# 

# ABSTRACT

Time series analysis describes, explains, predicts, and controls changes through the time of selected variables. Time-series analysis is a statistical method of analyzing data from repeated observations on a single unit or individual at regular intervals over a large number of observations. Time-series analysis can be viewed as the exemplar of longitudinal designs. The most widely employed approach is based on the class of models known as autoregressive integrated moving average linear regression,(ARIMA) models. regression models can address several major classes of research questions, including an analysis of basic processes, intervention analysis, and an analysis of the pattern of treatment effects over time. Technical aspects of ARIMA models are described.

# INDEX

**S.NO TITLE PAGE**

**ABSTRACT 3**

**LIST OF FIGURES 5**

1. **INTRODUCTION 6**
2. **SYSTEM ANALYSIS 7**

2.1EXISTING SYSTEM 7

2.2 PROPOSED SYSTEM 7

## 2.3 SYSTEM REQUIREMENTS 8

2.3.1 HARDWARE REQURIMENTS 8

2.3.2 SOFTWARE REQUIREMENTS 8

1. **METHODOLOGY 9** 
   1. SYSTEM ARCHITECTURE 9
   2. ALGORITHM DESCRIPTION 10
   3. MODULAR DESIGN 10
2. **IMPLEMENTATION 11** 
   1. WHAT IS PYTHON 11
   2. JUPYTER NOTE BOOK 12

## 4.3 MODULES INSTALLED IN PYTHON 13

4.3.1 NUMPY 16

4.3.2 PANDAS 16

4.3.3 SEABORN 16

4.3.4. MATPLOTLIB 17

4.3.5 SCIP 17

4.3.6 SKLEARN 17

### 4.3.7 USING PIP 18

1. **SAMPLE CODE 18**
2. **OUT PUT 22**

**CONCLUSION 25**

**BIBLIOGRAPHY 26**

**LIST OF FIGURES**

**FIG.NO.TITLE PAGE NO**

* 1. **Dataset 22**
  2. **Data before prediction 23**
  3. **prediction 24**

# 1.INTRODUCTION

We know that planning for the future is very necessary for every Business firm, and every govt. institute, every individual, and every country. Every family is also planning for his income expenditure. As like every business is planning for possibilities of its financial resources & sales and maximization of its profit.

A time series is a set of observations taken at specified times, usually at equal intervals.”

A time series may be defined as a collection of reading belonging to different periods of some economic or composite variables".

Time series analysis describes, explains, predicts, and controls changes through the time of selected variables. Time-series analysis is a statistical method of analyzing data from repeated observations on a single unit or individual at regular intervals over a large number of observations. Time-series analysis can be viewed as the exemplar of longitudinal designs. The most widely employed approach is based on the class of models known as autoregressive integrated moving average linear regression. regression models can address several major classes of research questions, including an analysis of basic processes, intervention analysis, and an analysis of the pattern of treatment effects over time.

# SYSTEM ANALYSIS

**2.1 Existing System**

* There are many other types of predictions available on the web for having a proper analysis for the requirement to be full filled.
* In time series problems, we expect observations close to each other in time to be more similar than observations far away, after accounting for seasonality

**2.2 Proposed System**

* 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
* In our prediction we’ve used and added different types of modules and graphs which weren’t in use previously.
* The graphs and predictions we’ve produced are interactive so that the users can also operate them.

**2.2 System Requirements `**

**2.3.1 Hardware Requirements**

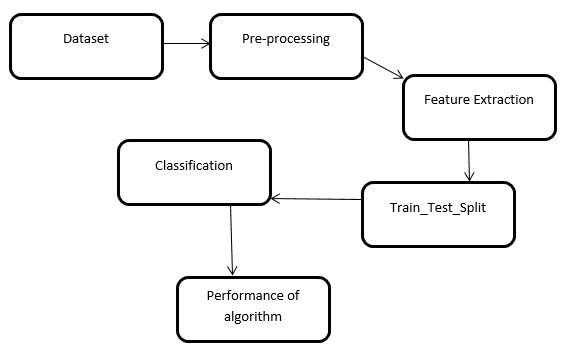
* Intel core i3
* 4 GB RAM
* Minimam storage

**2.3.2 Software Requirements**

* Windows 10 / Linux
* anaconda (Jupiter notebook)
* Ms excel
* Python version 3.11.1
* Any web browser

# 3.METHODOLOGY

**3.1 system architecture:**



**3.2 Algorithm Description :**

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

**Types of linear Regression:**

Linear regression can be further divided into two types of the algorithm:

**Simple Linear Regression:**

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

**Multi Linear Regression:**

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear

Regression.

* Here we use linear regression algorithm. The main concept of this algorithm to find predictive analysis.
* Linear regression is a type of supervised machine learning It deals with the two variables one is dependent variable and independent variable.
* For Linear Regression, we use the Mean Squared Error (MSE) cost function, which is the average of squared error occurred between the predicted values and actual values.
* The algorithm aimed to collect the number of likes, dislikes, and the view count of a trailer, release date, star ranking, and so on.
* Here we use linear regression algorithm because it is the best algorithm to give the correct reviews of movie.
* Main goal is to find the best fit line that means the error between predicted values and actual values should be minimized. The best fit line will have the least error.

**3.3Modular Design:**

**Uploding The Dataset:**

In any type of data analys project, the first step is gathering raw data. once this data has been gathered, it can then be cleaned, transformed, summarized, and visualized.in this project to we uploaded dataset using

df = pd.read\_csv("'C:\\Users\\REVALLA LOKESH'")

first we will collect the raw data an then it is uploaded by using csv file

**pre -processing:**

data preprocessing is a technique that is used to convert the raw data into a clean data set. in other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.by pre processing the data we will drop the some of the data like null values.

df.isnull().sum()

**splitting the data:**

split the data set into two pieces — a training set and a testing set. this consists of random sampling without replacement about 75 percent of the rows (you can vary this) and putting them into your training set. the remaining 25 percent is put into your test set.training data is the subpart of original data. testing data is used to find accuracy of the data from sklearn.model\_selection import train\_test\_split

# 4.IMPLEMENTATION

**4.1. What is Python?**

1. Python is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.
2. Python uses garbage collection and has dynamic typing. It supports a variety of programming paradigms, including procedural, object-oriented, and functional programming as well as structured programming (especially this). Due to its extensive standard library, it is frequently referred to as a "batteries included" language.
3. Python was created by Guido van Rossum in the late 1980s to replace the ABC programming language, and it was originally made available as Python 0.9.0 in 1991.
4. Python frequently causes programmers to fall in love due to the enhanced productivity it offers. The edit-test-debug cycle is extraordinarily quick because there is no compilation step.
5. Python programs are simple to debug since a segmentation failure is never caused by a bug or incorrect input. Instead, the interpreter raises an exception when it finds a mistake. The interpreter prints a stack trace if the application doesn't catch the exception.
6. Setting breakpoints, evaluating arbitrary expressions, inspecting local and global variables, stepping through the code one line at a time, and other features are all possible with a source level debugger. Python's ability to perform introspection is demonstrated by the debugger, which is developed in Python.

Python provides many useful features to the programmer. These features make it most popular and widely used language. Listed below are few-essential features of Python:

* + - Easy to use and Learn
    - Expressive Language
    - Interpreted Language
    - Object-Oriented Language
    - Open Source Language
    - Extensible
    - Learn Standard Library
    - GUI Programming Support
    - Integrated
    - Embeddable
    - Dynamic Memory Allocation

Python has wide range of libraries and frameworks widely used in various fields such as machine learning, artificial intelligence, web applications, etc. We define some popular frameworks and libraries of Python as follows.

* + - Web development (Server-side) - Django Flask, Pyramid, CherryPy
    - GUIs based applications - Tk, PyGTK, PyQt, PyJs, etc.
    - Machine Learning - TensorFlow, PyTorch, Scikit-learn, Matplotlib, Scipy, etc.
    - Mathematics - Numpy, Pandas, etc.

**4.2 Jupyter Notebook:**

The Jupyter Notebook is an open source web application that used to create and share documents that contain live code, equations, visualizations and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the Python project, which used to have an Python Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the Python kernel, which allows to write the programs in Python, but there are currently over 100 other kernels that can also be used.

Jupyter Notebook can be installed by using either of the two ways described below: Using **Anaconda:**

Install Python and Jupyter using the Anaconda Distribution, which includes Python, the Jupyter Notebook, and other commonly used packages for scientific computing and data science. To install Anaconda, go through How to install Anaconda on windows? And follow the instructions provided.

**Using PIP:**

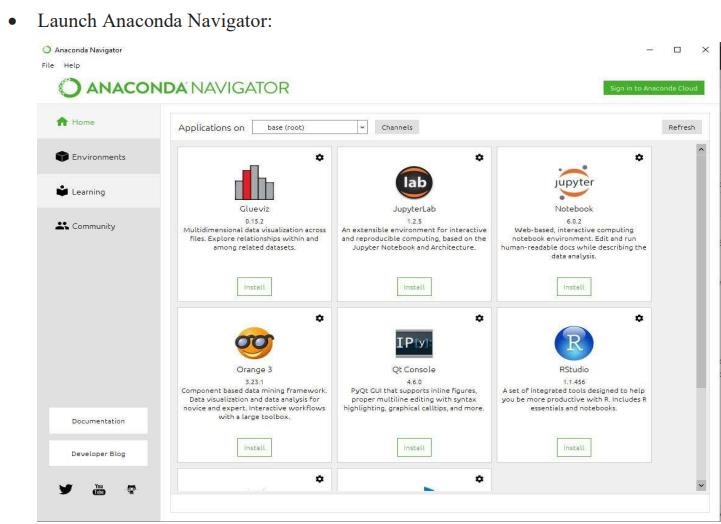
Install Jupyter using the PIP package manager used to install and manage software packages/libraries written in Python. To install pip, go through How to install PIP on Windows? and follow the instructions provided.

**Installing Jupyter Notebook using Anaconda:**

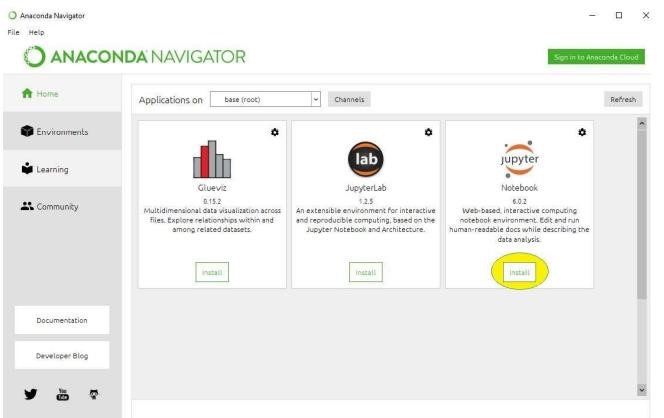
Anaconda is an open-source software that contains Jupyter, spyder, etc that are used for large data processing, data analytics, heavy scientific computing. Anaconda works for R and python programming language. Spyder(sub-application of Anaconda) is used for python.

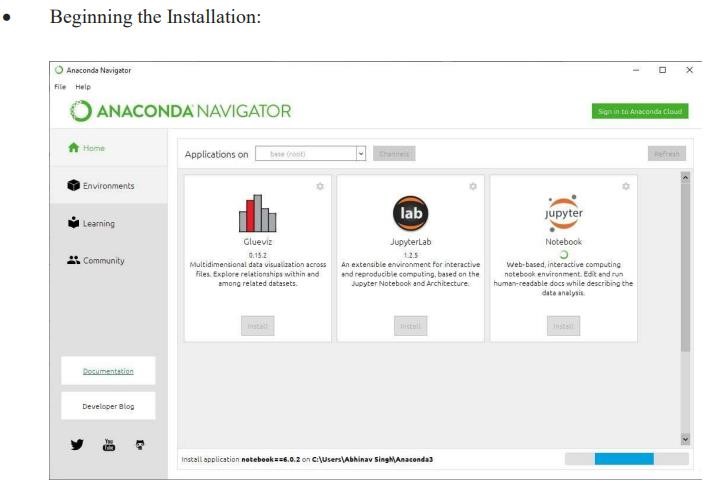
Opencv for python will work in spyder. Package versions are managed by the package management system called conda.

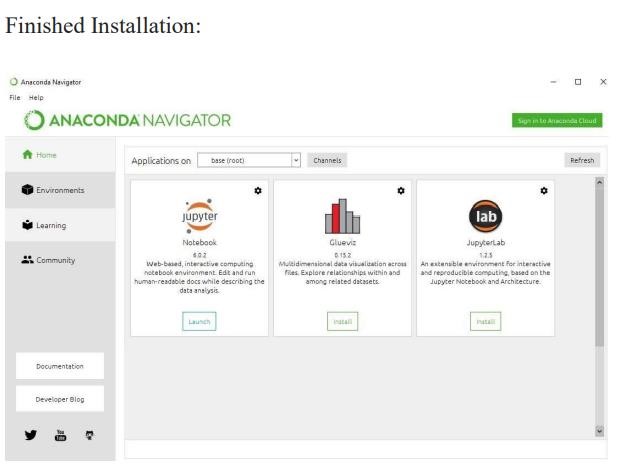
To install Jupyter using Anaconda, just go through the following instructions:

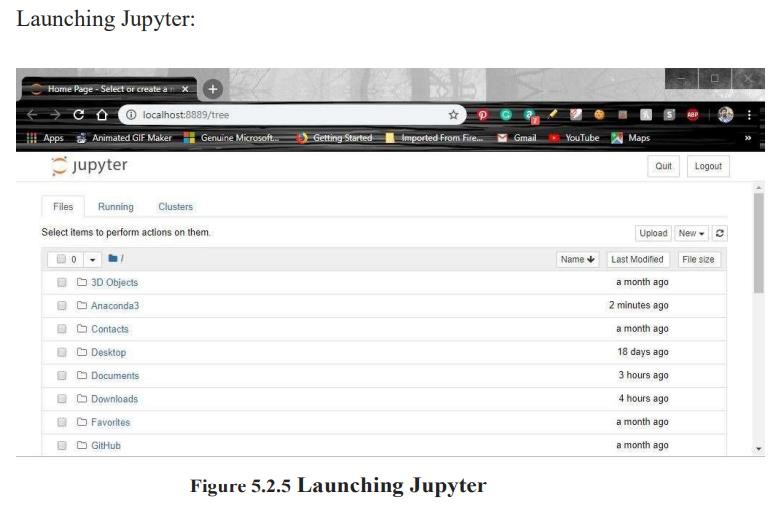


Click on the Install Jupyter Notebook Button:









**4.3. Modules:**

**4.3.1. Numpy:**

NumPy is a library that contains multidimensional array gadgets as well as a set of procedures for processing them. NumPy can be used to conduct arithmetic and scientific operations on arrays. NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

**4.3.2. Pandas:**

For the purpose of manipulating and analyzing data, the Python programming language has a software package called pandas. It includes specific data structures and procedures for working with time series and mathematical tables. It is free software distributed under the BSD license's three clauses. The word is derived from "panel data," a phrase used in econometrics to refer to data sets that contain observations for the same persons throughout a range of time periods. Python data analysis is a play on words in the name of the thing. When Wes McKinney worked as a researcher at AQR Capital from 2007 to 2010, he began creating the pandas that would eventually become famous.

**4.3.3. Seaborn:**

A Python library piled on top of matplotlib library that is freely available. For data visualization and exploratory data analysis, it's widely utilized. Data frames and the Pandas libraries are no problem for Seaborn. The graphs that are created can also be customized without difficulty. Seaborn is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and colour palettes to make statistical plots more attractive. It is built on the top of [matplotlib](https://www.geeksforgeeks.org/python-introduction-matplotlib/) library and also closely integrated to the data structures from [pandas.](https://www.geeksforgeeks.org/introduction-to-pandas-in-python/) Seaborn aims to make visualization the central part of exploring and understanding data. It provides dataset-oriented APIs, so that we can switch between different visual representations for same variables for better understanding of dataset.

**4.3.4. Matplotlib:**

Matplotlib is a graphical library for Python and its NumPy numerical arithmetic extension. It provides an object-oriented API for embedding visualizations in packages using popular GUI tool kits like as Tkinter, wxPython, Qt, or GTK. Matplotlib comes with a wide variety of plots. Plots helps to understand trends, patterns, and to make correlations. They’re typically instruments for reasoning about quantitative information.

**4.3.4. Scipy:**

SciPy is a systematic Python library that is open source and BSD-licensed for mathematics, science, and engineering. SciPy is a scientific computation library that uses [NumPy](https://www.w3schools.com/python/numpy/default.asp) underneath. SciPy stands for Scientific Python. It provides more utility functions for optimization, stats and signal processing. Like NumPy, SciPy is open source so we can use it freely. SciPy was created by NumPy's creator Travis Oliphant.

**4.3.4. Sklearn:**

Sklearn is the most usable and robust Python package for system learning. Through a Python consistency interface, it disseminates effective tools for device learning and statistical modelling, such as type, regression, clustering, and dimensionality reduction. scikit-learn is a [free software](https://en.wikipedia.org/wiki/Free_software) [machine learning](https://en.wikipedia.org/wiki/Machine_learning) [library](https://en.wikipedia.org/wiki/Library_(computing)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) [programming language.](https://en.wikipedia.org/wiki/Programming_language) It features various [classification,](https://en.wikipedia.org/wiki/Statistical_classification) [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support-vector machines,](https://en.wikipedia.org/wiki/Support_vector_machine) [random forests,](https://en.wikipedia.org/wiki/Random_forests) [gradient boosting,](https://en.wikipedia.org/wiki/Gradient_boosting) [*k*-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN,](https://en.wikipedia.org/wiki/DBSCAN) and is designed to interoperate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy) and [SciPy.](https://en.wikipedia.org/wiki/SciPy) Scikit-learn is a [NumFOCUS](https://en.wikipedia.org/w/index.php?title=NumFOCUS&action=edit&redlink=1) fiscally sponsored project.

**4.3.5. Using PIP:**

As a popular open source development project, Python has an active supporting community of contributors and users that also make their software available for other Python developers to use under open source license terms.

This allows Python users to share and collaborate effectively, benefiting from the solutions others have already created to common (and sometimes even rare!) problems, as well as potentially contributing their own solutions to the common pool.

Pip is the preferred installer program. Starting with Python 3.4, it is included by default with the Python binary installers.

**5**.**SAMPLE CODE**

import pandas as pd import numpy as np import os os.getcwd()

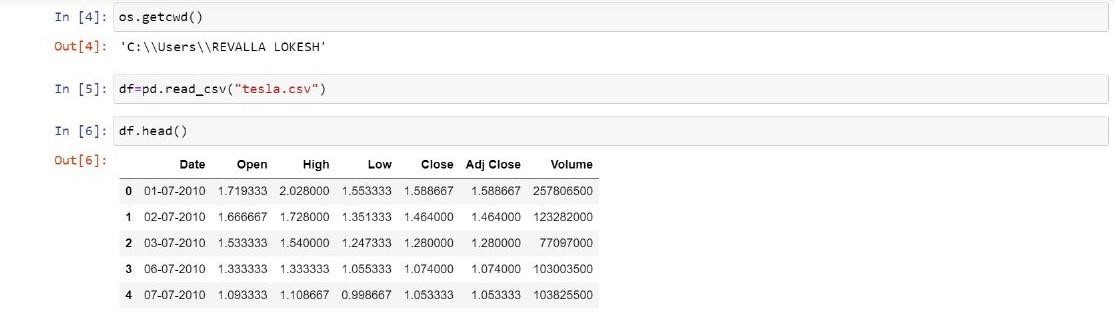
'C:\\Users\\REVALLA LOKESH'

df=pd.read\_csv("tesla.csv") df.head() import pandas as pd import numpy as np from sklearn.tree import DecisionTreeRegressor from sklearn.linear\_model import LinearRegression from sklearn.model\_selection import train\_test\_split import matplotlib.pyplot as plt plt.style.use('bmh') plt.figure(figsize=(16,8)) plt.title('netflix') plt.xlabel('days') plt.ylabel('Close price usd ($)') plt.plot(df['Close']) plt.show() df = df[['Close']] df.head(4)x\_future = df.drop(['Prediction'], 1)[:-future\_days] x\_future = x\_future.tail(future\_days) x\_future = np.array(x\_future)

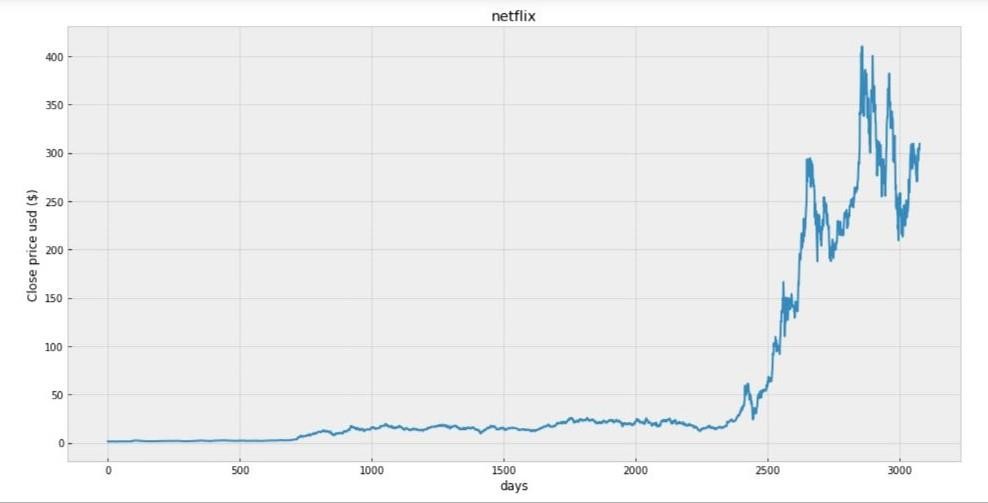
x\_future future\_days=25 df['Prediction']=df[['Close']].shift(-future\_days) df.tail(4) x = np.array(df.drop(['Prediction'], 1))[:-future\_days] y = np.array(df['Prediction'])[:-future\_days] x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.25) lr=LinearRegression().fit(x\_train,y\_train) x\_future = df.drop(['Prediction'], 1)[:-future\_days] x\_future = x\_future.tail(future\_days) x\_future = np.array(x\_future) x\_future lr\_prediction = lr.predict(x\_future) print(lr\_prediction) valid = df[x.shape[0]:] valid['Predictions'] = predictions plt.figure(figsize=(16,8)) plt.title('Model') plt.xlabel('days') plt.ylabel('close price usd($)') plt.plot(df['Close']) plt.plot(valid[['Close','Predictions']]) plt.legend(['Orig','Val','Pred'])

plt.show()

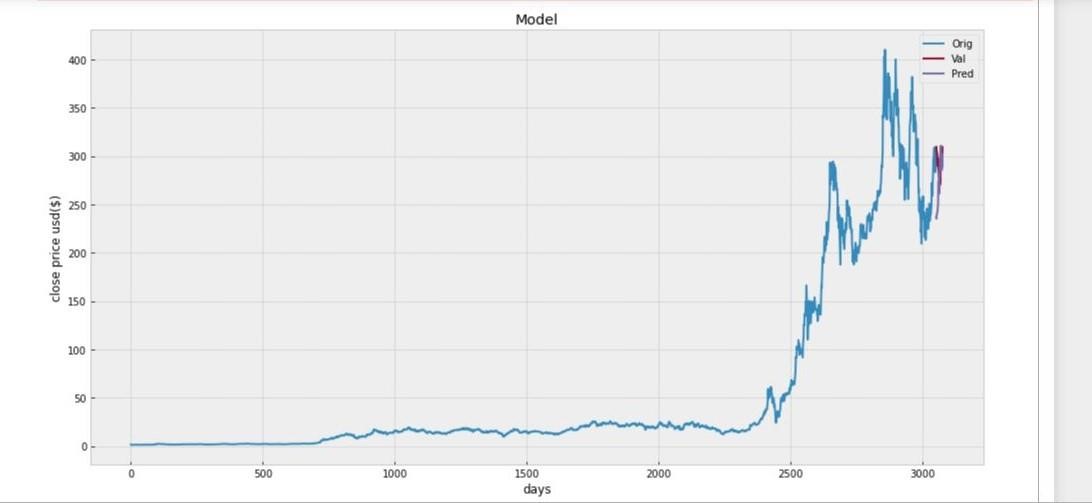
# 6.OUTPUT



**Data before prediction :**



**prediction :**



# CONCLUSION

Time series analysis is one of the most important aspect of data analytics for any large organization as it helps in understanding seasonality, trends, cyclicality and randomness in the sales and distribution and other attributes.

The task facing the modern time-series econometrician is to develop reasonably simple models capable of forecasting, interpreting, and testing hypotheses concerning economic data. . The challenge has grown over time the original use of time-series analysis was primarily as an aid to forecasting. As such, a methodology was developed to decompose a series into a trend, a seasonal, a cyclical, and an irregular component. Uncovering the dynamic path of a series improves forecast accuracy.. Using the time-series methods, it is possible to decompose this series into the trend, seasonal, and irregular components.

Time series analysis is one of the most important aspect of data analytics for any large organization as it helps in understanding seasonality, trends, cyclicality and randomness in the sales and distribution and other attributes.

# BIBLIOGRAPHY

**Sites Referred**

1. [https://www.python.org](https://www.python.org/)
2. <https://www.w3schools.com/python/python_ml_getting_started.asp>
3. <https://www.tutorialspoint.com/machine_learning_with_python/index.htm>
4. <https://www.freecodecamp.org/learn/machine-learning-with-python/>
5. <https://www.tableau.com/learn/articles/time-series-analysis>
6. https://www.kaggle.com/

**Books Referred**

* 1. Python Machine Learning (Sebastian Raschka)
  2. Introduction to Machine Learning with Python (Andreas C. Muller/Sarah Guido)