

STOCK-RETURN-PREDICTION

A PROJECT REPORT

Submitted by

V.Abhi raghava(RA2111047010146)

KG.Sandeep(RA2111047010147)

S.Ismiya(RA2111047010148)

M.Harini(RA2111047010149)

J.Akash (RA2111047010150)

Under the guidance of

Dr.R.ATHILAKSHMI

(Assistant professor,Department of computational intelligence)

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY BONAFIDE CERTIFICATE

Certified that this project report titled “STOCK RETURN PREDICTION” is the bonafide work of “ABHI RAGHAVA[RA2111047010146]”, who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature of student

signature of mam

INTRODUCTION ABOUT THE PLATFORMS WORKED ::

1.KNN :

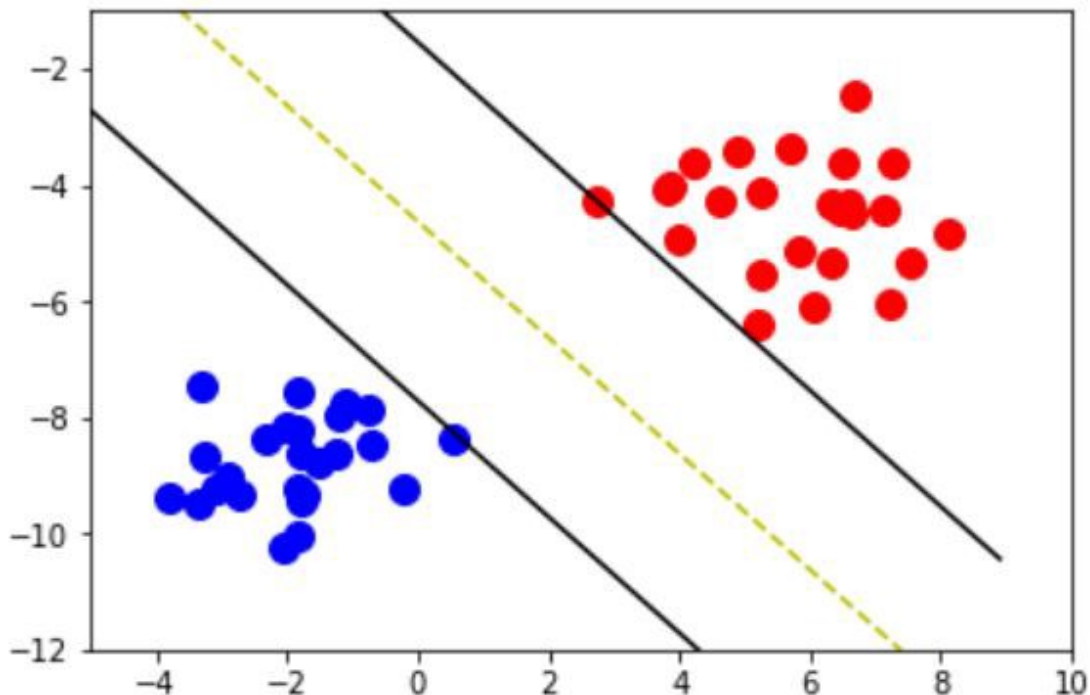
kNN Is a Supervised Learner for Both Classification and Regression. Supervised machine learning algorithms can be split into two groups based on the type of target variable that they can predict: Classification is a prediction task with a categorical target variable

The k-nearest neighbor algorithm is imported from the scikit-learn package.

- 1. Create feature and target variables.*
- 2. Split data into training and test data.*
- 3. Generate a k-NN model using neighbors value.*
- 4. Train or fit the data into the model.*
- 5. Predict the future.*

2.SVM :

Support Vector Machine (SVM) is a supervised machine learning algorithm capable of performing classification, regression and even outlier detection. The linear SVM classifier works by drawing a straight line between two classes.



3.GUASSIAN-PROCESS:

Gaussian Processes (GP) are a generic supervised learning method designed to solve regression and probabilistic classification problems.

The advantages of Gaussian processes are:

- *The prediction interpolates the observations (at least for regular kernels).*
- *The prediction is probabilistic (Gaussian) so that one can compute empirical confidence intervals and decide based on those if one should refit (online fitting, adaptive fitting) the prediction in some region of interest.*

The disadvantages of Gaussian processes include:

- *They are not sparse, i.e., they use the whole samples/features information to perform the prediction.*
- *They lose efficiency in high dimensional spaces – namely when the number of features exceeds a few dozens.*

4.ADABOOST:

An AdaBoost classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases.

5.TREE-REGRESSION :

A regression tree is built through a process known as binary recursive partitioning, which is an iterative process that splits the data into partitions or branches, and then continues splitting each partition into smaller groups as the method moves up each branch.

6.QDA :

Quadratic Discriminant Analysis (QDA) is a generative model. QDA assumes that each class follow a Gaussian distribution. The class-specific prior is simply the proportion of data points that belong to the class. The class-specific mean vector is the average of the input variables that belong to the class.

TOOLS USED

1. Visual Studio Code:

Visual Studio Code is a code editor in layman's terms. Visual Studio Code is "a free-editor that helps the programmer write code, helps in debugging and corrects the code using the intellisense method ". In normal terms, it facilitates users to write the code in an easy ma

2. Web Browsers :

Any web browser will suffice.

3. GitHub:

GitHub Inc. is a Git-based version control STOCK-RETURN-PREDICTION. It's primarily utilised in computer programming. It has all of Git's distributed version control and source code management features, as well as those of its own

INTRODUCTION

>>Forecast stock prices using machine learning approach. A time series analysis. Employ the Use of Predictive Modeling in Machine Learning to Forecast Stock Return. Approach Used by Hedge Funds to Select Tradeable Stocks.

PROBLEM STATEMENT ::

>> Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.

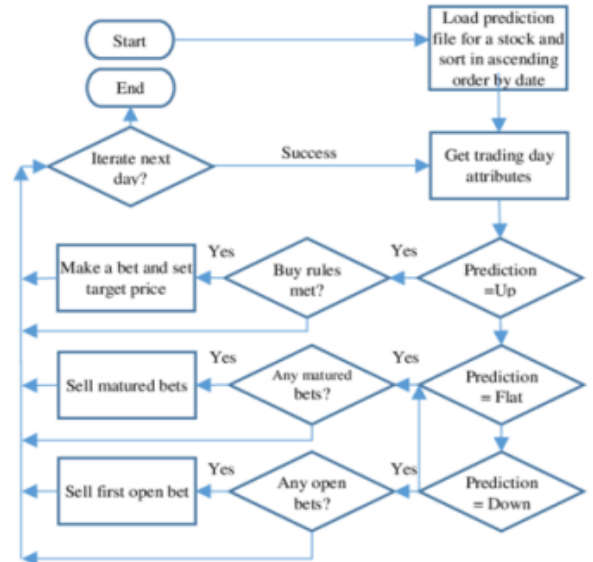
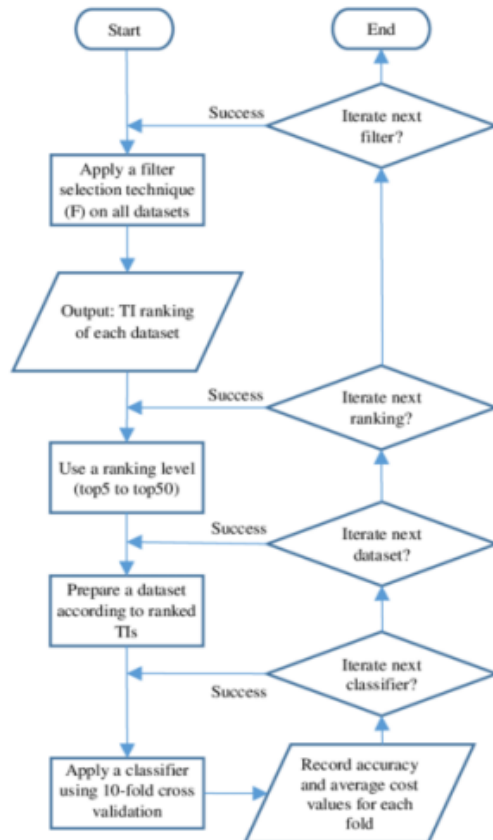
Used data sets :

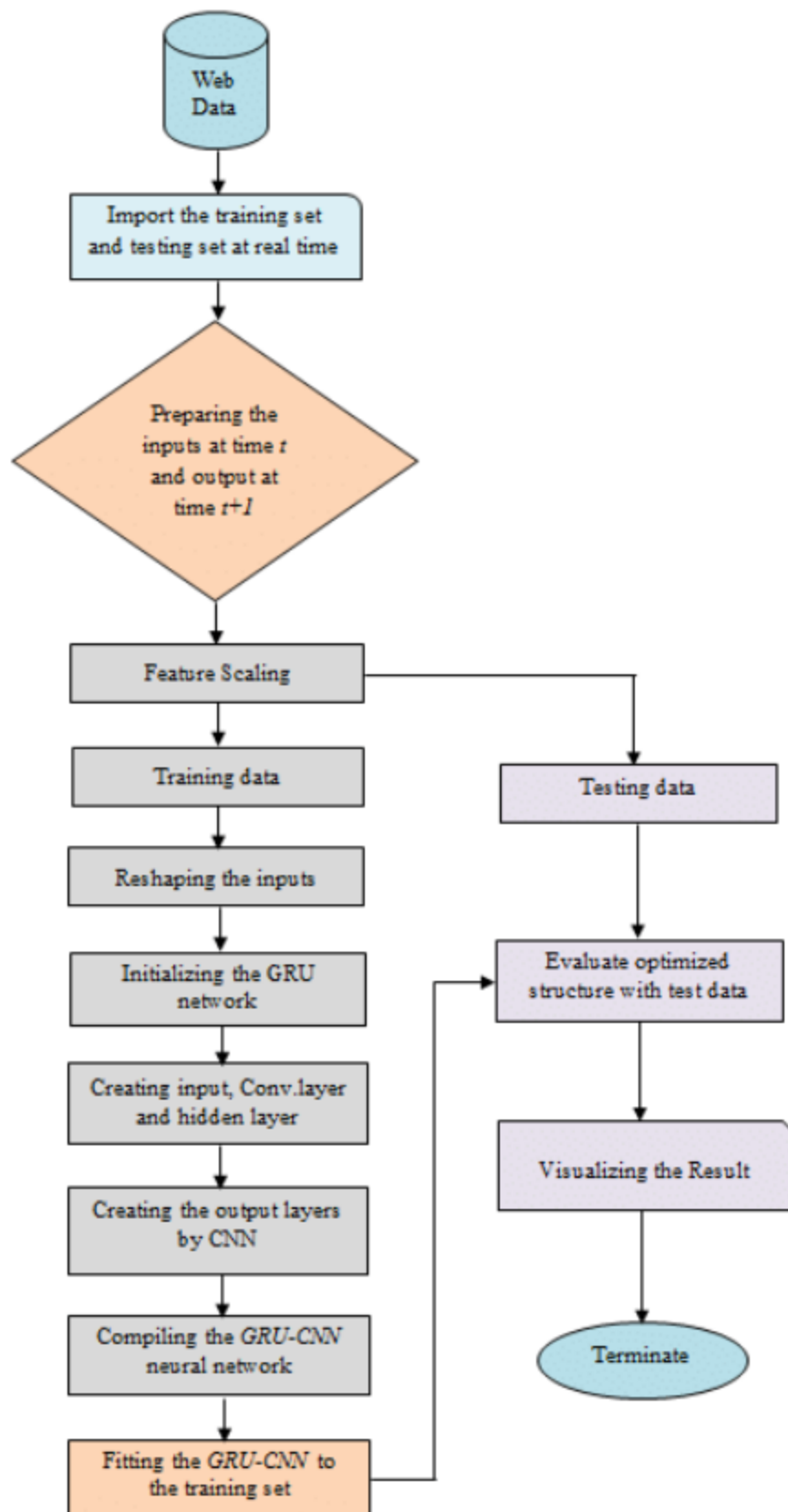
(from past 4 years crypto data)

1) https://raw.githubusercontent.com/kennedyCzar/STOCK-RETURN-PREDICTION-USING-KNN-SVM-GUASSIAN-PROCESS-ADABOOST-TREE-REGRESSION-AND-QDA/master/_SCRIPT/___pycache___/DataCollector.cpython-36.Pyc

2) https://raw.githubusercontent.com/kennedyCzar/STOCK-RETURN-PREDICTION-USING-KNN-SVM-GUASSIAN-PROCESS-ADABOOST-TREE-REGRESSION-AND-QDA/master/_SCRIPT/___pycache___/PredictiveModel.cpython-3.6.pyc

Design architecture:





IMPLEMENTATION>>

Downloading the Stock Prices Dataset

Summary	Chart	Conversations	Statistics	Historical Data	Profile	Financials	Analysis	Options	Holders	Sustainability
Time Period: Mar 03, 2019 - Mar 03, 2022 ▾ Show: Historical Prices ▾ Frequency: Daily ▾ Apply										
Currency in USD Download										
Date	Open	High	Low	Close*	Adj Close**	Volume				
Mar 02, 2022	388.93	389.22	375.21	380.03	380.03	5,349,400				
Mar 01, 2022	391.60	395.00	383.71	386.24	386.24	3,290,400				

Loading the Stock Prices Dataset

```
1 import pandas as pd
2 stock_data = pd.read_csv('./NFLX.csv', index_col='Date')
3 stock_data.head()
```

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-03-04	359.720001	362.250000	348.040009	351.040009	351.040009	7487000
2019-03-05	351.459991	356.170013	348.250000	354.299988	354.299988	5937800
2019-03-06	353.600006	359.880005	351.700012	359.609985	359.609985	6211900
2019-03-07	360.160004	362.859985	350.500000	352.600006	352.600006	6151300
2019-03-08	345.750000	349.920013	342.470001	349.600006	349.600006	6898800

Plotting the High and Low points of Netflix stock over 3 years, we see the below graph.

EVALUATION:

The code is for the above graphs::

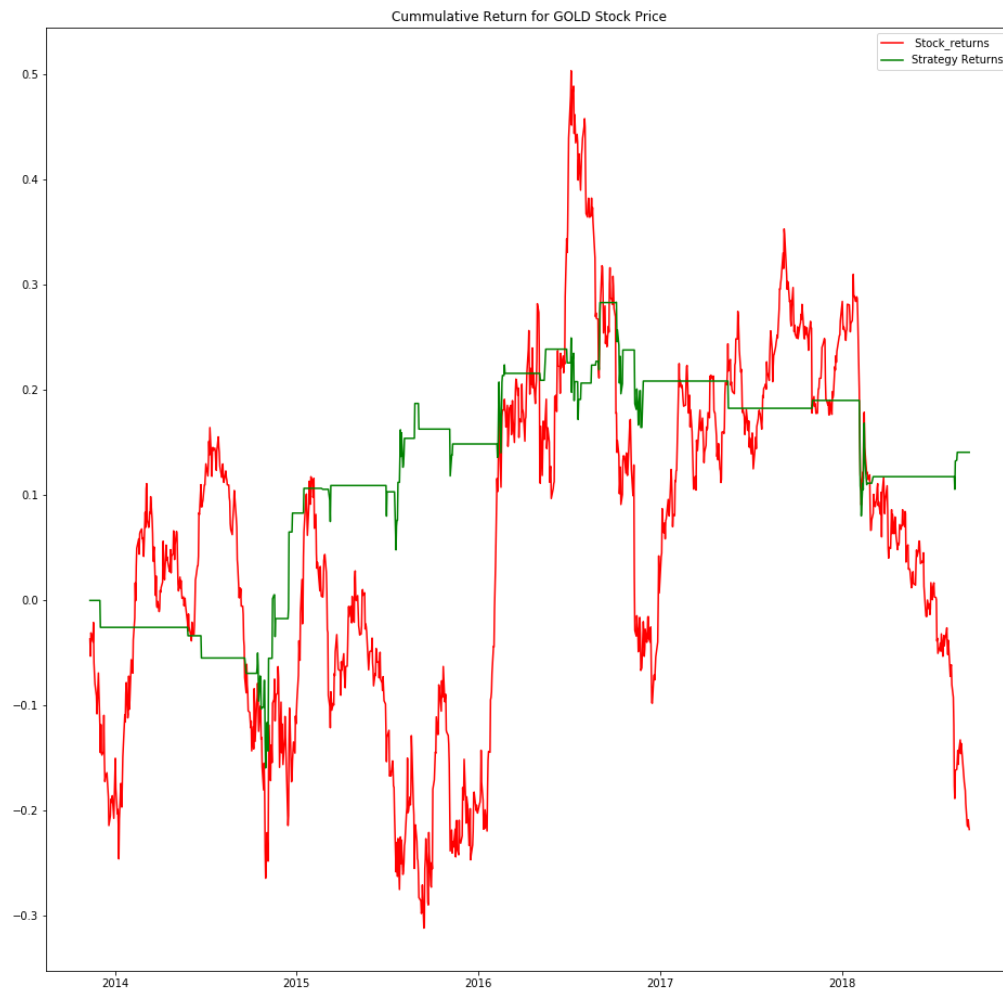
```
1 import matplotlib.dates as mdates
2 import matplotlib.pyplot as plt
3 import datetime as dt
4
5 plt.figure(figsize=(15,10))
6 plt.gca().axis.set_major_formatter(mdates.DateFormatter('%Y-%m-%d'))
7 plt.gca().axis.set_major_locator(mdates.DayLocator(interval=60))
8 x_dates = [dt.datetime.strptime(d,'%Y-%m-%d').date() for d in stock_data.index.values]
9
10 plt.plot(x_dates, stock_data['High'], label='High')
11 plt.plot(x_dates, stock_data['Low'], label='Low')
12 plt.xlabel('Time Scale')
13 plt.ylabel('Scaled USD')
14 plt.legend()
15 plt.gcf().autofmt_xdate()
16 plt.show()
```

Steps after this:

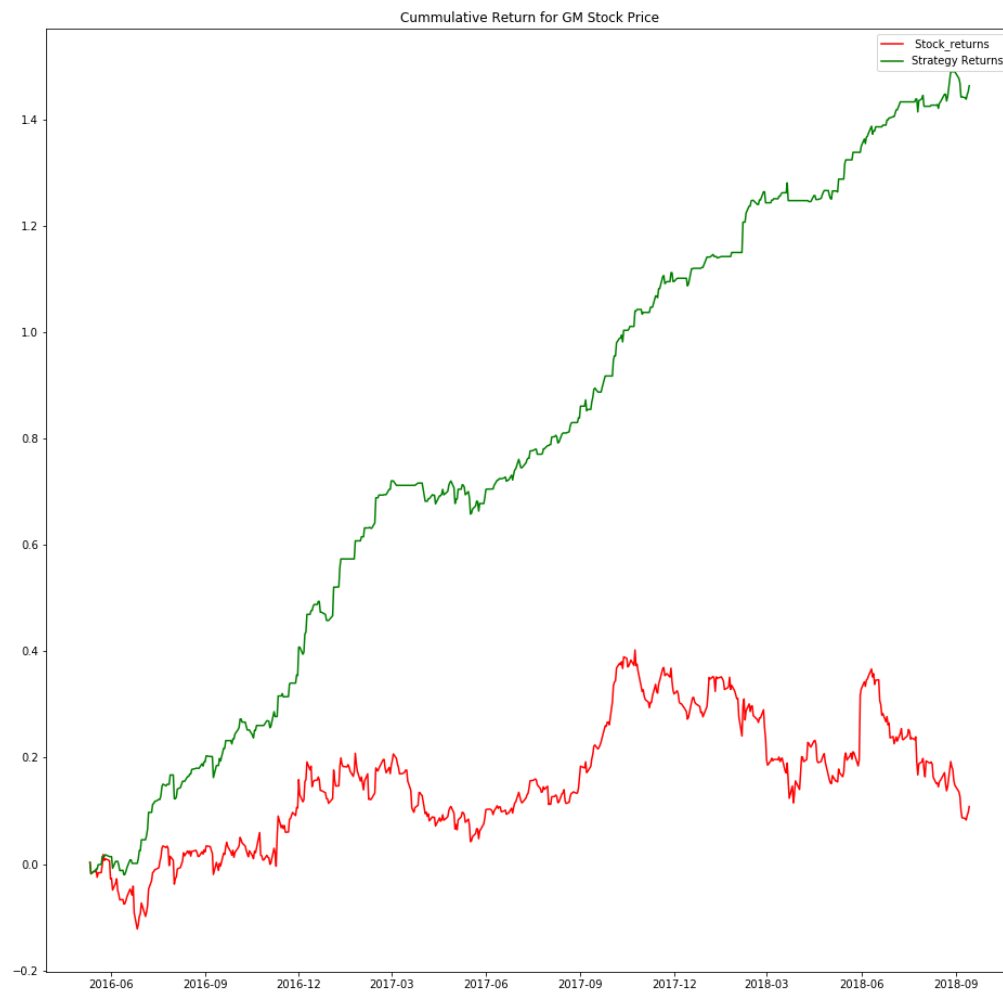
- 1.Importing the Libraries**
- 2.Data Preprocessing**
- 3.Train and Test Sets for Stock Price Prediction**
- 4.Building the LSTM model**

Types of stocks and their shares from past 4 years ::

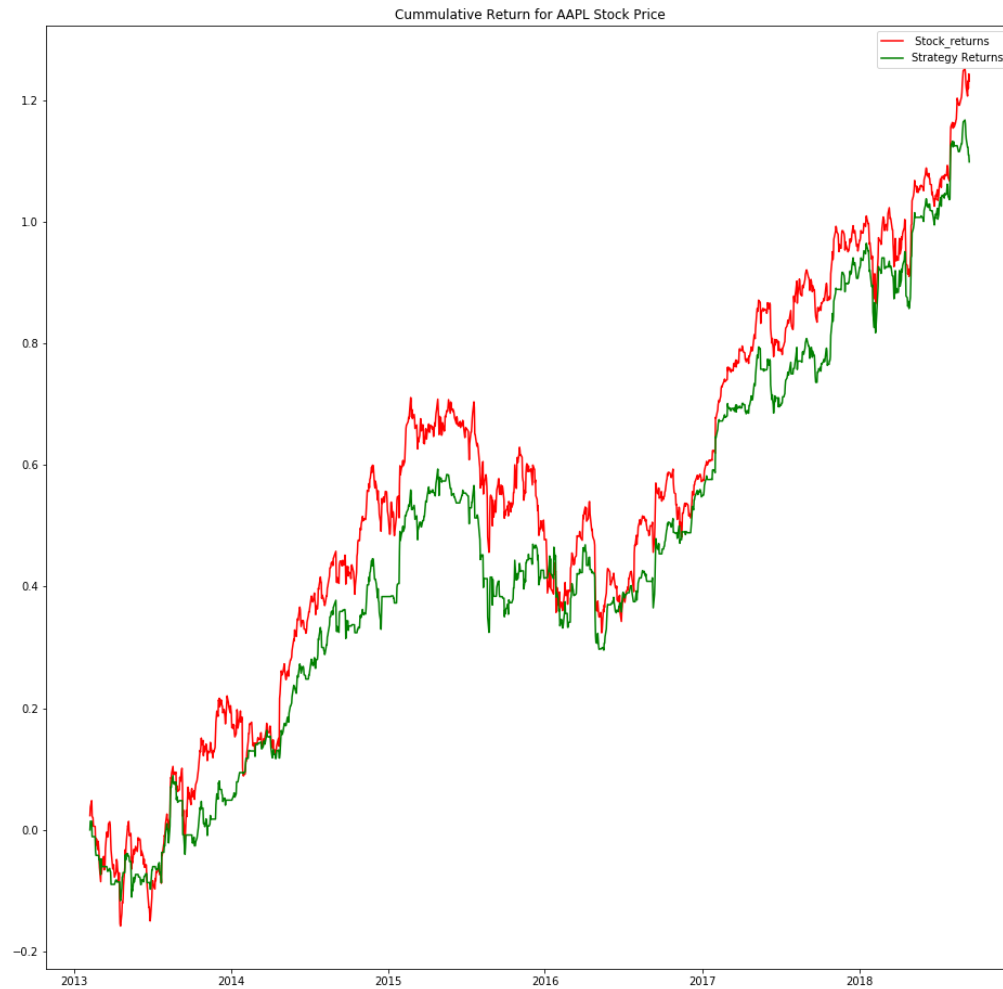
Gold Stock Retuns :



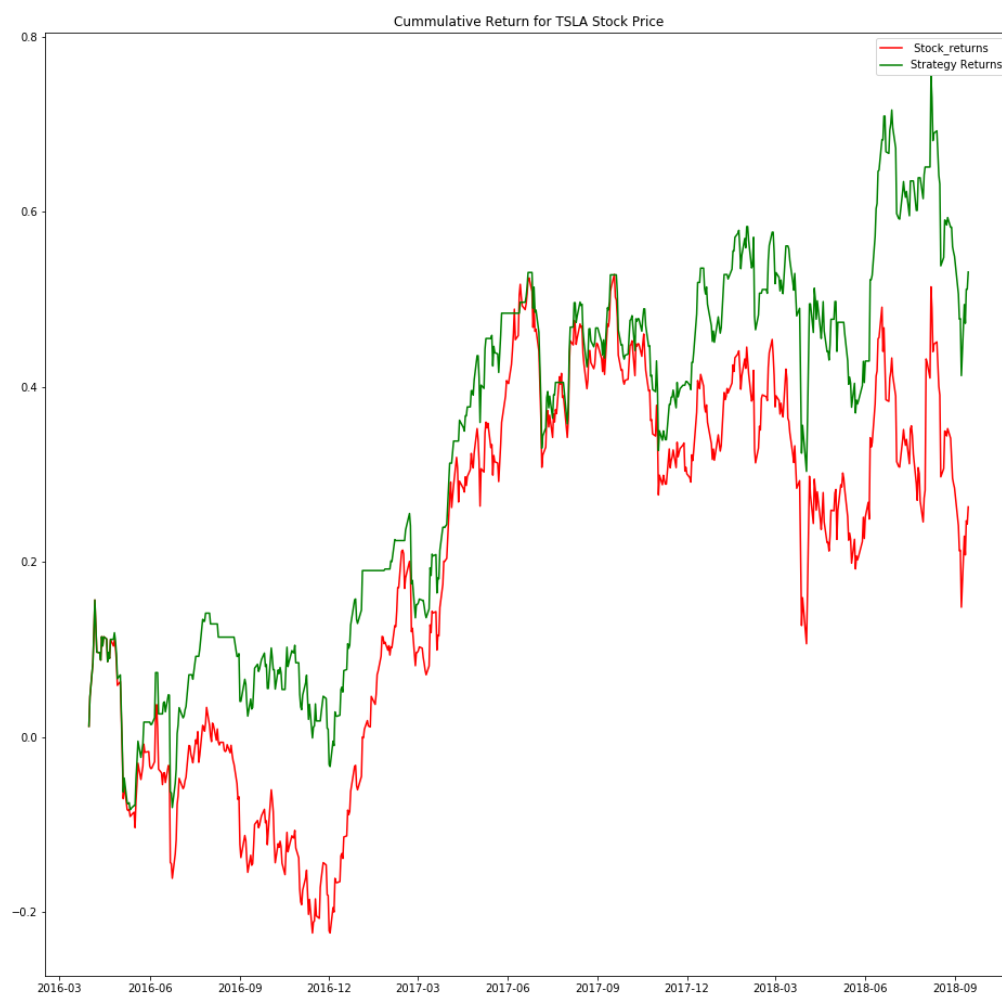
General Motors stock returns :



Apple stock returns :

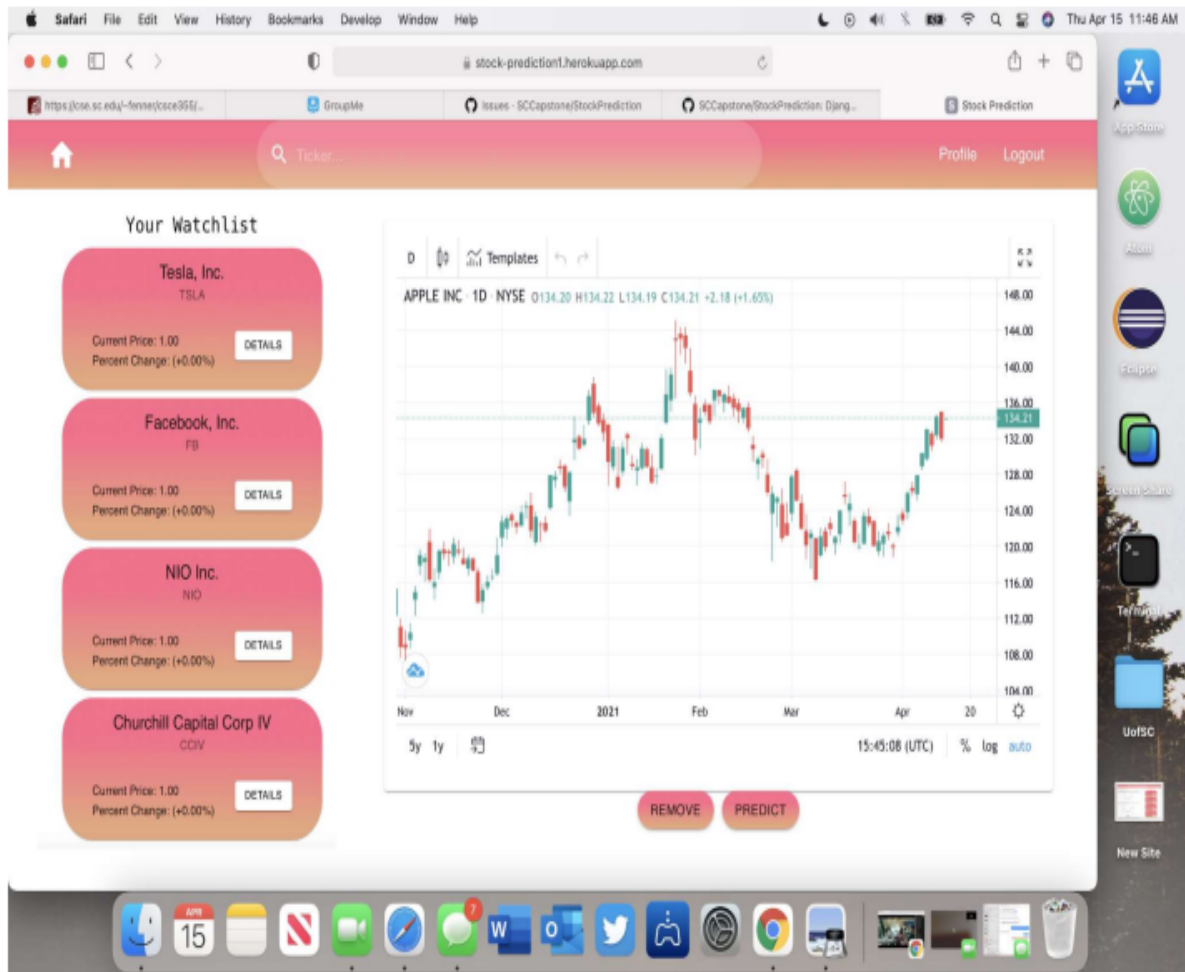


Tesla Stock Returns :

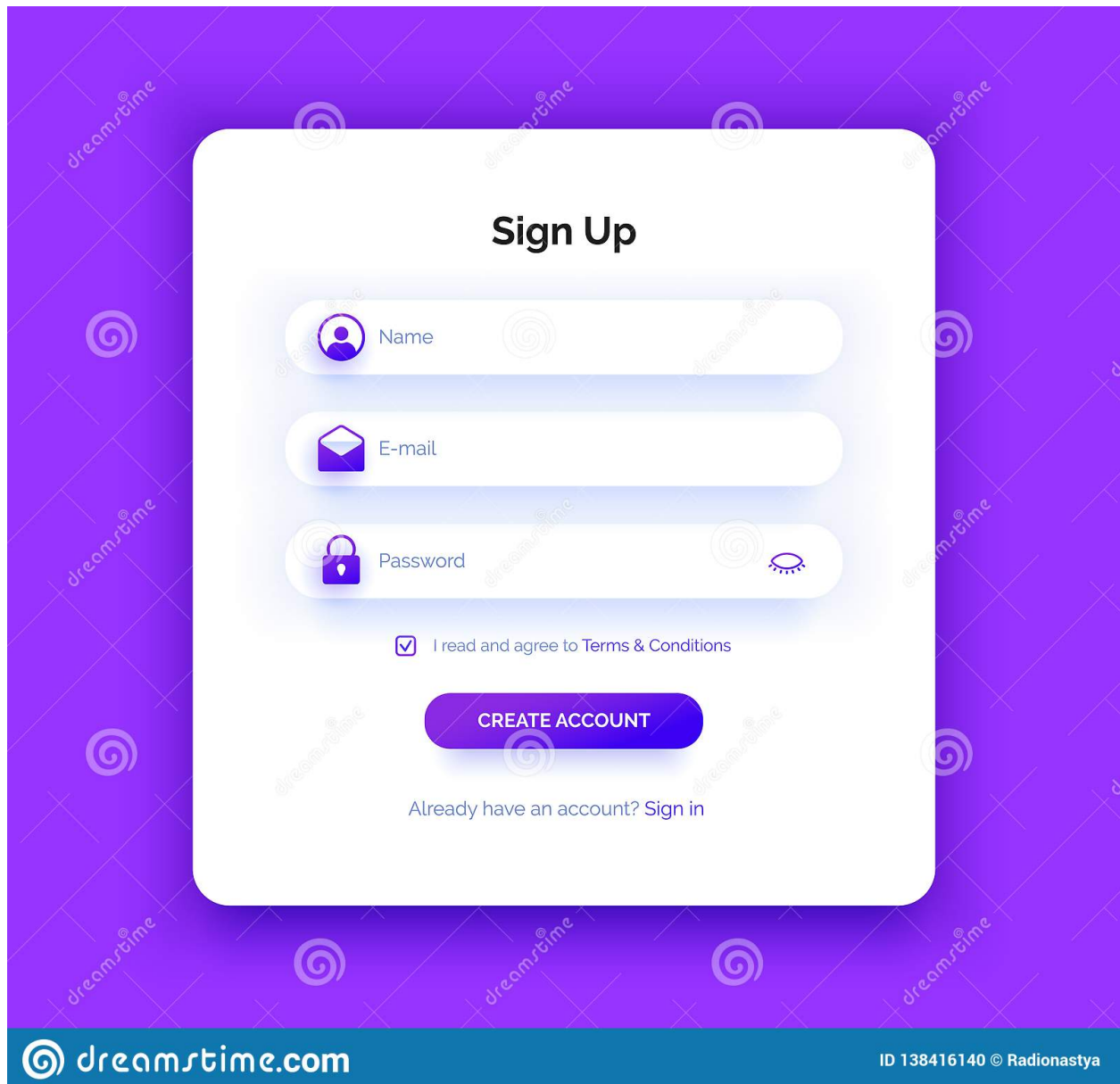


HOME PAGE


This is the home page of our web application for stock return prediction. The user can navigate to signup pages for the stocks and company respectively. He/she can also navigate to respective login pages. The user can also explore the stocks available for buy but only after login.






Signup page:



Sign Up

 Name


 E-mail

 Password 

☒ I read and agree to Terms & Conditions

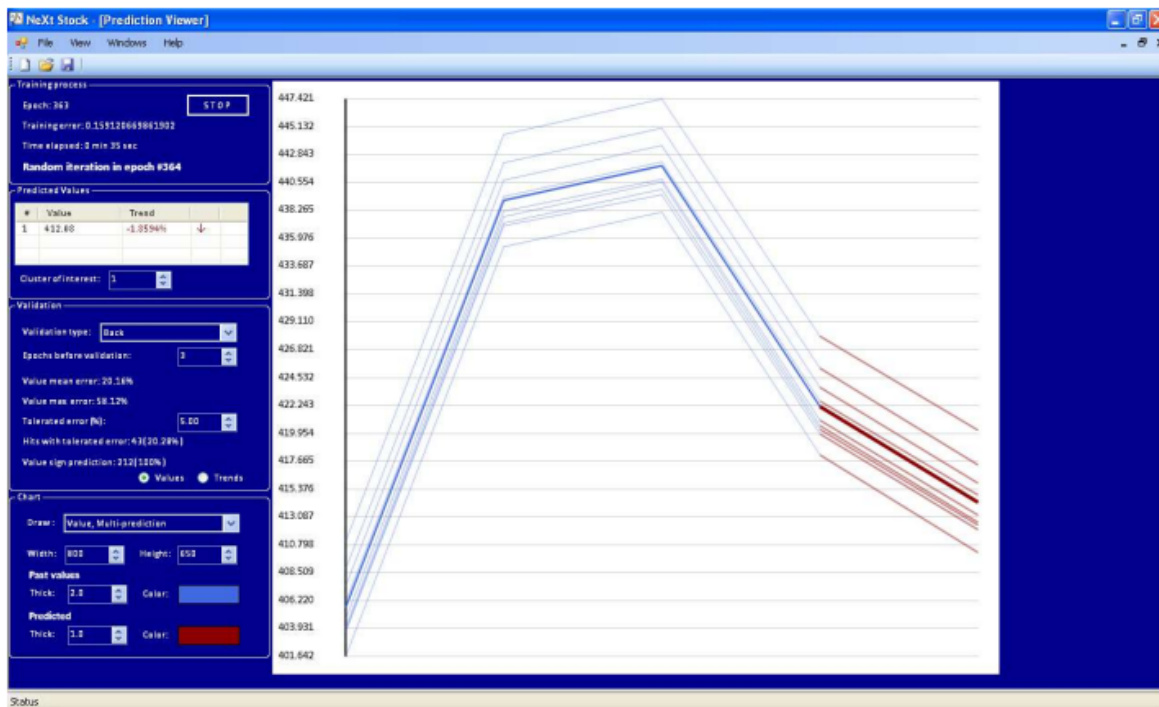
CREATE ACCOUNT

Already have an account? [Sign in](#)

 dreamstime.com

ID 138416140 © Radionastya

page interfreence:



Conclusion >>

>> You must note that this strategy is trading is a low frequency approach and this fits to make steady income over a period of time.

For high Frequency Trading the result of the return is quite high.

GOLD happens to give the most return on applied strategy(As shown in the graphs above).

Also worthy of mention is the fact that, Random Forest Classifier + PCA in most cases performed better for stocks prices with both unsteady and steady rise. Followed Next to Adaboost, then Gradientbost Classifier.

In any case, the performance of an algorithm depends on the structure of the underlying prices. Its behaviour over a time series.

For different stocks different agorithm perform best.....

