Prediction of Netflix Stock Prices using Machine Learning

Shikha Verma[1] , Arti[2], Shiwani[3]

[1]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt
[2]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt
[3]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt

ABSTRACT

The primary objective of this research is to predict stock prices of Netflix for analyzing profit on day closing. The research predicts stock market as profit in form of a Chart using Python. The attributes include in the model are Date, Open, Close, High, Low, Volume and Adj Close. The profit of various years are calculated and shown in the form of graphs.

Data Collected from 2018 to 2022, I have divided dataset into two parts, training set is defined as 80% and testing set is 20 %.Final prediction is to be done in testing part only. Then with the help of graph this is shown as profit v/s year.The findings imply that NFLX performs well. Machine learning approaches can be used to forecast the index.

KEYWORDS: Stock Prediction, NFLX, Machine Learning

INTRODUCTION

a. Netflix:

Netflix is a combination of two words Net (Internet) and Flix (Flick used as an abbreviation for movie/film). **In 1997, Netflix**, in full **Netflix, Inc.**, media-streaming and video-rental company founded by American entrepreneurs Reed Hastings and Marc Randolph in Los Gatos California. In 1999, Netflix

started its business by offering an online subscription service through the Internet.Netflix started a DVD-by-mail rental service which provided an online catalogue of movies. Subscribers chose movies and television shows from the Company's website, and the shows were then mailed to them in the form of DVDs, along with prepaid return envelopes, from one of the company's more than 100 distribution locations.

2.   PROPOSED METHOD

This section explains the six-step proposed method to predict the stock trend based on profit for four years.

2.1   Data Collection

The dataset used for this work was collected from Kaggle site for four years (from 5 Feb., 2018 to 4 Feb. 2022) Shown in Fig1.The data consist of 1009 instances and 7 features: date, the highest price of the day, the lowest price of the day, open price, close price, volume and adjacent close price.

Fig 1 Data collection from 2018 to 2022

	Date	Open	High	Low	Close	Adj Close	Volume
0	2018-02-05	262.000000	267.899994	250.029999	254.259995	254.259995	11896100
1	2018-02-06	247.699997	266.700012	245.000000	265.720001	265.720001	12595800
2	2018-02-07	266.579987	272.450012	264.329987	264.559998	264.559998	8981500
3	2018-02-08	267.079987	267.619995	250.000000	250.100006	250.100006	9306700
4	2018-02-09	253.850006	255.800003	236.110001	249.470001	249.470001	16906900
...	...	...	...	...	...	...	...
1004	2022-01-31	401.970001	427.700012	398.200012	427.140015	427.140015	20047500
1005	2022-02-01	432.959991	458.480011	425.540009	457.130005	457.130005	22542300
1006	2022-02-02	448.250000	451.980011	426.480011	429.480011	429.480011	14346000
1007	2022-02-03	421.440002	429.260010	404.279999	405.600006	405.600006	9905200
1008	2022-02-04	407.309998	412.769989	396.640015	410.170013	410.170013	7782400

1009 rows × 7 columns

2.2   Data cleaning

After data collection, data cleaning is done which dealt with missing data, duplicate data, and filtering out poor data.The downloaded datawas not contained any missing or duplicate values so no need to clean the data.

SRNO.	Attribute	Type	Work
1	DATE	Numerical	which contains all the dates between start date and end date
2	HIGH	Numerical	which describes the highest value the stock had in a previous year
3	LOW	Numerical	is quite the contrary to HIGH and resembles the lowest value the stock had in previous year
4	OPEN	Numerical	is the value of the stock at the very beginning of the trading day
5	CLOSE	Numerical	stands for the price at which the stock is valued before the trading day closes
6	VOLUME	Numerical	tells you how many shares of that particular stock were traded that day
7	ADJ. CLOSE	Numerical	closing price adjusted for splits anddividend distributions

2.3   Data Selection

At this stage, Data required for analysis was chosen and extracted from the dataset shown in Fig 3. Table 1 shows the kind and description of the seven (7) attributes in the Netflix stockdataset.

Volume and Adj Close columns have been dropped with the help of drop command, So we are using only 5 columns: Date, High, Low, Open and Close.

```
df2=df.drop(['Adj Close','Volume'], axis = 1)
df2
```

	Date	Open	High	Low	Close
0	2018-02-05	262.000000	267.899994	250.029999	254.259995
1	2018-02-06	247.699997	266.700012	245.000000	265.720001
2	2018-02-07	266.579987	272.450012	264.329987	264.559998
3	2018-02-08	267.079987	267.619995	250.000000	250.100006
4	2018-02-09	253.850006	255.800003	236.110001	249.470001
...	...	...	...	...	...
1004	2022-01-31	401.970001	427.700012	398.200012	427.140015
1005	2022-02-01	432.959991	458.480011	425.540009	457.130005
1006	2022-02-02	448.250000	451.980011	426.480011	429.480011
1007	2022-02-03	421.440002	429.260010	404.279999	405.600006
1008	2022-02-04	407.309998	412.769989	396.640015	410.170013

1009 rows × 5 columns

Fig 3. Extract Data from Dataset

2.4   Data Transformation

Data consolidation is another term for it. The chosen data is translated into

forms that can be used for data mining at this stage. The datasets were scaled to fit the model's tolerances and saved in the Commas Separated Value (.CVS) file format.

In this Dataset, record of each year is kept separately and is analyzed using python libraries.

2.5   Data Mining Stage

The data mining stage was divided into three phases. At each phase all the algorithms were used to analyze the stock datasets. The testing method adopted for this research was percentage split that train on a percentage of the dataset, crossvalidate on it and test on the remaining percentage. Thereafter interesting patterns representing knowledge were identified. Fig 5 shows splitting data into training and testing.

Fig 5 Training and Testing data

```
# Training and testing
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test=train_test_split(X,y, test_size=0.2)
print("shape of X_train", X_train.shape)
print("shape of y_train", y_train.shape)
print("shape of X_test", X_test.shape)
print("shape of y-test", y_test.shape)

shape of X_train (807, 7)
shape of y_train (807,)
shape of X_test (202, 7)
shape of y-test (202,)
```

Now you can see the size of dataset has been decreased as I have divided data into 80% training data and 20% Testing Data.

2.6   Applying machine learning algorithm

In this paper, **KNeighbors Regression** machine learning algorithm is

year	Date	Open	High	Low	Close	month	day	Profit
2020	2020-12-31	567.979980	575.369995	541.000000	556.549988	12	31	55.030029
2018	2018-12-31	421.380005	423.209991	413.079987	418.970001	12	31	32.529999
2021	2021-12-31	692.349976	700.989990	686.090027	691.690002	12	31	27.869996
2019	2019-12-31	382.769989	385.989990	378.799988	385.029999	12	31	17.639984

implemented. The final data set that will be fed to ML algorithm consisted of 8 features and maximum profit in each year as shown in fig.

3.   Results and Discussion

This section shows the results obtained from predicting the trend of NFLX stocks of four years by the selected model defined by section 2. The stock trend of each year shown as:
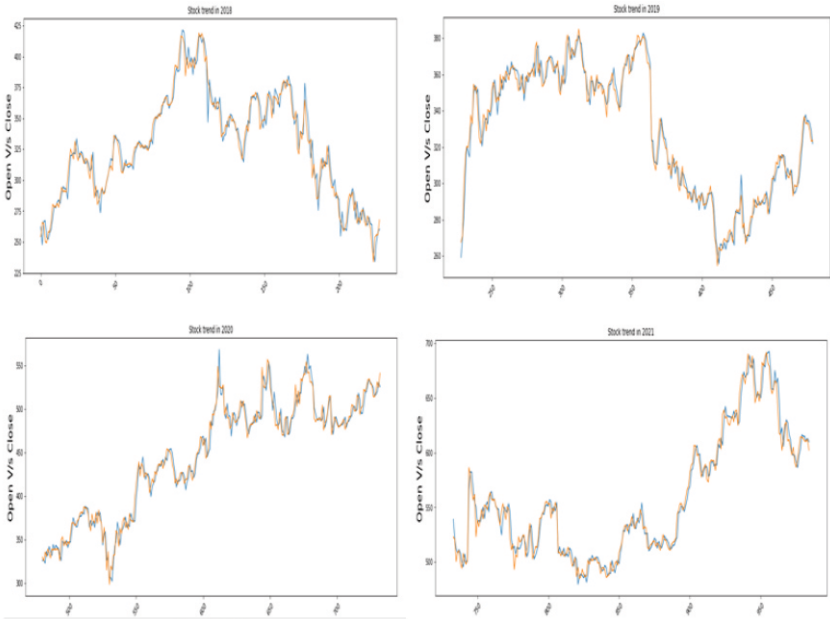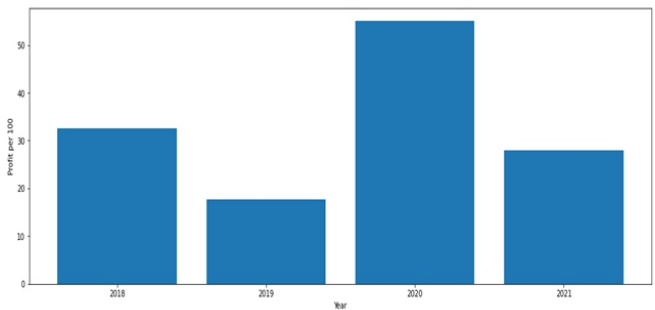

Fig. Stock trend of each year (2018-2021)

Finally, the bar graph shows the maximum profit made by NFLX in each year.

4. CONCLUSION AND FUTURE SCOPE:

In this paper, ML algorithms are used to predict the stock trend of NFLX by using technical indicators. The overall accuracy we obtained is fairly good enough i.e. 99% since the stock trend could be affected by many random factors other than news and price information. In this research paper, we end up with only one dataset of NFLX having a small number of instances (only from 2018 to 2021). In future projects, we extend our analysis on multiple datasets with many machine learning algorithms.

5.   REFERENCES:

1.   Khan W, Malik U, Ghazanfar MA, Azam MA, Alyoubi KH, Alfakeeh AS (2019) Predicting stock market trends using machine learning algorithms via publicsentiment and political situation analysis. SoftComput.https://doi.org/10.1007/s00500-019- 04347-y

2.   Khan, W., Ghazanfar, M.A., Azam, M.A. et al. Stock market prediction using machine learning classifiers and social media, news. J Ambient Intell Human Comput(2020).

3.   Neha, & Rohan Lal. (2019). Prediction Analysis of Stock Price using Machine Learning Approach. Retrieved October 02, 2020, fromhttps://www.academia.edu/43807018/Prediction_Analysis_of_Stock_Price_using_Machine_Learning_Approach

4.   Budak, C., Goel, S., & Rao, J. (2016). Fair and Balanced? Quantifying Media Bias Through Crowdsourced Content Analysis. SSRN Electronic Journal. doi:10.2139/ssrn.2526461

5.   Kanade, P. A. (2020). Machine Learning Model for Stock Market Prediction. International Journal for Research in Applied Science and Engineering Technology, 8(6), 209-216.doi:10.22214/ijraset.2020.6030