DOMAIN: APPLIED DATA SCIENCE

PROJECT NAME: STOCK PRICE PREDICTION

Problem Statement: To Build a predictive model to forecast stock prices based on historical market data, assisting investors in making well-informed decisions and optimizing their investment strategies.

Problem definition: The problem is to build a predictive model that forecasts stock prices based on historical market data. The goal is to create a tool that assists investors in making well-informed decisions and optimizing their investment strategies. This project involves data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

Understanding the Problem:

To solve this problem effectively, we crucially understand its various components:

Data Collection

- Obtain historical stock market data, including price, volume, and other relevant features.
- Explore sources like financial APIs, market data providers, or public datasets.

-Dataset Link: https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset (this dataset is given in our project).

Data Preprocessing

- Clean and preprocess the data to handle missing values, outliers, and inconsistencies.
- Convert timestamps to date-time objects for time-series analysis.

Feature Engineering

- Create meaningful features that can potentially influence stock prices, such as movingaverages, technical indicators, and sentiment scores from news or social media.

Model Selection

- Choose appropriate machine learning or deep learning algorithms for time-series forecasting. Options may include:
  - LSTM (Long Short-Term Memory) networks
  - ARIMA (Auto Regressive Integrated Moving Average) model

-Prophet

- Gradient Boosting Regressors

Training

- Split the dataset into training and validation sets.
- Train the selected models on historical data, optimizing hyperparameters where necessary.

Evaluation

- Evaluate model performance using metrics like Mean Absolute Error (MAE), Mean SquaredError (MSE), or Root Mean Squared Error (RMSE).
- Compare the performance of different models to select the best-performing one.

Proposed Design

Data Collection

- Utilize financial APIs (e.g., Alpha Vantage, Yahoo Finance) to fetch historical stock data for thetarget company.
- Collect data for relevant indices (e.g., S&P 500) to include market trends.

Data Preprocessing

- Handle missing data through imputation or removal.
- Detect and treat outliers that might skew the model.
- Normalize or scale features as necessary.

Feature Engineering

- Compute common technical indicators like Moving Averages, Relative Strength Index (RSI),and Bollinger Bands.

- Incorporate external data sources, such as news sentiment scores, economic indicators, orsector-specific data.
- Consider lag features to capture temporal dependencies.

Model Selection

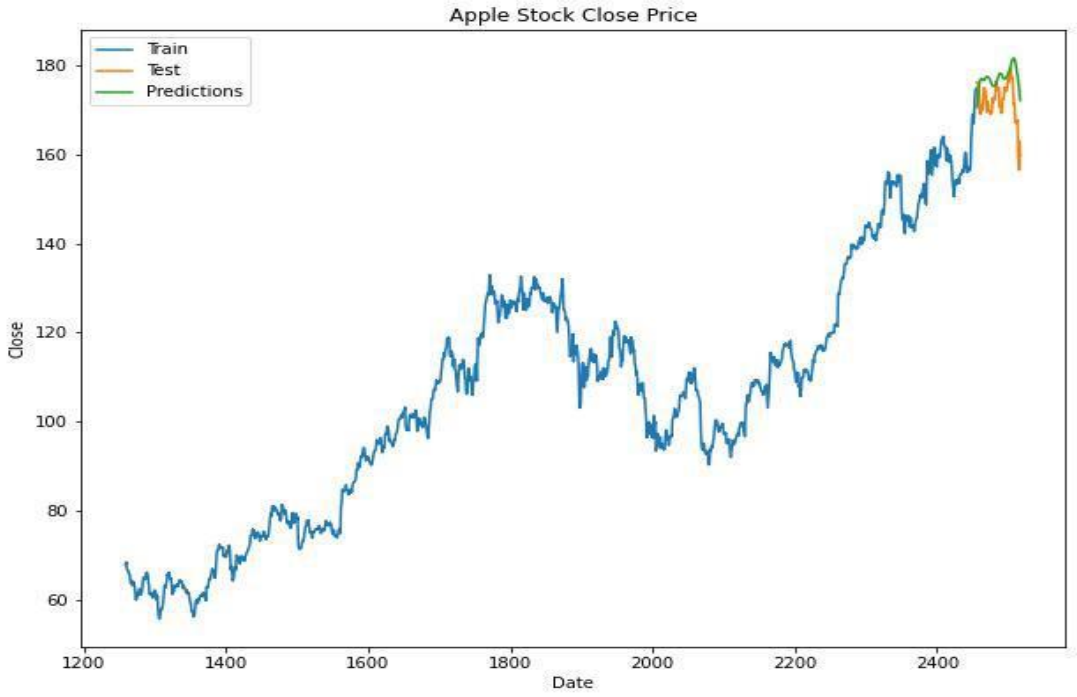
- Experiment with various forecasting models, including LSTM, ARIMA, and Prophet.
- Optimize model hyperparameters through techniques like grid search or Bayesianoptimization.
- Implement ensembling methods if necessary.

Training

- Split the data into training, validation, and test sets, ensuring a chronological order.
- Implement cross-validation to assess model stability.
- Train models with different feature sets and compare their performance.

Evaluation

- Evaluate models using appropriate metrics (e.g., MAE, MSE, RMSE).
- Visualize predicted vs. actual stock prices to understand model behavior.
- Conduct sensitivity analysis to assess model robustness.



This is an apple stock close price derived from https://images.app.goo.gl/NPHsuq6gdeafYZ3F8

Conclusion

In conclusion, this project aims to develop a predictive model for stock price forecasting, supporting investors in making informed decisions and optimizing their investment strategies.The proposed design encompasses data collection, preprocessing, feature engineering, modelselection, training, and evaluation. By following this structured approach, we aim to create a valuable tool for the financial community.

Timeline

1.Data Collection and Preprocessing:

- Gather historical stock data preprocess it.

2.Feature Engineering:

- Create relevant features for modeling.

3.Model Selection:

- Experiment with various forecasting models.

4.Training and Optimization:

- Train and fine-tune selected models.

5.Evaluation and Reporting:

- Evaluate model performance and generate a report.

This proposed timeline is subject to adjustments based on the complexity of the dataset andmodeling efforts.

By following this structured plan, we aim to create a powerful tool that assists investors in theirdecision-making process and helps optimize their investment strategies based on historical market data. This project requires a multidisciplinary approach, combining expertise in finance, data science, and machine learning to achieve its objectives.