732A92 Text Mining

CLASSIFYING STOCK PRICE MOVEMENTS BASED ON 8-K SEC FILINGS

November 25, 2019

Name: Julius Kittler
Student ID: julki092
Linköping University

Abstract

For over a century, fingerprints have been an undisputed personal identifier. Recent court rulings have sparked interest in verifying unique

Contents

List of Figures

1 Introduction

Forecasting stock prices has been a relevant problem since the existence of publicly traded companies. Today, it is a more relevant problem than ever before because we have the technological infrastructure to build automatic trading systems and thereby put our forecasts to practice. We can not only obtain massive amounts of data relevant for forecasts via APIs but also execute trades via APIs. With commission free trading having become the industry standard in the U.S., the latter services are even offered for free, for instance by the commission free algorithmic trading API Alpaca [1].

This text mining project aims to explore whether filings of the U.S. Securities and Exchange Commission (SEC) can be used to classify if the price of a stock will decrease, remain unchanged or increase. The results will be evaluated critically from the perspective of a trader, with the question in mind whether the classifications could actually be put to practice in an automatic trading system.

1.1 SEC Filings

The SEC is a government agency in the United States with the mission to "protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation" [2]. An important task of the SEC is to ensure that publicly traded companies inform their shareholders and the public about their business.

For instance, the SEC requires companies to publish their quarterly and annual results, and inform shareholders about certain relevant events. For each of these purposes, companies have to file specific documents. For instance, the annual report corresponds to the 10-K, the quarterly report corresponds to the 10-Q and another report for specific relevant events corresponds to the 8-K filing.

Importantly, SEC filings are actively used by traders when making investment decisions. Many trading platforms such as Webull and thinkorswim also provide traders with the recent SEC filings of any tradable company (along with other information such as fundamental data, news data and historical prices). SEC filings are interesting for stock price forecasting because they are standardized, publicly accessible for free and because they contain relevant, objective and generally accurate information.

1.2 8-K Filings

Companies need to publish an 8-K filing for major events relevant for their business. Such events might be a change in the board of directors, a delisting from a stock exchange and a merger or acquisition [3]. Importantly, 8-K filings are generally due within four business days after the event [4].

Because 8-K filing correspond to major events for the company and because they need to be published shortly after an event occurred, 8-K filings seem interesting for predicting short-term volatility in the stock market. Moreover, the important information in 8-K filings is generally represented in form of text data, whereas other filings such as the annual report often focus on numerical data represented in tabular form.

TODO: EXAMPLE

Figure 1 shows a boat.

1.3 Research Questions

There are mainly three research questions that this report aims to address. The focus is not only on the stock price classification itself but importantly also on understanding how the model works. Assuming that we do not know much about SEC filings, we would like the model to give insights about how it is using the text data of the SEC filings for the classification and in which cases the model is more or less reliable.

1. Can we successfully forecast stock prices based on 8-K filings? If so, how well?

2. Which text features are most important when forecasting stock prices based on 8-K filings?

3. In which scenario do the forecasts perform best (e.g. type of the 8-K filing, industry of the company, other metrics of the company)?

Figure 1: A boat.

2 Theory

Previous work in predicting stock prices with 8-K filings.

2.1 Stock Price Forecasting in General

tbd

2.2 Stock Price Forecasting with Text Data

tbd

2.3 State of the Art Classification with Text Data

tbd

3 Data

3.1 Retrieval

Data sources, Time period, Exchange Market, Resolution

3.2   Processing

3.3   Descriptive Statistics

3.3.1   General

3.3.2   Target Variable

3.3.3   Feature Variables

4   Method

4.1   Evaluation Metric

4.2   Baseline Models

4.3   Advanced Models

4.4   Hyperparameter Tuning

Train vs. test

4.5   Feature Relevance

5   Results

6   Discussion

7   Conclusion

8   References

radioactivedecay2 DLS1

References

[1]   *Alpaca - Commission-Free API First Stock Brokerage*, en. [Online]. Available: `https://alpaca.markets` (visited on 11/23/2019).

[2]   *SEC.gov | What We Do.* [Online]. Available: `https://www.sec.gov/Article/whatwedo.html` (visited on 11/23/2019).

[3]   *SEC.gov | Form 8-K.* [Online]. Available: `https://www.sec.gov/fast-answers/answersform8khtm.html` (visited on 11/23/2019).

[4]   W. Kenton, *8-K (Form 8k)*, en. [Online]. Available: `https://www.investopedia.com/terms/1/8-k.asp` (visited on 11/23/2019).