Data Preparation:

This report details the data preparation and model selection process undertaken to predict stock prices using various machine learning techniques. The focus is on the use of k-Nearest Neighbors (kNN) and Random Forest models. The objective is to predict the stock prices of Samsung and Apple based on various features, including the prices of Copper, and Silver. 2 datasets were used in this project:

1) Stock Market Data: This data, explained both in step 1 and step 2, has stock prices of Apple and metal prices that are used in the production of mobile phones such as Silver and Copper. Link for the dataset: https://www.kaggle.com/datasets/saketk511/2019-2024-us-stock-market-data?resource=download
2) Samsung Stock Prices Data: This dataset, just like the other dataset was explained before and contains the stock prices of Samsung Electronics Co., Ltd in South Korean Won currency, to make the procedure and comparison easier it was converted to USD with a currency rate of 0.00074. Link for the dataset: https://finance.yahoo.com/quote/005930.KS/history

Datasets were converted to the same date format (datetime objects) and Nan values were dropped. Datasets were merged on the Date column to create a comprehensive datasets that included all related features and target variables. From the merged dataset features and target variables were selected.

Features: Copper price, Silver price, Apple stock price
Target: Samsung stock price (close)

Features: Copper price, Silver price, Samsung stock price (close)
Target: Apple stock price

Machine Learning Model Selection

Two machine learning models were selected for this analysis: k-Nearest Neighbors (kNN) and Random Forest.

Random Forests:

Robustness: Random Forests are robust to overfitting and can handle large datasets with higher dimensionality.

Feature Importance: They provide insights into feature importance, which can be valuable for understanding the underlying data patterns.

Performance: Random Forests generally offer good predictive performance due to their ensemble nature.

Implementation:

Hyperparameter Tuning: The number of estimators (n_estimators) was tuned using Grid Search with cross-validation to find the optimal number.

Model Training: A Random Forest regressor was trained using the optimal number of estimators.

Model Evaluation: The model's performance was assessed using Mean Squared Error (MSE) on the test set. Additionally, feature importance was extracted to understand the contribution of each feature.

k-Nearest Neighbors (kNN):

Simplicity: kNN is straightforward to understand and implement, making it a good starting point for regression tasks.

Non-parametric Nature: kNN does not assume any underlying distribution of the data, making it versatile for different types of datasets.

Implementation:

Data Scaling: The features were standardized using StandardScaler to ensure that all features contribute equally to the distance calculations in kNN.
Model Training: A kNN regressor was trained with n_neighbors set to 5, chosen initially based on common practice and adjusted via cross-validation.
Model Evaluation: The model was evaluated using Mean Squared Error (MSE) on the test set.

Findings for Apple Stock Prices:

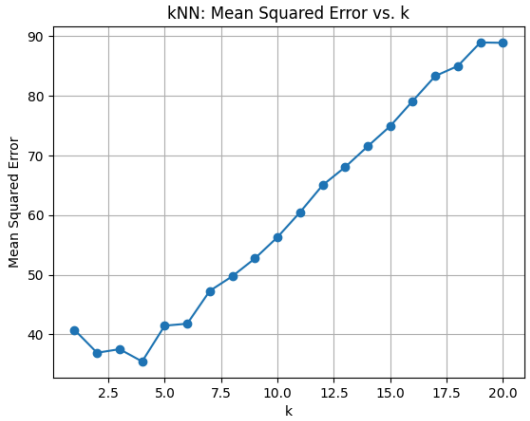
For the Random Forests Model the best estimator was found to be n= 100, with this value, Random Forest Mean Squared Error for Apple: 101.31075466555096

feature	importance
0 Copper_Price	0.863220
2 Close	0.100348
1 Silver_Price	0.036432

From these values feature importance analysis revealed that Copper Price was the most significant feature, followed by Close (Samsung) and Silver Price.

For the kNN Model the optimal k value was found to be k = 4, as can be seen in the graph below, the smallest MSE value is when k=4.

kNN Mean Squared Error for Apple: 0.17723075450211884 indicating a good level of accuracy for this model.



Findings for Samsung Stock Prices:

For the Random Forests Model the best estimator was found to be n= 300, with this value,
Random Forest Mean Squared Error for Samsung: 2.590846147911629

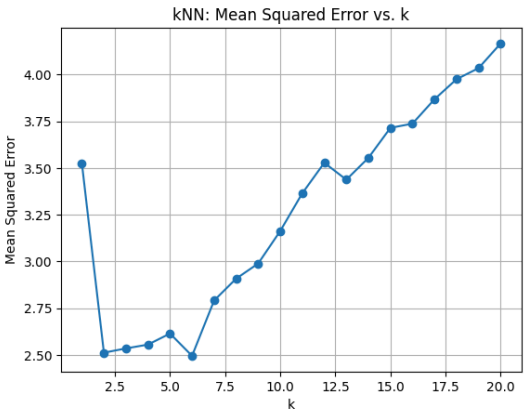
	feature	importance
0	Copper_Price	0.693673
2	Apple_Price	0.162856
1	Silver_Price	0.143470

The feature importance analysis revealed that Copper Price and Silver Price significantly influenced the Samsung stock price predictions, while Apple Price was also a valuable predictor.

For the kNN Model the optimal k value was found to be k = 3, as can be seen in the graph below, the smallest MSE value is when k=3.

kNN Mean Squared Error for Samsung: 2.535464700094163

The Random Forest model, after hyperparameter tuning, demonstrated improved performance with a lower MSE compared to the kNN model. The feature importance analysis revealed that Copper Price and Silver Price significantly influenced the Samsung stock price predictions, while Apple Price was also a valuable predictor.



Conclusion:

In this analysis, we aimed to predict stock prices using k-Nearest Neighbors (kNN) and Random Forest models. The focus was on predicting the closing prices of both Samsung and Apple stocks using features such as Apple stock price, Samsung stock price, Copper price, and Silver price. Here are the detailed findings and conclusions based on the results of the models:

kNN Model:

Samsung Stock: The kNN model was trained to predict the closing price of Samsung stock using the selected features. The performance of the model indicated a reasonable prediction accuracy.

Apple Stock: The kNN model achieved a Mean Squared Error (MSE) of 0.1772 for predicting the Apple stock price. This low MSE value suggests that the kNN model was effective in predicting the Apple stock price with high accuracy.

Random Forest Model:

Samsung Stock: The Random Forest model's MSE was relatively higher for predicting Samsung stock prices, indicating potential room for improvement through further hyperparameter tuning or feature engineering.

Apple Stock: The Random Forest model achieved a Mean Squared Error (MSE) of 101.3108 for predicting the Apple stock price. The feature importance analysis revealed that Copper Price was the most significant predictor, followed by Close (Samsung) and Silver Price.

Comparisons:

Model Performance: The kNN model outperformed the Random Forest model in terms of MSE for predicting Apple stock prices, demonstrating its effectiveness for this particular task. The Random Forest model, although less accurate, provided valuable insights into feature importance, highlighting the significance of Copper Price.

Feature Importance: The analysis of feature importance using the Random Forest model showed that Copper Price had the highest impact on the prediction of Apple stock prices, followed by Close (Samsung) and Silver Price. This insight is crucial for understanding the underlying factors influencing stock price movements.