

Using MD&A of company's SEC filing to predict future stock price

CS410 Final Project Proposal (Individual project)

Project coordinator/Author: Dmitry Villevald (dmitryv2@illinois.edu)

For my final course project I would like to choose a free topic (Option 5) and explore the impact of the information extracted from the Management's Discussion and Analysis (MD&A) section of the company's SEC filing on the future stock price.

Q: What are the names and NetIDs of all your team members? Who is the captain?

A: This is **individual** project by Dmitry Villevald (dmitryv2)

Q: What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

A: I would like to explore how the changes in a sentiment of the company's Management's Discussion and Analysis (MD&A) impact the future stock prices. MD&A section is an unaudited section of quarterly and annual SEC filings where the company's management discusses the current status of a company and, more importantly, the future risks and opportunities. I want to explore if investors overreact to changes in management sentiment in the long run. This knowledge would be important for investors trying to decide what to do with a company stock when management sentiment changes.

My plan is, first, to pull 10Q and 10K SEC filings of the companies - members of SP&500 or Dow Jones index - over a few recent years from SEC EDGAR database, parse the data and extract MD&A sections. Then, I plan to determine a sentiment score of each MD&A document using the vocabulary of positive and negative words (for example, Loughran and McDonald sentiment word lists). Third, I plan to select the events with significant changes in sentiment scores and see how these changes are correlated with the changes in company's stock price in a quarter following the filing and its announcement (using Yahoo Finance data stock price data or similar sources) adjusted for market or sector/industry returns.

If the stock markets are efficient then one would not expect a significant investors' over- or under-reaction to changes in sentiment which I want to confirm. The performance metric could be the confusion matrix showing the relationships between the significant

(i.e. higher, in absolute value, than a certain threshold value) changes in MD&A sentiment score - positive or negative - and the significant - positive or negative - next-quarter company's stock returns.

Q: Which programming language do you plan to use?

A: Python

Q: Please justify that the workload of your topic is at least $20 \cdot N$ hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

A: I plan to spend about 3 hours to write and test the script which will download, parse and clean MD&A data. Then it would take another 7-10 hours to actually perform these tasks (i.e. downloading, parsing and cleaning the data.) Another 2-5 hours would be spent on extracting the sentiment from MD&A sections and identifying the cases when sentiment score experiences significant changes. About 2 hours will be spent on exploring the relationships between the stock prices and sentiment scores and drawing the conclusions. Finally, I expect to spend about 5-6 hours to document my process/findings and to create a demo/presentation.