Stock Price Prediction using Machine Learning

AMAN KHARWAL / NOVEMBER 14, 2020 / MACHINE LEARNING / 27

Predicting the stock market is one of the most important applications of Machine Learning in finance. In this article, I will take you through a simple Data Science project on Stock Price Prediction using Machine Learning Python.

SEMRUSH

A Complete Toolkit for Backlink Analytics →

At the end of this article, you will learn how to predict stock prices by using the Linear Regression model by implementing the Python programming language.

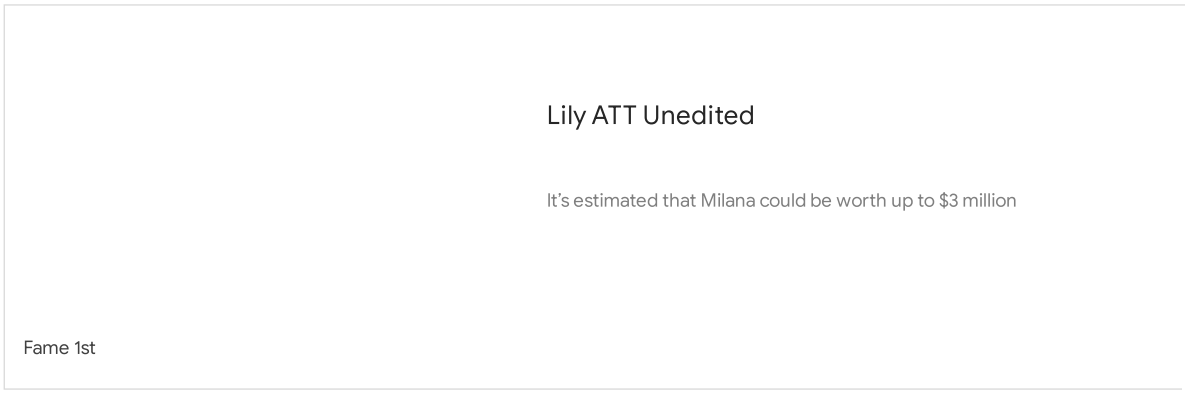
Also, Read – Machine Learning Full Course for free.

Stock Price Prediction

Predicting the stock market has been the bane and goal of investors since its inception. Every day billions of dollars are traded on the stock exchange, and behind every dollar is an investor hoping to make a profit in one way or another.

Entire companies rise and fall daily depending on market behaviour. If an investor is able to accurately predict market

movements, he offers a tantalizing promise of wealth and influence.



Today, so many people are making money staying at home trading in the stock market. It is a plus point for you if you use your experience in the stock market and your machine learning skills for the task of stock price prediction.

Let's see how to predict stock prices using Machine Learning and the python programming language. I will start this task by importing all the necessary python libraries that we need for this task:

```
import numpy as np
import pandas as pd
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

stock price prediction.py hosted with ♥ by GitHub view raw

Data Preparation

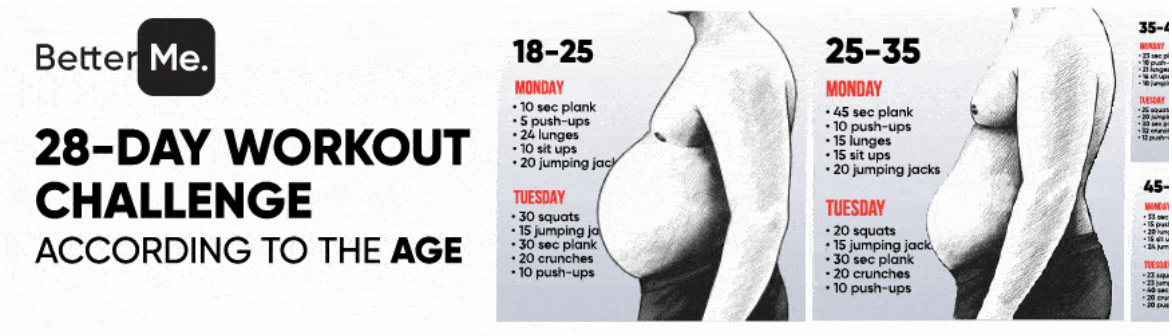
In the above section, I started the task of stock price prediction by importing the python libraries. Now I will write a function that will prepare the dataset so that we can fit it easily in the Linear Regression model:

```
def prepare_data(df,forecast_col,forecast_out,test_size):
    label = df[forecast_col].shift(-forecast_out) #creating new column called label with the
    X = np.array(df[[forecast_col]]) #creating the feature array
    X = preprocessing.scale(X) #processing the feature array
    X_lately = X[-forecast_out:] #creating the column i want to use later in the predicting
    X = X[:-forecast_out] # X that will contain the training and testing
    label.dropna(inplace=True) #dropping na values
    y = np.array(label)  # assigning Y
    X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size=test_size, random_st
```

```
10
11      response = [X_train,X_test , Y_train, Y_test , X_lately]
12      return response
```

stock price prediction.py hosted with ❤ by GitHub view raw

You can easily understand the above function as I have narrated the functioning of every line step by step. Now the next thing to do is reading the data:




```
df = pd.read_csv("prices.csv")
df = df[df.symbol == "GOOG"]
```

Now we need to prepare three input variables as already prepared in the function created in the above section. We need to declare an input variable mentioning about which column we want to predict. The next variable we need to declare is how much far we want to predict.

And the last variable that we need to declare is how much should be the size of the test set. Now let's declare all the variables:

```
forecast_col = 'close'
forecast_out = 5
test_size = 0.2
```

Applying Machine Learning for Stock Price Prediction

Now I will split the data and fit into the linear regression model:

```
X_train, X_test, Y_train, Y_test , X_lately =prepar
learner = LinearRegression() #initializing linear re
```

```
3
4 learner.fit(X_train,Y_train) #training the linear re
```

Now let's predict the output and have a look at the prices of the stock prices:

```
1  score=learner.score(X_test,Y_test)#testing the linear regression model
2  forecast= learner.predict(X_lately) #set that will contain the forecasted data
3  response={}#creting json object
4  response['test_score']=score
5  response['forecast_set']=forecast
6
7  print(response)
```

stock price prediction.py hosted with ❤ by GitHub view raw

{'test_score': 0.9481024935723803, 'forecast_set': array([786.54352516, 788.13020371, 781.84159626, 779.65508615, 769.04187979])}

So this is how we can predict the stock prices with Machine Learning. I hope you liked this article on Stock Price prediction using Python with machine learning by implementing the Linear Regression Model. Feel free to ask your valuable questions in the comments section below.

Aman Kharwal

I'm a writer and data scientist on a mission to educate others about the incredible power of data📈.

ARTICLES: 1376

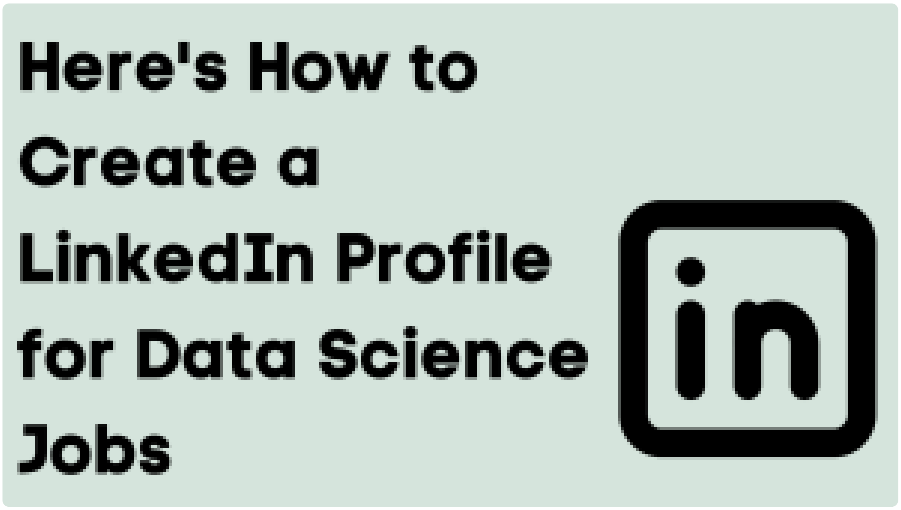
PREVIOUS

NEXT

Recommended For You

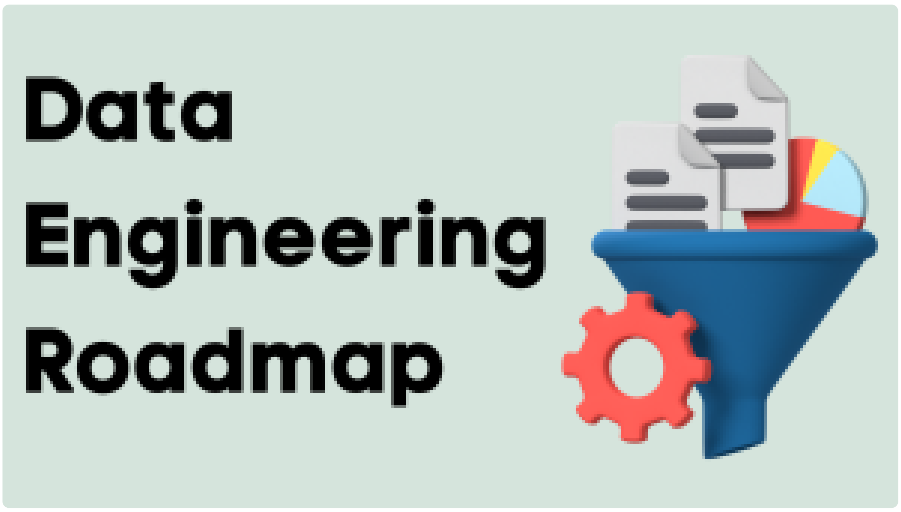
Data Science Project Ideas for 2023

February 14, 2023 / 2 Comments



Here’s How to Create a LinkedIn Profile for Data Science Jobs

February 7, 2023



Data Engineering Roadmap

January 31, 2023

Decision Tree Algorithm in Machine Learning

January 17, 2023

27 Comments

azizulhtuhin

DECEMBER 6, 2020 / 1:00 PM REPLY

can you tell me where is prices .csv file?

Aman Kharwal

DECEMBER 6, 2020 / 1:20 PM REPLY

You can download the latest data from yahoo finance

Rahul Patil
JANUARY 5, 2021 / 12:41 PM

Could you please share the sample Prices.CSV file or please share the navigation steps to download from yahoo finance.

Aman Kharwal
JANUARY 5, 2021 / 1:45 PM

https://thecleverprogrammer.com/2021/01/05/bitcoin-price-prediction-with-python/
here you will find all the steps to download data.

Vicky
JANUARY 26, 2021 / 10:52 AM REPLY

what should we on yahoo finance to get prices.csv dataset?

vicky
JANUARY 26, 2021 / 10:53 AM REPLY

what should we search on yahoo finance to get prices.csv dataset?

Aman Kharwal
JANUARY 26, 2021 / 10:59 AM REPLY

Go to Yahoo finance and search for the company, then click on the historical data and then click on download

Vicky
JANUARY 26, 2021 / 3:35 PM

I am unable to find prices.csv file at company search. Could you please share a dataset link. I would be grateful. Thanks in advance

Aman Kharwal
JANUARY 26, 2021 / 4:20 PM

prices.csv is just the name of the file: https://query1.finance.yahoo.com/v7/finance/download/INR=X?period1=1580035828&period2=1611658228&interval=1d&events=history&includeAdjustedClose=true

abdulwassey
JANUARY 27, 2021 / 4:20 PM REPLY

Hi Aman, I have got the output but with a very different test score.

{‘test_score’: 0.639145178346672, ‘forecast_set’: array([73.37040254, 73.12634778, 73.16456803, 73.20017668, 73.1776807])}

Also can you tell why we are taking below given point as it is giving me error:

df = df[df.symbol == “GOOG”]

Aman Kharwal
JANUARY 27, 2021 / 4:52 PM REPLY

maybe you are using a new dataset

RAMANDEEP SINGH BEDI
JANUARY 29, 2021 / 6:04 PM REPLY

can you please tell me what is your input and output data column

Aman Kharwal
JANUARY 29, 2021 / 6:18 PM REPLY

Close column is the input variable, which indicates close prices

vbasheer
JANUARY 30, 2021 / 10:59 AM REPLY

1.#calling the method were the cross validation and data preperation is in
X_train, X_test, Y_train, Y_test , X_lately =prepare_data(df,forecast_col,forecast_out,test_size)
learner = LinearRegression() #initializing linear regression model

learner.fit(X_train,Y_train) #training the linear regression model

ValueError: Found input variables with inconsistent numbers of samples: [246, 244] (i am getting this error when i ran above code… could you please solve for me

Aman Kharwal
JANUARY 30, 2021 / 11:07 AM REPLY

Check the dataset you are working with

mohammad basheer
JANUARY 30, 2021 / 12:28 PM

hey thanks, it worked there were some null values worked after deleting it

Rajeev
FEBRUARY 10, 2021 / 7:50 PM REPLY

@vbasheer how did you resolve the error could you please lete me know

vbasheer
JANUARY 30, 2021 / 12:32 PM REPLY

these are my results
{‘test_score’: 0.9132328868016113, ‘forecast_set’:
array([14733.28834587, 14678.01387179, 14455.85132032,
14320.59050617,
14044.5766888])}

i think test score is okay but i dont understand forecast set

Aman Kharwal
JANUARY 30, 2021 / 12:45 PM REPLY

Great

Deepti
FEBRUARY 4, 2021 / 11:08 AM REPLY

Hi am a beginner, want to know which tool to use? Spyder or Jupyter or Pycharm?

Aman Kharwal
FEBRUARY 4, 2021 / 11:23 AM REPLY

For any task where most of your work is related to analysis and visualization, you can use Jypyter notebook or Google Colab there. And for other tasks like GUI and logical problem solving you can use VS Code or any other IDE.

Aditya Astonkar
FEBRUARY 7, 2021 / 8:15 AM REPLY

hey please tell me why we are taking below given point as it is giving me error:

df = df[df.symbol == “GOOG”]

Aman Kharwal
FEBRUARY 7, 2021 / 9:00 AM REPLY

I think you have not downloaded the csv file

surbhi
MAY 15, 2021 / 9:08 PM REPLY

What is the meaning of this line-
df = df[df.symbol == “GOOG”]

Aman Kharwal
MAY 15, 2021 / 10:18 PM REPLY

GOOG is the financial symbol of stock prices of Google

priyanshu
AUGUST 16, 2022 / 11:23 PM REPLY

THE LINK THAT YOU SHARED https://query1.finance.yahoo.com/v7/finance/download/INR=X?period1=1580035828&period2=1611658228&interval=1d&events=history&includeAdjustedClose=true DOES NOT CONTAIN ATTRIBUTE ‘SYMBOL’
AttributeError: ‘DataFrame’ object has no attribute ‘symbol’
and without df=df[df.symbol=”GOOG”] it is is giving result as follow
‘test_score’: 0.6391451783466715, ‘forecast_set’: array([73.37040254, 73.12634778, 73.16456803, 73.20017668, 73.1776807])}

Aman Kharwal

AUGUST 16, 2022 / 11:26 PM REPLY

please follow this new article on Stock price prediction: https://thecleverprogrammer.com/2022/01/03/stock-price-prediction-with-lstm/

Leave a Reply

Enter your comment here...

FACEBOOK INSTAGRAM MEDIUM LINKEDIN

Copyright © Thecleverprogrammer.com 2023