**Machine Learning Algorithms(ECM2002)**

**FINAL PROJECT REPORT**

**STOCK PRICE PREDICTION BY**

**K-Nearest Neighbors (KNN) Algorithm In Python**

****

**SUBMITTED BY:**

**ABHYUDAY SINGH**

**MANDLOI *(18BLC1159)***

***IN GUIDANCE OF:***

***PROF. Omprakash sahu***

INDEX

1. Abstract…………………………………………………………………….3
2. INTRODUCTION………………………………………………………….3
3. LITERATURE REVIEW………………………………………………….4
4. *DATA*………………………………………………………………………….5
5. GRAPHICAL REPRESENTATION OF DATA…………………….6

## k-Nearest Neighbours………………………………………………..6

1. IMPLEMENTATION OF STRATEGY.……………………………..7
2. CODE………………………………………………………………………….13
3. Graph Prediction for different listed companies:……….13
4. CONCLUSION……………………………………………………..14

**Abstract:**

**Stock prices prediction is interesting and challenging research topic. Developed countries' economies are measured according to their power economy. Currently, stock markets are considered to be an illustrious trading field because in many cases it gives easy profits with low risk rate of return. Stock market with its huge and dynamic information sources is considered as a suitable environment for data mining and business researchers. In this paper, we applied**

**k- nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a sample of some major companies listed on the share market exchange to assist investors, management, decision makers, and users in making correct and informed investments decisions. According to the results, the k NN algorithm is robust with small error ratio; consequently, the results were rational and also reasonable. In addition, depending on the actual stock prices data; the prediction results were close and almost parallel to actual stock prices.**

**INTRODUCTION:**

**A stock market, equity market or share market is the aggregation of buyers and sellers of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange, as well as stock that is only traded privately, such as shares of private companies which are sold to investors through equity crowdfunding platforms. Investment in the stock market is most often done via stockbrokerages and electronic trading platforms. Investment is usually made with an investment strategy in mind.**

**Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physhological, rational and irrational behaviour, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.**

**Can we use machine learning as a game changer in this domain? Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we didn’t see before, and these can be used to make unerringly accurate predictions.**

**For trading and investing purposes in stock market, trader and investor uses different strategy and technical analysis, in this project. K-NN machine learning algorithm is used to predict stock prices which work more precisely for large number of data.**

**LITERATURE REVIEW:**

**Financial services companies are developing their products to serve future prediction. There are a large amount of financial information sources in the world that can be valuable research areas, one of these areas is stock prediction and also called stock market mining. Stock prediction becomes increasingly important especially if number of rules could be created to help making better investment decisions in different stock markets.**

**The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction. Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect**

**it.**

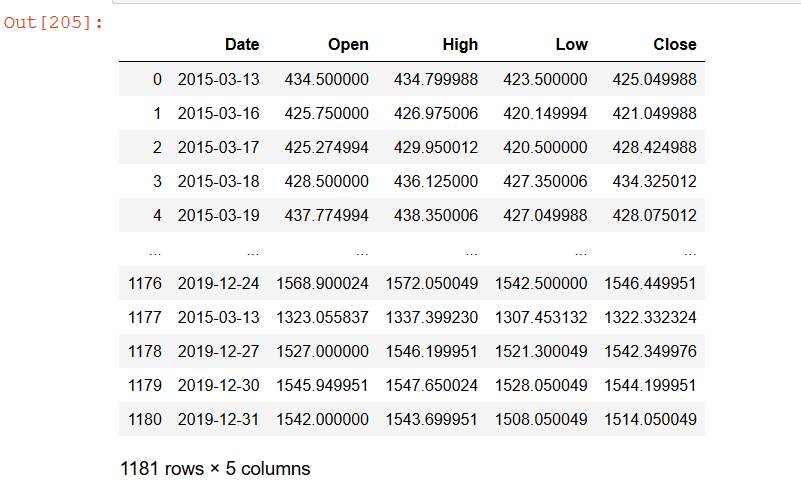
**The technical and fundamental analysis is done by sentimental analysis process. The use of machine learning and artificial intelligence techniques to predict the prices of the stock is an increasing trend. More and more researchers invest their time every day in coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model. Due to the vast number of options available ,there can be n number of ways on how to predict the price of the stock, but all methods don’t work the same way here are always other factors that influence the price of the stock, such as sentiments of the investor, public opinion about the company, news from various outlets, and even events that cause the entire stock market to fluctuate.**

***DATA :*** ****

**Data is taken from:**

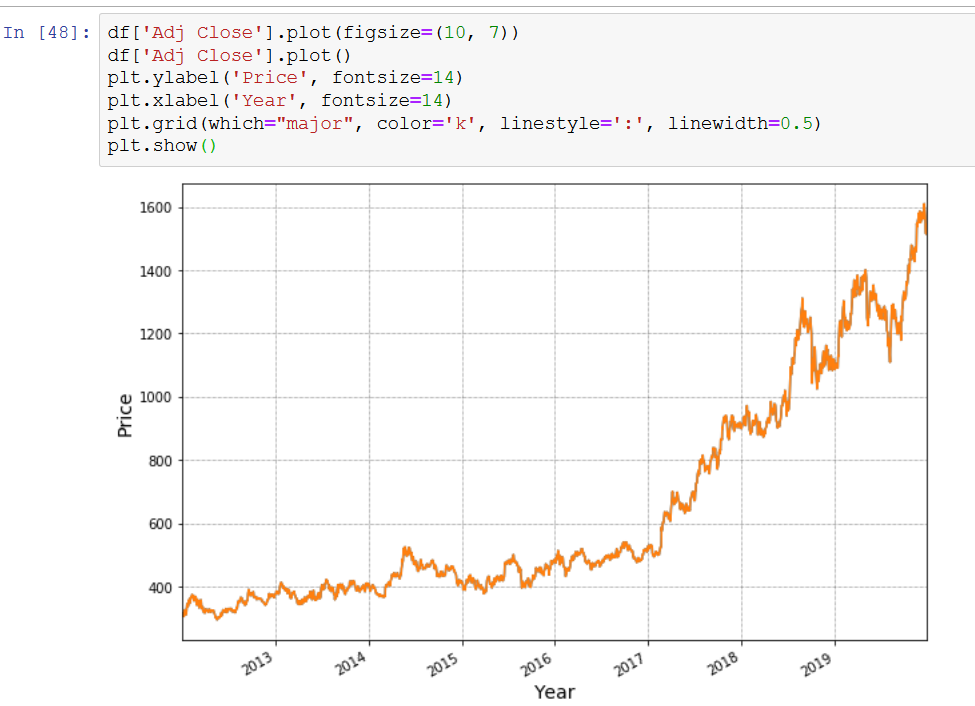
[**https://in.finance.yahoo.com/quote/RELIANCE.NS/history?period1=1325376000&period2=1577836800&interval=1d&filter=history&frequency=1d**](https://in.finance.yahoo.com/quote/RELIANCE.NS/history?period1=1325376000&period2=1577836800&interval=1d&filter=history&frequency=1d)

**It is data of “Reliance Industries India”.**

****

**At respective date, ‘open’ is price at which market opens, ’high’ is highest price of a share, ‘low’ is lowest price per share & ‘close’ is at which market closes.**

**GRAPHICAL REPRESENTATION OF DATA:**

****

## **k-Nearest Neighbours**

**ML algorithm that one can use here is kNN (k nearest neighbours). Based on the independent variables, kNN finds the similarity between new data points and old data points.**

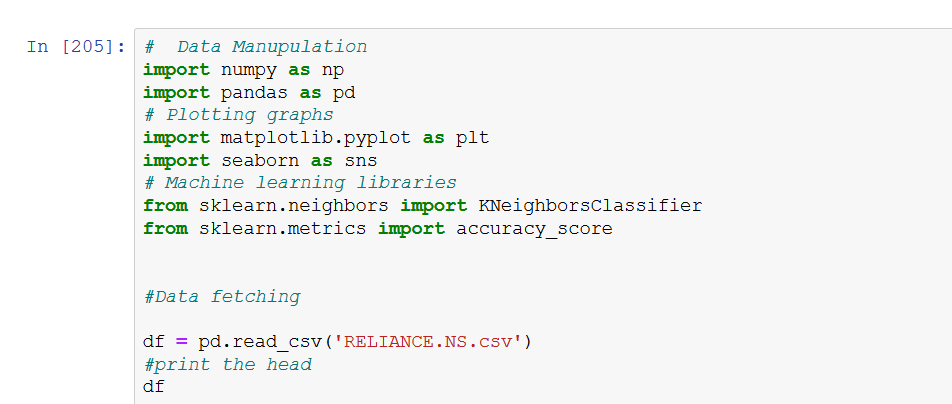
**K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. The following two properties would define KNN well −**

* **Lazy learning algorithm − KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.**
* **Non-parametric learning algorithm − KNN is also a non-parametric learning algorithm because it doesn’t assume anything about the underlying data.**

**Now, let us understand the implementation of K-Nearest Neighbor’s (KNN) in Python in creating a trading strategy.**

**IMPLEMENTATION OF STRATEGY**

1. **IMPORTING LIBRARIES AND FETCHING DATA:**

**We will start by importing the necessary libraries required to implement the KNN Algorithm in Python. We will import the numpy libraries for scientific calculation. Next, we will import the matplotlib.pyplot library for plotting the graph. We will import two machine learning libraries KNeighborsClassifier from sklearn. neighbors to implement the k-nearest neighbors vote and accuracyscore from sklearn.metrics for accuracy classification score. **

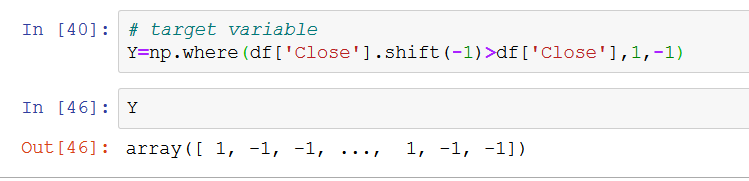
**2.) Define predictor value:**

**Predictor variable, also known as an independent variable is used to determine the value of the target variable. We use ‘Open-Close’ and ‘High-Low’ as a predictor variable. We will drop the NaN values and store the predictor variables in ‘X’.**

****

## **4. Define Target Variables**

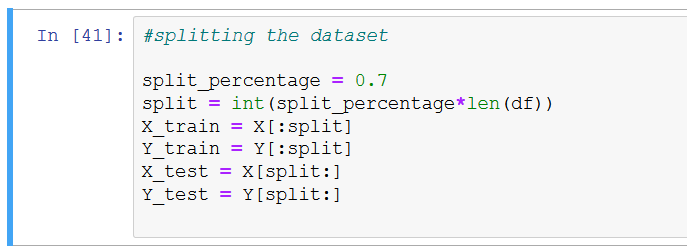
**The target variable, also known as the dependent variable is the variable whose values are to be predicted by predictor variables. In this, the target variable is whether stock price will close up or down on the next trading day. The logic is that if the tomorrow’s closing price is greater than today’s closing price, then we will buy the shares , else we will sell the share. We will store +1 for the buy signal and -1 for the sell signal. We will store the target variable in a variable ’Y’.**

****

## **5. Split the Dataset**

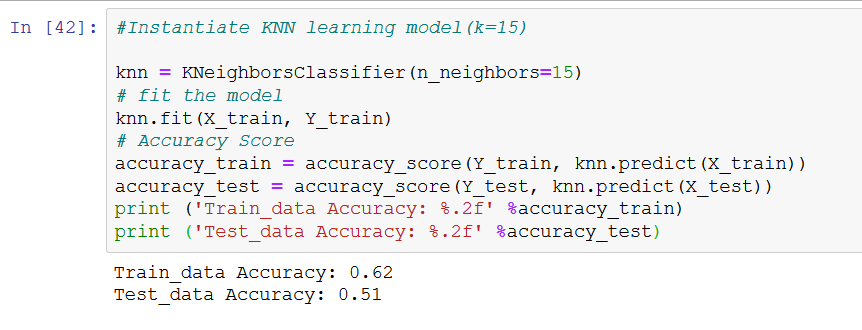
**Now, we will split the dataset into training dataset and test dataset. We will use 70% of our data to train and the rest 30% to test. To do this, we will create a split parameter which will divide the dataframe in a 70-30 ratio. You can change the split percentage as per choice, but it is advisable to give at least 60% data as train data for good results.**

**‘Xtrain’ and ‘Ytrain’ are train dataset. ‘Xtest’ and ‘Ytest’ are test dataset.**

****

**6. Instantiate KNN Model**

**After splitting the dataset into training and test dataset, we will instantiate k-nearest classifier. Here we are using ‘k =15’, you may vary the value of k and notice the change in result. Next, we fit the train data by using ‘fit’ function. Then, we will calculate the train and test accuracy by using ‘accuaracy\_score’ function.**

****

**Here, we see that an accuracy of 51% in a test dataset which means that 50% of the time our prediction will be correct.**

## **7. Create trading strategy using the model:**

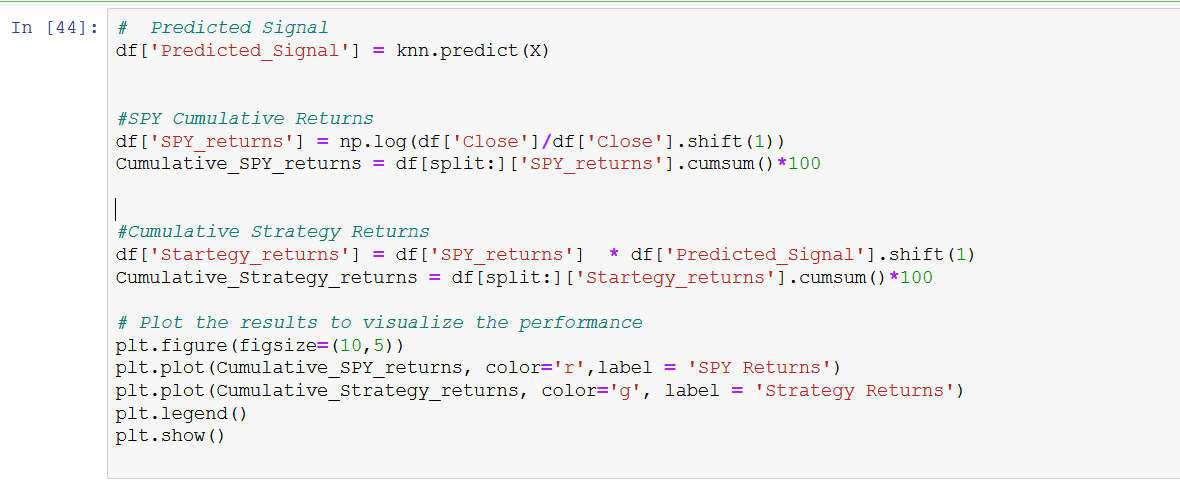
**Our trading strategy is simply to buy or sell. We will predict the signal to buy or sell using ‘predict’ function. Then, we will calculate the cumulative shares returns for test dataset. Next, we will calculate the cumulative strategy return based on the signal predicted by the model in the test dataset. Then, we will plot the cumulative share returns and cumulative strategy returns and visualize the performance of the KNN Algorithm.**

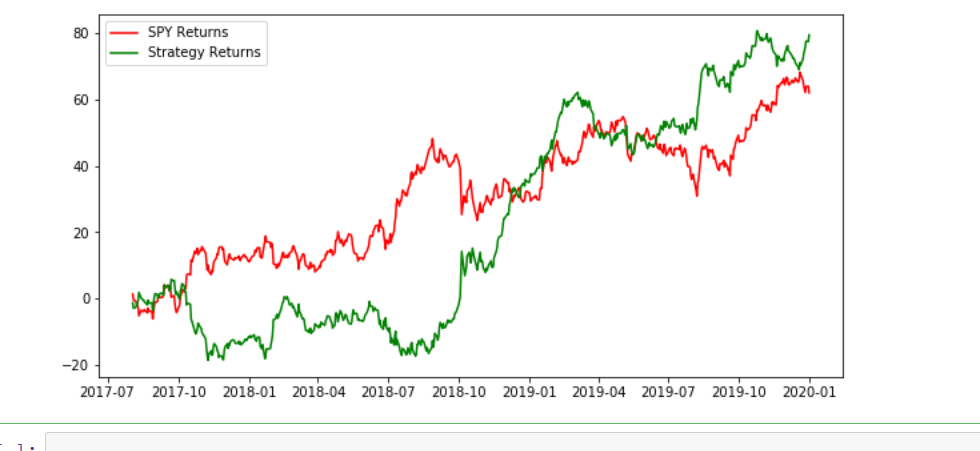
***CUMULATIVE RETURN is special technical indicator which is used, which gives value based on our prediction , comparing the closing price of two respective date, in this way it goes over complete data.***

***->Cumulative\_spy\_returns give the actual data of test data.(“RED”)***

***->Cumulative\_strategy\_returns give the predicted data which is calculated by our strategy.(“GREEN”)***

***Now both are compared by plotting with year as index(x-axis).***

****

**Output:(for testing set)**

**≈78%%**

**≈62%**

**This is clear from the graph that cumulative stock returns from 01-Jan-2012 to 01-Jan-2020 are around 62% and cumulative strategy returns in the same period are around 78%.**

**Training Set: 01-Jan-2012 to 01-Jan-2017**

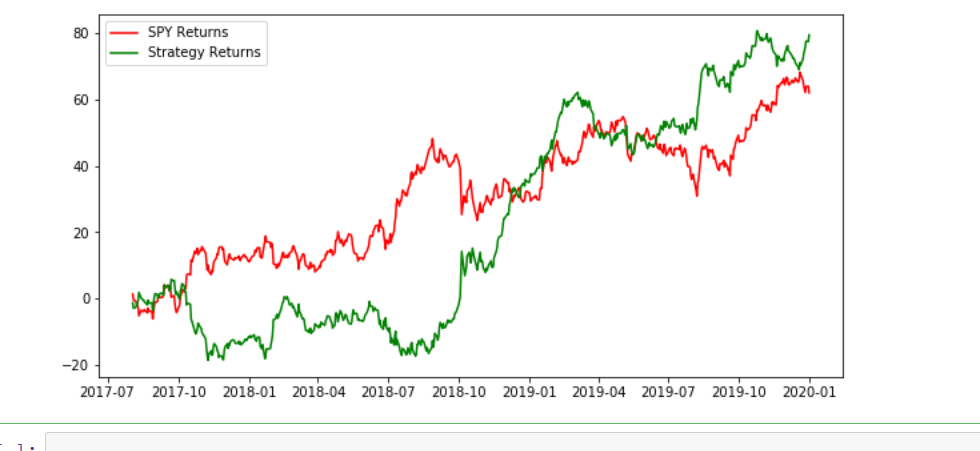
**Testing Set: 01-Jan-2017 to 01-Jan-2020**

**CODE:**

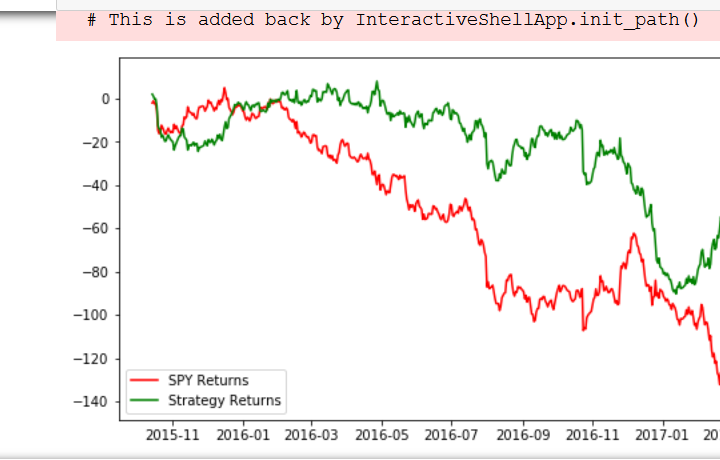
[**Untitled6 (1).ipynb**](Untitled6%20(1).ipynb)

**Graph Prediction for different listed companies:**

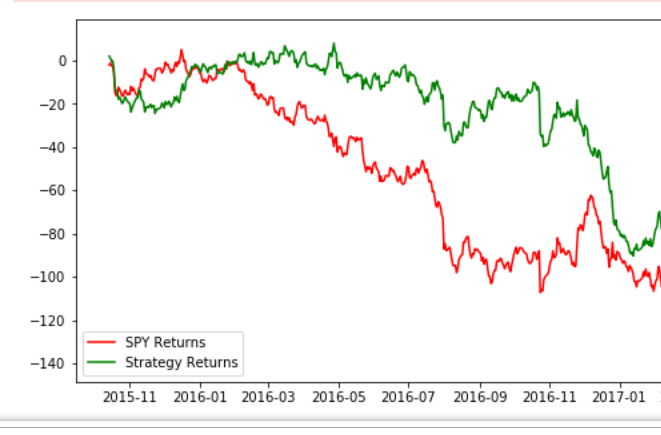
**1.)Reliance:**

****

**2.)TATA MOTORS:**

****

**3.)SBI INDIA:**

****

**We can see that there is very less difference between actual and predicted price.**

**CONCLUSION:**

**It is impossible to predict 100% correct price of stock, by using ML algorithm (K-NN) , we tried to predict the share price with accuracy of 51% , which is more precise for large dataset , we also can use another algorithms , strategy and technical indicator for getting more accurate result .**

**This project covered how machine learning could be used to predict stock prices which may be daunting as it might seem but with the right technique it could be accomplished.**