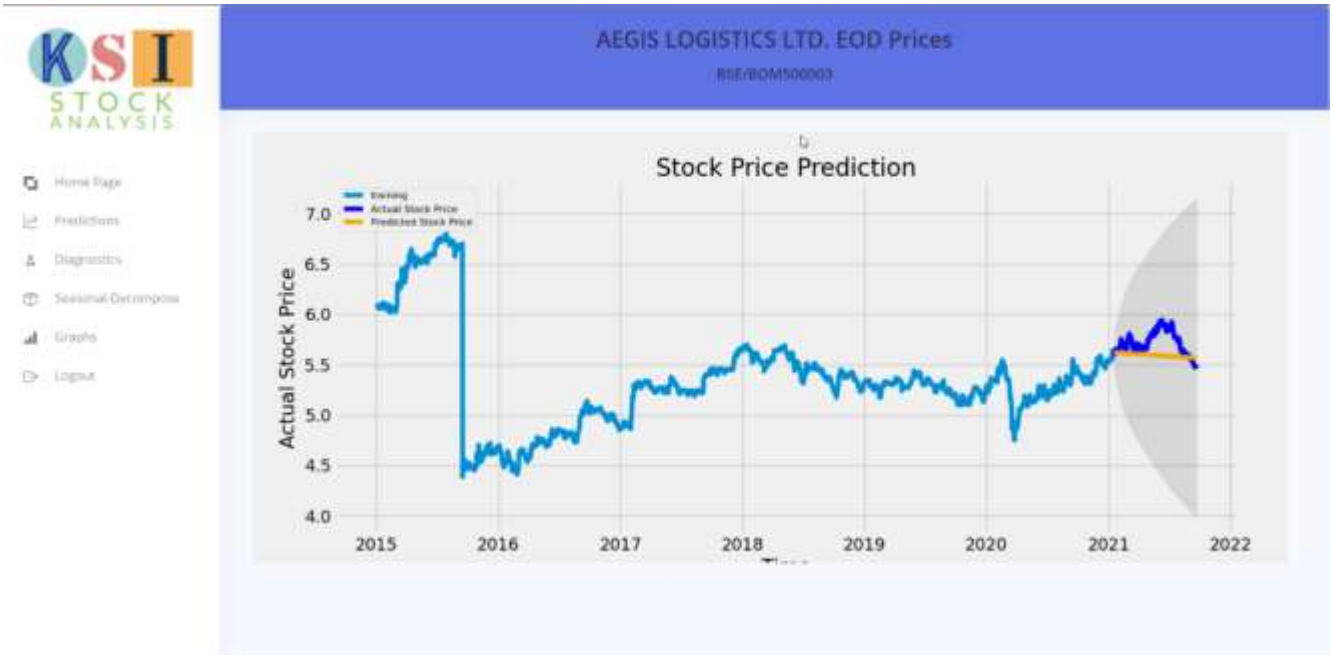
Prediction: Omansh Enterprises Ltd.

Prediction: Maharashtra Seamless Ltd.

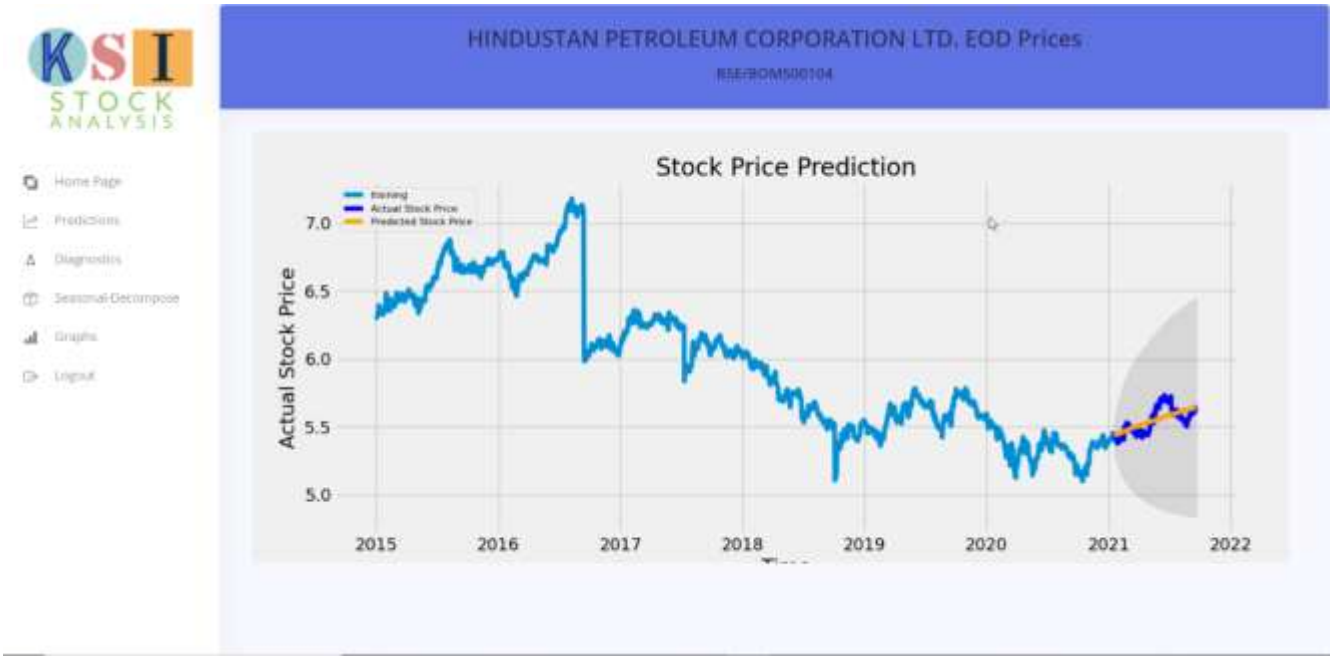


Prediction: Frontline Business Solutions Ltd.

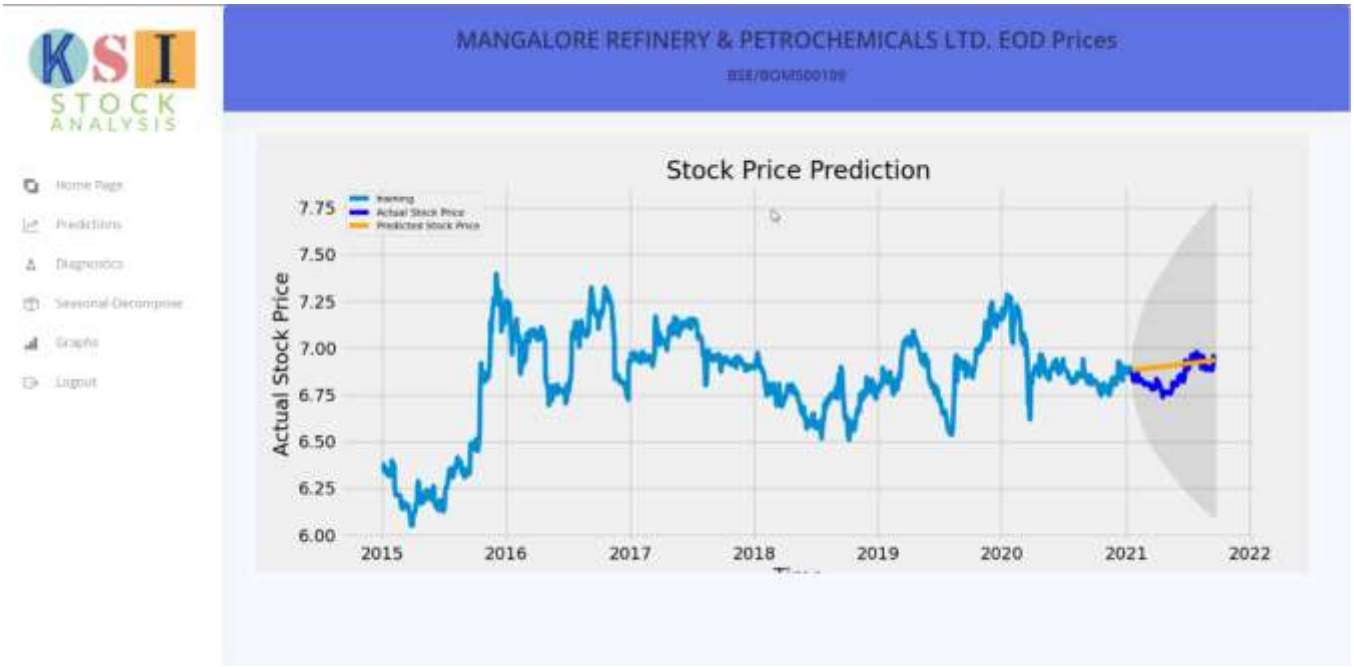
Prediction: Aegis Logistics Ltd.



Prediction: Hindustan Petroleum Corporation Ltd.

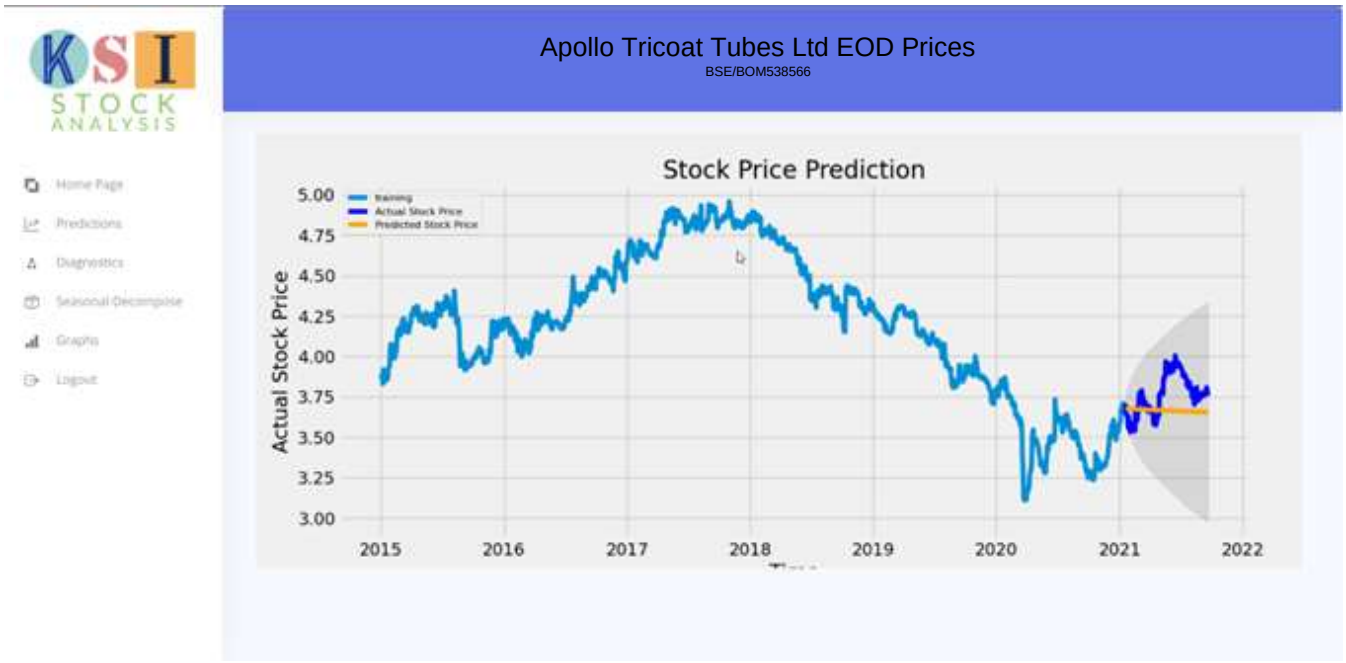


Prediction: Mangalore Refinery and Petrochemicals Ltd.

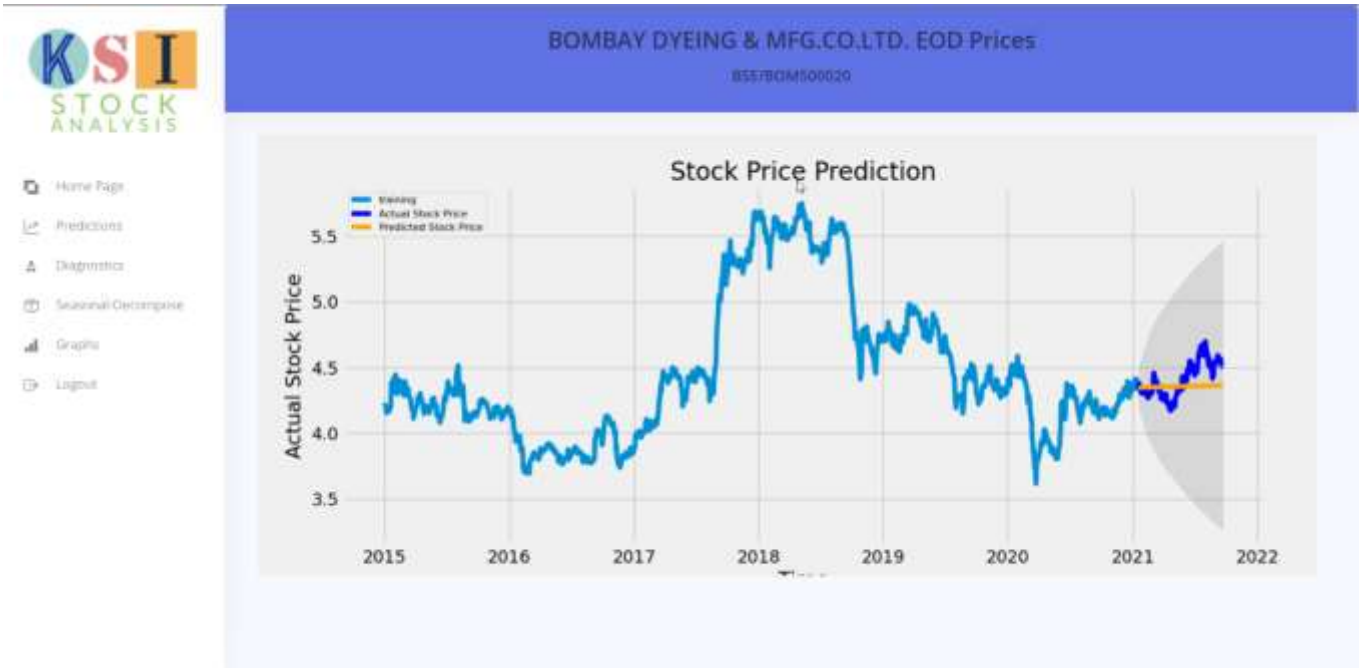


Prediction: The Phoenix Mills Ltd.

Prediction: Apollo Tricoat Tubes Ltd



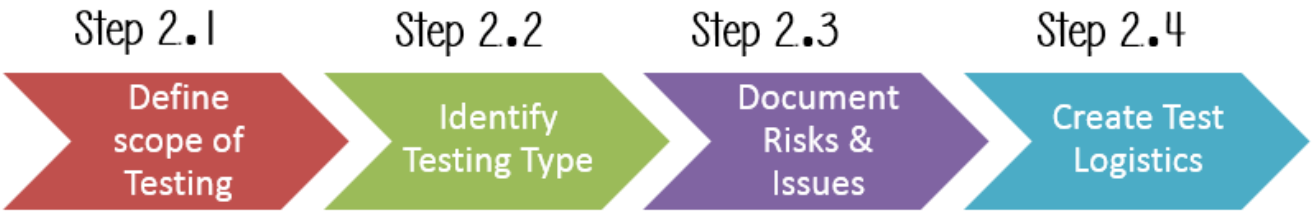
Prediction: Bombay Dyeing and MFG.CO.LTD



Testing

Testing Strategy

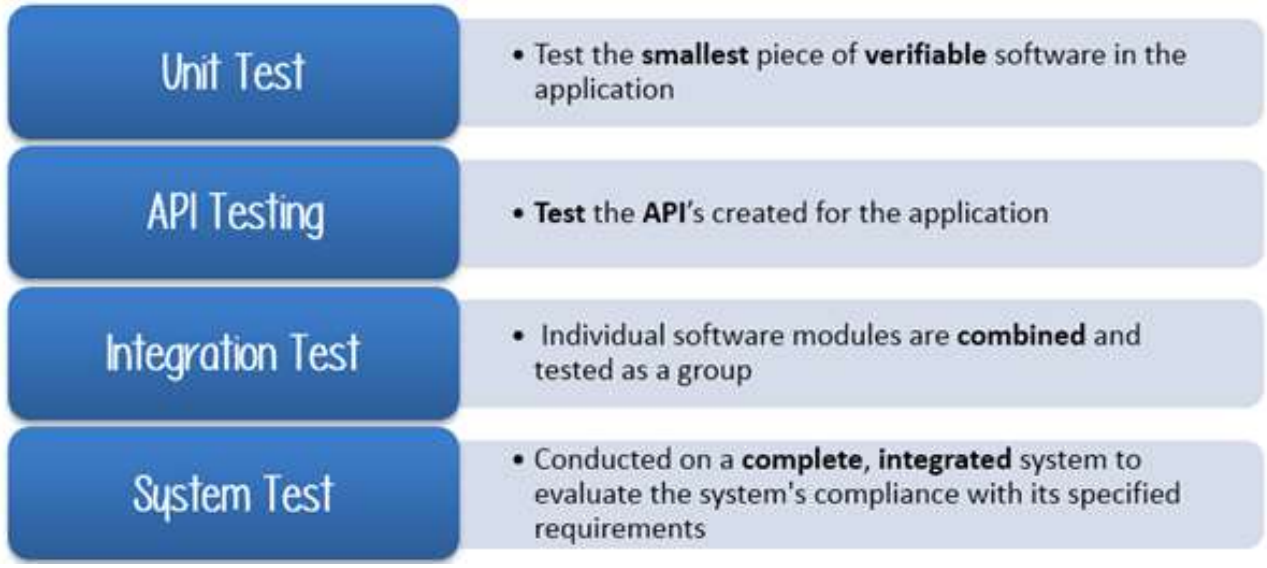
The scope of testing is to clear any runtime errors and also remove all clutter and have an efficient code.



Testing Objective

The objective is to achieve the minimum or no errors and execution of the analysis smoothly and efficiently.

Test Criteria



Resource Planning

Resource planning included division of work with a factor of skills. The front end is developed in Django and ARIMA model is coded in Python.

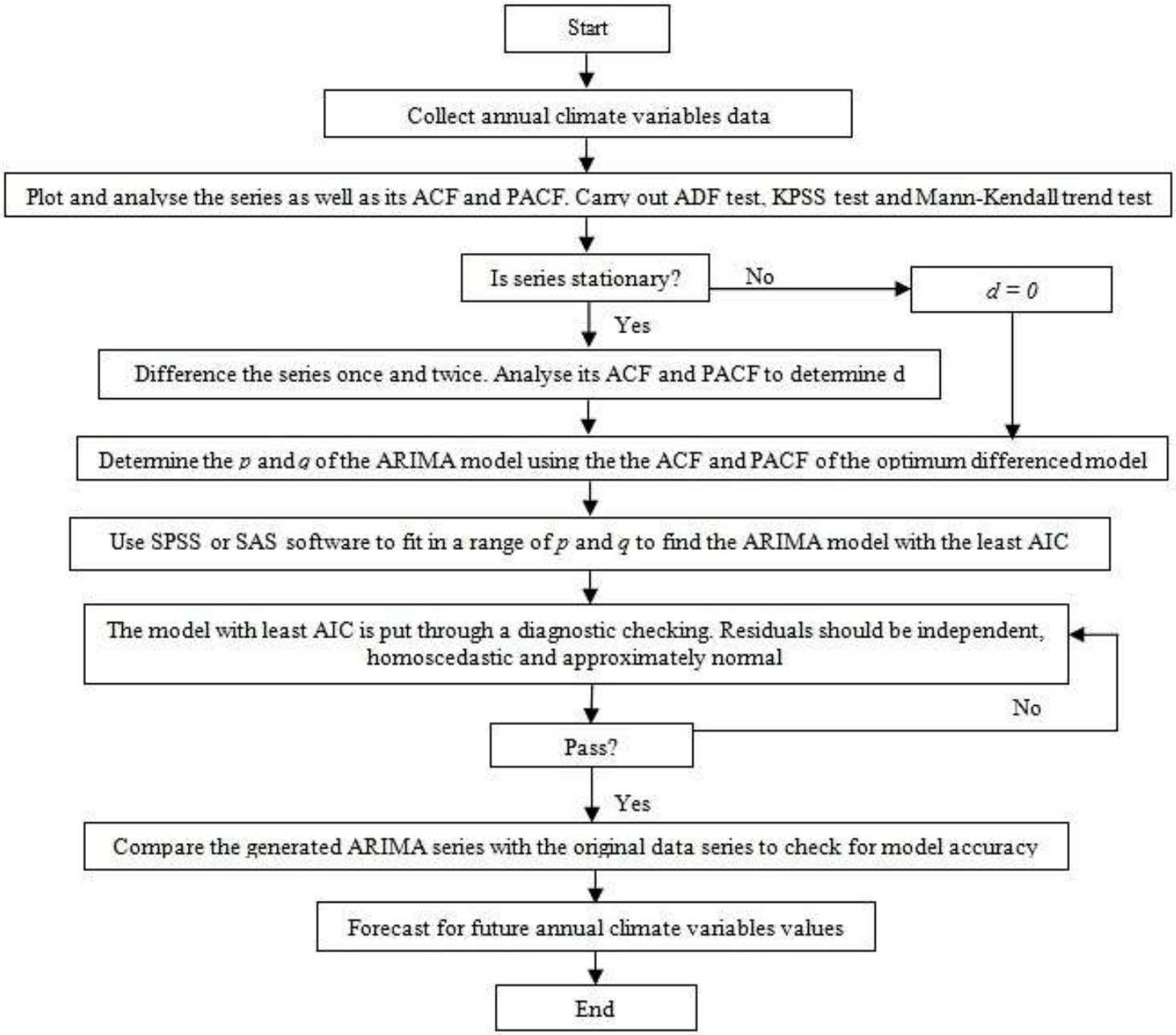
Testing Plan

The test plan consisted of unit testing and integration testing where the model code was first tested on Google Collaboratory platform and then added to the code of the website.

Also, unit testing of the website was performed before conducting integration testing. After the ARIMA model was properly integrated with the API and was well established with the website, integration testing was performed and was run on a virtual local environment.

1. Functionality Testing
   All the functionality of the code is tested so that it I properly directed.
2. Usability Testing
   The interface is user friendly and easily understandable.
3. Web UI Testing
   Its testing is done so check its compatibility with other platforms.
4. Performance Testing
   The testing is conducted to have the least computational time and get responses and insights quickly and easily.
5. Security Testing
   All user password authentication and other such validation of constraints are checked and verified for optimal security.

Testing Method



Program Implementation

Steps for ARIMA implementation

The general steps to implement an ARIMA model are –

- **Load the data:** The first step for model building is of course to load the dataset.

- **Pre-processing:** Depending on the dataset, the steps of pre-processing will be defined. This will include creating timestamps, converting the type of date/time column, making the series univariate, etc.

- **Make series stationary:** In order to satisfy the assumption, it is necessary to make the series stationary. This would include checking the stationarity of the series and performing required transformations, i.e., bringing the p-value < 0.5.

- **Determine d-value:** For making the series stationary, the number of times the difference operation was performed will be taken as the d value.

- **Create ACF and PACF plots(optional):** This is the most important step in ARIMA implementation. ACF PACF plots are used to determine the input parameters for our ARIMA model.

- **Determine the p and q values:** Read the values of p and q from the plots in the previous step.

- **Fit ARIMA model:** Using the processed data and parameter values we calculated from the previous steps, fit the ARIMA model.

- **Predict values on validation set:** Predict the future values.

- **Calculate RMSE:** To check the performance of the model, check the RMSE value using the predictions and actual values on the validation set.

Using Auto ARIMA:

Although ARIMA is a very powerful model for forecasting time series data, the data preparation and parameter tuning processes end up being really time consuming. Before implementing ARIMA, you need to make the series stationary, and determine the values of p and q using plots. Auto ARIMA makes this task really simple for us as it eliminates steps 3 to 6 we saw in the previous section. Below are the steps you should follow for implementing auto ARIMA:

- **Load the data:** This step will be the same. Load the data into your notebook
- **Pre-processing data:** The input should be univariate, hence drop the other columns
- **Fit Auto ARIMA:** Fit the model on the univariate series
- **Predict values on validation set:** Make predictions on the validation set
- **Calculate RMSE:** Check the performance of the model using the predicted values against the actual values

We completely bypassed the selection of p, d and q making it easier.

Program Evaluation

We can evaluate an ARIMA model by preparing it on a training dataset and evaluating predictions on a test dataset.

This approach involves the following steps:

- Split the dataset into training and test sets.
- Walk the time steps in the test dataset.
- Train an ARIMA model.
- Make a one-step prediction.
- Store prediction; get and store actual observation.
- Calculate error score for predictions compared to expected values.
- We can implement this in Python as a new standalone function called evaluate_arima_model () that takes a time series dataset as input as well as a tuple with the p, d, and q parameters for the model to be evaluated.

The dataset is split in two: 90% for the initial training dataset and the remaining 10% for the test dataset.

Each time step of the test set is iterated. Just one iteration provides a model that you could use to make predictions on new data. The iterative approach allows a new ARIMA model to be trained each time step.

A prediction is made each iteration and stored in a list. This is so that at the end of the test set, all predictions can be compared to the list of expected values and an error score calculated. In this case, a mean squared error score is calculated and returned.

Limitations and Future Extensions

Limitations:

- The computation takes long for more historical data of a given company.
- It is not optimized for data with seasonality, i.e., data containing variable values at same time cannot be computed.
- There are some limited features in Web Application for now.

Future work:

Following is the task which can be implemented in future:

- Data analysis for seasonal data.
- Data formatting for better and narrower prediction/absolute output.
- The computation time can be reduced significantly with better modelling and computational power.
- Adding of more features and functionalities.

Accelerates the Hunt for Successful Algorithmic Trading Plans:

Machine Learning is for implementation to accelerate the search for successful Algorithmic Trading Strategies. Since it offers an automated approach, it is quite a bit better than the manual procedure. All these Algorithmic Trading Strategies help dealers by optimizing their profits and mimicking risks. Anyway, there is a competitive advantage in case you've automation to support you for any job. As an example, many strategies make use of Machine Learning for maximizing algorithms, such as linear regressions, deep learning, neural networks, and so forth.

Conclusion and References

Conclusion:

For as long as we have been recording data, time has been a crucial factor. In time series analysis, time is a significant variable of the data. What sets time series data apart from other data is that the analysis can show how variables change over time. Additionally, time series data can be used for forecasting—predicting future data based on historical data. Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Time series forecasting is part of predictive analytics. Time series analysis is not a new study, despite technology making it easier to access. Time series analysis includes many categories or variations of data, analysts sometimes must make complex models. Thus, the program helps to provide a more insightful and learned decision among the market.

References:

https://docs.djangoproject.com/en/3.2/

https://scikit-learn.org/stable/

https://plotly.com/

https://github.com/

https://pypi.org/project/pmdarima/

https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.ARIMA.html

https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Bhaskaracharya National Institute for Space Applications and Geo-informatics
MeitY, Government of India
Phone: 079 - 23213081 Fax: 079 - 23213091
E-mail: info@bisag.gujarat.gov.in, website: https://bisag-n.in/

ISO 9001:2008
ISO 27001:2013
CMMI LEVEL-5

Report Verification Procedure

Date:23/09/2021

Project Name: Stock Market Prediction Time Series Analysis – ARIMA model

Student Name & ID: 1. Ishan Pansuriya - M1
2. Ketul Patel - M1
3. Smit Borasaniya - M1

	Soft Copy	Hard Copy
Report Format:	☐	☐
Project Index:	☐	☐

Sign by Training Coordinator **Sign by Project**