

IE 582 Term Project: Stock Market Forecasting Challenge (Fall '23)

Date: November 23, 2023

1. Introduction

In the era of big data, extracting meaningful patterns from vast datasets is a valuable skill across disciplines, from engineering to social sciences. The financial markets, particularly Borsa İstanbul, offer a complex and rich area for data-driven exploration. This project aims not only to challenge you but also to enhance your prediction skills using statistical learning and data mining techniques, with a focus on real-world financial analysis applications.

2. Data

You will find the past data in the **Project section** on Moodle.

The data includes hourly average prices of selected stocks from Borsa İstanbul, presented in separate csv files for each time interval, encompassing about 4 years of data in long format. Make sure to understand the data format, including timestamps and stock identifiers.

3. Description

Your task is to forecast the next day's hourly average prices for 30 companies. You are expected to provide 10 predictions per day, corresponding to the hours from 9 AM to 6 PM. Feel free to use any external data like Google Trends to enhance your predictions.

This project is structured as a competition, similar to platforms like [Kaggle](#). We've developed a submission system via Google Forms, and you can track your progress on Google Sheets. More details on this system will follow next week.

Goal

Develop a forecasting product that can be applied in various business scenarios, such as inventory management or service level optimization. Accurate forecasts assist businesses in reducing waste and understanding risk implications associated with uncertainty.

4. Evaluation

Submit your predictions for the day (day t) before 9:00 of the same day. You can do submission for the day t starting from 18:00 of day $t-1$ until 9:00 of day t . We will consider only your latest submission for that day. You must make predictions for each company and hour, totaling 300 predictions.

We will calculate the Weighted Mean Absolute Percentage Error (WMAPE) for your predictions per stock. Your rank will be determined based on your average rank over all 30 stocks. No submissions will result in zero points. Make the most of this experience!

Understanding WMAPE: Weighted Mean Absolute Percentage Error

What is WMAPE? The Weighted Mean Absolute Percentage Error (WMAPE) is a measure used to evaluate the accuracy of forecast models. Unlike traditional Mean Absolute Percentage Error (MAPE), WMAPE accounts for the varying sizes of the data points, making it a more accurate reflection of forecasting errors in scenarios where data points have different weights or importance.

WMAPE Formula The WMAPE is calculated using the following formula:

$$\text{WMAPE} = \frac{\sum_{i=1}^n |A_i - F_i|}{\sum_{i=1}^n |A_i|} \times 100$$

where:

- A_i = Actual value for the i^{th} data point
- F_i = Forecasted value for the i^{th} data point
- n = Total number of data points

Interpretation

- A WMAPE of 0% indicates perfect forecasting.
- Lower WMAPE values indicate higher accuracy of the predictions.
- This metric provides a normalized error rate across all predictions, allowing for a fair comparison between different forecasting models or different data sets.

Importance in Our Project

For our project, WMAPE will be used to evaluate the accuracy of your daily stock price predictions. Understanding and minimizing WMAPE in your forecasts will be crucial for success in this competition.

5. Timeline

You will receive your system access credentials soon. The coming weeks will serve as a trial period (next week, we will announce once the evaluation system is ready). Use this time to familiarize yourself with the system, build your prediction pipeline, and ask any questions. Scores will be reset on December 25th, 2023, marking the start of the actual project phase. The submission system closes on January 12th, 2024.

6. Deliverables

Towards the end of the semester (date to be announced, likely post-finals), you will present your work. Your report should be structured as follows:

1. **Introduction:** Problem description, proposed approach summary, and data analysis.
2. **Related Literature:** Summarize any relevant literature.
3. **Approach:** Detailed explanation of your methodology.
4. **Results:** Discuss your findings.
5. **Conclusions and Future Work:** Summary of findings, approach evaluation, and potential improvements.
6. **Code:** Link to your GitHub repository.

Reports should be concise, well-structured, and reflective of your process and findings. Your reports are expected to follow a journal article structure (i.e. no codes in the report - instead pseudocodes if necessary). Some links that can help are:

1. [How to Review a Research Paper](#)
2. [How to Write a Scientific Article](#)
3. [Writing a Journal Article](#)
4. [How to Write a Journal Article](#)

Contact Information

For queries or clarifications, please reach out to through discussion forum.

We encourage creativity and innovative approaches in this project and look forward to seeing your solutions to this real-world problem in financial data analysis.