SDA FINAL PROJECT

KADRI & ANDREI

18.03.2023

Project overview

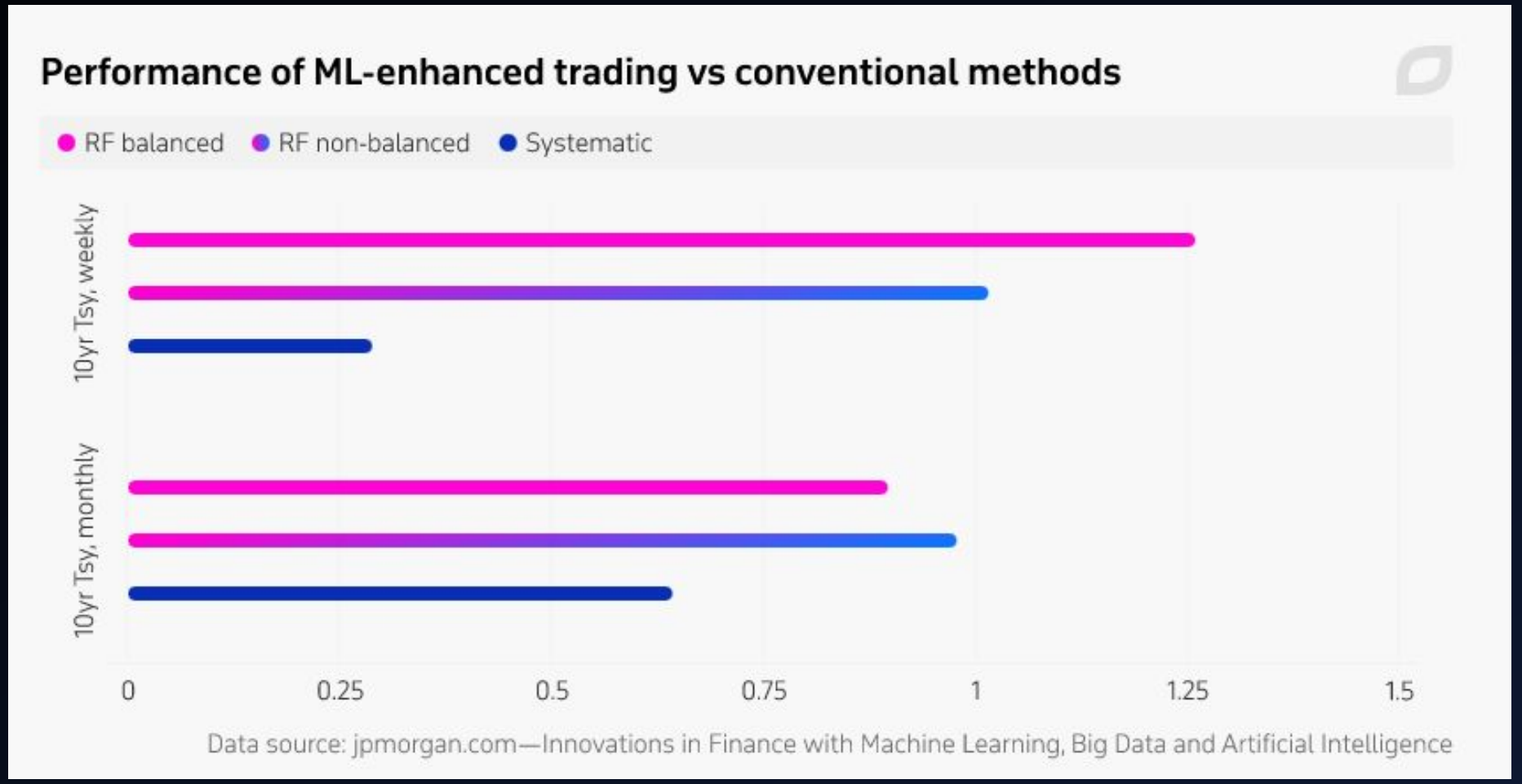
- Project name: Microsoft Stocks Price Prediction
- Project goal: Predict Microsoft stock price as accurate as possible using different models
- Dataset: Microsoft stock price from 05.03.2018 till 03.03.2023
  From Yahoo Finance

- **Github link**

Business problem solution

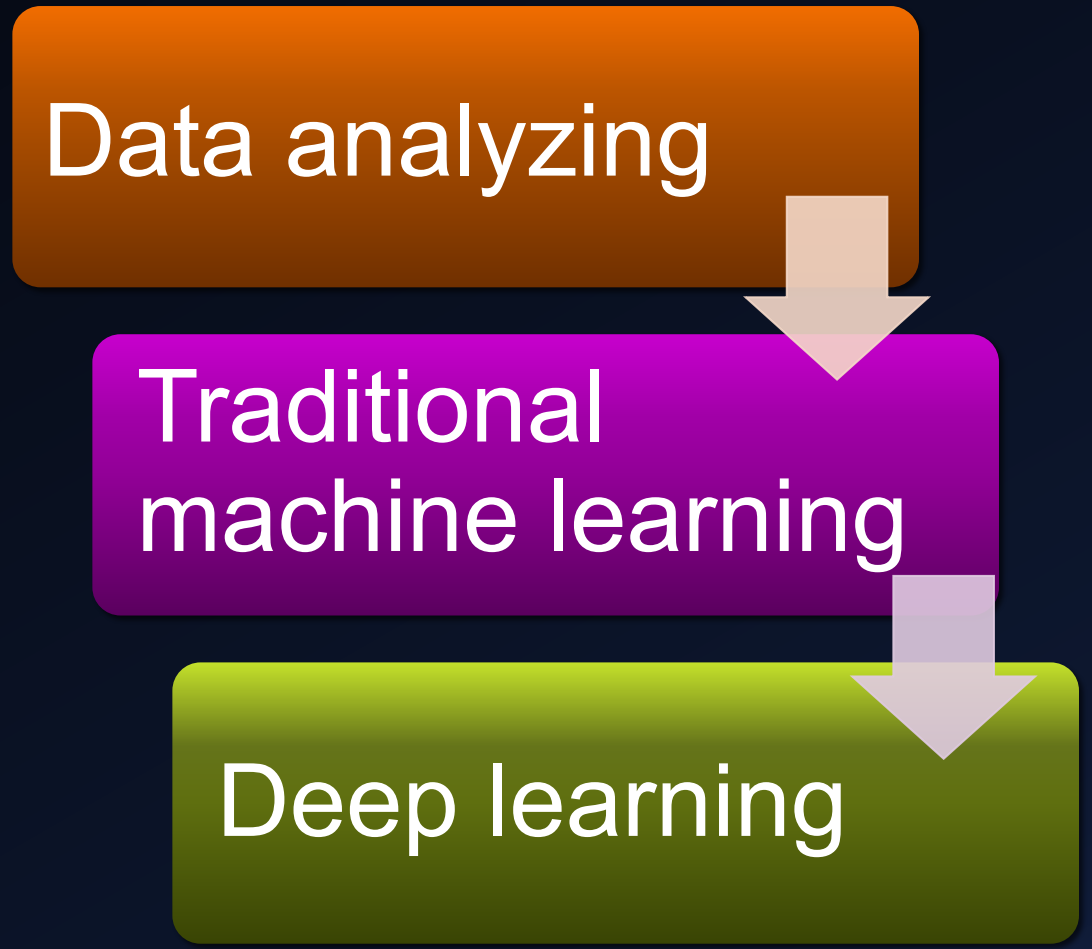
- Company’s stock analysts are spending time and resources on getting first evaluations on possible future stock prices of different companies.
- Having an automatized solution that gathers possible future evaluations on different stock prices will help analysts filter the stocks for deeper analysis.
- Algorithms can better analyze complex sets of historical data, discover hidden relationships between data sets, make forecasts, and learn along the way to become even more accurate.

Business problem solution



Project structure

- Processing of data set and data visualisation
- Traditional machine learning models building
- Deep learning models building and fine tunings



PROCESSING OF DATA SET AND DATA VISUALISATION

Dataframe info and preprocessing

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   Date       1259 non-null   object 
 1   Open       1259 non-null   float64
 2   High       1259 non-null   float64
 3   Low        1259 non-null   float64
 4   Close      1259 non-null   float64
 5   Adj Close  1259 non-null   float64
 6   Volume     1259 non-null   int64  
dtypes: float64(5), int64(1), object(1)
```

- No non-null values
- Date is an object, changed it correct form

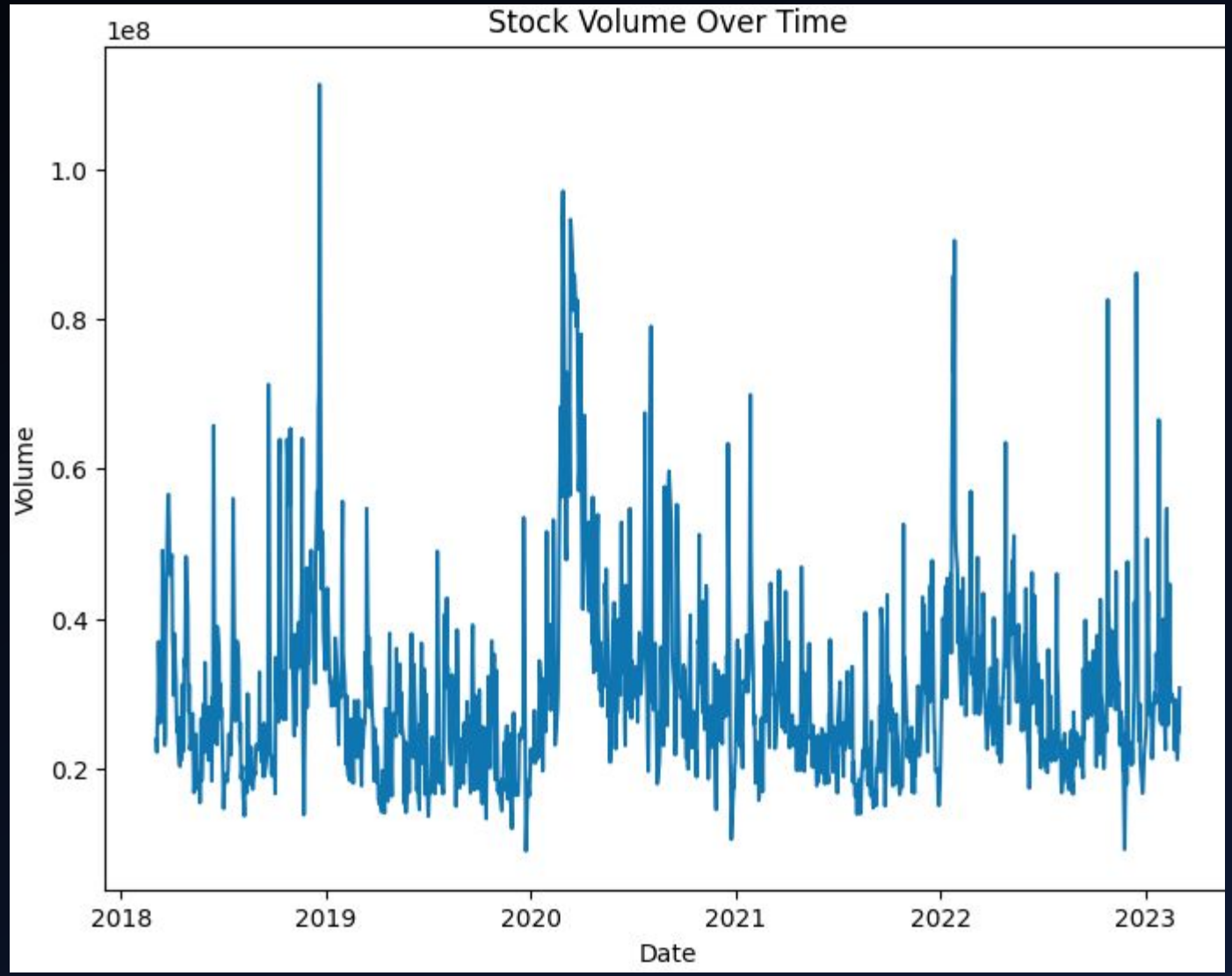
Date	Date	Open	High	Low	Close	Adj Close	Volume	Year	Month	Day	Weekday name
2018-03-05	2018-03-05	92.339996	94.269997	92.260002	93.639999	88.375061	23901600	2018	3	5	Monday
2018-03-06	2018-03-06	94.339996	94.489998	92.940002	93.320000	88.073044	22175800	2018	3	6	Tuesday
2018-03-07	2018-03-07	93.160004	93.940002	92.430000	93.860001	88.582703	26716100	2018	3	7	Wednesday
2018-03-08	2018-03-08	94.269997	95.099998	93.769997	94.430000	89.120636	25887800	2018	3	8	Thursday
2018-03-09	2018-03-09	95.290001	96.540001	95.000000	96.540001	91.112022	36937300	2018	3	9	Friday

- Added new columns: Year, Month, Day, Weekday name
- To increase data, analyze it more deeply

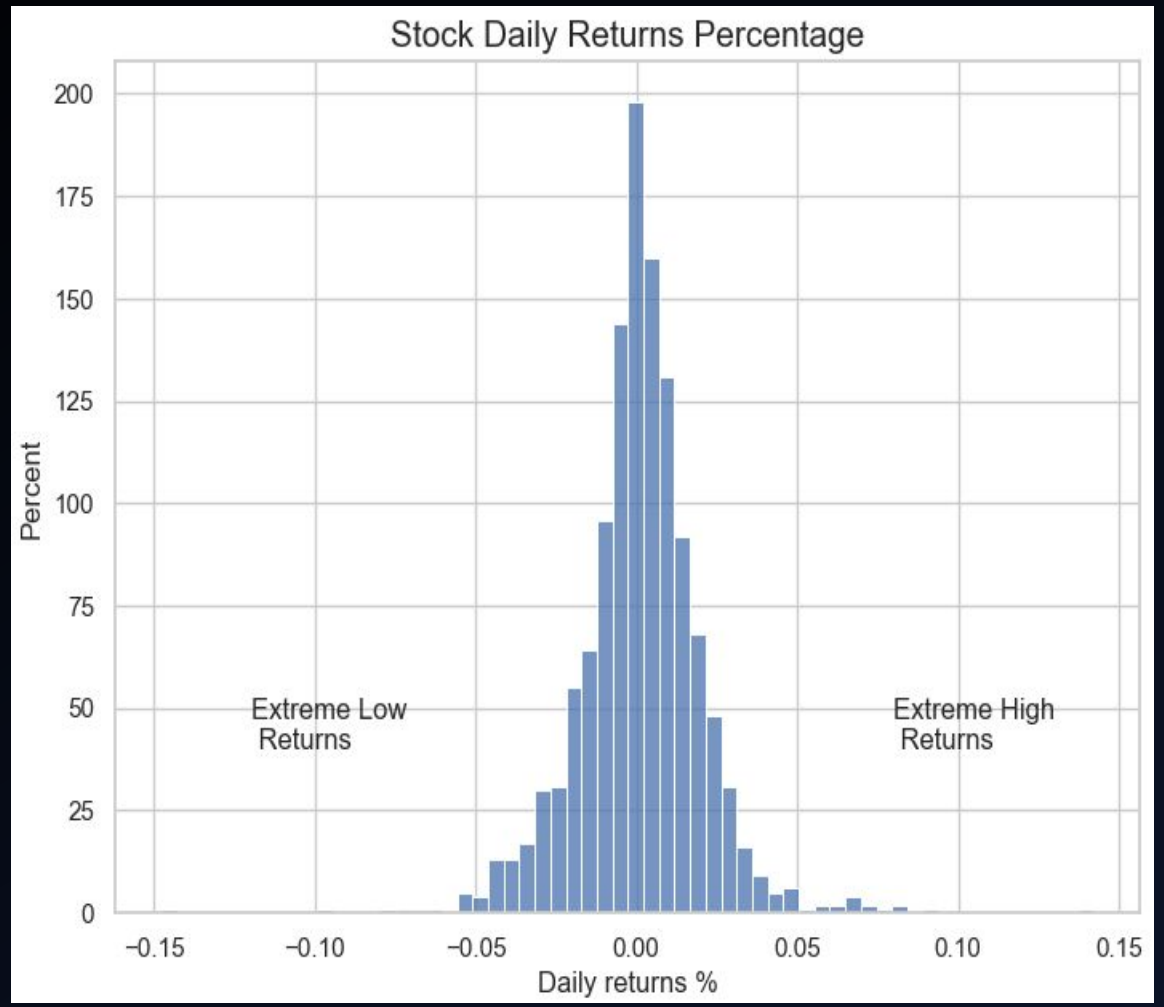
Data visualizations

Stock closing price overview for all periods of the dataset - future we will be predicting this parameter in our models

Checking the overall stocks volumes in the past 5 years

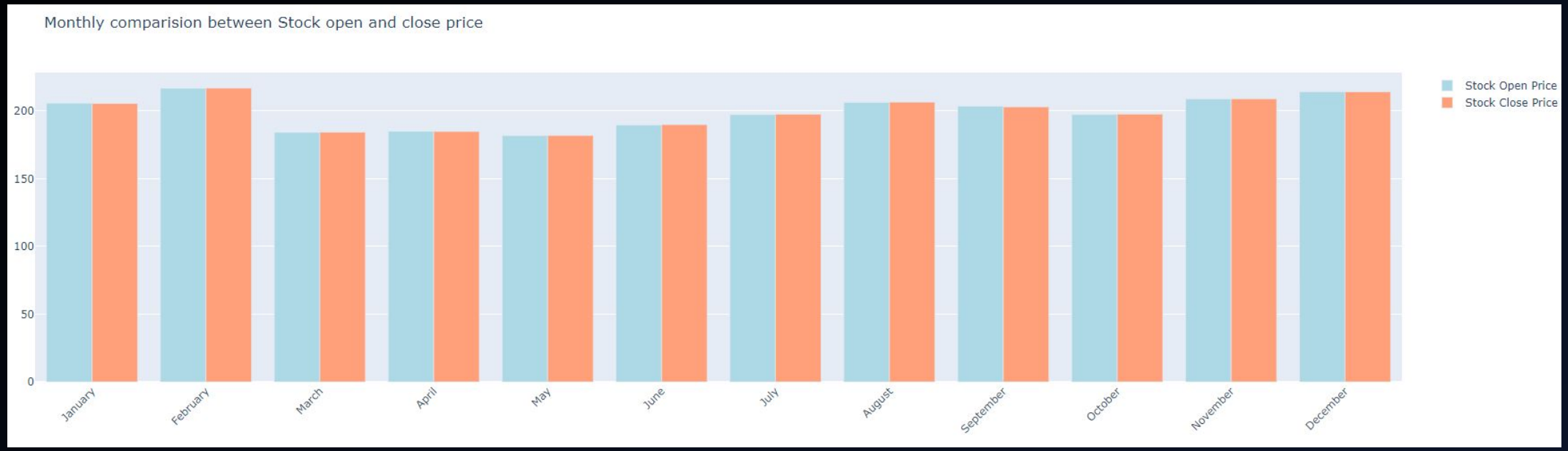


Checking stock daily returns percentage

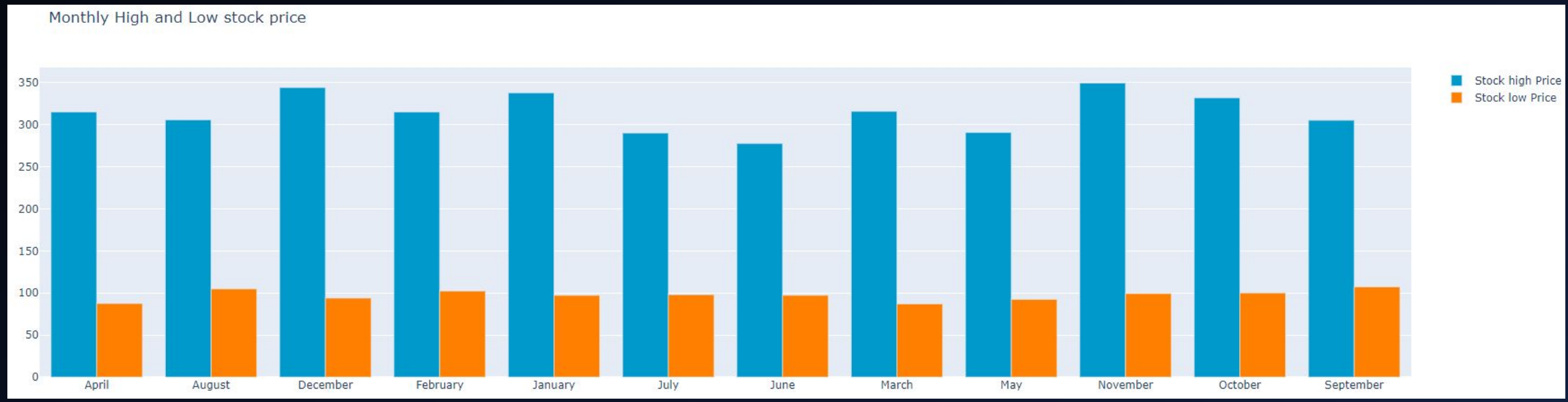


Comparing open and close prices for the all period

Checking the open/close mean price monthly comparison

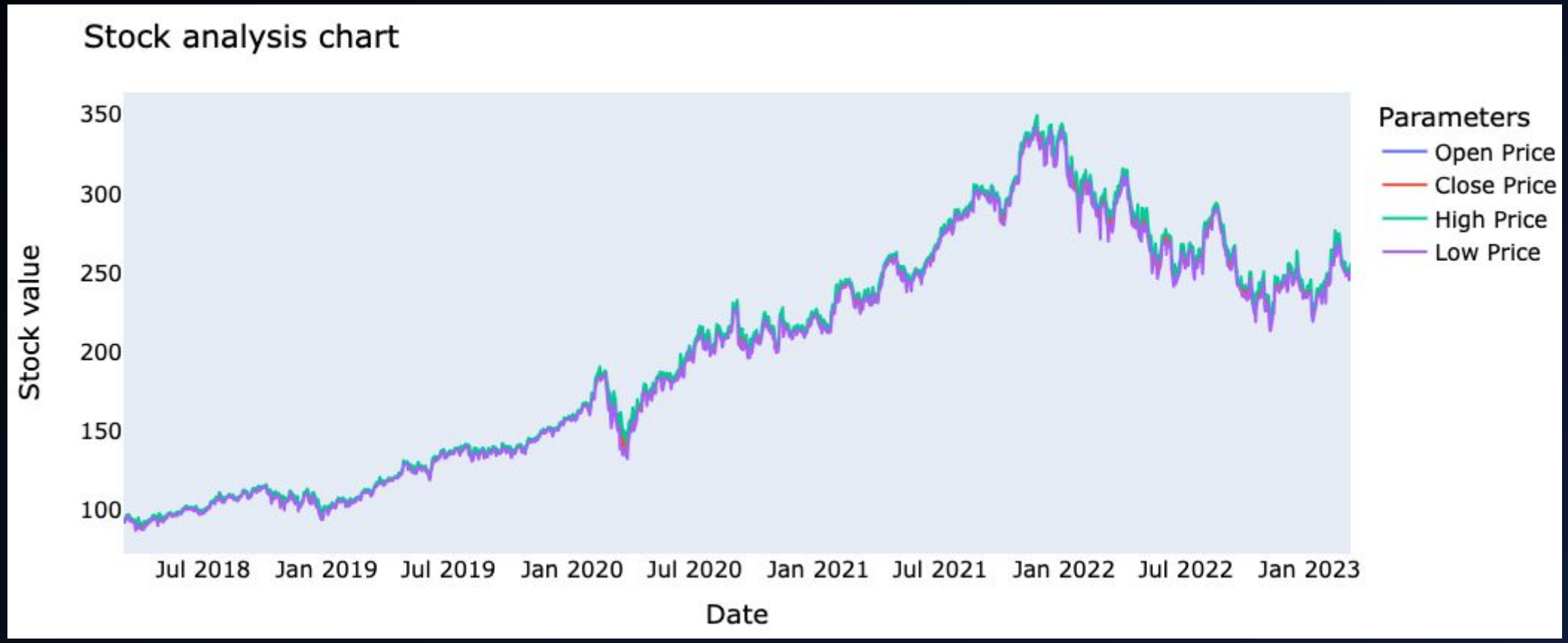


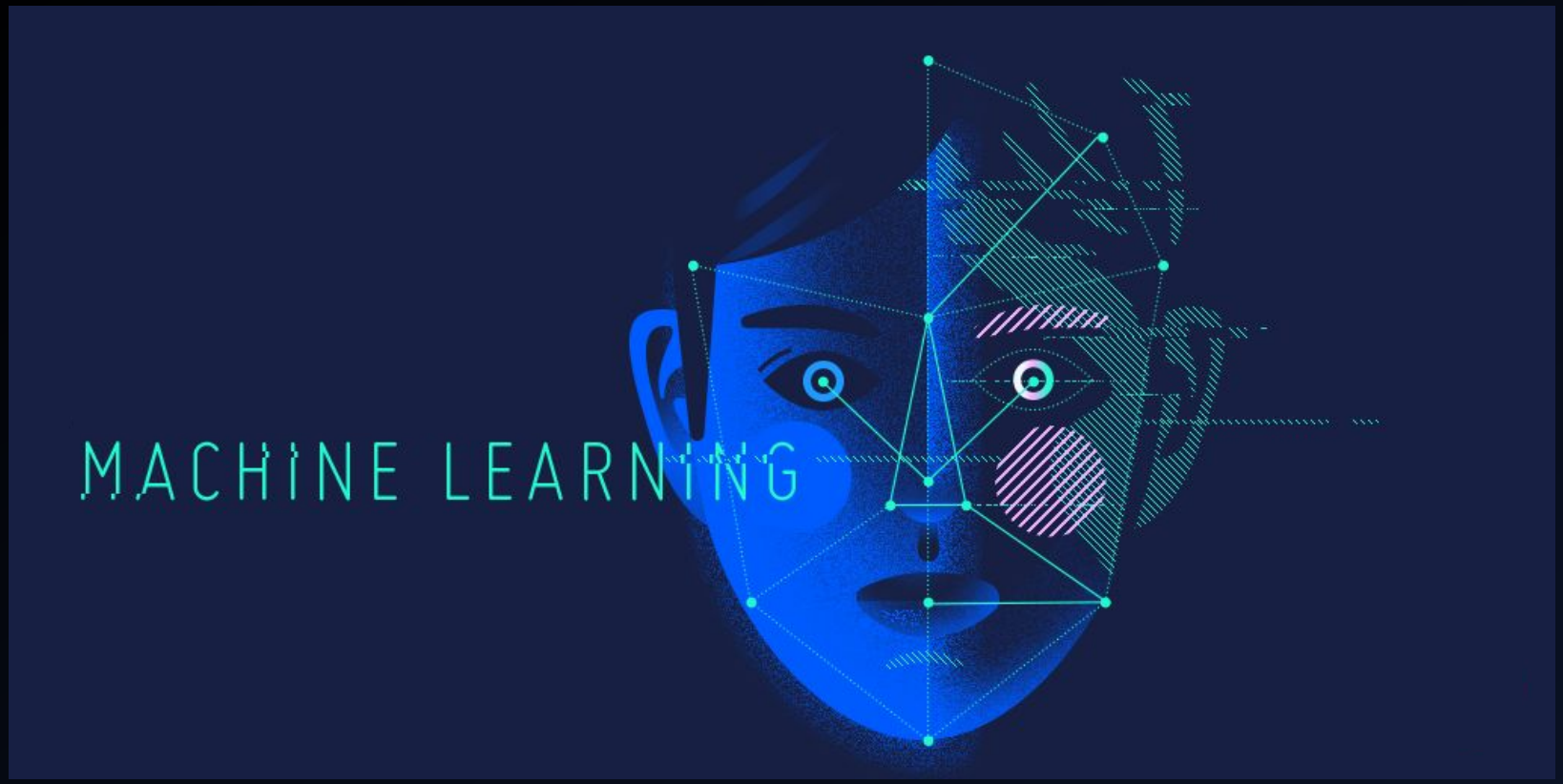
Checking the low/high price month comparison



STOCK ANALYSIS CHART

Trend comparison between stock open price, close price, high price and low price



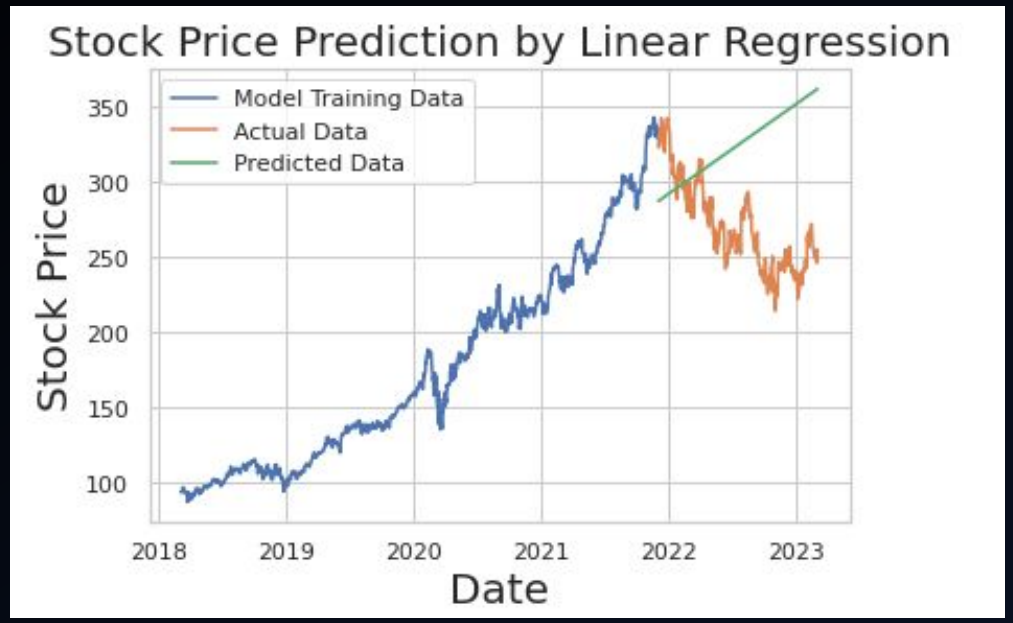


TRADITIONAL MACHINE LEARNING MODELS

Traditional machine learning

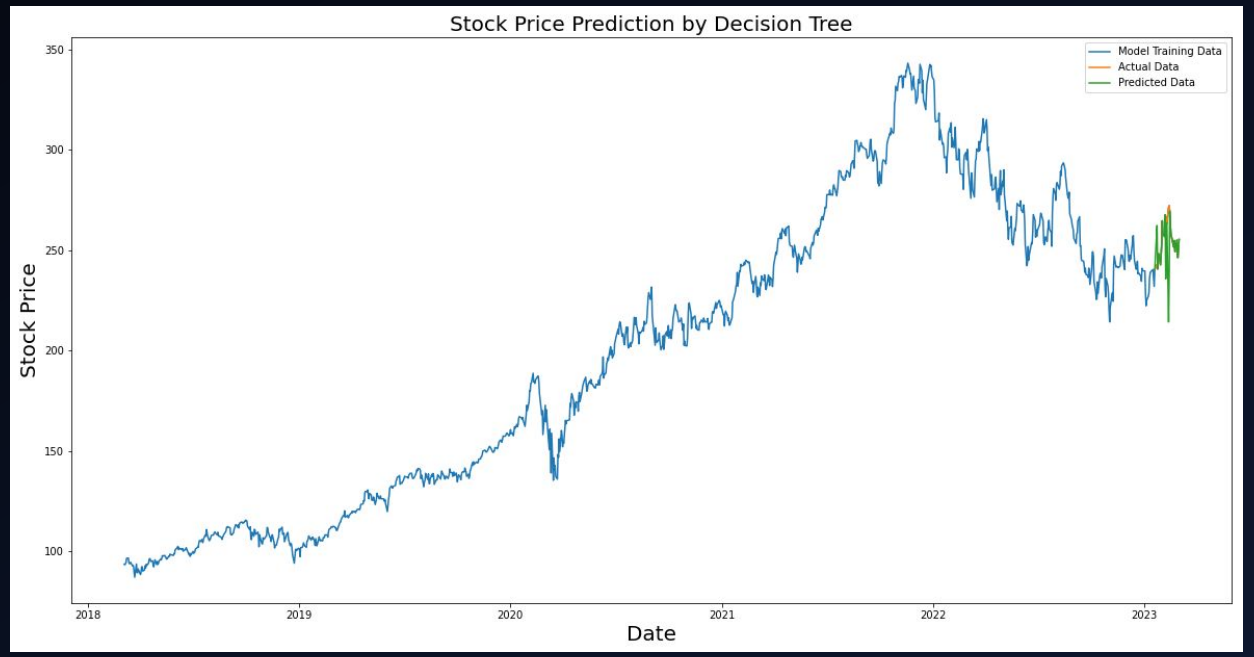
- **Linear Regression:** This algorithm is widely used for predicting numerical values and establishing relationships between variables. It works by finding the best-fit line that describes the relationship between the independent variables (input) and dependent variable (output) in a dataset.
- **Random Forest:** This algorithm is particularly effective at achieving high accuracy with large datasets and is commonly used in stock prediction for regression analysis, which involves identifying relationships among multiple variables.
- **Decision Tree:** Algorithm that recursively divides data into subsets based on significant features to create a tree-like model of decisions. It's commonly used for both regression and classification tasks and can handle both numerical and categorical data. While decision trees are easy to interpret, they may overfit noisy or complex data, leading to less accurate predictions compared to other algorithms.
- **K-nearest Neighbor:** This algorithm uses a computationally expensive, distance-based approach to predict the outcome of an event based on the records of the most similar historical situations, referred to as “neighbors.”

Linear Regression



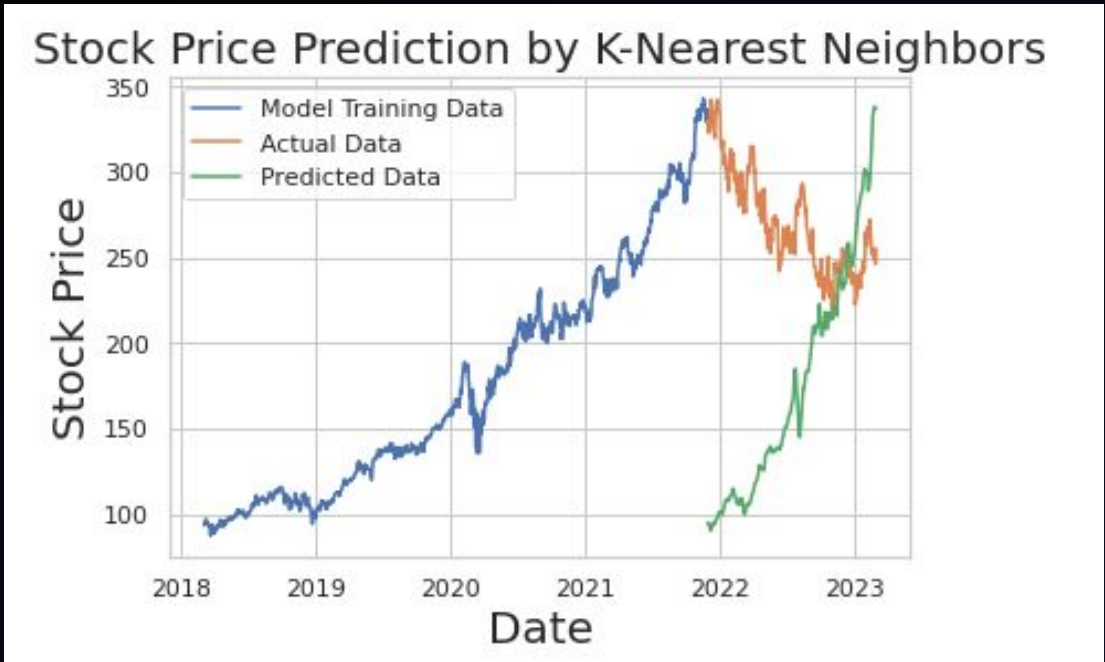
- Root Mean Squared Error: 65.4910
- Mean Squared Error: 5358.71745
- R^2 score: -5.0737

Decision Tree

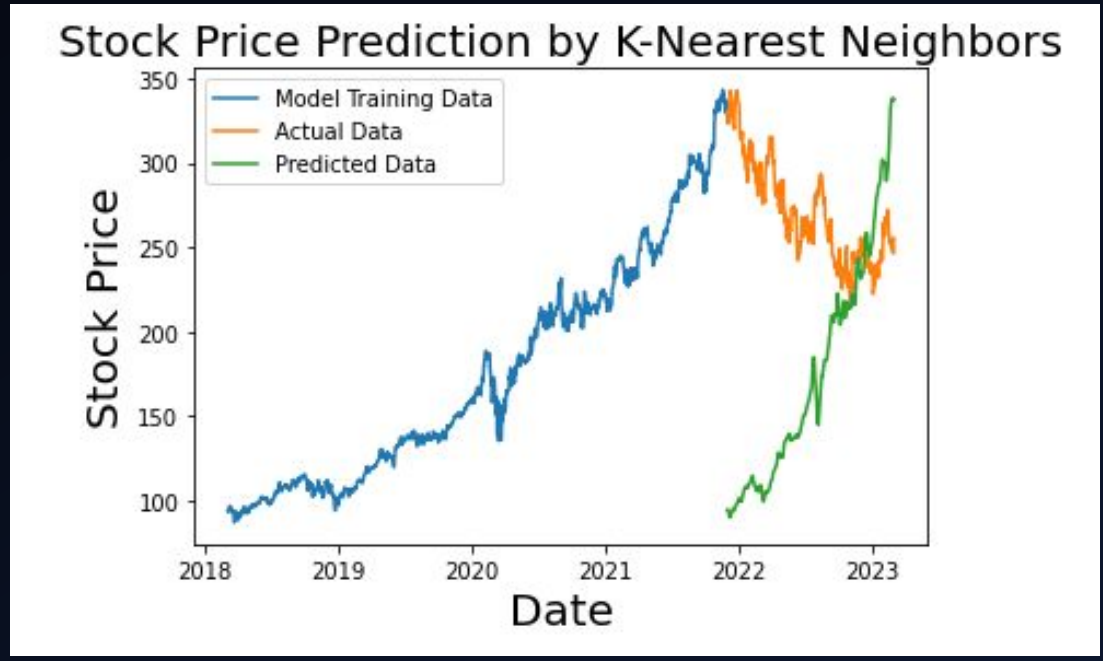


- Root Mean Squared Error: 22.6996
- Mean Squared Error: 515.27069
- (R^2) Score: 0.9041

K-Nearest Neighbours



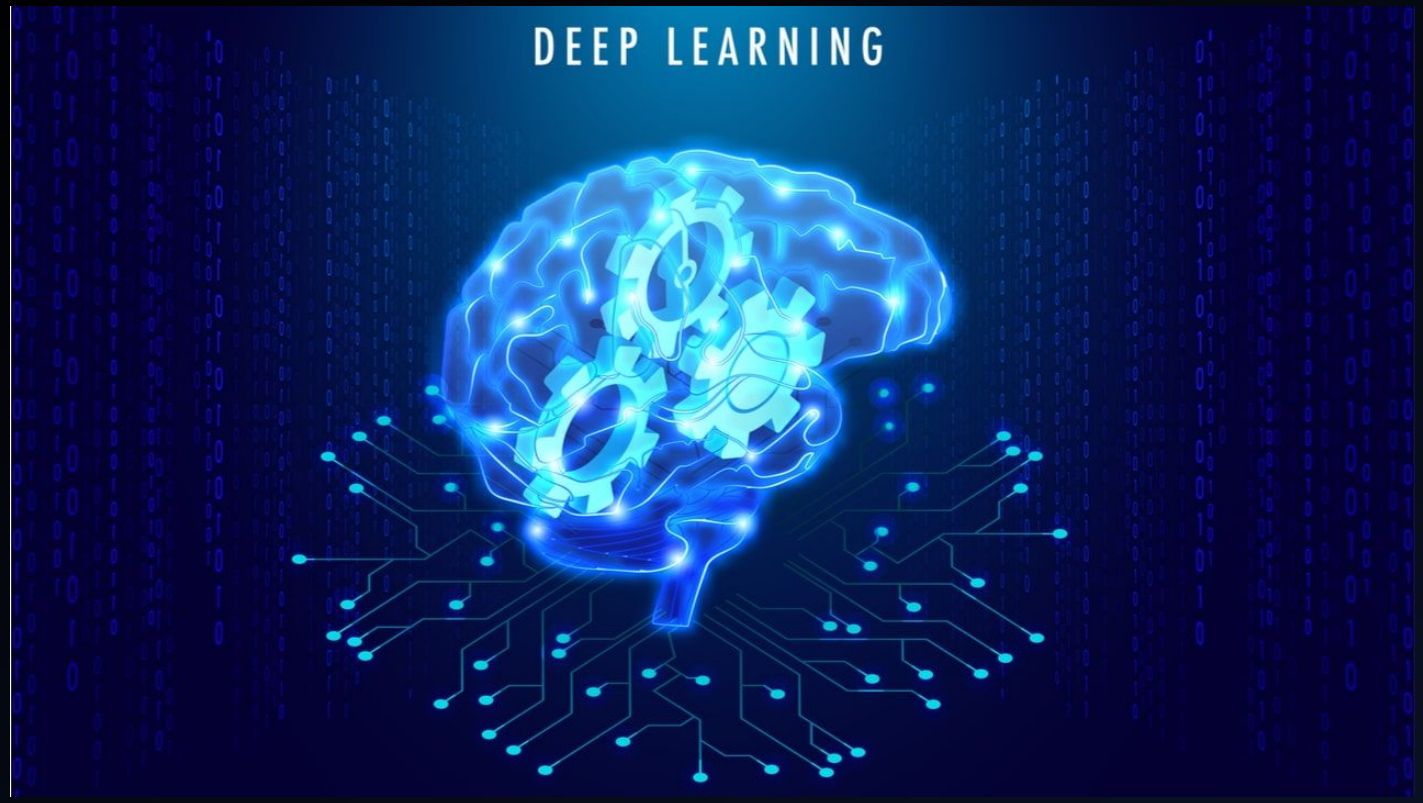
- Root Mean Squared Error: 132.0688
- Mean Squared Error: 17442.17266
- R^2 score: -18.5818



- Root Mean Squared Error: 132.1640
- Mean Squared Error: 17467.3337
- R^2 score: -18.559

Random Forest

- Mean Absolute Error: 2.427
- Mean Squared Error: 11.7059
- Root Mean Squared Error: 3.4214
- (R^2) Score: 0.9978
- Train Score : 99.93% and Test Score : 99.78% using Random Tree Regressor.
- Accuracy: 98.79 %.

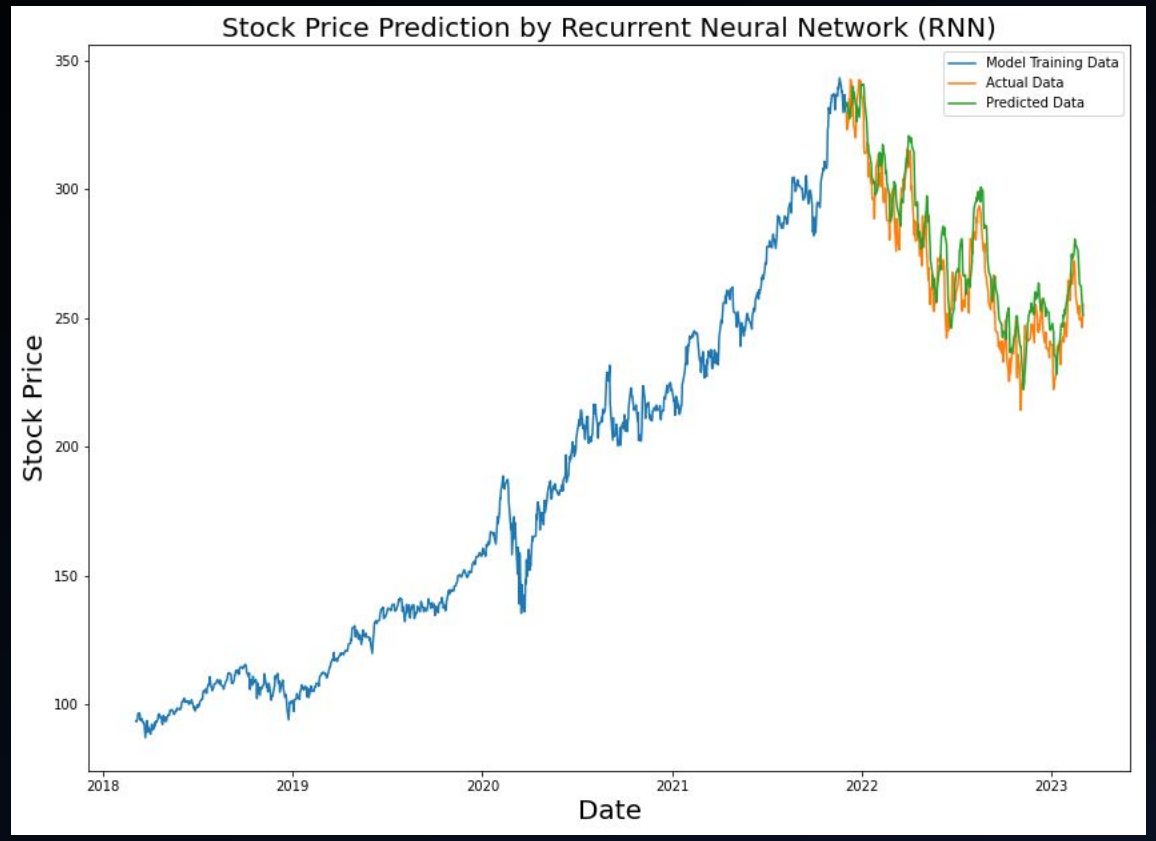


DEEP LEARNING MODELS

Deep learning models

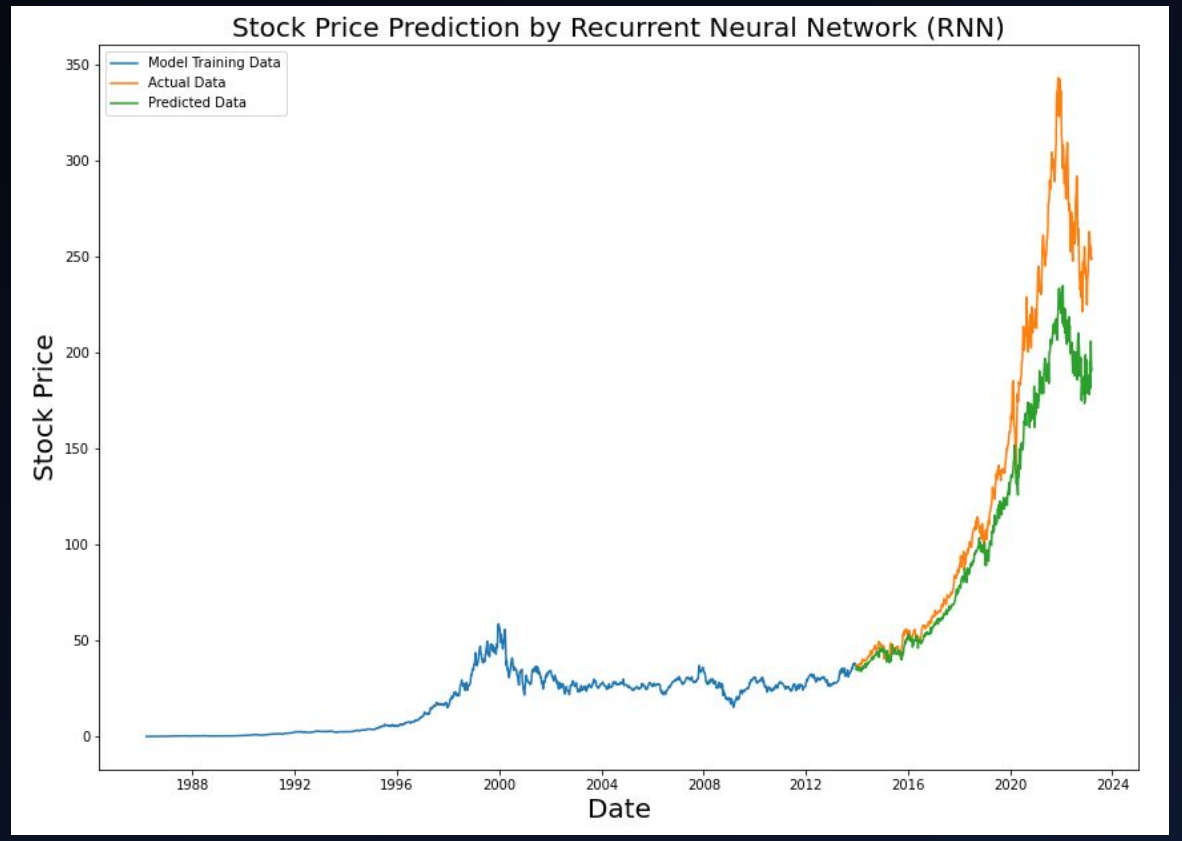
- **Long short-term memory (LSTM):** Many experts currently consider LSTM as the most promising algorithm for stock prediction. It's a type of RNN, but it can process both individual data points and more complex sequences of data, making it well-suited to handle non-linear time series data and predict highly volatile price fluctuations.
- **Recurrent neural networks (RNN):** A specific type of ANN where each processing node also functions as a "memory cell," enabling it to retain relevant information for future use and send it back to previous layers to improve its output.

RNN Model



Optimizer = Adam, Epochs = 10

- Root Mean Squared Error: 20.7347
- R^2 score: 0.5127



Optimizer = Adam, Epochs = 20

- Root Mean Squared Error: 39.7099
- R^2 score: 0.8055

LSTM

Optimizer = Adam, Epochs = 10

- Root Mean Squared Error: 11.4625
- R^2 score: 0.85108
- Max Error: 26.6943
- Variance: 0.8625

Optimizer = Adam, Epochs = 10

- Root Mean Squared Error: 29.4447
- R^2 score: 0.8931
- Max Error: 99.1555
- Variance: 0.9367

Conclusion

- Based on ML models results it can be concluded that the Random Forest model performs the best.
- In comparison, the Linear Regression model has a very high RMSE and a negative R^2 score, indicating a poor fit for the data. The K-Nearest Neighbors model also performs poorly, with a high RMSE and a negative R^2 score.
- The Decision Tree has a lower RMSE than Linear Regression and K-Nearest Neighbors, but it still performs worse than the Random Forest. Decision Tree has lower R^2 score of the Random Forest.
- Deep learning models have better performance rates, although need to consider that the models are overfitting
- The results showed LSTM model outperformed the RNN model.
  The LSTM model has lower RMSE and maximum error values for both datasets. Also, LSTM model has higher R^2 scores, indicating that it explains a greater portion of the variance in the target variable.
- Further research is needed to refine the models and to better understand their limitations and potential use cases.

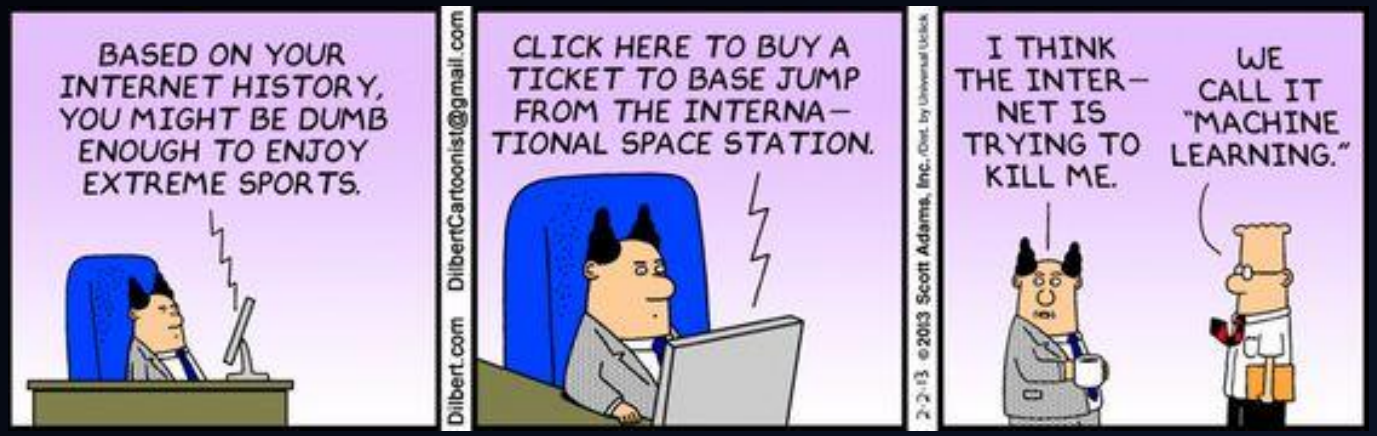
Challenges

- Overfitting
- Limited predictability: The stock market is influenced by a multitude of factors. It's challenging to predict the stock market's behavior accurately, and predictions may not always be correct.
- Ethical considerations: It may raise ethical concerns related to the potential impact on investors, the economy, and society.

Summary

- The LSTM model can be adjusted by modifying parameters.
- However, relying solely on LSTM predictions may not be sufficient to determine whether stock prices will rise or fall. External factors can have a significant impact on stock prices and are often impossible to predict.
- Time series forecasting is a fascinating field! While there is truth to the notion that it can be difficult, mastering the fundamental techniques can make it easier.
- We are curious about how LSTM performs on other types of time series problems and encourage you to experiment with it yourself!

THANK YOU!


BASED ON YOUR INTERNET HISTORY, YOU MIGHT BE DUMB ENOUGH TO ENJOY EXTREME SPORTS.
Dilbert.com DilbertCartoonist@gmail.com
CLICK HERE TO BUY A TICKET TO BASE JUMP FROM THE INTERNA-TIONAL SPACE STATION.
2-2-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick
I THINK THE INTER-NET IS TRYING TO KILL ME.
WE CALL IT "MACHINE LEARNING."