Financial Forecasting through Predictive Analytics:
Insights into Netflix's stock price evolution

Submitted by Group 15

Team members:
Joshua Benezer
Lavansh Choudhary
Prarthana Mudkavi
Vanshika Singh
Divya Gupta

Introduction

The introduction of the report focuses on a comprehensive study using predictive analytics to forecast the stock price movement of leading streaming service provider Netflix. The study examines more than two decades of historical data from January 2, 2003 to February 23, 2034 in detail. The aim of the research remains to identify the hidden patterns and insights that are important for the prediction of stock price.

Predictive analysis, an analysis of the quantity with the aim of finding the most important point after our exploratory analysis of the data. In addition to adjusted closing prices, the stock dataset includes various factors such as evaluating effects such as stock splits and dividends to highlight different stock characteristics, opening, closing highs and lows etc.

To get a nuanced understanding of market dynamics, to decide on important issues, attributes play an important role. Integration of machine learning models, like random forest models, highlight the research methodology. Unlike other forecasting models, it takes traits that interact with stock prices to predict future prices. By analyzing Netflix's historical data, the model spins out subtle patterns and trends that might be missed by a human analyst's eye, making it more significant for future stock movements The introduction justifies that in-depth analysis of data-driven insights provides a competitive edge in stock market forecast prediction. Generally, forecasting results guide us in decoding the nuances of stock trading, reinforcing them with supporting research evidence.

Data Background

The historical data of Netflix, a leading streaming giant, is multifaceted, catering to its stock market performance over the last two decades. The data encompasses categorical as well as statistical attributes, generating it to be the right concoction for analysis. The data entails several key attributes that offer insights into Netflix stock's daily performance. These attributes include:

Data Attributes:

Date: The date feature is important for time series analysis, which helps track a stock's performance over a long period of time.

Open: This represents the opening price of a stock on any trading day, and is the starting point for the daily price movement.

High: The highest price a stock reaches on a trading day, meaning the highest market price of the day.

Low: This displays the stock's lowest price on the trading day, which reflects the lowest market price.

Close: The closing price is important because it represents the final valuation of the stock at the end of the trading day.

Adj Close: This adjusted price determines any splits or shares of the stock, and provides an indication of the net value of the stock.

Data Preprocessing

An essential first step in guaranteeing the precision and dependability of the analysis is data preprocessing. Among the crucial preprocessing actions are:

Date Conversion: In order to properly analyze time series data and perform chronological sorting and analysis, the "Date" column must be converted to a date type.

Binary target variable: It provided a straightforward but perceptive look at the daily movement of stocks. It indicated whether the closing price of the stock had increased above its opening price.

Statistical Imputation Method: The median of the corresponding column was used to fill in the missing values to address the problem of missing data. This methodology minimizes the potential bias or inaccuracies that missing data may introduce while preserving the integrity of the dataset.

The study provides important insights into the financial dynamics of the streaming service company by laying a solid foundation for using predictive analytics to forecast stock price movements through these specific data attributes and thorough preprocessing steps.

Research question

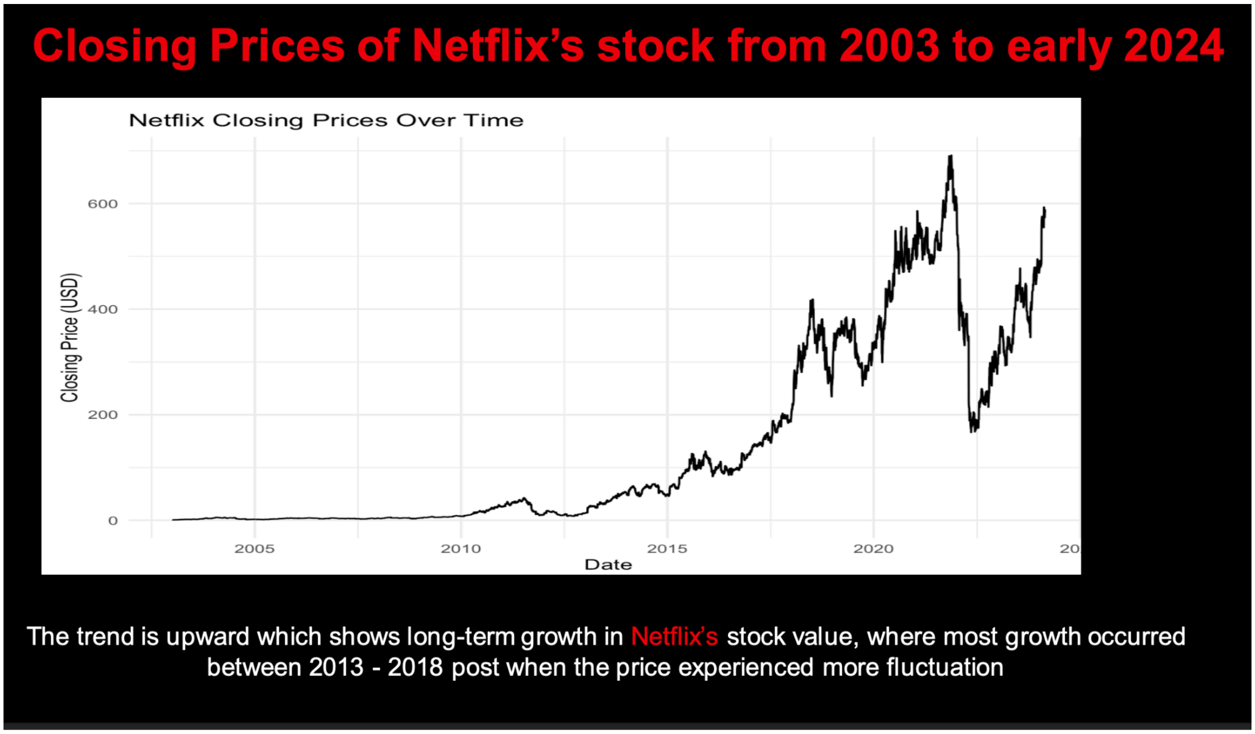
The core research question of the study is: **"Can we predict the direction of Netflix's stock closing price based on the attributes (opening price, high, low, and volume of trading)?"** This question aims to explore the predictive power of these specific stock attributes in determining the future closing price direction of Netflix's stock.

The question comes from the field of financial analysis. Accurately predicting stock prices has great value for investors, analysts and financial planners. By focusing on the opening price, daily high, daily low and trading volume, the study will determine if these data points, which show different parts of the stock's performance each day, can provide useful clues about its closing direction.

The opening price shows where the stock started for the day. It provides a baseline to measure the day's results against. The highest and lowest prices show the stock's volatility within the day's trading. The trading volume offers additional context, as high volumes can indicate strong interest or sentiment about the stock, potentially correlating with significant price movements.

By analyzing these attributes, the study aims to uncover patterns or trends that could suggest a predictive relationship, enabling more informed decision-making regarding stock trades. The research question drives the analytical approach, focusing on data-driven insights to enhance understanding of stock market dynamics, particularly for Netflix's stock.

Visualization of Target Outcome: Closing prices



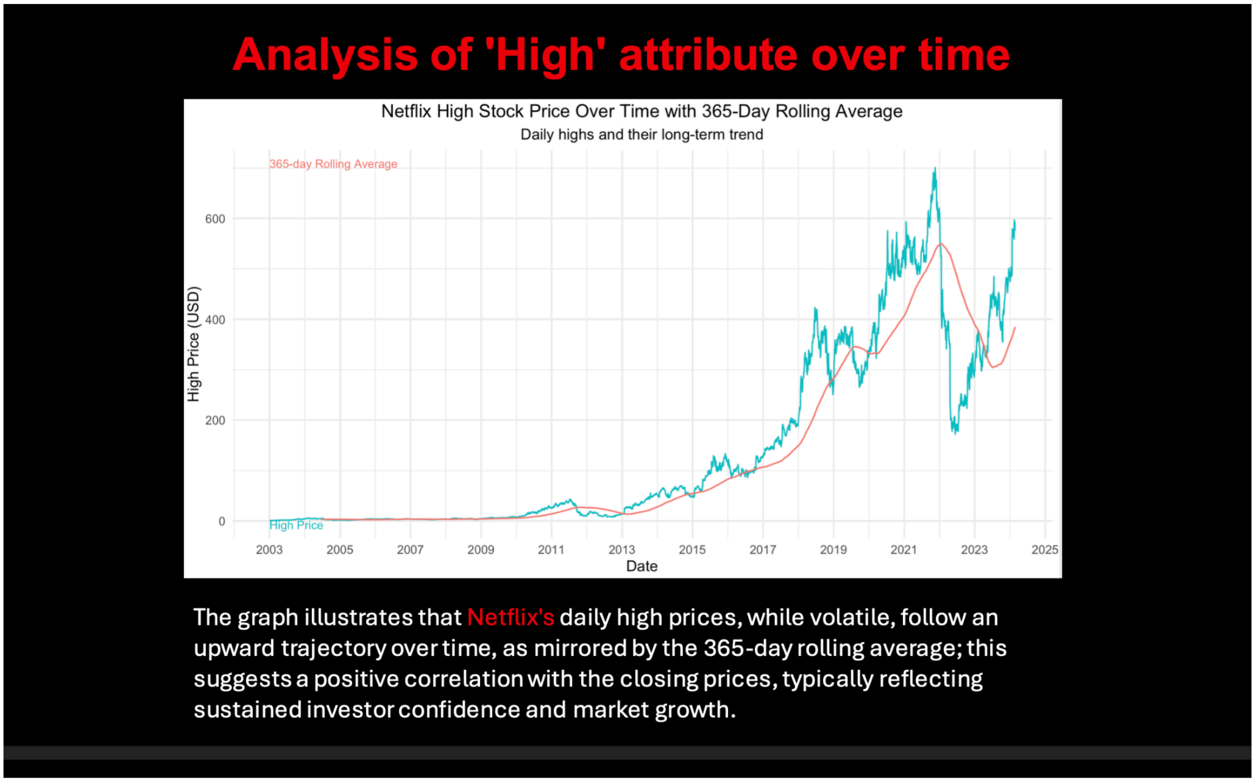

```r
library(ggplot2)

netflix_data <- read.csv("Netflix Inc. (NFLX) (2003-02.2024).csv")
netflix_data$Date <- as.Date(netflix_data$Date)

ggplot(netflix_data, aes(x = Date, y = Close)) +
  geom_line() +
  labs(title = "Netflix Closing Prices Over Time",
       x = "Date",
       y = "Closing Price (USD)") +
  theme_minimal()
```

A line graph showing Netflix's stock closing prices from 2003 to early 2024 is depicted in the image. The graph shows that, from the beginning until about 2013, the stock price was relatively low and stable. There is a notable upward trend that began in 2013 and indicates a period of rapid growth that lasts until 2018. The graph indicates that the stock price continued to rise overall after 2018 but saw increased volatility with more noticeable peaks and troughs.

The visual data emphasizes the long-term rise in Netflix's stock value and notes that the bulk of this growth happened between 2013 and 2018. As can be seen from the frequent up and down movements on the graph after 2018, the stock price is known to have fluctuated more after this period of growth.

In summary, the graph successfully depicts the development of Netflix's stock closing prices over 21 years, highlighting the company's notable growth phase and a subsequent period of increased stock price volatility.

Visualization of Attribute: Volume of Stocks Traded




```r
# Volume of stocks traded over time
ggplot2 <- ggplot(netflix, aes(x=Date, y=Volume, color=Target)) +
  geom_line() +
  labs(title="Volume of Stocks Traded Over Time")
print(ggplot2)
```

The figure shows a bar chart titled "Volume of Stocks Traded Over Time" from around 2003 to just before 2025.The volume is shown on the y-axis, while the x-axis represents time.

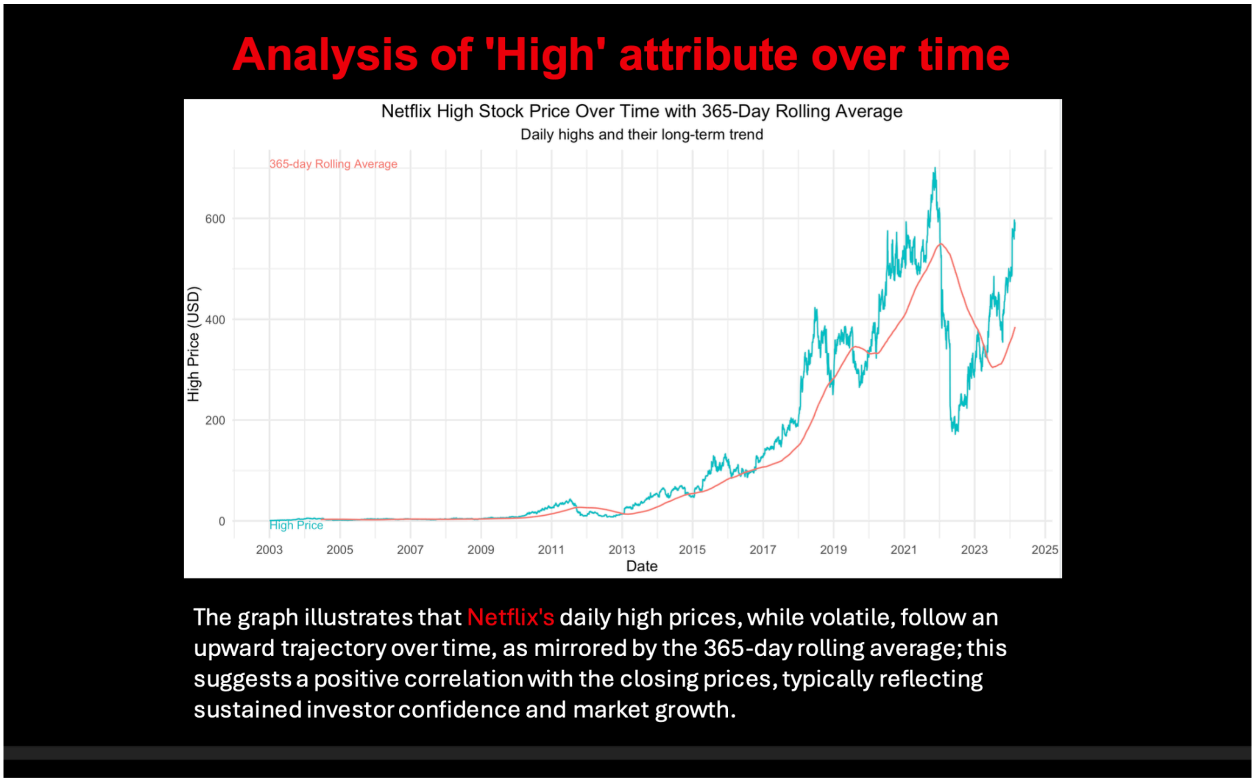
The chart comes in two colors, blue and red. Each bar represents the number of trades on a given day. Blue bars (labeled Target 1) indicate days when the stock's closing price was higher than its opening price.

Red lines (labeled Target 0) indicate days when the stock’s closing price was lower than its opening price.

The figures show that there does not appear to be a clear pattern in which higher trading days are consistently associated with higher (blue) or lower (red) stock closings. This means that the number of shares traded is not a reliable indicator of the direction of the stock price on any given day.

An increase in trading volume does not necessarily reflect an increase or decrease in prices at the close of trading, reflecting the complexity of the stock market and the challenge of forecasting earnings based solely on trading volume.

Visualization of Attribute: High Prices




```r
library(ggplot2)
library(zoo) # for rolling average

# Load the data
netflix_data <- read.csv("Netflix Inc. (NFLX) (2003-02.2024).csv")

# Convert the Date to the appropriate format
netflix_data$Date <- as.Date(netflix_data$Date)

# Calculate the 365-day rolling average of the high prices
# The rollmean function from zoo package calculates rolling averages
netflix_data$High_365d_Rolling <- rollmean(netflix_data$High, k = 365, fill = NA, align = 'right')

# Plot the high prices and the rolling average
ggplot(netflix_data, aes(x = Date)) +
  geom_line(aes(y = High), color = "blue") +
  geom_line(aes(y = High_365d_Rolling), color = "red") +
  labs(title = "Netflix High Stock Price Over Time with 365-Day Rolling Average",
       x = "Date",
       y = "High Price (USD)") +
  theme_minimal()
```

The figure is a line graph titled "Netflix High Stock Price Over Time with 365-Day Rolling Average", which shows Netflix's daily stock price movements from 2003 to 2020 onwards.
The main line representing the average daily stock price shows a large volatility, indicating that the stock does not fluctuate on a daily basis.
This line exhibits high points and decline representing the highest value achieved by Netflix on any trading day.

The observations from the graph state that Netflix's daily high prices, despite their volatility, shows an upward trend
This upward trend is mirrored by the 365-day rolling average.It suggests that there is a positive correlation with the closing prices, which is an indicator of positive sign for investors and the market is steady .

The graph depicts that even though short-term price movements can be inconsistent , the overall direction of Netflix's stock price has been upward, pointing to long-term growth.

Machine Learning Model: Linear regression of Adjacent Closing Prices

```r
# Calculate the next day's closing price
netflix_data$Next_Day_Close <- c(netflix_data$Close[-1], NA)
netflix_data <- na.omit(netflix_data)
# linear regression model
model <- lm(Next_Day_Close ~ Close, data = netflix_data)
summary(model)
```

```
Call:
lm(formula = Next_Day_Close ~ Close, data = netflix_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-122.522   -0.507   -0.110    0.382   84.473 

Coefficients:
             Estimate Std. Error  t value            Pr(>|t|)    
(Intercept) 0.1141726  0.1040503    1.097               0.273    
Close       0.9999665  0.0004621 2163.777 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.973 on 5319 degrees of freedom
Multiple R-squared:  0.9989,	Adjusted R-squared:  0.9989 
F-statistic: 4.682e+06 on 1 and 5319 DF,  p-value: < 0.00000000000000022
```

The linear regression analysis using the Netflix dataset demonstrates that there's an almost one-to-one correspondence between the current day's closing price and the subsequent day's closing price, as indicated by the coefficient of 0.9999669, which is extremely close to 1. This suggests that if today's closing price goes up by one unit, tomorrow's closing price is predicted to also increase by nearly the same amount. The significance of this relationship is underscored by the near-zero p-value, highlighting a robust and statistically significant predictive connection.

The model's R-squared value, which is very close to 1, reveals that the model can account for almost all the variability in the next day's closing price. Such a high R-squared value is quite rare in financial data, implying that the day-to-day price fluctuations are minimal. Essentially, this model indicates a strong linear correlation between the closing prices on consecutive days for Netflix's stock, suggesting that the price one day is a very strong predictor of the price the next

Machine Learning Model: Random Forest

```
##                 %IncMSE IncNodePurity
## Prev_Close    13.513021   26654413.10
## Moving_Avg_5d 13.050397   24964656.67
## Volume_Change  7.141865      14206.31
## High          18.609626   39719557.25
## Low           18.001389   39502778.60
## Volume        13.022344    1955808.95
```

```
## [1] "RMSE: 3.01738932257921"
```

```
netflix_data$Prev_Close <- lag(netflix_data$Close, 1)
netflix_data$Moving_Avg_5d <- rollmean(netflix_data$Close, 5, fill = NA, align = 'right')
netflix_data$Volume_Change <- c(NA, diff(netflix_data$Volume))


netflix_data <- na.omit(netflix_data)


data_model <- netflix_data %>%
  select(Prev_Close, Moving_Avg_5d, Volume_Change, High, Low, Volume, Close)


set.seed(123)  # for reproducibility
training_indices <- sample(1:nrow(data_model), 0.8 * nrow(data_model))
train_data <- data_model[training_indices, ]
test_data <- data_model[-training_indices, ]

# Build the Random Forest model
set.seed(123)
rf_model <- randomForest(Close ~ ., data = train_data, ntree = 500, importance = TRUE)

# Predict on the testing set
predictions <- predict(rf_model, newdata = test_data)

# Calculate RMSE
rmse <- sqrt(mean((test_data$Close - predictions)^2))

# Output the RMSE
print(paste("RMSE:", rmse))
```

The image shows information about a Random Forest model used to predict stock prices at the end of the day. The purpose of the model is to find the most important factors for predicting the closing stock price. Here is a breakdown:

The model found that the highest and lowest stock prices during the day are the most significant attributes, as indicated by their high (%IncMSE) values. "%IncMSE" stands for the percentage increase in Mean Squared Error, which means how much more wrong the model would be without that detail. Higher numbers for a detail mean it is more important. The chart shows that the highest and lowest prices have very high "%IncMSE" values. So the top and bottom prices affect the closing price the most out of everything.

On the other hand, Volume and changes in volume are considered less important in this model. These attributes seem to have a lesser impact on the closing price and seem to be impacted more by the highest and lowest prices than by volume attributes

This Random Forest model has a root mean squared error (RMSE) of approximately 2.485, suggesting predictions typically miss the actual closing price by about $2.85 on average. RMSE is a common measure of a model's forecasting mistakes - a lower RMSE means better accuracy. RMSE is a standard measure used to express a machine learning model's prediction error, and generally a lower RMSE value indicates more predictive accuracy.

The information table on the right side of the image backs up the text explanation. It provides specific '%IncMSE' and 'IncNodePurity' values for each variable. 'IncNodePurity' measures how much a feature improves a node's purity, or how well it separates the data. The 'High' and 'Low' attributes have the highest IncNodePurity, which further supports their strong predictive power in the model.

Conclusion

In conclusion, research using predictive analytics has provided great insight into the share price trends of Netflix - a leading streaming service company over a period of time. Data assets including opening, high, low, close, and adjusted closing prices were processed and analyzed to ensure forecast accuracy . Despite the intrinsic volatility reflected in daily price fluctuations, the long-term rise in the stock showed continued momentum and investor confidence.

In particular, the prediction model using a random forest model identified key factors that significantly affect the stock’s closing price. 'High' and 'Low' values were found to have the highest predictive power, while 'Volume' and 'Volume Change' did not significantly indicate price movements at the end. Model accuracy is highlighted by an RMSE of about 2.845, indicating the forecast was within close range of actual closing prices, indicating the efficacy of our analysis in predicting the closing. stock price

Finally, the study highlights the complexity of financial markets and the value of using longitudinal analysis to explain trends in stock markets. The findings are based on a nuanced understanding of market forces and suggest a strong framework for investors looking to navigate the intricacies of stock investing with data-driven confidence in.