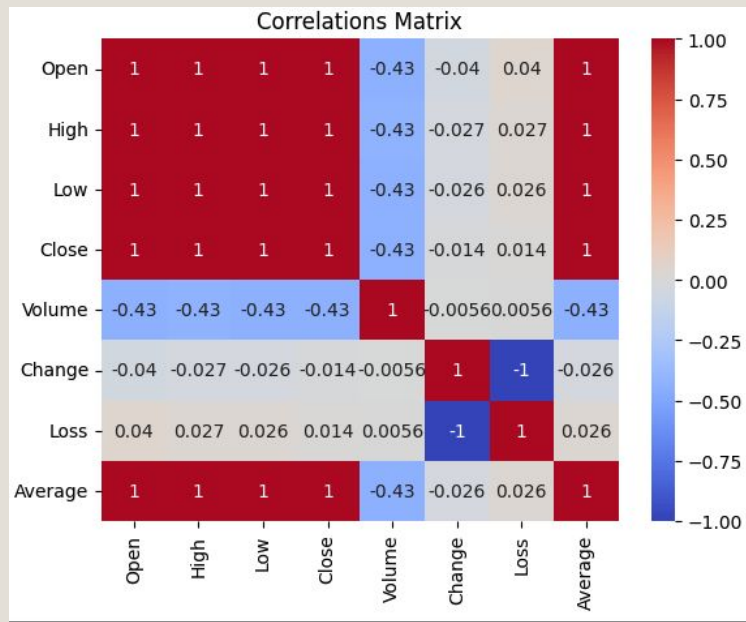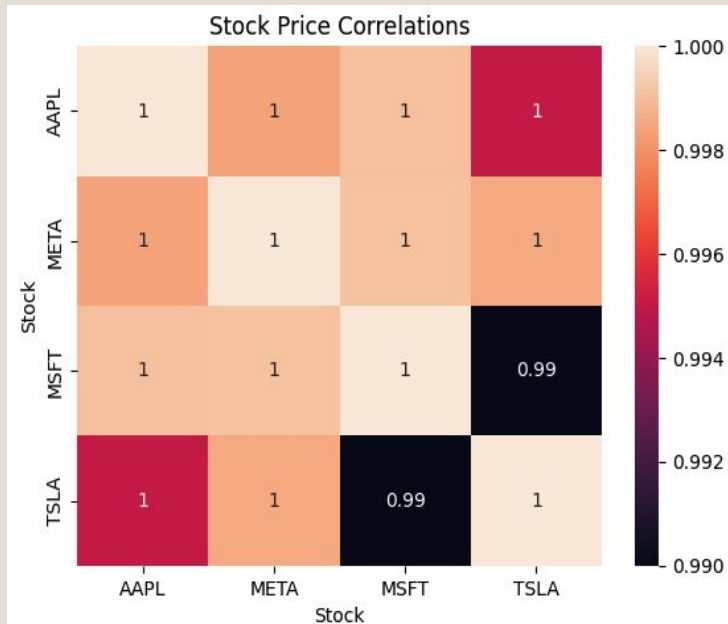# Stock Price Prediction
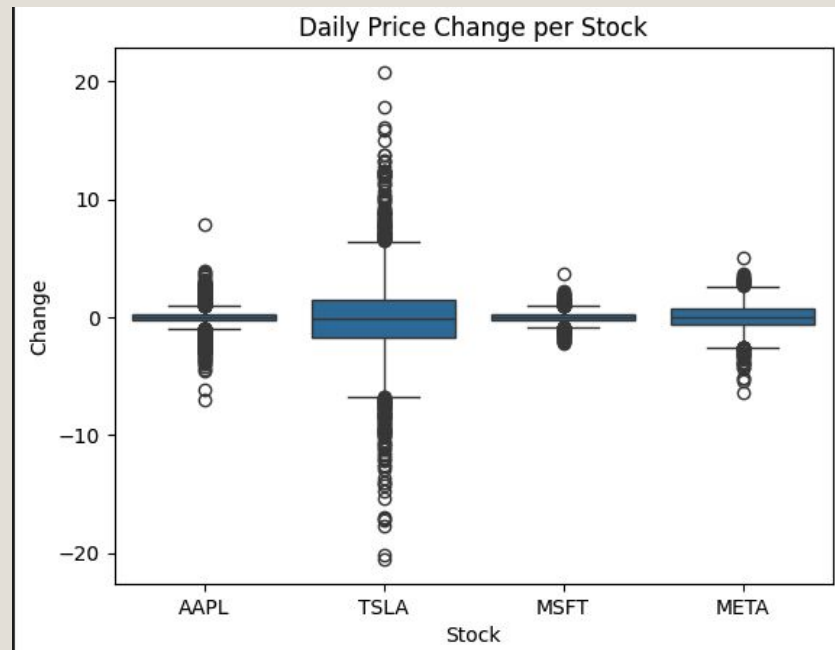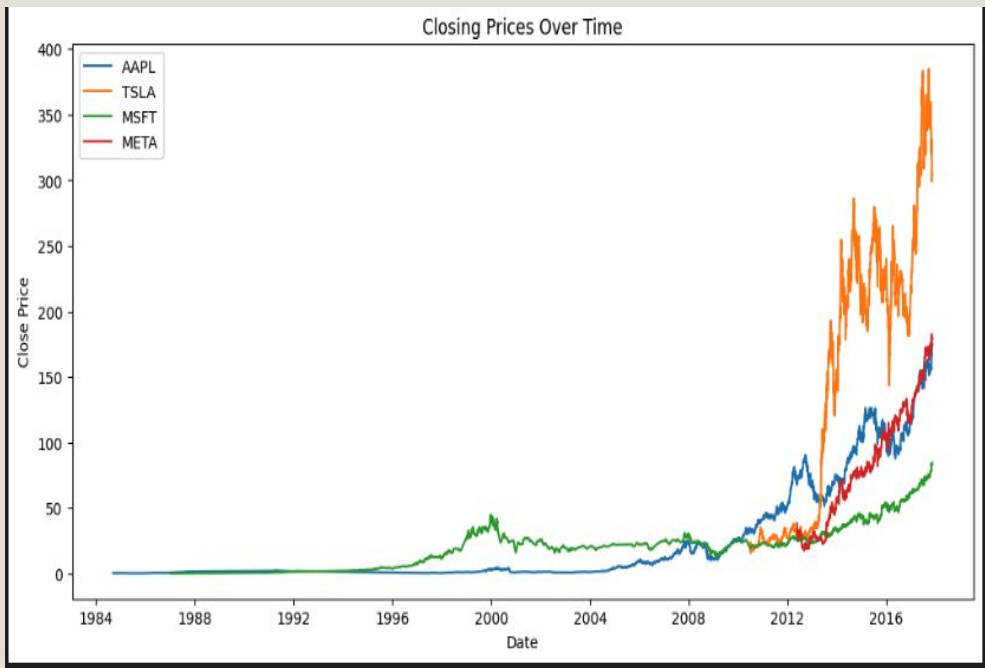
TABLE OF CONTENTS

PROBLEM IDENTIFICATION

❖   In this project, the goal is to build a regression and classification models that can predict the next day's closing price of a stock based on historical trading data such as open, high, low, close prices, volume, and daily returns.

❖   Stock prices are influenced by numerous factors and are inherently noisy and volatile, making accurate prediction challenging. However, even modestly successful models can provide insights into market trends and inform better decision-making for investors, analysts, and automated trading systems.

Dataset

❖   The dataset contained 8326 rows and 8 columns

❖   It includes features such as Open, High, Low, Close prices, Volume, Change, Daily Return, and Cumulative Return.

❖   The target variable is 'Target'

❖   https://www.kaggle.com/datasets/sanamps/stock-price-prediction-with-rnn

ANALYSIS

Closing Prices Over Time

Daily Price Change per Stock

MODELING RESULTS

Model Evaluation Results with hyperparameters

Models	R²	MAE	RMSE
Gradient boosting regression	0.78	21.79	46.92
k-Nearest Neighbors	0.77	22.68	47.78
Random forest regression	0.77	22.41	47.80

Model Evaluation Results with hyperparameters

Models	Accuracy	Precision	Recall	F1 Score
Random Forest Classifier	0.66	0.82	0.45	0.58
XGBoost Classifier	0.69	0.83	0.53	0.64
LGBM Classifier	0.67	0.81	0.50	0.61
Gradient Boosting Classifier	0.70	0.66	0.87	0.75

CONCLUSION

❖ Applying GridSearchCV to optimize hyperparameters led to noticeable improvements across most models

❖ The tuned Gradient Boosting Classifier achieved the highest F1 score (0.75), strong recall (0.87), and competitive precision (0.66).

❖ Incorporating more complex financial indicators, macroeconomic variables, and sentiment analysis from news or social media could provide deeper insights into market behavior.

❖ Leveraging time-series-based models such as Long Short-Term Memory (LSTM) may help capture temporal dependencies in stock movements more effectively.