A PROJECT ON

"FOREX ANALYSIS"

SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE COURSE OF
DIPLOMA IN BIG DATA ANALYTICS FROM CDAC

SUNBEAM INSTITUTE OF INFORMATION TECHNOLOGY
Phase 2, Hinjewadi Rajiv Gandhi Infotech Park,
Hinjawadi, Pune,
Maharashtra 411057

SUBMITTED BY:

Komal Prakash Bhandarkar

UNDER THE GUIDANCE OF:
Mr. Girish Gaikwad
Faculty Member
Sunbeam Institute of Information Technology, PUNE.

CERTIFICATE

This is to certify that the project work under the title 'Forex Analysis' is done by Komal Prakash Bhandarkar in partial fulfillment of the requirement for award of Diploma in Big Data Analytics.

Mr. Girish Gaikwad
Project Guide

Mrs . Pradnya Dindorkar
Course Co-Coordinator

Date:

ACKNOWLEDGEMENT

A project usually falls short of its expectation unless aided and guided by the right persons at the right time. We avail this opportunity to express our deep sense of gratitude towards Mr. Nitin Kudale (Center Coordinator, SIIT, Pune) and Mrs. Pradnya Dindorkar (Course Coordinator, SIIT ,Pune) and Project Guide Mr. Girish Gaikwad.

We are deeply indebted and grateful to them for their guidance, encouragement and deep concern for our project. Without their critical evaluation and suggestions at every stage of the project, this project could never have reached its present form.

Last but not the least we thank the entire faculty and the staff members of Sunbeam Institute of Information Technology, Pune for their support.

Komal Prakash Bhandarkar

DBDA February 2020 Batch,

SIIT Pune

TABLE OF CONTENTS

1. Introduction

Predicting the forex asset has been the bane and goal of investors since its existence. Everyday billions of dollars are traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Entire companies rise and fall daily based on the behavior of the market. Should an investor be able to accurately predict market movements, it offers tantalizing promises of wealth and influence. It is no wonder then that the Forex Market and its associated challenges find their way into the public imagination every time it misbehaves. The 2008 financial crisis was no different, as evidenced by the flood of films and documentaries based on the crash. If there was a common theme among those productions, it was that few people knew how the market worked or reacted. Perhaps a better understanding of Forex market prediction might help in the case of similar events in the future. The cloud deployment is also available on https://www.kaggle.com/hiteshshebe/forex-analysis .

1.1. Statement of the problem

Forex market is very vast and difficult to understand. It is considered too uncertain and predictable due to huge fluctuation of the market. Forex market prediction task is interesting as well as divides researchers and academics into two groups, those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the the market absorbs it by correcting itself, thus there is no space for prediction.Investing in a good Asset but at a bad time can have disastrous result, while investing in an Asset at the right time can bear profits. Financial investors of today are facing this problem of trading as they do not properly understand as to which Asset to buy or which Asset to sell in order to get optimum results. So, the proposed project will reduce the problem with suitable accuracy faced in such a real time scenario.

1.2 Relevant current/open problems

1. Data-are-humongous, nowadays we are seeing a rapid-explosion of numerical-Asset quotes and textual-data. They are provided from all different-sources.

2. Demand forecasts are important since the basic op management process, going from the vendor raw-materials to finished goods in the customers' hands, takes some time. Most firms cannot-wait for demand to elevate and then give a reaction. Instead, they make-up their mind and plan according to future demand so-that they can react spontaneously to the customer's order as they arrive.

3. Generally, demand forecasts-lead to good-ops-and great-levels of customer satisfaction, while bad forecasts will definitely-lead to costly ops and worst-levels of customer satisfaction.

4. A confusion for the forecast is the horizon, which is, how distant in the future will the forecast project? As a simple rule, the further into the future we see, the blurry our vision will become -- distant forecasts will be inaccurate that short-range forecasts.

1.3 Technical analysis

In contrast to fundamental analysis, technical analysis does not try to gain deep insight into a company's business. It assumes the available public information does not offer a competitive trading advantage. Instead, it focuses on studying a company's historical Asset/Currency price and on identifying patterns in the chart. The intention is to recognize trends in advance and to capitalize on them.

1.4 Goal

The goal was to build a system capable of the following tasks:

1. Collecting fundamental and technical data from the internet The system should be able to crawl specific websites to extract fundamental data like news

articles and analyst recommendations. Furthermore, it should be able to collect technical data in the form of historical Asset prices.

2. Simulating trading strategies The system should offer ways to specify and simulate fundamental and technical trading strategies. Additionally, combining the two approaches should be possible.

3. Evaluating and visualizing trading strategies The system should evaluate and visualize the financial performance of the simulated strategies. This allows a comparison to be made between technical,fundamental and the combined approaches.

1.5 Project Goals and Scope

Despite its prevalence, Forex Market prediction remains a secretive and empirical art. Few people, if any, are willing to share what successful strategies they have. A chief goal of this project is to add to the academic understanding of Forex market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial crisis. The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies. It is important here to define the scope of the project. Although vital to any investor operating in the real world, no attempt is made in this project at portfolio management. Portfolio management is largely an extra step done after an investor has made a prediction on which direction any particular Asset will move. The investor may choose to allocate funds across a range of assets in such a way to minimize his or her risk. For instance, the investor may choose not to invest all of their funds into a single Assets lest that Asset take an unexpected rate. A more common approach would be for an investor to 3 invest across a broad range of forex based on some criteria he has decided on before . This project will focus exclusively on predicting the daily trend (price movement) of top 3 assets. The project will make no attempt to decide how much money to allocate to each prediction. More so, the project will analyze the accuracies of these predictions.

1.6 Overview of proposed solution approach

1. Basically the main objective of this project is to collect the Forex Asset information Since 1995 till date and then accordingly predict the results for the predicting what would happen next. So for we are going to use of two well-known techniques Machine Learning and data mining for Forex market prediction. Extracting useful information from a huge amount of data set and data mining is also able to predict future trends and behaviors through Time Series Techniques and advanced machine learning algorithms. Therefore, combining both these techniques could make the prediction more suitable and much more reliable

2. As far as the solutions for the above problems, the answer depends on which way the forecast is used for. So the procedures that we will be using have proven to be very applicable to the task of forecasting the rate of stock in a stock market. Many techniques, which can prove useful for forecasting-problems, have shown to be inadequate to the task of demand forecasting in the share market.

1.7 Novelty/Benefits:

The rich variety of on-line information and news make it an attractive resource from which one can get data. Forex market predictions can be aided by data mining and analysis of such financial information. Numerical Asset quotes collected from rapidapi.com available in an organized manner with rest api service but we have to apply **http script** to extract data into the file. websites released daily updated data.

2. Product Overview and Summary

2.1 Purpose

The aims of this project are as follows:
1. To identify Seasonal variation in Asset rate using EDA
2. To generate the pattern from a large set of data of the Forex market for prediction of Euro Currency movements.
3. To predict whether the Euro price is going to increase or decrease or steady in the next few days.
4. To provide analysis for users through CLI application The project will be useful for investors to invest in the stock market based on the various factors. The project target is to create an application that analyzes previous Forex data of companies and implement these values in a data mining algorithm to determine the value that particular Asset will have in the near future with suitable accuracy. These predicted and analyzed data can be observed by individuals to know the financial status of companies and their comparisons. It can be very useful to even researchers, market makers,general people.

2.2 Scope

Forex market includes daily activities like updating Asset prices, keeping track of how many units sold.The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives.
The Asset values of Forex Commodity depend on many factors, some of them are:
1. Demand and Supply of Commodity of a company is a major reason price change in Forex Asset. When Demand Increases and Supply is less, price rises. And vice versa. As well as the Asset country GDP also Affects the Increase or decreased the rate of that Asset.
2. Main Strength in hands of Assets buyer. Popularity of a Asset can affect buyers. Like if any good news of an Asset, may result in the rise of Asset price. The Asset value depends on other factors as well, but we are taking into consideration only these main factors.

3. Requirements and Feasibility study

3.1. Feasibility Study

Simply put, the Forex market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The Forex market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases.So, there is a factor which causes people to buy and sell. It has more to do with emotion than logic. Because emotion is unpredictable, Forex market movements will be unpredictable. It's futile to try to predict where markets are going. They are designed to be unpredictable.There are some fundamental financial indicators by which an Asset value can be estimated. Some of the indicators and factors are: Price-to-Earning (P/E) Ratio,Price-to-Earning Growth (PEG) Ratio, Price-to-Sales (P/S) Ratio, Price/Cash Flow (P/CF) Ratio and Price-to-Book Value (P/BV) Ratio.

Some of the parameters are available and accessible on the web but all of them aren't. So we are confined to use the variables that are available to us. The proposed system will not always produce accurate results since it does not account for human behaviors. Factors like change in country GDP ,political leadership, internal matters, strikes, protests, natural disasters, change in the authority cannot be taken into account for relating it to the change in Forex market by the machine.The objective of the system is to give an approximate idea of where the Forex market might be headed. It does not give a long term forecasting of a Asset value.There are way too many reasons to acknowledge for the long term output of a current Asset rate.Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

3.2. Requirement Analysis

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements.

These requirements are listed below:
1. Functional Requirements

2. Non-Functional Requirements

3.2.1 Functional Requirements

Functional requirements are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

1. The system should be able to generate an approximate Asset rate.
2. The system should collect accurate data from the rapidapi.com consistent manner.
3. The prediction shall abide by the following functional requirements:
    i. The charts and comparison of the Assets would be done only on the latest data Forex market data.
    ii. The user can look at previous data Information which was collected.
    iii. The user can also be recommended on the basis of the trending Asset which would require the data regarding the stocks.

3.2.2 Non-Functional Requirements

Non-functional requirement is a description of features, characteristics and attributes of the system as well as any constraints that may limit the boundaries of the proposed system. The non- functional requirements are essentially based on the performance, information, economy, control and security efficiency and services.

Based on these the non-functional requirements are as follows:
1. The system should provide better accuracy.
2. The system should have a simple interface for users to use.
3. To perform efficiently in a short amount of time.

1. Reliability:

The reliability of the product will be dependent on the accuracy of the dataset of purchase, how much Asset was purchased, high and low value range as well as opening and closing figures. Also the asset data used in the training would determine the reliability of the software.

2. Maintainability:

The maintenance of the product would require training of the software by recent data so that their commendations are up to date. The database has to be updated with recent values.

3. Portability:

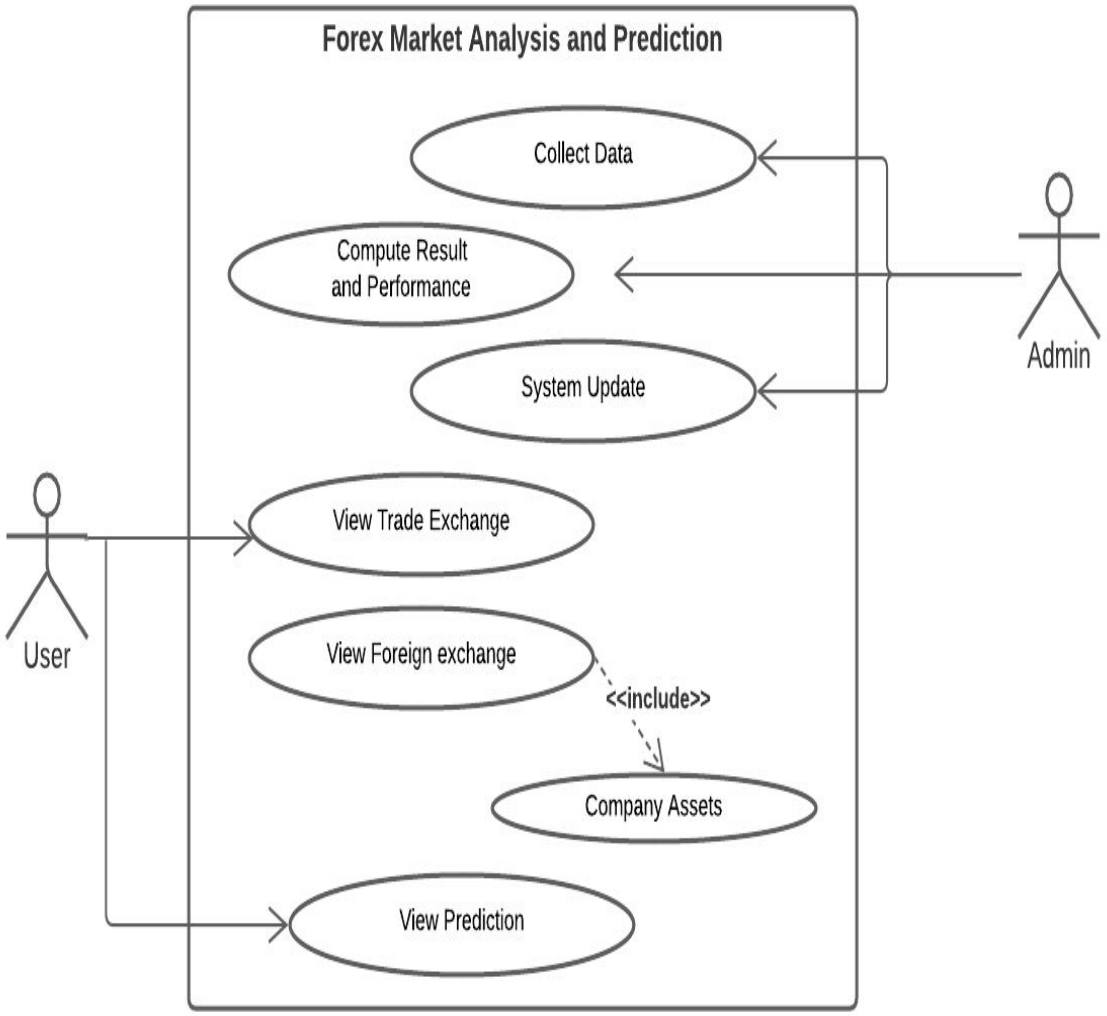
The Complete Project is completely Reliable. We can Deploy projects in cloud platforms like Kaggle,AWS,GCP,Azure etc.

4.Data Integrity:

The data integrity is high because we collect real time data of the Forex market of the last 25 year which provide the accurate result of our models. All used models Provide better accuracy due to data integrity.

4. System Design and Architecture

4.1 Use Case Diagram



1-use case index

Use case id	Use Case Name	Primary actor	Scope	Complexity	Priority
1	Collect Data	Admin	In	High	1
2	Compute Result And Performance	Admin	In	High	1
3	System Update	Admin	In	High	1
4	View Trade Exchange	User	In	Medium	2
5	View Foreign Exchange	User	In	Medium	2
6	Company Asset	User	In	Medium	2
7	View Prediction	User	In	Hard	1

2-Use case description:

Use case ID:1
Use case name: Collect data
Description: Every required data will be available in Forex exchange. System will be able to collect the data for the system.

Use case ID:2
Use case name: Compute result and performance
Description: Prediction result will be handled and generated by System. The system will be built, through which the result of prediction and system performance will be analyzed.

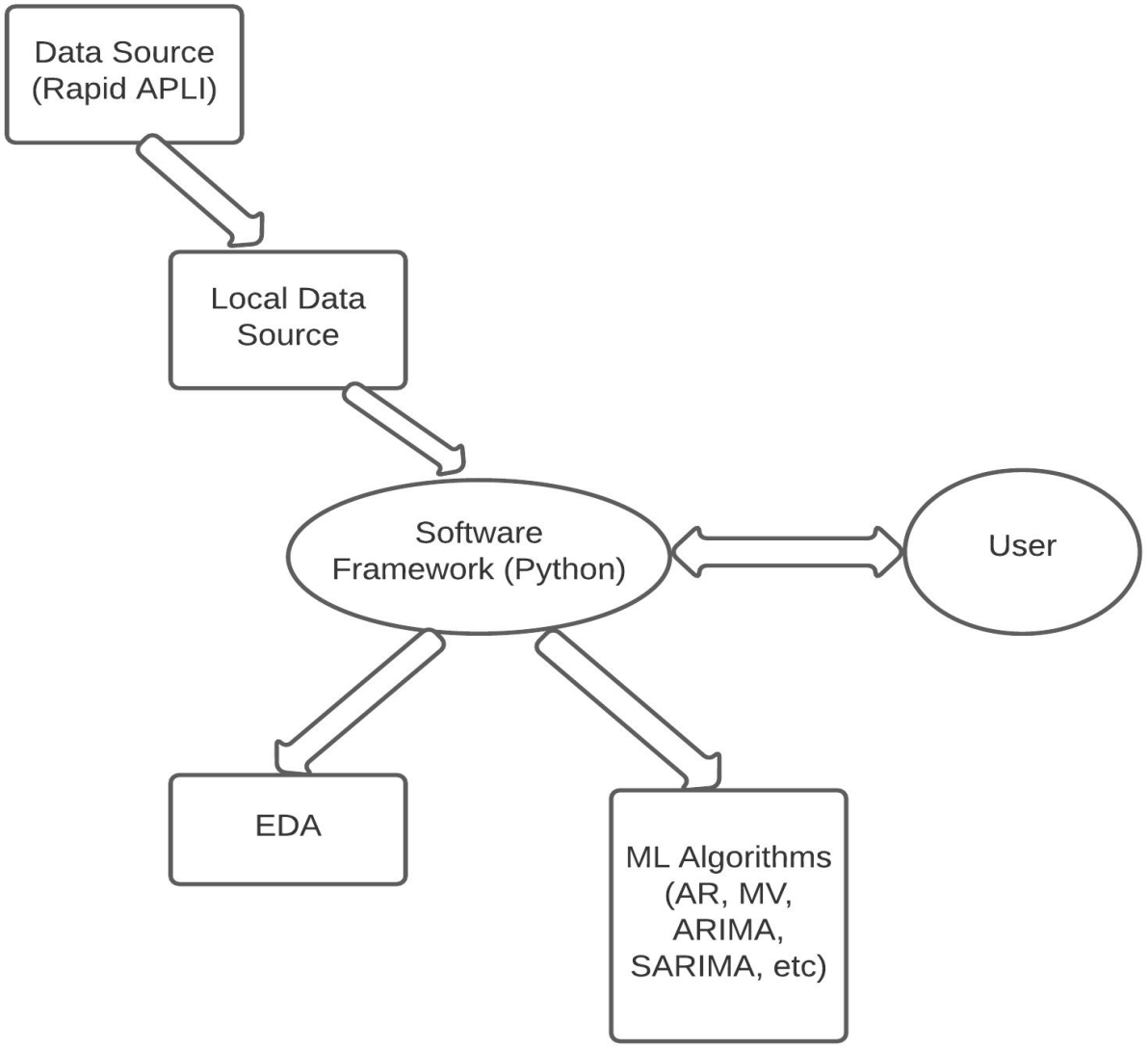
Use case ID: 3
Use case name: System update
Description: With the change of market and technology regular update of system is required. Beside there the predicted result of Forex exchange and their actual price will be updated by system automatically on a regular basis.

Use case ID: 4
Use case name: View traded exchange
Description: Asset trading which is held at Forex exchange can be viewed by users.

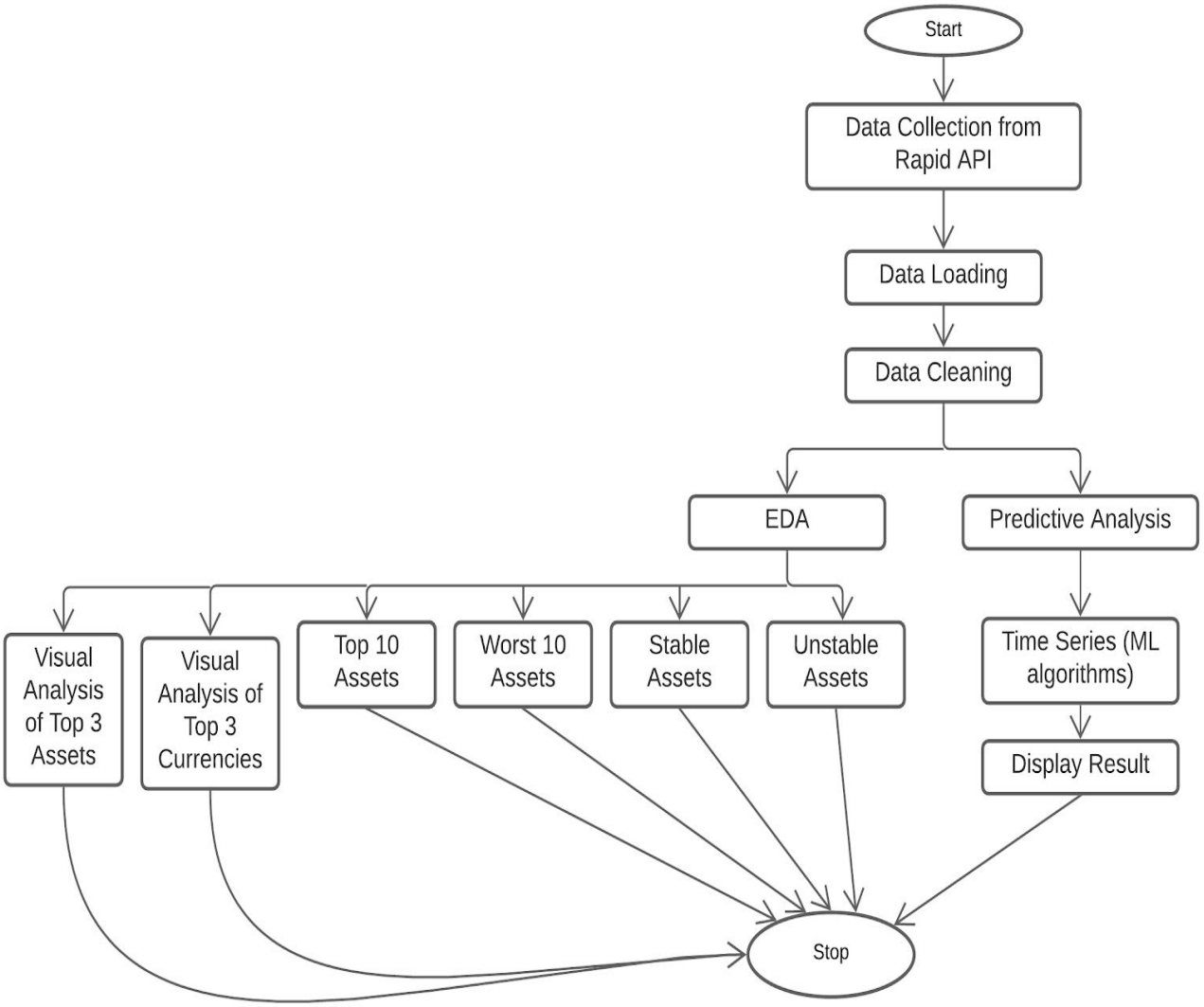
Use Case ID: 5
Use Case Name: Forex Asset
Description: It is an extended feature of view traded exchange. This includes the Asset value of a particular Forex Commodity.

Use Case ID: 6
Use Case Name: View prediction
Description: This use case is most important in the whole project. The key feature of This project is to predict the Asset value of the Forex Market. Thus, this will be an available notebook file and viewer can observe them.

4.2. System Flow diagram



4.3 CONTROL FLOW DIAGRAM



4.4 Proposed Algorithms:

1. **AR**: Autoregression modeling centers on measuring the correlation between observations at previous time steps (the lag variables) to predict the value of the next time step (the output).

If both variables change in the same direction, for example increasing or decreasing together, then there is a positive correlation. If the variables move in opposite directions as values change, for example one increasing while the other decreases, then this is called negative correlation. Either way, using basic statistics, the correlation between the output and previous variable can be quantified.

The higher this correlation, positive or negative, the more likely that the past will predict the future.

Since this correlation is between the variable and itself at previous time steps, it is referred to as an autocorrelation.

In addition, if every variable shows little to no correlation with the output variable, then it's likely that the time series dataset may not be predictable. Which is also useful for deep learning training.

2. **Naive Persistence Model**: The naive algorithm uses the value at the current time step (t) to predict the expected outcome at the next time step (t+1).

Here we transform our ‘’ dataset into a supervised learning problem, we can achieve this by creating a lag feature. In this transformed dataset, values at (t) are predictors (X) and values at (t+1) are the target variable (Y).

3**. Moving Average**: A moving average term in a time series model is a past error (multiplied by a coefficient).

Let wt~iidN(0,σw2), meaning that the *wt* are identically, independently distributed, each with a normal distribution having mean 0 and the same variance.

The 1st order moving average model, denoted by MA(1) is:

$$xt=\mu+wt+\theta 1wt-1$$

The 2nd order moving average model, denoted by MA(2) is:

$$xt=\mu+wt+\theta 1wt-1+\theta 2wt-2$$

The qth order moving average model, denoted by MA(q) is:

$$xt=\mu+wt+\theta 1wt-1+\theta 2wt-2+...+\theta qwt-q$$

4**. ARIMA** (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.

5**. SARIMA**: An extension to ARIMA that supports the direct modeling of the seasonal component of the series is called SARIMA. Seasonal Autoregressive Integrated Moving Average, or SARIMA, method for time series forecasting with univariate data containing trends and seasonality.

- The limitations of ARIMA when it comes to seasonal data.
- The SARIMA extension of ARIMA that explicitly models the seasonal element in univariate data.
- How to implement the SARIMA method in Python using the Statsmodels library.

  It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

  A seasonal ARIMA model uses differencing at a lag equal to the number of seasons (s) to remove additive seasonal effects. As with lag 1 differencing to remove a trend, the lag s differencing introduces a moving average term. The seasonal ARIMA model includes autoregressive and moving average terms at lags.

How to Configure SARIMA

Configuring a SARIMA requires selecting hyperparameters for both the trend and seasonal elements of the series.

Trend Elements

There are three trend elements that require configuration.

They are the same as the ARIMA model; specifically:

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.

Seasonal Elements

There are four seasonal elements that are not part of ARIMA that must be configured; they are:

- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

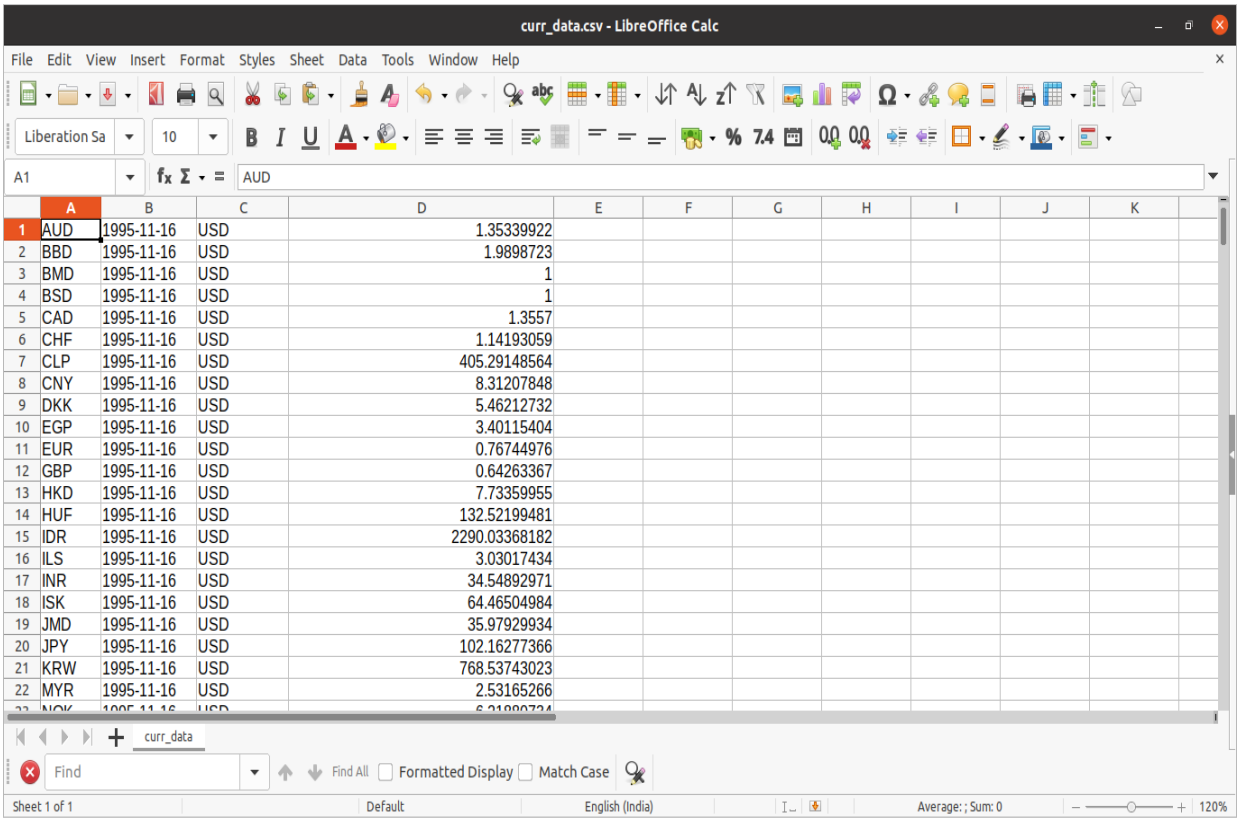
Together, the notation for an SARIMA model is specified as:

SARIMA(p,d,q)(P,D,Q)m

5.FINDINGS AND CONCLUSIONS

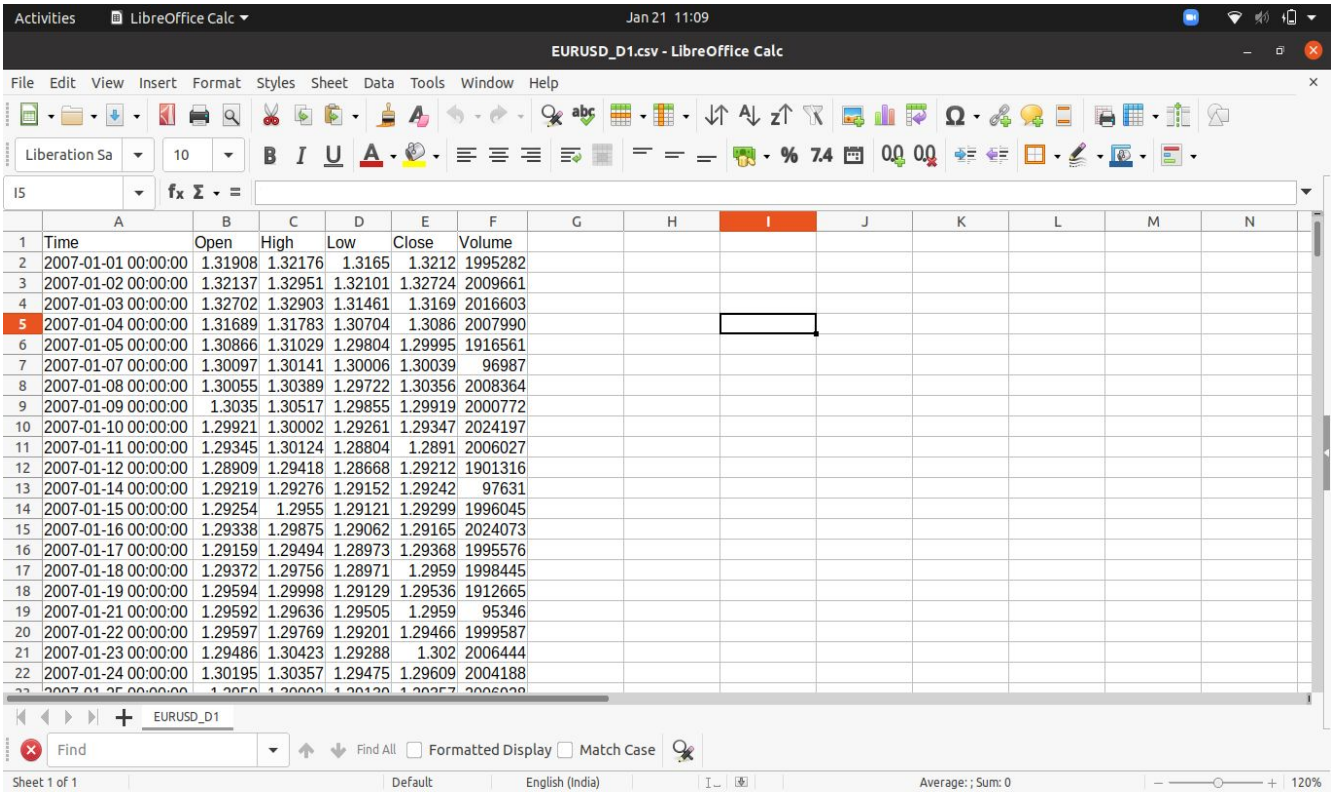
The system evaluation on the Asset from Forex Exchange is carried out. For a given day's open index, week's high, week's low,Weeks average our forecaster will forecast the approx Asset Rate value for a particular trading day. Our predictive model is evaluated on the Forex market on the financial historical stock data over the training period of sept 1995 to till today The data is collected from the REST api's web sites https://www.rapidapi.com . The data is

collected once in a day. The Asset value corresponding to each trading day were downloaded from https://www.rapidapi.com .The accuracy of the system is measured as the percentage of the predictions that were correctly determined by the system. For instance, if the system forecasts an upward trend and the index indeed goes up, it is supposed to be correct, otherwise, if the index goes down or remains stable for an uptrend, it is assumed to be wrong.Following Forex dataset is taken as sample training data of Euro_data over the period of 20 year.



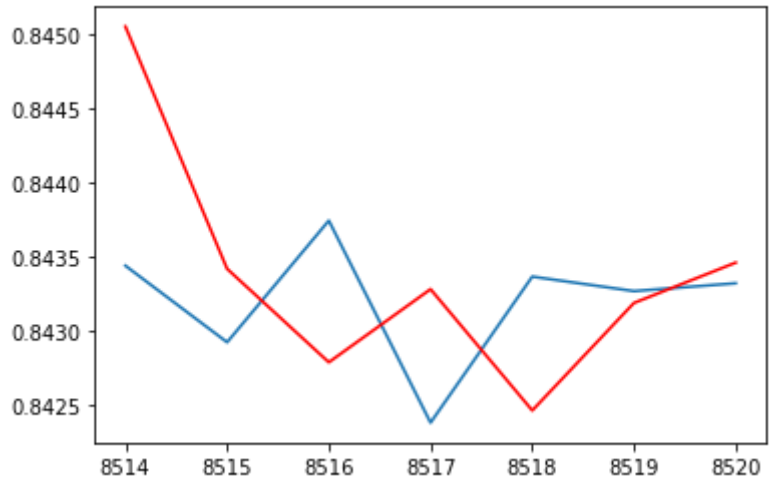
	A	B	C	D
1	AUD	1995-11-16	USD	1.35339922
2	BBD	1995-11-16	USD	1.9898723
3	BMD	1995-11-16	USD	1
4	BSD	1995-11-16	USD	1
5	CAD	1995-11-16	USD	1.3557
6	CHF	1995-11-16	USD	1.14193059
7	CLP	1995-11-16	USD	405.29148564
8	CNY	1995-11-16	USD	8.31207848
9	DKK	1995-11-16	USD	5.46212732
10	EGP	1995-11-16	USD	3.40115404
11	EUR	1995-11-16	USD	0.76744976
12	GBP	1995-11-16	USD	0.64263367
13	HKD	1995-11-16	USD	7.73359955
14	HUF	1995-11-16	USD	132.52199481
15	IDR	1995-11-16	USD	2290.03368182
16	ILS	1995-11-16	USD	3.03017434
17	INR	1995-11-16	USD	34.54892971
18	ISK	1995-11-16	USD	64.46504984
19	JMD	1995-11-16	USD	35.97929934
20	JPY	1995-11-16	USD	102.16277366
21	KRW	1995-11-16	USD	768.53743023
22	MYR	1995-11-16	USD	2.53165266

Which is extracted and calculated on the basis of a dataset called Curr_data.csv. Through feature engineering and various time series analytics techniques Predictions using Various time series algorithms like arima,sarima,auto regression.



	A	B	C	D	E	F
1	Time	Open	High	Low	Close	Volume
2	2007-01-01 00:00:00	1.31908	1.32176	1.3165	1.3212	1995282
3	2007-01-02 00:00:00	1.32137	1.32951	1.32101	1.32724	2009661
4	2007-01-03 00:00:00	1.32702	1.32903	1.31461	1.3169	2016603
5	2007-01-04 00:00:00	1.31689	1.31783	1.30704	1.3086	2007990
6	2007-01-05 00:00:00	1.30866	1.31029	1.29804	1.29995	1916561
7	2007-01-07 00:00:00	1.30097	1.30141	1.30006	1.30039	96987
8	2007-01-08 00:00:00	1.30055	1.30389	1.29722	1.30356	2008364
9	2007-01-09 00:00:00	1.3035	1.30517	1.29855	1.29919	2000772
10	2007-01-10 00:00:00	1.29921	1.30002	1.29261	1.29347	2024197
11	2007-01-11 00:00:00	1.29345	1.30124	1.28804	1.2891	2006027
12	2007-01-12 00:00:00	1.28909	1.29418	1.28668	1.29212	1901316
13	2007-01-14 00:00:00	1.29219	1.29276	1.29152	1.29242	97631
14	2007-01-15 00:00:00	1.29254	1.2955	1.29121	1.29299	1996045
15	2007-01-16 00:00:00	1.29338	1.29875	1.29062	1.29165	2024073
16	2007-01-17 00:00:00	1.29159	1.29494	1.28973	1.29368	1995576
17	2007-01-18 00:00:00	1.29372	1.29756	1.28971	1.2959	1998445
18	2007-01-19 00:00:00	1.29594	1.29998	1.29129	1.29536	1912665
19	2007-01-21 00:00:00	1.29592	1.29636	1.29505	1.2959	95346
20	2007-01-22 00:00:00	1.29597	1.29769	1.29201	1.29466	1999587
21	2007-01-23 00:00:00	1.29486	1.30423	1.29288	1.302	2006444
22	2007-01-24 00:00:00	1.30195	1.30357	1.29475	1.29609	2004188

- The red line indicate Forecasted values
- Blue line indicates Actual values.




```
==========================================================================================================
auto regression model accuracy measure
----------------------------------------------------------------------------------------------------------
mape=0.031509124639991275
me=-0.014031172711939228
mae=0.02827369007578921
mpe=-0.015599619381225695
rmse=0.035563771028891134
corr=0.2467351064886135
minmax=0.031259062039052155
==========================================================================================================
==========================================================================================================
moving avg model accuracy measure
----------------------------------------------------------------------------------------------------------
mape=1.0002836865547002
me=-0.8436193759874283
mae=0.8436193759874283
mpe=-1.0002836865547002
rmse=0.8436208632413661
corr=-0.7970966476661303
minmax=1.0002836865547005
==========================================================================================================
==========================================================================================================
ARIMA model accuracy measure
----------------------------------------------------------------------------------------------------------
mape=0.0027087931780031094
me=-7.067502603089041e-05
mae=0.0023970128936628213
mpe=-7.299412585211595e-05
rmse=0.0035530986697691388
==========================================================================================================
==========================================================================================================
SARIMA model accuracy measure
----------------------------------------------------------------------------------------------------------
mape=0.0066703030513751566
me=0.005576874742922288
mae=0.005615578887983056
mpe=0.006624329451493553
rmse=0.007180399779732414
==========================================================================================================
```

5.1 CONCLUSION

Evaluating the Forex market prediction has at all times been tough work for analysts. Thus, we attempt to make use of vast written data to forecast the Forex market indices. If we join both techniques of textual mining and time series analysis the accuracy in predictions can be achieved. SARIMA Model is qualified to forecast Euro Currency upcoming trends. Financial analysts, investors can use this prediction model to make trading decisions by observing market behavior.

5.2 FUTURE WORK

1.Need to make a model generic which calculates the Predicted value with all Asset of Forex Market.

2. More work on refining key phrases extraction will definitely produce better results. Enhancements in the preprocessor unit of this system will help in improving more accurate predictability in the stock market.

3.Need to build Dashboard for daily monitoring Purpose .