

Predicting Next-Day Stock Open Prices using Machine Learning and News Sentiment

Sara Mezuri¹ and Alireza Golkarieh²

¹saramezuri@oakland.edu
²golkarieh@oakland.edu

This manuscript was compiled on October 6, 2024

Abstract

Problem Statement: The project aims to predict the next day's stock opening price using a combination of historical stock data and sentiment analysis on financial news articles.
Motivation Accurately predicting stock prices is crucial for investors and traders, as it enhances decision-making and profitability. By integrating sentiment analysis, the project aims to provide a more comprehensive model, offering deeper insights into how news events influence market trends.

Rho LaTeX Class © This document is licensed under Creative Commons CC BY 4.0.

1. Goal

Our goal is to build a predictive model that accurately forecasts the next day's stock opening price by leveraging historical stock data and sentiment analysis of news articles.

2. Intended Outcomes

1. Different functional predictive models capable of providing next-day stock opening prices.
2. Visualizations showing the impact of sentiment on stock price movements.
3. A GitHub repository containing all the codes.
4. A detailed report summarizing findings, challenges, and recommendations for future work.

3. Methodology

3.1. Data Preprocessing

- Collect historical stock data for all the stocks of Nasdaq-100 and S&P500 indices for the last 24 years, using TC 2000 and Bloomberg.
- Collect and clean news articles on the same stocks, for the last 24 years, using News API.
- Clean the data, preprocess, and handle any missing values (if any).

3.2. Feature Engineering

- Select relevant and independent features.
- Perform sentiment analysis on news data to create sentiment-based features.

3.3. Model Development

- Implement baseline models (e.g. Linear regression, ARIMA).
- Develop more accurate models using machine learning (e.g. Random Forest, LSTM).
- Train our models.
- Evaluate the models (using metrics such as MSE, MAE, and R-squared).
- Compare the performance of models with and without sentiment analysis.

4. Challenges and Scope

4.1. Key Challenges

- Processing and cleaning financial news text data.

- Combining structured (numerical) data and unstructured (textual) data effectively.
- Achieving reliable and interpretable predictions despite market unpredictability.

4.2. Scope and Boundaries

- In-Scope:
 - Daily stock data of the Nasdaq-100 and S&P500 stocks for the last 24 years.
 - Sentiment analysis of financial news articles related to chosen stocks.
 - Integration of historical and sentiment data for price prediction.
 - Implementation of various machine learning and deep learning models.
- Out-of-scope:
 - Intraday (minute-by-minute) stock price predictions.
 - Incorporating social media sentiment (e.g., Twitter).
 - High-frequency trading models or strategies.
 - Real-time predictions or live trading algorithms.

5. Tasks management

5.1. Timeline

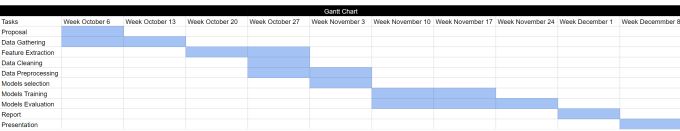


Figure 1. The Gantt Chart

5.2. Responsibilities

We intend for this project to be a collaborative effort, where we will work together on data selection, feature engineering, model development, report writing, and presentation.

Sara will take the lead on *developing the time series model and handling the theoretical aspects of the models*.

Alireza will take the lead on *model selection, code implementation, and debugging*.