PREDICTING THE FUTURE:
WHERE WILL STOCKS, CRYPTOS, AND COMMODITIES STAND BY 2026?

UNIVERSITY OF ARIZONA
COLLEGE OF INFORMATION SCIENCE

INFO 511 - FOUNDATIONS OF DATA SCIENCE

TEAM BYTE-SIZED INSIGHTS
RAFI LEVITON
PRAJWAL SATHYANARAYANA
SHASHWAT SINGH
VISHAL BHASHYAAM
LIKHITH RAMESH

INTRODUCTION

The financial markets have always been a cornerstone of economic activity, influencing investment decisions and driving global commerce. Their dynamic nature, encompassing stocks, cryptocurrencies, and commodities, presents opportunities and challenges for investors and institutions. The ability to predict future asset prices is a critical skill for financial analysts, enabling them to make informed decisions that can significantly impact their organizations' investment portfolios.

This project seeks to forecast the prices of major stocks, cryptocurrencies, and commodities by the end of 2026. By leveraging a dataset that spans daily price and volume information from 2020 to 2024, this study will examine historical trends and apply predictive modeling techniques to estimate future asset performance.

The analysis will use a linear regression model to capture the relationship between asset prices and time. Key evaluation metrics, including Mean Squared Error (MSE) and R-squared values, will ensure the accuracy and reliability of the predictions. This study also aims to test the widely accepted notion that the stock market grows at an average annual rate of 10% while comparing the performance of traditional assets, such as stocks, with emerging ones like cryptocurrencies and commodities.

Through this project, we aspire to contribute actionable insights to the financial domain, empowering stakeholders to navigate the complexities of the market confidently.

LITERATURE REVIEW

ASPECT	INSIGHTS	REFERENCES
Stock Market Growth	The stock market historically exhibits a consistent average annual return of ~10%	Sullivan, B. (2024, July 30). Average stock market return. Forbes. Retrieved from https://www.forbes.com/advisor/investing/average-stock-market-return/
Cryptocurrencies	High-risk, high-reward investments are influenced by macroeconomic factors and market sentiment.	Observations based on trends in the cryptocurrency market (2020–2024).
Commodities	Commodities like gold, oil, and natural gas serve as benchmarks for stability and hedging.	Observations from commodity markets (2020–2024).
Predictive Modeling	Linear regression is widely used for financial forecasting, with MAE and MSE as key metrics.	Chugh, A. (2020, December 8). MAE, MSE, RMSE, coefficient of determination, adjusted R squared - which metric is better? Medium. Retrieved from https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e
Model Evaluation Metrics	MAE focuses on average residuals; MSE penalizes larger errors. R-squared measures variation.	Chugh, A. (2020, December 8). MAE, MSE, RMSE, coefficient of determination, adjusted R squared - which metric is better? Medium. Retrieved from https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e
Emerging Asset Comparison	Comparative analysis between stocks, cryptocurrencies, and commodities is essential for strategy.	Sullivan, B. (2024, July 30). Average stock market return. Forbes. Retrieved from https://www.forbes.com/advisor/investing/average-stock-market-return/

METHODS

I. DATA

1. Data Information:

The dataset utilized in this project, titled "US Stock Market and Commodities Data (2020-2024)", comprises daily records of major stocks, cryptocurrencies, and commodities. It contains essential columns, including:

- Date: The trading date for each asset.
- Price: The opening price of the asset on the corresponding date.
- Volume: The number of units traded on that day.

This dataset provides a robust foundation for understanding historical trends and evaluating the performance of these assets over time.

2. Facts About the Data

- Period: January 2, 2020, to February 2, 2024.
- Assets Covered:
    - Stocks: Notable companies like Apple, Nvidia, Google, Meta, and indices like the S&P 500.
    - Cryptocurrencies: Bitcoin & Ethereum.
    - Commodities: Natural gas, crude oil, gold, silver, platinum, and copper.
- Key Metrics:
    - Price trends are reflective of economic events (e.g., the COVID-19 pandemic's market impact).
    - Volume changes highlight shifts in trading activity.

3. Key Insights in Data

- Stock Market Growth: A general upward trend aligns with the long-term market growth theory (~10% annual return).
- Cryptocurrency Volatility: High fluctuations indicate the potential for large gains or losses, necessitating detailed modeling.
- Commodities Trends: Steady but susceptible to geopolitical events and natural resource supply-demand dynamics.
- Volume vs. Price Relationship: Initial observations suggest a potential correlation between trading volume and price changes for some assets.

II. DATA WRANGLING

- Date Formatting: The date columns contain inconsistent values (2/2/2024 and 31-01-204). We standardized the formatting of the date columns and converted it to the correct type, datetime.
- Date Ordinal: A column named data_ordinal was created which numbers each date in the dataset from 0-1490. This column enables the creation of a linear regression model between asset prices and the ordinal date since a regression model requires numeric inputs.
- Removal of dates: The days representing February 1st and 2nd were removed from the data so the dates span from January 2nd, 2020 through January 31st, 2024.

III. EXPLORATORY DATA ANALYSIS

For exploratory data analysis, five functions were created that can be used to analyze and visualize asset prices over time. We will consider the Apple stock as an example for each function shown below.

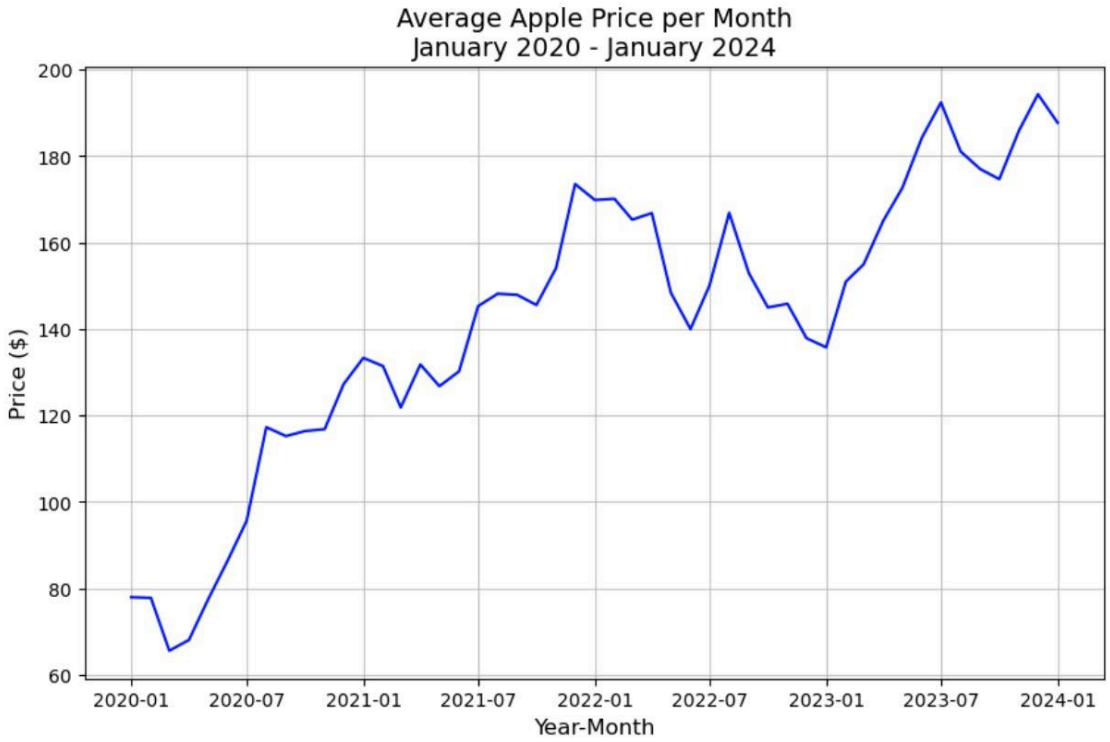
- **Asset summary stats:** This function takes an asset as an input and displays its highest, lowest, and average prices from the time period of the dataset. It also displays the dates where the highest and lowest prices occurred.

```
asset_summary_stats('Apple_Price')
✓ 0.0s

Summary of Apple Prices
January 2, 2020 - January 31, 2024

    Statistic  Apple Price ($)       Date
Highest Price           198.11 2023-12-14
 Lowest Price            56.09 2020-03-23
Average Price           141.88        NaT
```

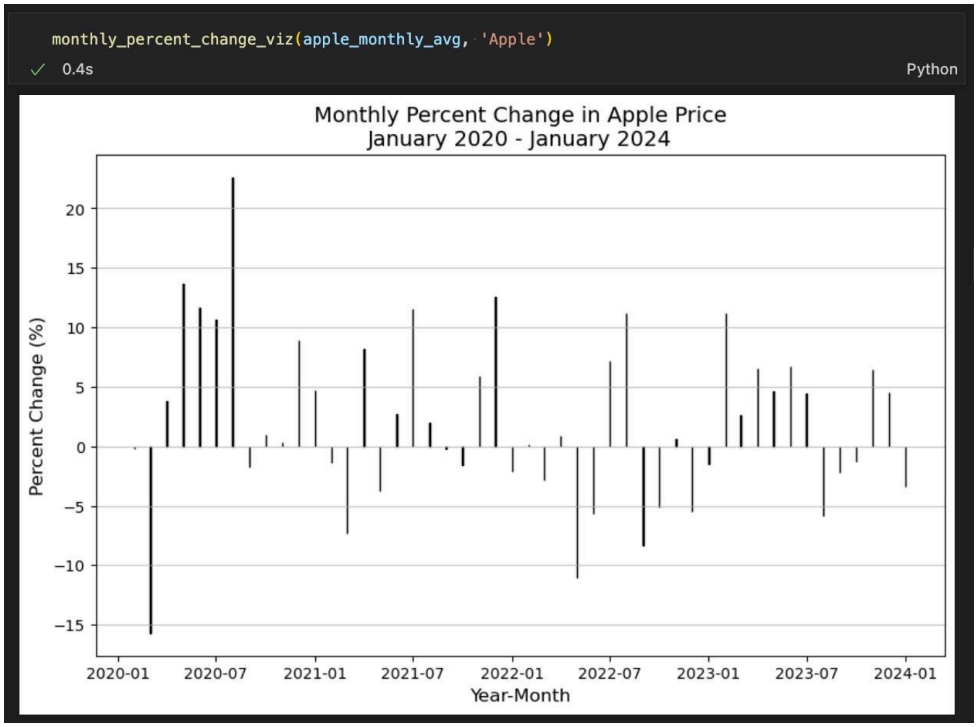
- **Avg price per month:** This function takes an asset as an input and finds the average price of that asset for each month in the dataset. It also creates a column called percent_change that calculates the price the asset has changed since the previous month and a binary column named positive_month that indicates if the asset price has increased that month. The function outputs the number of months where the asset price has increased and decreased as well as the monthly percent return.

```
avg_price_per_month('Apple_Price')
✓ 0.0s

   Year  Month  Avg_Price  Percent_Change Year_Month  Positive_Month
0  2020      1      77.98             NaN 2020-01-01               0
1  2020      2      77.82           -0.21 2020-02-01               0
2  2020      3      65.61          -15.69 2020-03-01               0
3  2020      4      68.10            3.79 2020-04-01               1
4  2020      5      77.41           13.67 2020-05-01               1

Number of months where Apple_Price increased: 28

Number of months where Apple_Price decreased: 20

The average monthly percent return of the Apple_Price between January 2020 and January 2024' is 2.09%
```

- **Avg price per month viz:** This function takes a data frame containing asset prices and an asset name and creates a line plot displaying the monthly average prices for that asset.

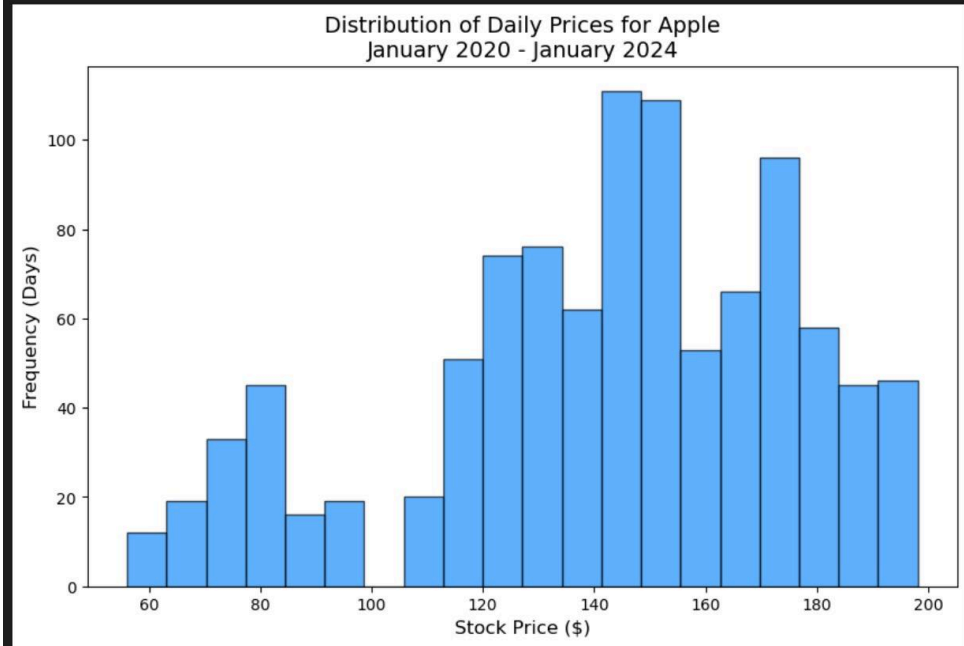


- **Monthly percent change viz:** This function takes a data frame containing assets and an asset name as an input and creates a bar plot displaying the monthly percent changes for that asset.



- **Price distribution viz:** This function takes a data frame containing assets and an asset as an input and creates a histogram displaying the frequency of monthly average prices for that asset.

```
price_distribution_viz(assets, 'Apple_Price')
✓ 0.1s                                                    Python
```



IV. REGRESSION MODEL

```
X = apple[['Date_Ordinal']]
y = apple['Apple_Price']

apple_model = LinearRegression()
apple_model.fit(X, y)

y_pred = apple_model.predict(X)
mse = round(mean_squared_error(y, y_pred), 2)
r2 = round(r2_score(y, y_pred),4)

print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')
✓ 0.0s
Mean Squared Error: 266.98
R-squared: 0.7654
```

- Linear Regression model fitted against the data and calculation of Mean Squared Error and R-squared for the Apple stock

```
X = sm.add_constant(X)

apple_model = sm.OLS(y, X).fit()

summary = apple_model.summary()
print(summary)
```

✓ 0.1s

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Apple_Price   R-squared:                       0.765
Model:                            OLS   Adj. R-squared:                  0.765
Method:                 Least Squares   F-statistic:                     3293.
Date:                Mon, 09 Dec 2024   Prob (F-statistic):          5.65e-320
Time:                        20:47:44   Log-Likelihood:                -4258.9
No. Observations:                1011   AIC:                             8522.
Df Residuals:                    1009   BIC:                             8532.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          90.8355      1.028     88.402      0.000      88.819      92.852
Date_Ordinal    0.0684      0.001     57.382      0.000       0.066       0.071
==============================================================================
Omnibus:                       12.786   Durbin-Watson:                   0.027
Prob(Omnibus):                  0.002   Jarque-Bera (JB):                8.904
Skew:                          -0.099   Prob(JB):                       0.0117
Kurtosis:                       2.585   Cond. No.                     1.72e+03
==============================================================================
```

Ordinary Least Square (OLS) Regression Results

- The model indicates that 76.5% of the variance in Apple's stock price can be explained by the date, with an intercept of $90.84 on January 2, 2020.
- The stock price shows a positive trend, increasing by approximately $0.07 per day, as evidenced by significant p-values (less than 0.05).
- The 95% confidence interval for the coefficient suggests a daily increase between $0.066 and $0.071, with the intercept for January 2, 2020, estimated to be between $88.85 and $92.87.

V. RESULTS

```
ordinal_day_to_predict = 502
print("Day 502 Date:", apple.loc[apple['Date_Ordinal'] == ordinal_day_to_predict, 'Date'].dt.strftime('%Y-%m-%d').values[0])

predicted_price = round(intercept + (slope * ordinal_day_to_predict),2)
print("Predicted Price of Apple Stock on May 18th, 2021 (Day 502): " + "$" + str(predicted_price))

actual_day502_price = apple.loc[apple['Date_Ordinal'] == 502, 'Apple_Price'].values[0]
print("Actual Price of Apple Stock on May 18th, 2021 (Day 502): " + "$" + str(actual_day502_price))
```

✓ 0.0s

```
Day 502 Date: 2021-05-18
Predicted Price of Apple Stock on May 18th, 2021 (Day 502): $125.17
Actual Price of Apple Stock on May 18th, 2021 (Day 502): $124.85
```

Predicted stock price V/S Actual stock price on day 502

```
start_date = datetime(2020, 1, 2)

date_1600_days_later = start_date + timedelta(days=1600)
print(date_1600_days_later)

ordinal_day_to_predict = 1600
predicted_price = round(intercept + (slope * ordinal_day_to_predict),2)
print("Predicted price of Apple stock on May 20th, 2024 (Day 1600): " + '$' + str(predicted_price))
print("Actual Apple stock price on May 20th, 2024 (Day 1600): $191.04")
```

✓ 0.0s

```
2024-05-20 00:00:00
Predicted price of Apple stock on May 20th, 2024 (Day 1600): $200.25
Actual Apple stock price on May 20th, 2024 (Day 1600): $191.04
```

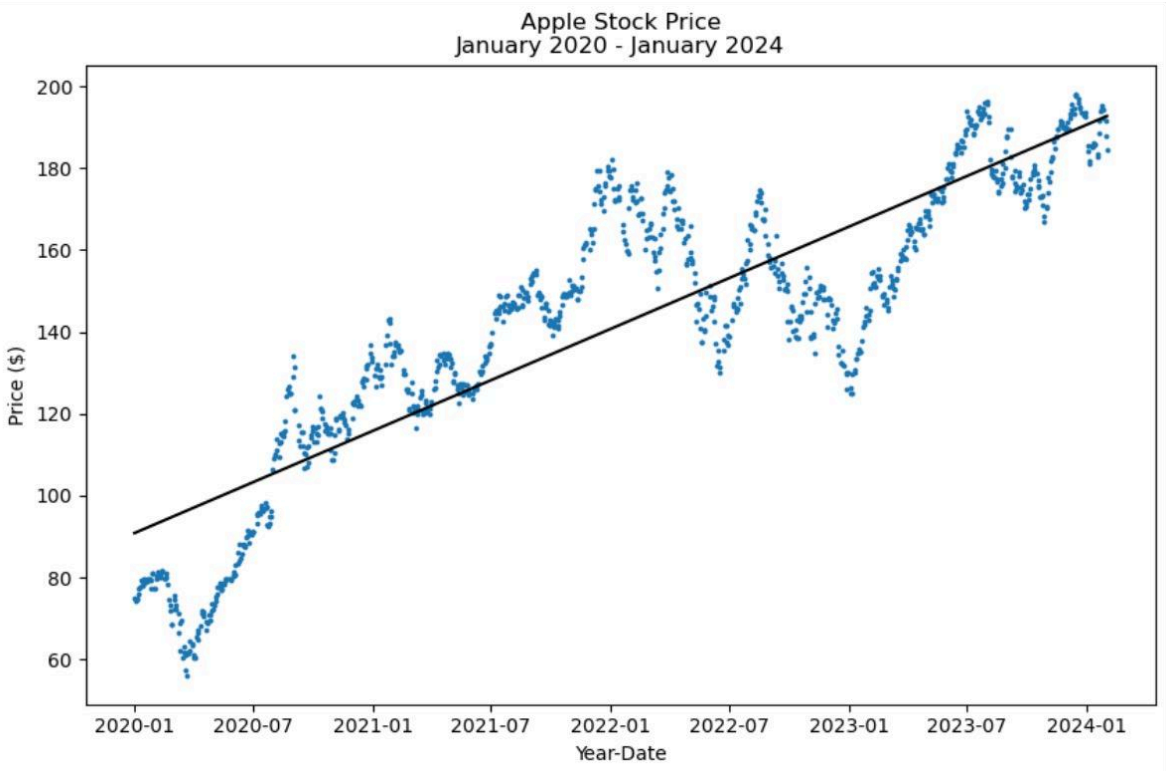
Predicted stock price V/S Actual stock price on day 1600

```
start_date = datetime(2020, 1, 2)

jan_2026 = start_date + timedelta(days=2192)
print(jan_2026)

ordinal_day_to_predict = 2192
predicted_price = round(intercept + (slope * ordinal_day_to_predict),2)
print("Predicted Price of Apple Stock on January 2nd, 2026 (Day 2192): " + '$' + str(predicted_price))
```

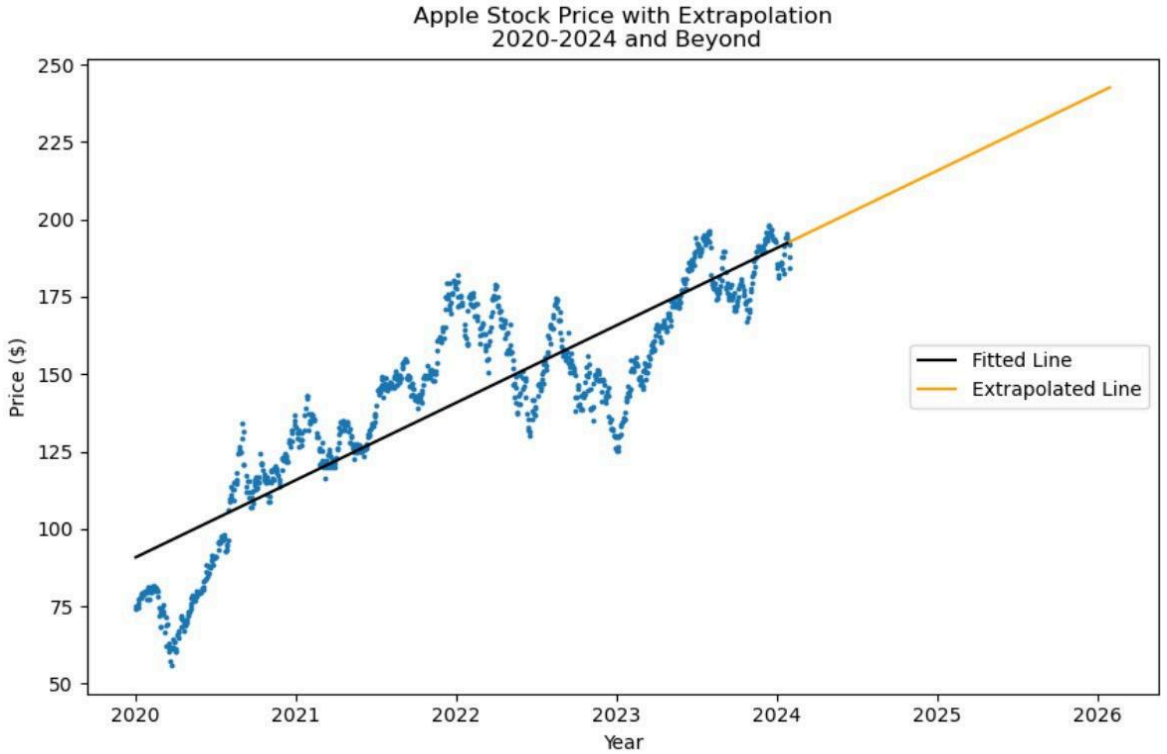
✓ 0.0s

```
2026-01-02 00:00:00
Predicted Price of Apple Stock on January 2nd, 2026 (Day 2192): $240.74
```

Predicted stock price of Apple for January 2, 2026



Visualization of Apple stock from January 2020 through January 2024



Linear regression and extrapolated lines predicting Apple stock price for 2025 and 2026

CONCLUSION & FUTURE WORK

- Apple's stock prices show a consistent positive growth trend over the analyzed period, increasing by approximately $0.07 per day.
- The regression model explains 76.6% of the variance in stock prices, demonstrating its effectiveness in capturing the relationship between time and price.
- Both the intercept and slope of the model are statistically significant (p-values < 0.05), confirming the robustness of the findings.
- The 95% confidence intervals suggest a reliable estimate for both the daily price increase and the starting price on January 2, 2020.
- The findings align with broader market trends, validating the utility of using historical data for understanding stock price dynamics.
- In the future, there is a possibility to explore the inclusion of economic indicators, company financials, and geopolitical events to enhance the predictive power of the model.
- Implement machine learning models, such as Random Forests or Neural Networks, to capture non-linear relationships and improve accuracy.

REFERENCES

Patel, D. (2024, September 1). *US stock market and commodities data (2020–2024)*. Kaggle. https://www.kaggle.com/datasets/muhammadehsan02/us-stock-market-and-commodities-data-2020-2024

Sullivan, B. (2024, July 30). *Average stock market return*. Forbes. https://www.forbes.com/advisor/investing/average-stock-market-return/

Chugh, A. (2020, December 8). *MAE, MSE, RMSE, coefficient of determination, adjusted R squared—Which metric is better?*. Medium. https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e