Project Charter

Leeds School of Business Senior Capstone

September 29, 2020

Royce Schultz, Andrew Yee, Chace Trevino, Tianwei Zhao

Revision History

Change Description	Sponsor Signature	Project Manager Signature	Date of Revision
Initial Project Charter	Diego García	Chace Trevino	9/29/20

Project Summary

Project Name: Research Project, Effects of News Media sentiment and Stock price fluctuations.

Problem Statement

This project will develop machine learning models and techniques to further the research done by the Leeds School of Business in applying machine learning to finance. Specifically, this project will apply NLP models to gauge sentiment in business TV corpora and correlate the sentiment to movements in stock prices. Additionally, this process will attempt to measure the effects of different news sources or political segments and explore how news media has changed over time.

While the main goal of the project is to correlate sentiment from the TV corpora with changes in stock prices, there are several other potential topics of interest. One of these is investigating the correlation of politics and economics, specifically how political news affects the stock market. Additionally, investigating and predicting stock reactions to business interviews on a minute to minute basis would be another topic of interest. Finally, time allowing, utilizing optical character recognition on the Moody's manuals would be an additional area of research.

Objectives

- Organize and analyze TV corpora of business and political news shows
- Evaluate sentiment in TV corpora and correlate it in movements in the stock market
- Investigate the impacts that political news has on the stock market
- Correlate minute by minute stock reactions to sentiment in interviews with corporate executives
- Utilize optical character recognition techniques to mine data from Moody's manuals

Successful Outcome Statement

- Create a system for downloading, storing, and reading data from Archive.org (but potentially additional sources).
- Use statistical methods to provide evidence for hypotheses relating to market sentiment and stock price.
- Use NLP techniques to analyze TV corpora to determine sentiment and other metrics
- Combine and deploy methods used into a data processing pipeline

Business Case

A successful model can predict stock price movements given data of live interviews, political segments, or business news, giving a stock trader the edge on their competition.

People

Primary Project Contacts

Role	Name	Email
Project Sponsor	Diego García	diego.garcia@colorado.edu
Teaching Assistant	Joewie Koh	Joewie.Koh@colorado.edu
Project Manager	Chace Trevino	Chace.Trevino@colorado.edu
Project Team Members	Royce Schultz, Andrew Yee, Chace Trevino, Tianwei Zhao	Tianwei.Zhao@colorado.edu, Royce.Schultz@colorado.edu, Andrew.Yee@colorado.edu,
Capstone Professor	Alan Paradise	alan.paradise@colorado.edu

Key Roles and Responsibilities

Role	Name	Responsibilities
Project Sponsor	Diego Garcia	• Provides vision and goals for team. • Requests changes to requirements and scope when circumstances change.
Project Manager	Chace Trevino	• Handle team/sponsor communications • Maintain project charter and change requests. • Create and track JIRA tasks • Track and manage project progress including deliverables and milestones • Write weekly status reports • Scrum Master
Project Keeper	Rotates Weekly	• Record meeting attendance • Collect weekly timesheets • Take meeting notes
Tech Lead	Royce Schultz	• Designs system specifications to implement scalable features within the project's scope. • Enforces code quality.
Testing Lead	Tianwei Zhao	• Manage testing of code. • Maintains CI/CD system.
Database Lead	Andrew Yee	• Manage database design and implementation • Maintain database integrity

---

Project Scope

In-Scope Statement

Throughout the time working on this project we will help further the research of the correlation between sentiment in business TV corpora and movements in stock prices. If positive results are made, our project will deliver effective methods to process TV corpora. Also, this research could be published by our sponsor: Diego Garcia, Chair of the Finance Division at CU Boulder which will detail the relationships we have discovered. Anyone who keeps up with financial research could be affected by this in learning to what extent outside forces affect the stock market.

Our area of expertise is software systems. We will create a software system that is robust, scalable, maintainable, and flexible for many use cases.

Provide specific detail on what the project will deliver. Provide information on the departments and the functional areas that will be impacted by the project.

Out-Of-Scope Statement

Our sponsor has stated that this is an empirical research project which means that our project goals could change as we continue to observe our work. In saying this, our project boundaries are very loose since we will most likely explore any new, interesting insights that come our way. At the same time, we need to remember that we are here as software developers. Therefore, we want to make sure that this is not just financial research but a project in which we are maintaining a software system that implements ML, NLP, and data processing. We have our main goal to gain insight into the correlation of sentiment in TV corpora and stock price movement and we do not want to leave any stone unturned in pursuing

this. The only similar research out there is in looking at the sentiment in business text corpora and stock price movement which is why we are sticking to our medium of TV.

Provide clarification of the project boundaries. As applicable, it may be helpful to identify related work that is addressed by other projects or initiatives as a means of preventing scope creep and project collisions.

Requirements

- Create database structure to store scraped web data from multiple sources
- Implement web scrappers to extract data for NLP processing
- Clean and preprocess data for analysis and future processing
- Analyze data collected and correlate it with stock market changes
- Mature database and data processing pipeline to maintain up-to-date metrics
- Create tests to maintain code with CI/CD pipeline

Constraints

- Personal team compute is limited, so project team will utilize CU Boulder compute cluster.
- Team members are taking additional courses which will limit the total hours that can be spent per week.
- Due to COVID 19, team can not hold in-person meeting, have to cooperate remotely.

Risks

- Remote work due to COVID 19, may affect team productivity.
- The project team consists of 4 team members, which was originally 5. This could potentially negatively impact the total output of the team.

Schedule and Budget

Schedule

https://thebuffsofwallstreet.atlassian.net/jira/software/projects/CAP/boards/1/roadmap

Project Phase	Deliverables/Milestones	Start	Finish
Initiate and Plan	• Set up consistent meetings with sponsor • Set up communication and administrative processes • Write project charter • Determine long term plans and goals	9/8/20	12/7/20
Initial Processing and Investigation	• Clean up and organize corpora • Bring in metadata (in particular stock price data and timestamps) • Explore which companies do the TV networks choose to interview	11/1/20	1/11/20
Analysis and Execute	• Investigate correlations and regressions of sentiment and stock market movement • Apply models to current interviews and evaluate if our past research can indicate what happens	1/11/20	4/22/20
Close	• Present project results to sponsor • Assist in paper writing if necessary	4/23/20	4/29/20

Labor Estimates *

Activity	Hours per Week per Person	Hours per Week
Administration	2	8
Meetings	2	8
Coding	4	16
Testing	1	4

Research	2	8
Review	1	4
Total	12	48

*Provided for context and for budgeting purposes

Performance Metrics

- Commits and pull requests to the code base.
- Reviewing other pull requests.
- Resolving GitHub issues.
- Completing Jira tasks.
- Writing and editing project documentation.
- Writing administrative documents (Status reports, time sheets, project charter, etc.)