

Project Description

Department of Economics, University of Copenhagen

Summer 2023

On August 15 (8.00 PM), you need to hand in your project description on Absalon. You only need to hand in one project description per group, and you should bring it to your last supervision meeting.

Your project description needs to answer all questions below.



1. Group number

Group 12

2. Names of the students in the group

- Søren Mingon
- Sebastian Faurby

3. What is your research question?

Can sentimental analysis of news-headlines by LLM (Large Language Model) improve predictability in a machine learning model when trying to predict returns in the stock market?

The purpose is to compare machine learning models using sentimental data only with models using financial data only and models using a combination of both.

4. What kind of data are you planning on using? How will you get access to these data?

We are planning on using data of different types and from different source. Here is a summary:

- Stock-data: End-of-day stock data, fundamental data and technical indicators.
 - Data collection through APIs
- Stocks related news: Day-to-day news headlines for the relevant stocks.
 - Data collection through APIs
- FAMA French 3 Factors model (daily):
 - Data collection through web scraping

5. What will your data analysis be like? Will you use machine learning? How?

The analysis will consist in the following parts:

- ChatGPT (LLM: GPT-3.5-turbo) will be used to make a sentimental analysis of the news headlines and classify them in positive, negative or neutral.
- We will be using machine learning to predict end-of-day returns given stock-data and sentimental day as features. This will be a time-series analysis, where the objective is to feed the model with the preceding 30-days news-headlines, fundamental data and technical indicators in order to predict returns for the next day.

The overall objective is to compare the accuracy of a model, where the models will be separated in three categories:

- 1) ML-models trained with all data available (sentimental- and financial data)
- 2) ML-models trained with sentimental data only
- 3) ML-models trained with financial data only

We will use ElasticNet Regression with a weighted combination of L1 and L2 regularization.

Furthermore, Cross validation will be applied using *sliding window* method.

6. Have you already identified other papers within this area that you can use in a literature review? If so, name a few and explain what they do in one sentence only.

- Lopez-Lira, A., & Tang, Y. (2023, May 11). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *SSRN*, Volume 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4412788

About: This paper examines the potential of chatGPT and other large language models in predicting stock market returns and sentiment analysis of news-headlines.

7. How do you 'contribute' to the literature?

We plan to contribute to the literature by exploring the predictive power of LLM with complementary information (stock data and FAMA French) and thereby explore the possibility of enhancing existing models.