ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CREDIT SUISSE

A thesis presented by
Siqi Dai

Measuring Reputational Risk by External Data

Submitted to attain the degree of
MASTER OF SCIENCE of ETH ZURICH
(M. sc. ETH Zurich)

Department of Management, Technology, and Economics (D-MTEC)
Swiss Federal Institute of Technology (ETH) Zurich

Supervised by

Prof. Didier Sornette (Chair of Entrepreneurial Risks, D-MTEC, ETHZ)
Makbule Asir (Data Scientist, IA Data Analytics, CREDIT SUISSE)
Wael Malek (Director, IA Data Analytics, CREDIT SUISSE)

Dec. 2020

Acknowledgements

Time passed by so fast, I still remember the first day I left from Shanghai and arrived in Zurich for my Master's journey in ETH. At that time, I was impressed deeply by the beauty of Zurich town. After two years and 2 months, now, I am going to complete my study program in D-MTEC with the end point of this six-month thesis project, in the unique period of COVID 19. Even though the pandemic forces me to work from home mostly, it also makes every online meeting and face-to-face discussion more valuable. As a student in risk management field, COVID 19 provides me opportunities to learn from real-time data, numbers, medias, charts, and so on with rationality.

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Dr. Didier Sornette for his guidance during my study and this project. His immense knowledge, scientific thinking, and passion for working inspired me a lot, and will continue to inspire me in the future. Every discussion with him not only solved my problem in my project, but also solved the questions happen in my life. He taught me academic expertise and also shared me the philosophy and methodologies of daily life. For example, "Simplicity is the ultimate sophistication" is a quote from Leonardo da Vinci and shared by Prof. Sornette to guide me how to write this thesis. I feel so lucky to work under the supervision of such a professional and patient professor.

Next, my gratitude also goes to my co-supervisors, Makbule Asir and Wael Malek from Credit Suisse for their guidance of this project. I appreciate a lot their suggestions and help during the whole of my master's thesis. They accepted me as a thesis student in data science group to join in this interesting topic and also provided me chances to discuss with other expertise from CS. Their business mind and creative thinking encourage me to study more.

In addition, my thanks have to go to RepRisk AG, Peter Mitic, and Moritz Sudhof. They selflessly shared their methodology and useful materials to me for academic use. They answered my question emails very patiently and encouraged me for my work. Their kindness and selfless impressed me a lot.

Moreover, I would like to thank my friends in Zurich. They are the ones who really accompany me in the days of laughing, crying, anxiety, happiness, and etc; because of them, my life in Zurich becomes much more colorful and enjoyable. I hope they will follow their hearts and achieve their success sincerely.

Last but not least, I want and I have to say thank you, to my dearest mom and dad, who always stand by my side and support every decision I make. Because of their support, I am more dared to purse my goals and dreams. I am the only child of my parents, a child who grew up in love. They love me very much but they are also very strict with me. If I have to say the most thankful persons in my life, the honor must go to them.

Contents

Abstract

This thesis aims to measure reputational risk with external data while developing a practical rule-based measurement model of reputation derived from sentiment of economical and financial newspaper articles from January 2008 to November 2020. We define two reputation indexes based on big amounts of financial news articles from Googlenews and Reuters, which are collected by our creative scrapers. We compare the predictive accuracy of different sentiment analysis models based on the results from recent works and find the results highlight the gains from combining existing lexicon with heuristic rules. We creatively introduce the idea from behavior economics by using Cumulative Prospect Theory model to amplify the impacts of articles in negative sentiment. Then after dealing with the missing data by linear interpolation, we correlate sentiment and reputation to emphasize the "memory" of reputation by Autogressive model. Lastly, we provide a user-friendly visualization application of reputation measurement and compare the reputation trends among giant financial institutes, which also represents the reputation changes of top-level in the banking industry. Finally, we do the event analysis to display the robustness of the model.

Keywords: Reputational risk, reputation, sentiment analysis, lexicon, machine learning, nlp, memorized reputation, scrapers, financial institutes, banking, Credit Suisse

Nomenclature

Symbols

S	Sentiment
O	Opinion
R	Reputation
C	Compound Score
D	Document
tf	Term Frequency
idf	Inverse Document Frequency
w	Weighting
u	Utility Function
g	Weighting Function
Rep	Reputation
AS	Average Sentiment

Indicies

i	identifier
t	timestamp
G	target of the opinion
H	opinion holder the opinion
title	title of each article
paras	paragraphs of each article
m	organization

Acronyms and Abbreviations

G-SIBs	List of Global Systemically Important Banks
NLP	Nature Language Processing
ML	Machine Learning
TF-IDF	Term Frequency-Inverse Document Frequency

AR	Autoregressive model
GG	Googlenews
TT	Twitter
RT	Reuters
MTEC	Management, Technology, and Economics
ETH	Eidgenössische Technische Hochschule
IA	Internal Audit
CS	Credit Suisse

Chapter 1

Introduction

1.1 Motivation

Harrison in 2008 said that the most valuable asset in the capital economy is not cash, stock, or buildings, but trust. Reputation is the belief and trust that a variety of people have for your organization and they expect the same attribute in the future [20]. Customers and stakeholders get emotionally and rationally attached to an organization when they decide who to work for, what to buy, sell, invest, and supply [13]. A good reputation encourages shareholders to invest in a company and correlates with superior overall returns; it helps the company to attract and retain talent and limit personnel turnover. Thus it is important a sense of responsibility towards the community for a company to build a good reputable business. In the words of Banjamin Franklin:

"It takes many good deeds to build a good reputation, and only one bad one to loose it."

Companies build reputation over many years and their reputation can be disrupted within seconds. Although a shortage of cash can bring a company to its knees, it is more frequently a loss of reputation that deals the final blow [37]. Importantly, once reputation is compromised, the process of rebuilding it may be costly and lengthy, and in worst-case scenarios, reputational capital[1] (a function of benefits gained and costs avoided) may never be recovered [8].

Over the past decade, interest in reputational risk in financial institutions has grown after the occurrence of some prominent examples of reputational and operating losses due to reputational risk events, like financial scandals, internal frauds, large lawsuits, money laundering, and so on. In a recent survey of financial services institutions, more respondents cited reputational risk than any other risk class as the greatest potential threat to their firm's market value [30]. But the attention devoted to managing reputational risk is very recent and the management of reputational risk should not feature when there is a reputation crisis. While tools and techniques proliferate for managing monetary risks, the art of protecting reputations is poorly developed and understood (Economist Intelligence Unit, 2005). Reputation Risk is generally agreed as a more elusive risk category compared to other risk categories, because of the difficulty of identifying reputation changes and quantifying the risk, with expecting the industry to further develop techniques for managing all aspects of these risks [26].

Traditional reputational risk management is based on internal data with company self-disclosures. It is now well-accepted that self-reported information is not always reliable data, especially when

[1] Blanc (2016) defines reputational capital as the total sum of a company's relationships with its stakeholders.

it comes to risks [32]. The importance of introducing external data and information in the process of risk management has been discussed a lot in recent studies.

Additionally, reputation risk has gained new prominence across industries in the age of the Internet [31]. Business is at constant scrutiny by social media and there is no hiding place [25]. Internet significantly eroded public trust in large corporations and financial institutions in particular, such that events, which in the past would not have been significant beyond the direct cost, can now turn into a reputational nightmare[31]. Good risk management strives to identify potential risks before materialization in order to either avoid or minimize the exposure of a firm to these risks. It thus has become more difficult to manage reputation and thereby increases the wavelength of reputation risks [4].

Obviously, managing reputation and reputational risk are essential to an organization. Corporates are constantly confronted with the need to measure and manage corporate reputation [17]. Different stakeholder groups have different expectations and thus management of reputation risk becomes crucial, difficult, and delicate [13]. Protecting and maintaining a good reputation is one of the risk manager's most important but most difficult tasks.

1.2 Aim of the Work

The understanding that drives reputational losses in the banking industry is unknown and the need for empirical studies is noticeable [9]. Considering the value of reputation and the high cost of reputational risk, the aim of this thesis is trying to measure reputational risk in banking by a rule-based model that systematically monitors two online external data sources and identifies reputational risk by these two indexes. We approach this problem, measuring reputation in a short time time-scale, indirectly by examining the sentiment in the financial news on a daily basis from the Internet and linking reputation with sentiment by the "memorized reputation" method. Our model mainly integrates Sentiment Analysis in Natural Language Processing, Lexicon Method, Logistic Regression in Classification, and Behavior Economics Theory. At present, our model only supports two data sources and one language - English. However, by verifying the feasibility of the methodology in this study, the model can be extended to include more data sources and languages. Our news corpus consists of economic and financial news articles from major newspapers from January 2018 to October 2020. Our index specifically relies on extracting sentiment and calculating the daily reputation for each organization from these news articles using computational analysis.

This thesis project is research and development oriented. The major contribution of our thesis is that it provides a unique rule-based reputation measurement model with a big amount of external data in the banking sector, and it visualizes the results in Tableau by user-friendly designed interfaces, which are also easy for banks to integrate the models in their current Microsoft Office system. Firstly, we consider the sources of external data and figure out creative ways to gather valuable data sets. Specifically, we collect the related external data not only for Credit Suisse, but also for big names in the banking sector. Second, we identify the methodology based on the previous work and our own research and then construct our unique model in 4 main modules: data acquisition module (scrapers and parsers) to monitor and download real-time data from the Internet, rule-based sentiment module to extract sentiment for each piece of information, reputation module to calculate daily reputation, and visualization module to show the reputation trends in a user-friendly way. Based on our sentiment model, we construct a time-series measurement of the public sentiment of an organization and its short-term reputation. Specially, we calculate the index of reputation for each of the large sets of financial news dating back to 2018. We then aggregate the individual organization scores into two daily time-series indexes. One possible application of our reputation measures is to monitor the reputational risk of any organization from the news data sources and help the risk management for that organization.

Except for the whole methodology we studied and the deployment of the model by python and Tableau, the methods we used to scrape these data sets are also useful in other practical projects. The process of collecting data is hard and tricky, but we finally found ways to deal with the problems, and the APIs we constructed to scrape related data can be directly used by banks or other organizations to identify other value of the data and develop more useful commercial applications. In the sentiment module, we have tried two possible ways: lexicon-based method and machine learning method. The lexicon-based method is finally be chosen as the better performance of it. Even though we only generate two indexes by two external data sources, Googlenews and Reuters; we still built the related modules for Twitter and collected a big amount of Tweets. More details will be discussed in the next chapters.

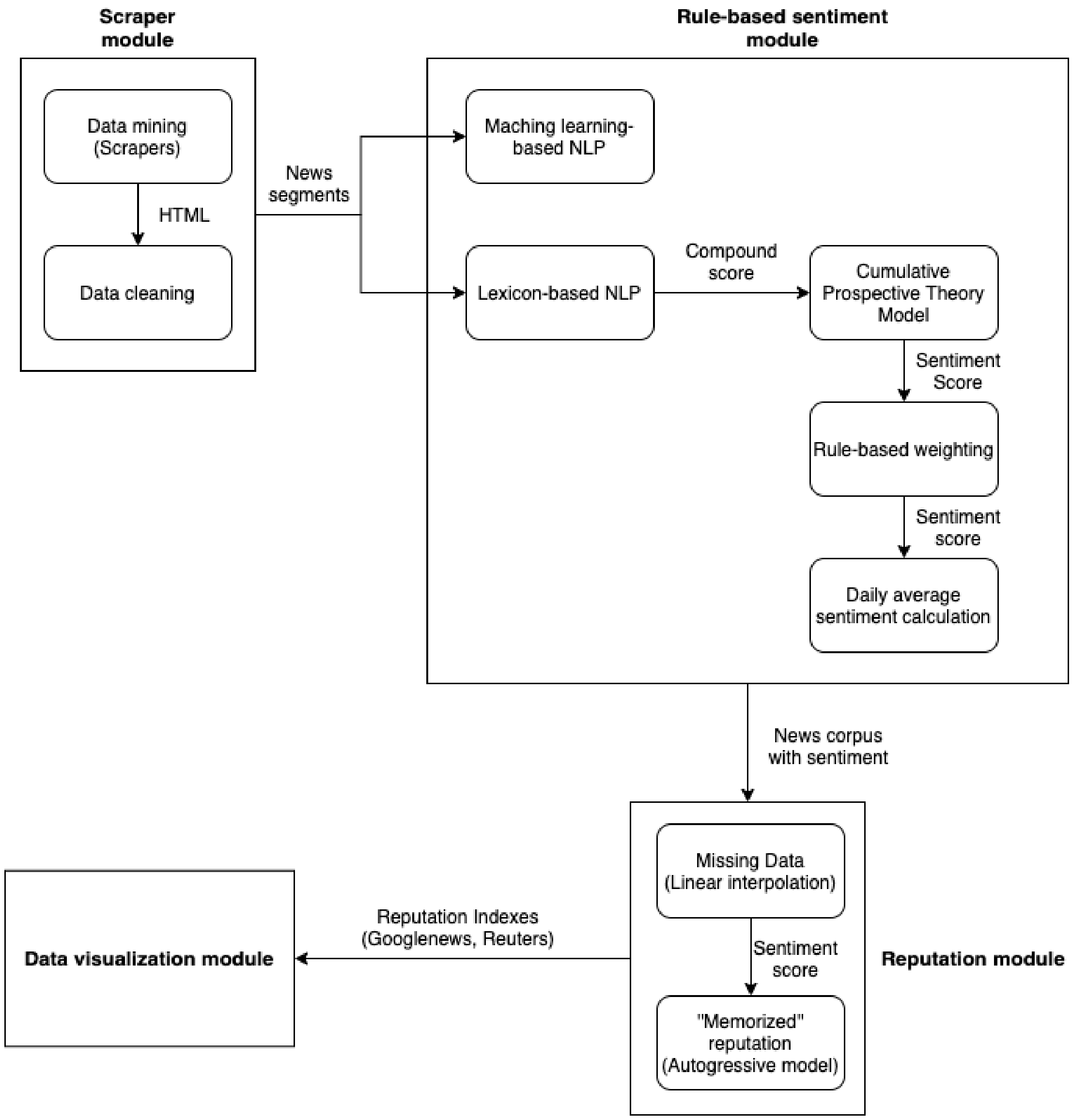

Figure 1.1: Model Structure Overview.

1.3 Thesis Structure

This paper is organized as follows. In chapter 2, we describe an overview of canonical works and techniques of research on links about reputation, sentiment, and risk management. Next chapter, we provide the important information of the data selection and data processing in this paper. Then we discuss the methodology used in the study, where significant definitions are given. Results, testing of the methodology, and conclusion are presented in the next chapter. The thesis closes with some discussion and concluding comments.

Chapter 2

Literature Review

In this chapter, we give a brief summary of the existing literature focusing on measuring and managing reputational risk. In this paper, we describe the previous literature about sentiment analysis firstly and literature about reputational risk then.

Known to Mukherjee 2014, there are two traditional ways of measuring reputation. One assigns a monetary valuation using market capitalization or return on assets. Another uses a relative approach of valuation as intellectual capital using an internal performance scorecard and other indices. The first one has been used in many earlier papers and is based on stock market reactions due to event study. The second one has not been used extensively due to its very nature being subjective.[1] As the Internet becomes much more popular, recent studies try to measure reputational risk based on the online media information feeds like Twitter, Google, Facebook, and so on. Our research is started from such studies.

2.1 Sentiment Analysis

Sentiment Analysis, or Opinion Mining, is a sub-field of Natural Language Processing (NLP) that tries to identify and extract opinions within a given text.[2] Importantly, the aim of sentiment analysis is to gauge the attitudes, sentiments, and emotions, but not meanings, of the speaker/writer based on the computational treatment of subjectivity in the text.

Go (2009) [2] from Stanford compares the accuracy of machine learning algorithms when trained with emotion data. Most importantly, they generate and open-source their training data for developers to train their own model, which is called sentiment 140. From the result of this paper, we try the SVM and Maximum Entropy method and finally choose the Maximum Entropy.

Pang (2008) [5] gives important definitions of sentiment analysis and covers techniques that promise to directly enable opinion-oriented information-seeking systems. They also summarized broader issues regarding privacy, manipulation, and economic impact that opinion-oriented information-access services will bring to. A similar definition is given by Liu (2015) [21], which defines the useful background and definitions in the sentiment analysis essentials with teaching lectures and slides.

Cha (2010) [22] illustrate an in-depth comparison of the influence of indegree, retweets, and men-

[1] These results are from Mukherjee 2014
[2] Parul Pandey, Simplifying Sentiment Analysis using VADER in Python, medium.com

tions in Tweets. The results reflect that popular users with high indegree are not necessarily influential in terms of spawning retweets or mentions; most influential users can hold significant influence over a variety of topics; influence is gained through concerted effort like limiting tweets to a single topic. The results from Cha help us to define the "influence" of each tweet in our model.

Nakov (2013) and his team [28] joined the SemEval 2013 and studies sentiment analysis in social media like Twitter. They generated their Twitter training dataset by Amazon Mechanical Turk along with additional test sets of Twitter and SMS messages.

Hutto (2014) [14] study and provide a rule-based lexicon for general sentiment analysis and compare its effectiveness with other typical state-of-practice benchmarks. Based on these, they then provide the idea that combing the lexical features with consideration for five general rules that embody grammatical and syntactical conventions for expressing and emphasizing sentiment intensity. This lexicon is used in our lexicon-based NLP model.

Cha (2010) [22] illustrate an in-depth comparison of the influence of indegree, retweets, and mentions in Tweets. The results reflect that popular users with high indegree are not necessarily influential in terms of spawning retweets or mentions; most influential users can hold significant influence over a variety of topics; influence is gained through concerted effort like limiting tweets to a single topic. The results from Cha help us to define the "influence" of each tweets in our model.

Swanepoel (2017) [8] proposes a reputational measurement matrix to measure and assess reputational risk nationally and internationally with the comparison of four key aspects ('who', 'where', 'what', and 'how').

Shapiro (2020) [3] study the state-of-the-art text sentiment analysis tools with a new time-series measure of economic sentiment derived from economic and financial newspaper articles. They compare the predictive accuracy of a large set of sentiment analysis models by different lexicons and trained machine learning model with articles that have been rated by humans on a positivity/negativity scale.

2.2 Reputational Risk

The concept of reputational risk is relatively new - it has been around for about a decade and only more seriously examined in the last 2-3 years [1]. Reputational risk has been the subject of significant attention in both academic literature and the financial press, yet direct evidence of reputational losses at financial firms has been limited [16]. Most of these studies focus on estimating the extent of reputational losses as market reaction to the operational loss announcement running and even study.

The first definition of reputational risk is from the Board of Governors of the Federal Reserve System (2004): "Reputational risk is the potential that negative publicity regarding an institution's business practices, whether true or not, will cause a decline in the customer base, costly litigation, or revenue reductions". In general, reputation risk is any risk that can potentially damage the standing or estimate of an organization in the eyes of third-parties [23]. Walter (2006) [36] defines the reputational risk for banking and financial industries as "the possibility of loss in the going-concern value of the financial intermediary - the risk-adjusted value of expected future earnings". Basel Committee, which gives the guidelines for banks, defines reputational risk as a negative belief by the stakeholders which can affect a bank's performance.

Reputational risk management is the management of factors that are a source of reputation because reputation is, to a large extent, a perception which forms outside of the company [35].

Perry (2005) [16] measure reputational losses by examining a firm's stock price reaction to the announcement of a major operational loss event and find the relation among market values, external events, internal fraud, and power of shareholder rights.

Bebbington (2006) [15] explained that an organization's reputation comprises of elements: financial status, quality of employees, management quality, environmental, social, and governance (ESG) strategy, and the standard of the goods and services provided.

Micocci (2008) [23] measure reputational effects for financial institutions by examining a firms' stock price reaction to the announcement of particular operational loss events with the OpVar database[3].

Fiordelisi (2011) [9] empirically study the determinants of reputational loss following operational losses in banking. By estimating a large sample of banks in Europe and the U.S between 2003 and 2008, they provide evidence that the probability of reputational damage increases as profits and size increase; and a higher level of capital invested and intangibles reduce the probability of reputational damage.

Mukherjee (2014) [25] explore the definitions surrounding reputational risk and study the regulatory requirements for banks on the management of reputatioanl risk with examples of leading banks of the European Union.

Blanc (2016) [1] write a book of their studies in a groundbreaking approach to reputation risk management for organizations combining the qualitative and quantitative approaches to understanding and managing reputation risk. This book introduces the fundamental definitions of reputational risk and illustrates the importance of managing reputation. The quantitative approaches in this book stand in the view of a risk manager by using the data from Reprisk to generate the Risk Heat Map for the organizations.

Farha [31] propose measurement technique quantifies reputation risk by estimating the reputation risk impact as the difference between the actual market capitalization loss from an event, and expectation had the event not occurred. The most significant results for us are that more than half the events had a reputation risk impact, and the initial reputation of a firm was an important factor driving the loss, with the reputation risk losses more than doubling when an event happens to a firm with a strong brand.

Mitic (2018) [24] find the results that reputation risk can be measured in terms of a single index, arising from a data mining process directed at the opinions in a complex multi-agent network and the results of the measurement process can be expressed directly in monetary terms by finding a correlation between the daily changes in the index and in sales. These two results are the fundamental idea of our reputation measurement model with the rule-based sentiment.

[3]OpData dataset supplied by OpVantage

Chapter 3

Data Preparation

In this chapter, we explain why we choose the data sources and how we construct the scrapers to download the daily data for ten banks worldwide between 2018 and 2020. Different from other studies, we collect the historical and daily corpus of financial news articles by our scrapers. First, we describe the lexicon and training data for constructing the predictive model to predict the sentiment of news articles. Then we explain the raw textual article data to which we apply the reputational risk model.

It is now well-accepted that self-reported information is not reliable data, especially when it comes to risks [27]. We have always taken an outside-in approach to reputational risk, by analyzing information from publicly available sources that supply comments, articles, and reports. Therefore, the sources of our testing and real-time data are all external data from publicly available websites, including news channels and social media.

To gain a more comprehensive understanding of the reputation of Credit Suisse and the banking industry, we want to compare the reputation trends of Credit Suisse and those of other similar banks. Except for Credit Suisse, we select other nine banks in a similar level to Credit Suisse from the List of Global Systemically Important Banks (G-SIBs), including HSBC, UBS, Citigroup, Deutsche Bank, Bank of China, Bank of America, Goldman Sachs, Morgan Stanley, and UniCredit. G-SIBs is a list of global systemically important banks that are racked and labeled by several authorities as systemically important financial institutions, depending on the scale and the degree of influence they hold in global and domestic financial markets.[1]

3.1 Data Selection

All of our model components are data-based. The Sentiment module and reputation module are constructed by the developed NLP text sentiment analysis techniques. There are two general methodologies for quantifying sentiment in the text - lexical method and machine learning method, which are built on different kinds of training data sets. We have tested both methods in our scenarios, but finally we decide to use lexicon-based method in this paper. We compare several official and common-used lexicon in lexical method to decide the suitable lexicon to use in our model and domain. We have also searched for and tested several existed training data sets in the machining learning method under our financial domain. Because of the high time & money cost of constructing a new domain-specific data set, creating the designed labeling training data sets is

[1] List of systemically important banks, https://en.wikipedia.org/wiki/List_of_systemically_important_banks

not the task of this work. However, the performance of sentiment analysis highly depends on the domain and size of the data set. Constructing a complex corpus in the domain is the most efficient way to improve the accuracy of sentiment prediction. Therefore, if the accuracy of sentiment analysis is the key priority, we highly recommend customizing the domain-specific training data or domain-specific lexicon. Even though it is time-consuming with high costs, it is still the most sufficient way to improve the performance of the sentiment analysis model. In this work, we mainly focus on the exploration and exploitation of the whole methodology of reputational risk measurement without studying the improvement of the performance of the sentiment prediction.

The labeled training data sets and lexicons are the fundamental components for constructing the sentiment analysis model, which helps the machine to learn the features automatically and predict the sentiment. The news feeds data, we used to predict sentiment and measure reputation, are collected by our own valuable scrapers. We have spent more than 1.5 months to generate such pairs of scrapers as the existed news sets are too expensive for our master thesis use goal and most websites have very complex anti-spider mechanisms to prevent the data downloading. For example, Twitter has updated its anti-spider mechanisms at least three times during Jun. 2020 to Oct. 2020, which leads our Twitter scraper cannot get any Tweets and requires us to find the updated algorithm and solutions. Thus, the scrapers and data we collected are also valuable for other study or practical application.

3.1.1 Lexicon in Lexical method

The traditional and easy but efficient method used in NLP is known as the Lexical methodology. This approach relies on pre-defined lists of words, called lexicons or dictionaries, with each word assigned a score for the emotion of interest [3]. Normally the scores of words are gained by lots of manual experiments and averaging the results finally. For example, one kind of lexicon is the valence lexicon, which contains a list of words with each assigned a score to indicate the degree of its positiveness or negativeness. There are a number of realization of lexical approach, which are popularly used in the academic and also real life, like Loughran and McDonald (2011), Harvard General Inquirer (GI) dictionary, Hu and Liu (2004), and so on. However, the lexical method is still a static rule-based scoring system basically.

We studied many pieces of paper and experiments about the commonly-used lexicons, and pleasantly found one newly-published paper from Shapiro (2020) [3] has the same results as us, which studied the different lexicons' performances in financial and economics domain. We finally choose and test positively an existed and powerful sentiment lexicon, VADER [14], to construct the sentiment analysis part in our model. VADER (Valence Aware Dictionary and Sentiment Reasoner) is an open-sourced lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.[2] From Hutto [14], we can know that the VADER lexicon performs as well as (and in most cases, better than) eleven other highly regarded sentiment analysis tools, including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms. The results of our own practical experiments are same to the results from Shapiro: the ML approach performs far worse than the lexical models, likely because the limited size of the training set for the unspecific domain.

Shapiro and Moritz (2020)[3] shared their own created lexicon for financial and economic domains, which is built based on VADER and included more words from financial news. We also tried the lexicon and compared the results with original VADER. However, in our case, the prediction performance of the two lexicons are very closed, so we continue to use VADER in the sentiment prediction process. Thank them for sharing the lexicon with us selflessly.

[2]VADER-Sentiment-Analysis, https://github.com/cjhutto/vaderSentiment

3.1.2 Training Data in Machine Learning

Recently, another approach employs machine learning (ML) to construct complex models for probabilistically predicting the sentiment of a given set of text. An ML predictive model is typically estimated/trained on a large training set of texts containing a mapping between textual utterances and sentiment ratings assigned by humans [3]. But importantly, the performance of the ML approach is only as good as its training set (that is, a data set pairing text with sentiment labels). Especially for the ML model involving deep learning, typically requires much larger training sets. To improve the performance of the prediction by applying different algorithms will help, but can be very limited. Significantly, a trained ML model is very sensitive for the domain associated with the training set. For example, “liability” generally expresses negative meaning in the real world, but in finance or economics, it is neutral. Thus, without the representative training set, the predictions will be less accurate.

We tried the machine learning approach based on different classifier algorithms and compared the testing results with Alec[2]. We accordingly choose machine learning on Maximum Entropy and SVM classification algorithms combined with TF-IDF feature engineering to construct our sentiment analysis model. In order to train a classifier, supervised learning generally requires hand-labeled training data and the performance of the classifier is highly based on the training data. The training data set in this study is Sentiment140 [2] from Stanford University [3], which is designed for the projects to discover the sentiment of a brand, product, or a topic on Twitter or longer articles. We have searched for different existed training sets and wanted to find the most suitable one, but most of the existed labeled training data are not good enough for our problem, as the lack of training data in the financial domain and the non-enough size of some training sets. So Sentiment140 is still the best choice for us to study this method. To improve the prediction performance, it is better to manually label the specific training data sets and train the classifiers by the domain-designed training sets.

As Sentiment140 is designed for the social media domain, not for our financial domain, the trained classifier has more inaccurate prediction results than the lexicon-based method, which sometimes can be a major disaster, especially for predicting important extreme cases like financial scandals. Due to the difficulties to customize an extensive training data set and the labor-intensive problem to manually create a comprehensive lexicon, we finally decided to deploy the lexical approach in our sentiment module.

3.1.3 Selection of External Data Sources - News Articles

After building the sentiment analysis model, we would like to screen the news to predict the sentiment and measure the organization’s reputation, on a daily basis, in English. Identifying the proper external data sources are one of our project objects. We have chosen three media channels to collect the historical data sets and daily data sets for each bank in the selected list; they are Twitter (social media), Googlenews (news aggregator), and Reuters(news media).

Twitter(TT) is a popular microblogging service where users create status messages (called “tweets”).[4] Private individuals, organizations, companies and the mass media can publish Tweets on Twitter to distribute short text messages. The Tweet has several characters like texts, comments, likes, retweets, hashtags, and so on. These characters can help us to search for the related Tweets and calculate the influence of the tweets and the contributors.

Googlenews(GG) is a news aggregator application, which presents a continuous flow of articles

[3] http://help.sentiment140.com/for-students

[4] Alec Go, Richa B., etc., Twitter Sentiment Classification using Distant Supervision

organized by thousands of publishers and magazines.[5] The news on Googlenews does not come from Google itself, but is compiled by computers using an algorithm from a large number of news sources, like Bloomberg, CNN, and so on. The sequence of the articles on the website is decided by several factors, like the popularity of the news in its original website, relevance or truthfulness of the articles, user personalization, and so on.

Reuters(RT) is an international business and financial news organization with a long history.[6] Reuters employs some journalists and publishes its articles everyday in different topics like finance, society, sports, and etc. Compared with Googlenews, which acts like an average of the news from many medias , the opinions and articles from Reuters are more obvious and clear.

We also maintain a blacklist with the media channels which cannot be reached by machine scrapers in general. For example, Bloomberg is on the blacklist as it doesn't allow to read the body text a without subscription. Until now, our blacklist only contains two channels: Bloomberg and Nasdaq. To ensure the model can continue to run when an unpredicted error comes, the time limitation for downloading each piece of news is 30 sec.

3.2 Data Preparation

The Vader lexicon and Sentiment140 are open-source data that can directly download from the websites. However, the most important data sets, historical and real-time data sets, can only be collected by official APIs (Application Programming Interface)[7] or scrapers. As we only use our data for this academic thesis and the official APIs are expensive for this project, we generate our own scrapers to collect the needed historical and real-time data online. Scrapping historical data is a much harder problem than downloading the daily data, because almost all of the information websites have very strict limitations on data collection and protection.

For each data source, two scrapers are created to download historical news and daily news separately. It is because the scraper for historical data only needs to be run once and the scraper for daily data should be run every day to renew the daily data corpus. Then the downloaded raw data will be cleaned, including filtering the un-efficient pieces, splitting the raw data into elements, encoding to the code standard "utf-8" (Universal Character Set/Unicode Transformation Format)[8], and removing the non-sense characters.

With the help of the RSS (Really Simple Syndication)[9], it is possible to extract the financial news of Googlenews for each organization on a daily basis. But importantly, the RSS of Googlenews only supports to reach the news "today" in the Internet time. For the historical news, we generate the scraper based on a python package "Googlenews". The variety of both historical and daily articles are formatted in HTML as well as the diversity of meta information. Each piece of Google news contains the title, publish date, media channel, and original link. By checking the reachability of the original link under the 30-second time limit, our crawlers access the original link and then save the body text of the news. Therefore, for each Google news, we finally have its title, publish date, media channel, original link, and body text. In the sentiment prediction module, we only use top 20 Googlenews and the newly-published ones everyday to calculate the daily sentiment, because Googlenews offers 100 popular news everyday and some of them are out-of-date.

Reuters only allows searching the news data for organizations in the date range of the last day, last week, last year, and past time. So our scrapers monitor Reuters every day and save the news of

[5]Wikipedia, Google News
[6]Wikipedia, Reuters
[7]https://en.wikipedia.org/wiki/API
[8]https://en.wikipedia.org/wiki/UTF-8
[9]https://en.wikipedia.org/wiki/RSS

"past day" as daily data. Historical articles are downloaded by setting the date range to past time in our scrapers. For every piece of Reuters news, we then get its title, publish date, link, and body text. Reuters offers small amounts of news everyday, so we use all of the articles in the sentiment prediction step.

Twitter API is the easiest and most beautiful method to scrape tweets. However, it only allows to freely download the limited amounts of tweets within 7 days. So we decided to create our own Twitter crawlers based on two open-source Twitter scrapers, twiter scraper [10], and twint[11]. Our Twitter crawler has no limit in the speed, amounts of tweets, or date range. The main idea is by simulating the swiping of the browser to get the JSON file, but without speed limit by not using the Middlewares like Selenium. Therefore for every tweet, we have the information about the text, publish time, id, username, replies, likes, retweets, hashtags, and mentions.

We test our methodology by analyzing the historical data of Googlenews, Reuters, and Twitter from 01.Jan.2018 to 30.Jun.2020. As discussed above, all the historical and daily data are scraped by our crawlers. The historical news data for some dates are missing randomly. First, it is because the crawlers cannot fully download all the historical data only by searching on the front page of the news website. To solve this, we create a date-checker to re-set the dates in the scrapers and re-run the scrapers. Besides, it is because some original news pages have already been no longer accessible, and this problem cannot be solved.

Our collection of historical and daily news data are valuable with practical and commercialized applications. In this work, we generate our model with real-time news data from the three sources on a daily basis. Based on our data and reputational risk model, there are more applications that can be developed. For example, further to predict the stock price of the organizations based on its short-term reputation.

[10]https://github.com/taspinar/twitterscraper
[11]https://github.com/twintproject/twint

Chapter 4

Methodology

This chapter provides important background information on reputational risk and the methodology used in our model. The methodology can be separated into three big parts: 1) train the natural language processing model (NLP) and heuristic-based lexicon model to prepare for the prediction of sentiment in the following parts; 2) scale sentiment scores by cumulative prospect theory model and average the scores by dynamic weightings; 3) and download the daily data to measure the reputation for each organization.

We first distinguish among reputation, sentiment, and opinion; and outline the basic definitions used throughout the analysis. Then we illustrate the approaches for sentiment analysis, including NLP algorithms and VADER lexicon. Next we explain our measurement of Reputational Risk (defined as Rep-Risk Measurement in the Table(4.1)).

Our methodology is closely related to recent work by Mitic (2018) [24] and Shapiro (2020) [3], who also measuring reputational risk by publicly available data sources. Known from Mitic, the basic idea of this methodology has been used by the reputation consultancy alva[1]. Mitic focuses on finding the correlation between the daily changes in the reputation index and in sales. Shapiro focuses on the economic sentiment derived from economic and financial newspaper articles. The current work focuses on measuring reputational risk by monitoring reputation trends derived from sentiment of financial newspaper, which are calculated by the Autoregressive model with the daily sentiment as the variable.

We tested our methodology by event study and case analysis. We find that the reputational events of banks are linked with the extreme results in our model, but not all extreme results are linked with reputational events. In addition, we create a user-friendly Tableau Dashboard for users to monitor the reputation trends and compare among different banks. More details will be discussed in the following sections.

4.1 Definitions

Sentiment in Cambridge Dictionary[2] is defined as a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something. The opinion is to indicate a broad context covering sentiment based on Liu(2015) [21].

[1] www.alva-group.com
[2] https://dictionary.cambridge.org/dictionary/english/sentiment

Reputation can be simply defined as a collection of opinions, but not only that. The Reputation of an Organization is influenced by its performance, policies, and people. Risk to reputation occurs where the organization fails to meet the expectations of a specific stakeholder group.[12] The reputation-related definitions are reflected in the following table, which is extended from Mitic(2018) [21] and Liu(2015) [12].

Table 4.1: Reputation-related definitions

Term	Definition
Sentiment	A view based on a feeling about a situation or a way of thinking about something
Opinion	Sentiment expressed by a holder of a target at a particular time
Reputation	Stakeholder perception of an organisation that can affect, positively or negatively, the business relationship between the stakeholder and the organisation
Reputation Risk (Rep-Risk)	The difference between stakeholder expectation and organisation performance
Reputation Event	An occurrence or action that affects Reputation
Rep-Risk Measurement	Numerical assessment of Reputation

The following definitions for sentiment, opinion, reputation are based on the definitions in Table(4.1). We only illustrate the fundamental definitions here; specific definitions of the terms in each step will be given in the methodology section.

Before re-scaling the sentiment, we define S, a standard measurement of sentiment, as a real number between -1 (most extreme negative) and +1 (most extreme positive).

$$S \in \mathbb{R} : -1 \leq S \leq 1 \tag{4.1}$$

where the threshold values are: positive sentiment ($0.05 \leq S$), neutral sentiment ($0 \leq S < 0.05$), and negative sentiment ($S < 0$).

Opinion, O, extends sentiment S to include: a unique identifier i, a timestamp t, a target G, an opinion's holder H, sentiment S.

$$O_{i,t} \ = \ \{G, H, S\} \tag{4.2}$$

Reputation at time t can be defined as the collection of opinions $\{O_{i,t}\}_{t \in T}^{i \in I}$, where T is a set of discrete time and I is a set of unique identifiers. The reputation of an organization G at time t, $R_G(t)$, can be then defined as some generic function f of the opinions.

$$R_G(t) \ = f(O_{i,t}) \tag{4.3}$$

where $i \in I$ and $t \in T$.

The function f, defined in this paper, is a AR(1) model[3] of past reputation and current sentiment.

To extend the definition of reputation $R_G(t)$ over the times in T, we then define the reputation $\hat{R}_G$, of the target G as the time series

$$\hat{R}_G \ = \{R_G(t)\}_{t \in T} \tag{4.4}$$

This definition is based on reputation measurements over an period. It is not sufficient to deal with cases where a potential opinion holder notes a small number (perhaps one only) of isolated comments, and formulates his/her own opinion based solely on that [24].

[3] Autoregressive model

4.2 Computational Methods for Sentiment Analysis

Predicting sentiment is essential for measuring reputation in our model. Sentiment expression is compositional and contextual. As know from Sharipo (2020) [3], general approaches for sentiment analysis emphasizes two key objectives in characterizing the sentiment of a given set of text: domain specificity and complexity. Domain refers to the subject matter of the corpus of text that one wants to analyze and complexity relates to all of the multifaceted aspects of a set of text beyond just the prevalence of particular words, which means words can be in different sentiment in different domain and the sentiment of the next word can be completely changed by the previous word. For example, the word "liability" is generally a negative word whereas it is neutral in the financial domain. Another example can be that "good" to "not good" changes the sentiment orientation totally.

Two NLP text sentiment analysis techniques are frequently used in recent studies, which are the lexical method and machine learning techniques. Even though lexicon-based method is a static method with requirements to update the lexicon words regularly, recent advances in lexical methods have focused on accounting for the contextual characteristics of words, which empower the sentiment analysis more. The Machine learning method is a recently developed approach that captures features and predicts sentiment by training on a big set of labeled texts containing a mapping between textual utterances and sentiment ratings assigned by humans. We firstly try the ML method and then use the lexical method. We finally decide to deploy the lexicon approach into our model so we first illustrate this method and then explain the machine learning method.

4.2.1 Heuristic-Rules Sentiment Lexicon Models - VADER

The pure lexicon-based measurement is simple and transparent. The positiveness of each news article is constructed by calculating the proportion of words in the article that are positive minus the proportion of words that are negative. This proportion calculation is mathematically equivalent to averaging the word-specific valence scores across all words in the article, where positive words are assigned a score of 1 and negative words a score of -1 [3]. But the pure lexicon-based sentiment analysis approaches often ignore important lexical features and general sentiment intensity differentials for features within the lexicon.

Thus, it is also important to account for the contextual characteristics of words. One simple solution is to build the models with defined simple heuristic rules which modify the sentiment scores of each word. The VADER is one application of such a solution. VADER adjusts each word's initial sentiment score (also called valence score) by a set of simple rules on its context within the sentence. VADER is a sentence-based sentiment classifier, which consists of a lexicon and a set of heuristic rules. A negative sentence with a negativity score is assigned by VADER by aggregating across negativity scores of words within the sentence.

VADER aims to classify the polarity of a piece of text as positive, negative, or neutral, but does not try to determine if a sentence is objective or subjective, fact, or opinion. VADER lexicon contains the top 400 most positive and negative social media text snippets with manual word ratings from -4 (extremely negative) to +4 (extremely positive), which are also called valence scores. VADER analyzes a piece of text by checking any of the words in the text are present in the lexicon. By adding the heuristic-rules, then Vader modifies each word's score by five rules related to negation (eg., being preceded by a negation word like "not"), punctuation (eg., the exclamation point), capitalization (eg., ALL-CAPS), following the contrastive conjunction word "but", and degree modifier such as "extremely", "slightly", etc.. For example, if a word is preceded within three words by a negation term, the word's valence score is multiplied by -0.74, indicating that negation reverses the valence of the word, though it reverses it by less than 100%. All the ratings of Vader, like both the initial unigram negativity labels (weights) and the scalar multiples associated with the rules, are obtained from a large-scale human rating process (using Amazon's Mechanical Turk). The detailed ratings of each heuristic rule are as following:

1. Negation: scalar multiple equals to -0.74, which means the valence score is multiplied by -0.74. We check 1 word, 2 words, and 3 words preceding the lexicon word positions. There is a trick in this situation that we don't use the valence of "no" as a lexicon item. Instead, we set it's valence to 0.0 and negate the next item which is preceded.

2. Punctuation: we only check the cases up to 4 of exclamation pints (amplifier equals to 0.292), with less than 3 question marks (amplifier equals to 0.18), with 3 more question marks (amplifier equals to 0.96). The final amplify for punctuation equals to amplify of exclamation points pluses amplify of question marks.

3. Capitalization: ALL-CAPS is to emphasize a word and increase the sentiment intensity. If the valence score of the laden word is negative, it will be adjusted by decreasing 0.733. If it is positive, it will be added 0.733.

4. Constractive conjunction: if the laden word is before the but, its sentiment will multiply 0.5; if it is after but, its sentiment will multiply 1.5.

5. Degree modifier (booster words): booster of increasing the sentiment intensity equals to 0.293 and booster of decreasing the sentiment intensity is -0.293.

After the heuristic-rules, for each piece of text, VADER produces four sentiment metrics, Positive, Neutral, Negative, and Compound. The Positive, Negative, and Neutral scores represent the proportion of text that falls in these categories. The Compound score (the most useful metric to measure the sentiment by Vader) is computed by summing the rating (valence score) of each word of the text and then normalizing the sum to be between -1 (most extreme negative) and +1 (most extreme positive). The normalization function is:

$$C = \frac{C'}{\sqrt{(C')^2 + 1}} \tag{4.5}$$

where C is the compound score with range of [-1, 1] and C' is the sum of the adjusted valence scores. The threshold values of compound score C are: positive sentiment ($C \geq 0.05$), neutral sentiment ($-0.05 < C < 0.05$), and negative sentiment ($C \leq -0.05$). In our model, as we would like to emphasize the negative effects of the articles with negative sentiment, we define our threshold values in this paper as: positive sentiment ($C \geq 0.05$), neutral sentiment ($0 < C < 0.05$), and negative sentiment ($C \leq 0$)

For example, the sentence "The food is good and the atmosphere is nice" has two words in the lexicon (good and nice) with ratings of 1.9 and 1.8 respectively. The example sentence is rated as 45% positive, 55% neutral, and 0% negative. The compound score of the example sentence is 0.69, which is the normalized sum of all of the lexicon ratings (1.9 and 1.8). This compound score is larger than 0.05 and indicates the example sentence is pretty positive.

Unfortunately, VADER is designed for the social media domain rather than the finance domain. In addition to domain specificity, the size of lexicon is also important for predicting sentiment. Constructing a new lexicon is very time-consuming and costly. So again, if accuracy is in high priority, a specific-defined lexicon is recommended.

Thanks to Adam Hale Shapiro, Moritz Sudhof, and Daniel Wilson, the authors of the paper "Measuring News Sentiment", they constructed a new lexicon based on Vader for financial domain news articles and shared it with me generously. They attempt to infer the sentiment orientation for all unique words in their full corpus of 238685 news articles. Their result is that the newly constructed news lexicon, which combined with the LM and HL lexicons and augmented with a negation rule yields the highest predictive accuracy for the test set of labeled news articles. As we don't have a labeled test set, so we can only decide the lexicon in our model based on manual event analysis with extreme points. We found the performance of VADER, augmented with five heuristic rules, is slightly more suitable for our cases. The compound scores from VADER will then be inputted into the next module to construct the indexes.

4.2.2 NLP Sentiment Models

Machine Learning (ML) techniques can potentially help to identify the contextual characteristics that contribute to the sentiment by learning sentiment weights with the ability to predict the sentiment of an entire expression. But the disadvantage of ML is the requirement of a large size of labeled training data sets in the related domain, which are very time-consuming and expensive.

It is necessary to apply data pre-processing techniques because the type of textual content that appear in different sources, like newspaper and social media, is referred to as unstructured text. The pre-processing techniques reduce the complexity of the documents to simplify data handing [19]. In this paper, we use the algorithm of transforming the documents from the unstructured text into the structured vector space model (VSM).

The fundamental concept of the representation of a document as a vector is considered it as a Bag-Of-Word (BOW) model, which means each document is represented by the words in which it contains [33]. For example, if we consider a corpus as a set of documents $D = \{d_1, d_2, d_3, ..., d_n\}$ and a dictionary as the set of words (terms) that appear in the corpus $A = \{a_1, a_2, a_2, ..., a_m\}$, then corpus D can be represented as a document-term matrix where rows and columns are indexed by the documents and the words respectively. Each element in the matrix is the weight of each word in the related document. Weights of each entry of w_{ij} can be determined by many approaches, like binary, TF, and so on.

$$D = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \cdots & a_m \end{array} \\ \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix} \end{array} \tag{4.6}$$

Based on the above document-term matrix, each document is mapped into an m-dimensional space:

$$\forall d_i \in D, 1 \leq i \leq n \tag{4.7}$$

$$\Phi : d_i \longmapsto \Phi(d_i) = (w_{i1}, w_{i2}, ..., w_{im}) \in \mathbb{R}^m \tag{4.8}$$

Then, the initial steps to data pre-processing of the financial news are listed below [10]:

1. Choose the desired scope and domain of the text to be processed like documents, paragraph, and sentences.
2. Tokenize: break the text into discrete words (token).
3. Remove stopwords: eliminate common words such as the, an ,a.
4. Normalize spelling: unify misspelings and other spelling variations into a single token.

5. As punctuation and upper cases are part of the heuristic rules, we didn't remove the punctuation or normalize case.
6. Since our data is scraped by our crawlers, so other detailed data cleaning steps are also required. For example, use regular expression to remove the strange characters or garbled characters.

We applied Porter stemming for removing various suffixes such as -ED, -ING, -ION, and so on from conflated words. In fact, this process will reduce the total number of features during the extraction that leads to a decrease in size and complexity of the document-term matrix [29].

As mentioned above, there are many methods to determine the weightings w in the document-term matrix. The main goal of a term-weighting method is to assign appropriate weights to terms in order to enhance the effectiveness of feature extraction [10]. Keeping the relevant features and eliminating the extraneous features together determine the effectiveness. As shown in Paltoglou and Thelwall, the term weighting methods based on BM25, a variant of the original TF-IDF, provide significant increases in the performance of sentiment analysis.

TF-IDF is a numerical statistic method that allows the determination of weight for each term (or word) in each document [7]. The method is often used in NLP or in information retrieval and text mining [18].

TF, Term Frequency, computes the number of repetitions (frequency) of a word (term) a in the document d [11].

$$tf(a,d) \quad = \frac{f_{a,d}}{\sum_{k=1}^{n} f_{k,d}} \tag{4.9}$$

where $f_{a,d}$ is the raw count of a term a in a document d, $f_{k,d}$ is the raw count of any term k in the document d, and the document d has n terms.

IDF, Inverse Document Frequency, determines the weight of rare words across all documents in the corpus.

$$idf(a,D) \quad = log(\frac{D}{df_a + 1}) \tag{4.10}$$

where D is the number of documents in the collection and df_a is the document frequency of term a in the collection. $df_a + 1$ is to prevent the denominator from 0.

TF-IDF determines the relative frequency of words in a specific document through an inverse proportion of the word over the entire document corpus [7].

$$w_{ij} \quad = tfidf(a_i, d_j, D) \quad = tf(a_i, d_j) \times idf(a_i, D) \tag{4.11}$$

As known from Alec Go, Max Entropy (MaxEnt) classification performs very well as it handles feature overlap better. We first try Max Entropy models, which are feature-based and the main idea behind which is that one should prefer the most uniform models that satisfy a given constraint [2]. Max Entropy method is the same as using Logistic Regression to find a distribution over the two classes.

$$P_{ME}(c|d,\lambda) \quad = \frac{exp[\sum_i \lambda_i f_i(c,d)]}{\sum_{c'} exp[\sum_i \lambda_i f_i(c,d)]} \tag{4.12}$$

where c is the class, d is the tweet, and λ is a weight vector. The weight vectors decide the significance of a feature in the classification [2].

We then use k-fold cross validation to select the most relatively accurate classifier. k is the single parameter in this procedure, which refers to the number of groups that a given data sample is to be split into. It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.[4] The chosen of k is a bias-variance trade-off as the computation complexity is crazily increasing with an increase of k. it is general to choose k as 5 or 10. In this paper, k equals 5. Cross Validation is a statistical method used to estimate the skill of machine learning models [6]. In most other regression procedures (e.g. logistic regression), there is no simple formula to compute the expected out-of-sample fit. Cross-validation is, thus, a generally applicable way to predict the performance of a model on unavailable data using numerical computation in place of theoretical analysis.[5]

We also consider a very popular developed transfer-learning model known as BERT, which is developed at Google by Devlin, Chang, Lee, and Toutanova (2019). BERT (Bidirectional Encoder Representations from Transformers) generates context-aware word and document embedding with a bidirectional approach. BERT is a very powerful model to incorporate context and sequential information in language modeling. However, as the lack of suitable training data for BERT to capture efficient features, the prediction accuracy of BERT is not enough for our model.

We hold the same opinion as in Shapiro (2020) [3]: lexical model for the analyses of financial news is preferred even though BERT seems promising. First, the size of training set is very limited, thus there is less confidence in the robustness and generalizability of BERT's performance across the entire news corpus than in that of the lexical model. Lexicon models are more common and easier to implement. Common critiques of ML methods are that they are "black boxes" and difficult to implement and replicate.

4.3 Cumulative Prospect Theory Model

As constructed by Tversky and Kahneman (1992), cumulative prospect theory treats gains and losses separately and the valuation rule is two-part cumulative functional with S shape. The five major phenomena of choice, which violate the standard model and set a minimal challenge that must be met by any adequate descriptive theory of choice [34]:

1. Framing effects: formally equivalent descriptions of a decision problem can elicit different responses, which means that the assumption that variations of form do not affect preference and choice is violated.
2. Nonlinear preferences: there is abundant experimental evidence for the notion that people overweight small probabilities and underweight larger ones.
3. Source dependence: people's willingness to bet on an uncertain event depends not only on the degree of uncertainty but also on its source. Evidence indicates that people often prefer a bet on an event in their area of competence over a bet on a matched chance event, although the former probability is vague and the latter is clear. (Heath and Tversky, 1991)
4. Risk seeking: people ofter prefer a small probability of winning a large prize over the expected value of that prospect. Risk seeking is prevalent when people must choose between a sure loss and a substantial probability of a larger loss.
5. Loss aversion: losses loom larger than gains (Kahneman and Tversky, 1984; Tversky and Kahneman, 1991). The observed asymmetry between gains and losses is far too extrme to be explanined by income effects or by decreasing risk aversion.

[4] https://machinelearningmastery.com/k-fold-cross-validation/
[5] Wikipedia, Cross Validation, https://en.wikipedia.org/wiki/Cross-validation_(statistics)

From Yang and Jiang (2014) [38], we can know that CPT mainly includes two parts, namely, the value function and the cumulative weighting function of probability.

The value (utility) function u is defined by Tversky and Kahneman as follows:

$$u(x) = \begin{cases} x^{\alpha}, & x \geq 0 \\ -\lambda(-x)^{\alpha}, & x < 0 \end{cases} \tag{4.13}$$

For $\alpha < 1$, the value function exhibits risk aversion over gains and risk seeking over looses. Furthermore, if λ, the loss-aversion coefficient, is greater than one, individuals are more sensitive to losses than gains.

The weighting function g proposed by Tversky and Kahneman is:

$$g(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}} \tag{4.14}$$

Prelec (1998) proposes an alternative specification for the weighting function with the nerly identical shape to that of Tversky and Kahneman's weighting functions:

$$g(p) = e^{-(-lnp^{\gamma})} \tag{4.15}$$

Tversky and Kahneman (1992) estimated $\alpha = 0.88, \lambda = 2.25, \gamma = 0.61$ for gains, and $\gamma = 0.69$ for losses. Camerer and Ho (1994) overall estimated $\alpha = 0.32, \gamma = 0.56$. Wu and Gonzalez (1996) estimated $\alpha = 0.48, \gamma = 0.74$ (as shown in Fig.4.1). The fact that these parameter estimates are robust to the use of different data from different tasks leads one to believe that these parameterizations are useful for estimation with non-experimental data.[6]

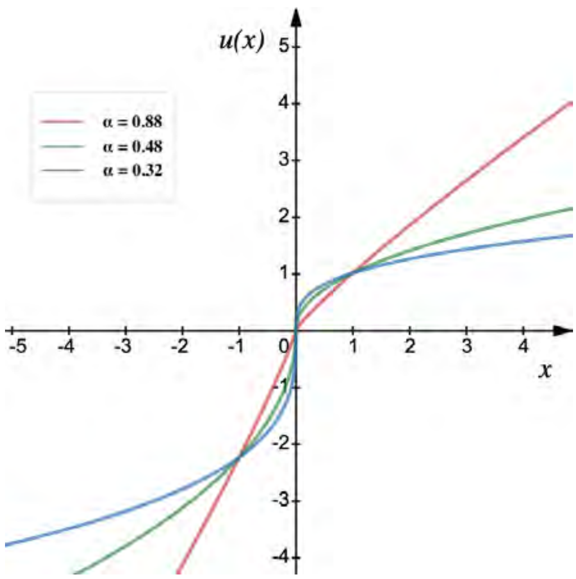

Figure 4.1: Tversky and Kahneman's value function with different α parameters.

In this paper, in order to amplify the negative scores and emphasize the negative effects, compound scores will be scaled by the cumulative prospective model. According to the parameterization results from the previous literature and the range of our sentiment sores, we choose $\lambda = 2.25, \alpha = 0.5$. We have tried two $\alpha = 0.5$ and $\alpha = 0.32$, the final outputs with $\alpha = 0.5$ are more suitable for our case.

[6]W. Neilson and J. Stowe, A Further Examination of Cumulative Prospect Theory Parameterizations, The Journal of Risk and Uncertainty, 24:1; 31-46,2002

4.4 Constructing Reputation Indexes

Based on the fundamental definition of sentiment and reputation in Mitic(2018), we defined the Reputation Index(RI), including Googlenews Reputation Index (gRI) and Reuters Reputation Index (rRI), to measure the short-term reputation of organizations based on the sentiment of daily financial news articles from the Internet. Different with previous studies, we collect the news data sets by our own-generated scrapers and we combined existed models like the Cumulative Prospective model and Autoregressive model into the procedures of reputation measurement; we developed the NLP model based on the existed lexicon and tested the results by event analysis and back-testing. As discussed in the previous sections, we have tried both the ML approach and the lexical-based approach to predict sentiment polarity, but we finally choose to use lexical-based approach as the sentiment analysis model in our model. Figure 4.2 provides an overview of the whole process of our reputation measurement model and figure 4.3 shows the methodology from the perspective of data flow.

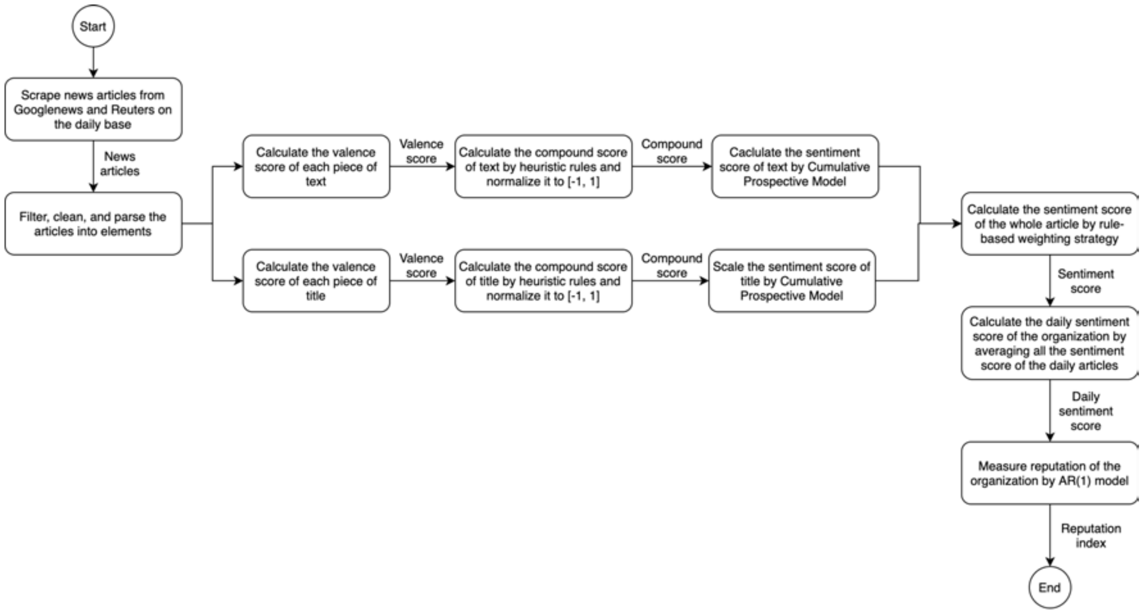

Figure 4.2: Methods and process approach of reputation measurement overview.

The process of reputation measurement is constructed in the following:

1. Data mining stage for extracting information from the publicly available media sources on a daily basis. As described in Chapter 3.1, for our selected banks, we download tweets from Twitter, news from Googlenews, and news from Reuters. Each item received is termed as ‘opinion’.

2. Data cleaning stage to parse the original HTML files and filter the non-efficient items. In this step, it is also important to structurelize the cleaned data into news segments.

3. For each opinion, predict the sentiment polarity by the heuristic-based lexicon model and get the compound score C, which is in the range of -1 (most extreme negative) to 1 (most extreme positive).

$$C \in \mathbb{R} : -1 \leq C \leq 1 \tag{4.16}$$

where the threshold values are: positive sentiment ($0.05 \leq C$), neutral sentiment ($0 \leq C < 0.05$), and negative sentiment ($C < 0$).

4. Scale each compound score by the cumulative prospect theory model and get the sentiment score S, which ranges from -2.25 to 1.

$$S(x) = u(x) = \begin{cases} x^{\alpha}, & x \geq 0 \\ -\lambda(-x)^{\alpha}, & x < 0 \end{cases} \tag{4.17}$$

where $x \in C, \alpha = 0.5, \lambda = 2.25$. Then the threshold values for sentiment are expended to: positive sentiment ($0.24 \leq S$), neutral sentiment ($0 \leq S < 0.24$), and negative sentiment ($S < 0$).

5. Special step only for opinions from Googlenews and Reuters as every piece of news articles contains one title and one body text with paragraphs . Average the sentiment of title (S_{title}) and the sentiment of paragraphs (S_{paras}) based on pre-defined rules to get the sentiment of the whole news (S).[7]

$$S(x) = w_{title} * S_{title}(x) + w_{paras} * S_{paras}(x) \tag{4.18}$$

The rules regarding to the weightings are (Table 4.2):

- Both S_{title} and S_{paras} are negative, and S_{title} is less than S_{paras}:
$w_{title} = 0.7$, and $w_{paras} = 0.3$, if $S_{title} \leq -0.5$, $S_{paras} \leq -0.5$, and $S_{title} \leq S_{paras}$
- Both S_{title} and S_{paras} are negative, and S_{title} is bigger than S_{paras}:
$w_{title} = 0.3$, and $w_{paras} = 0.7$, if $S_{title} \leq -0.5$, $S_{paras} \leq -0.5$, and $S_{title} \geq S_{paras}$
- S_{title} is negative and S_{paras} is not negative:
$w_{title} = 0.7$, and $w_{paras} = 0.3$, if $S_{title} \leq -0.5$ and $S_{paras} > -0.5$
- S_{title} is not negative and S_{paras} is negative:
$w_{title} = 0.3$, and $w_{paras} = 0.7$, if $S_{title} > -0.5$ and $S_{paras} \leq -0.5$
- S_{title} equals to zero and S_{paras} is not zero:
$w_{title} = 0$, and $w_{paras} = 1$, if $S_{title} = 0$ and $S_{paras} \neq 0$
- S_{title} is not zero and S_{paras} equals to zero:
$w_{title} = 1$, and $w_{paras} = 0$, if $S_{title} \neq 0$ and $S_{paras} = 0$
- Others:
$w_{title} = w_{paras} = 0.5$

Table 4.2: List of the rule-based weightings for sentiment calculation.

S_{title}	S_{text}	w_{title}	w_{paras}	conditions
Negative	Negative	0.7	0.3	$S_{title} \leq -0.5$, $S_{paras} \leq -0.5$, and $S_{title} \leq S_{paras}$
Negative	Negative	0.3	0.7	$S_{title} \leq -0.5$, $S_{paras} \leq -0.5$, and $S_{title} > S_{paras}$
Negative	Positive	0.7	0.3	$S_{title} \leq -0.5$ and $S_{paras} \geq 0.24$
Negative	Neutral	0.7	0.3	$S_{title} \leq -0.5$ and $-0.5 < S_{paras} < 0.24$
Positive	Negative	0.3	0.7	$S_{title} \geq 0.24$ and $S_{paras} \leq -0.5$
Positive	Positive	0.5	0.5	$S_{title} \geq 0.24$ and $S_{paras} \geq 0.24$
Positive	Neutral	0.5	0.5	$S_{title} \geq 0.24$ and $-0.5 < S_{paras} < 0.24$
Neutral	Negative	0.3	0.7	$-0.5 < S_{title} < 0.24$ and $S_{paras} \leq -0.5$
Neutral	Positive	0.5	0.5	$-0.5 < S_{title} < 0.24$ and $S_{paras} \geq 0.24$
Neutral	Neutral	0.5	0.5	$-0.5 < S_{title} < 0.24$ and $-0.5 < S_{paras} < 0.24$

6. The average sentiment for each organization will then be calculated on a daily base.For data source m on day t, call its average sentiment AS_t, ranging from [-2.25, 1].

$$AS(S_t) = \frac{\sum_{i=1}^{n} S_t}{n} \tag{4.19}$$

[7]For each opinion from Twitter, there is only one sentiment score - calculated by the text of the tweet.

where n is the total number of opinions from data source m on day t. The threshold values for average reputation index are: positive sentiment ($0.24 \leq AS$), neutral sentiment ($0 \leq AS < 0.24$), and negative sentiment ($AS < 0$).

7. The average sentiment score will be inputted into Autoregressive model, AR(1) model, to calculate the daily reputation for the organization m.

$$Rep_m(t) \ = \sum_{t^{'}<t} Rep_m(t^{'}) + a * AS_m(S_t) \tag{4.20}$$

We selected the starting points for all banks' reputation as 0. So negative reputation is showing by the points under 0 line and positive reputation is above the 0 line. However, except the exact reputation value points, the changes of reputational indexes are also significant information for reputational risk analysis. This will be discussed in the Event Analysis section.

The data preparation algorithms for Reuters and Googlenews are different as they provide different amounts of news every day. Googlenews is a news aggregator so it contains its own ranking algorithms according to the popularity of the news, the level of the published agency, and so on. So it always provides the news published "today" together with the news published "before" as long as they are still popular and highly discussed. According to our method, we only keep the top 20 news offered by Google and the newly-published news on that day. But we save all the news provided by Googlenews in the database. Reuters only provides the news it published every day, so we will save and use all the articles from Reuters.

We then provide a visualized application of these two Reputation Index, Googlenews Reputation Index and Reuters Reputation Index, investigating the news sentiment impact of reputation shocks. We also generated a series of Python APIs (Application Programming Interface)[8] for users to practically calling each module component. Therefore, each module and the functions inside each module can be used into other applications individually without big changes of the code.

[8]https://en.wikipedia.org/wiki/API

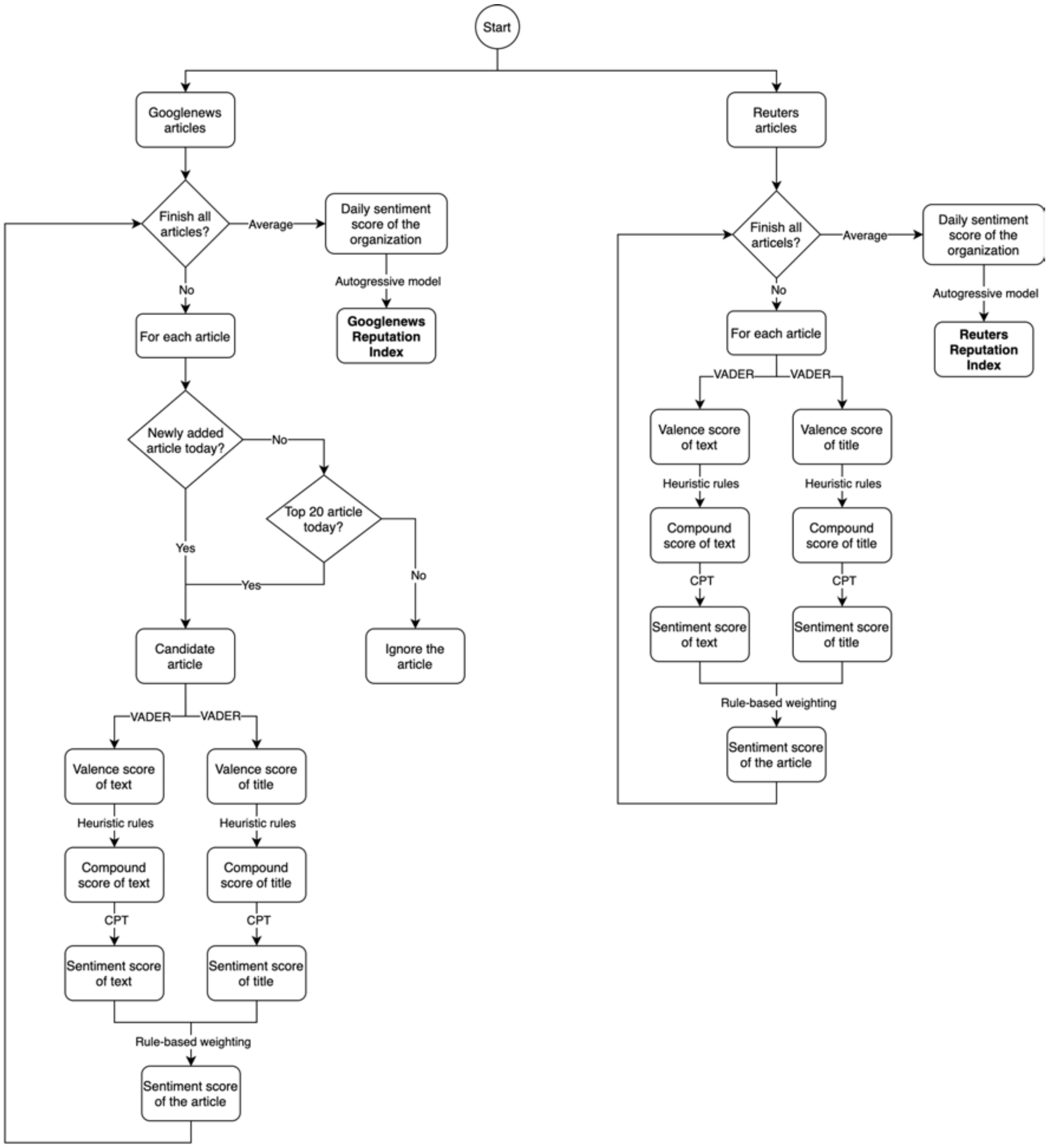

Figure 4.3: Methods and process approach of reputation measurement overview.

Chapter 5

Results

5.1 Empirical Results

We built two practical applications using these reputational indexes. First, we created the Tableau dashboard (as shown in figure 5.1) for users to monitor the whole reputation trends for each organization between Jan. 2018 to Oct. 2020. Second, we created the APIs in Python language to make the functions and modules to be much more easily callable.

In the Tableau Dashboard, The orange line represents the reputation from Reuters and the blue one is the reputation from Googlenews. Two triggers are important for users to monitor the organization's reputation: 1) the exact negative points, which represent the negative reputation at that time; 2) the huge decreasing gaps, especially same in both lines, can be the signal for some on-going reputational events. We will illustrate some typical cases later in the next section.

In figure 5.1, the left top is the reputation trends for all the selected banks, which represents the reputation changes of the top level in the banking industry. The left bottom is the reputation trends for Credit Suisse only. We put them here as we focus on the reputation trends of Credit Suisse and we wouldl liek to compare the indexes with the industry. We can see that some big decreased lines of the reputation of CS may not because of CS itself, but because of the decreasing in reputation of the banking industry. However, when the whole industry is quite flat, the big changes in CS prove the reputational events of CS. On the right side of the plane is the reputation indexes of the selected banks separately. From this image, we can easily compare the reputation of these banks. For example, Bank of China was the first one who suffered from the reputation decrease since Dec. 2019 and was followed by the same trends from the other banks. We analyzed the news during that time found almost all the articles were discussing the big shock about coronavirus starting from China and the block of Wuhan. Another example is that UniCredit always has the lowest reputation results as the big environment of the country's economy is not good enough when compared to other big names.

The distribution of Reputation Indexes 5.2 shows that reputation for banks, in general, is around neutral and slightly positive, but there are still few extreme reputation cases for them.

We can see from Figure 5.3 that the reputational index of Googlenews and of Reuters have quite similar trends for the banking industry. It is better to analyze reputation by combining the two indexes together at the same time. Not only the points below zero are important and great signals for relatively bad reputation, but also the similar decreasing changes of the two indexes are significant signals for reputational events. And there are four obvious big decreasing of reputation

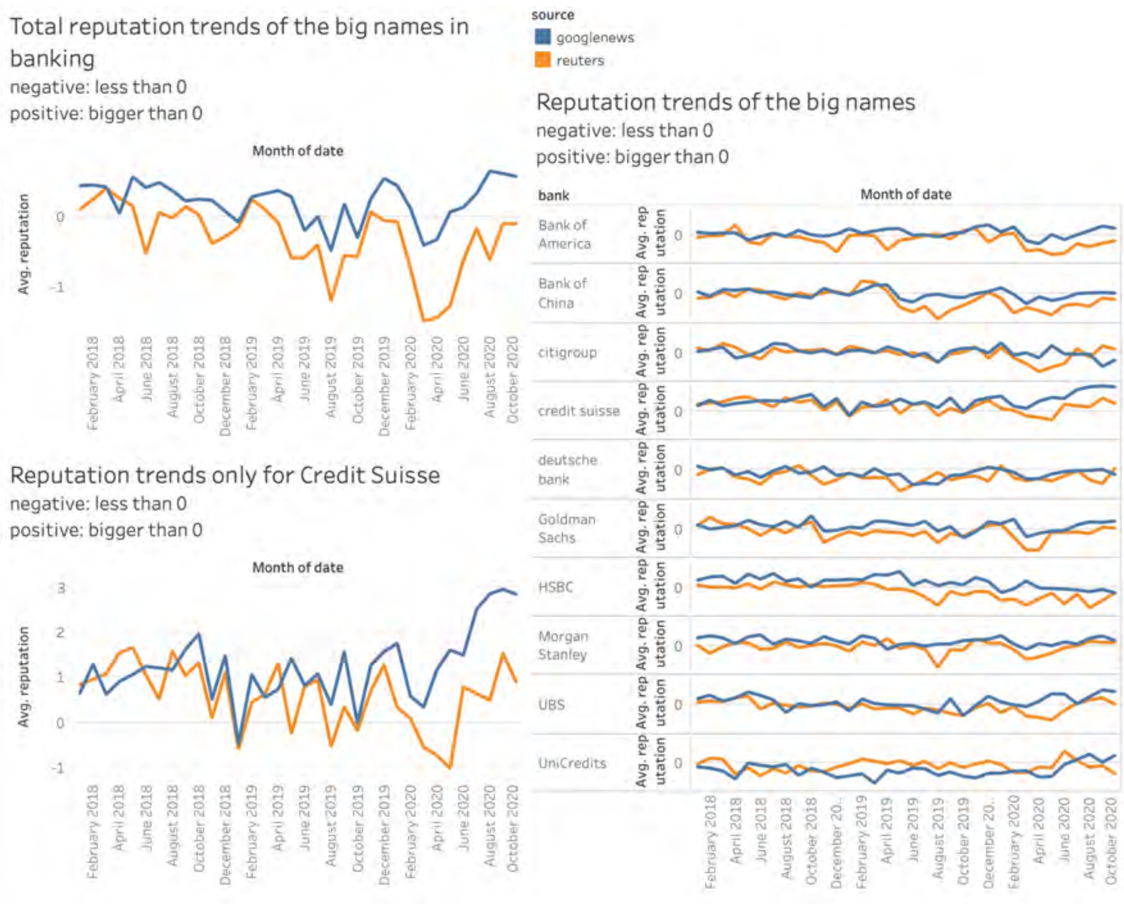

Figure 5.1: Tableau Dashboard for practical using.

for the big names. The biggest one happened around Jan. 2020 and continued to around May. 2020 as the coronavirus situation became worse since Jan. this year. We outputted all the news articles during that time and found most of the news were discussing about coronavirus. More details will be discussed with the example of Credit Suisse in the next section.

5.2 Event Analysis for Credit Suisse

We did the event analysis for credit suisse based on the historical famous reputation events from Wikipedia and the Internet. We find exact decrease trends during the time the reputational events happened and we manually analyzed all articles during those periods. We discuss these events in the order from the most recent to the oldest.

- CEO of Credit Suisse Tidjane Thiam exits after spying scandal on 7th Feb. 2020.

  We can see from figure 5.4, there is a very obvious decrease in reputation of Credit Suisse since 3rd Feb. 2020 to May 2020. Except for the quit of CEO, Credit Suisse has also been reported the big lawsuit in India since Mar. 2020. These two big reputational events, plus the effects of coronavirus, lead the reputation of CS to decrease to the lowest point.

  We outputted all the articles for Credit Suisse during that time and found more than 90% articles were talking about the scandal, the quits, the lawsuit, and coronavirus. However,

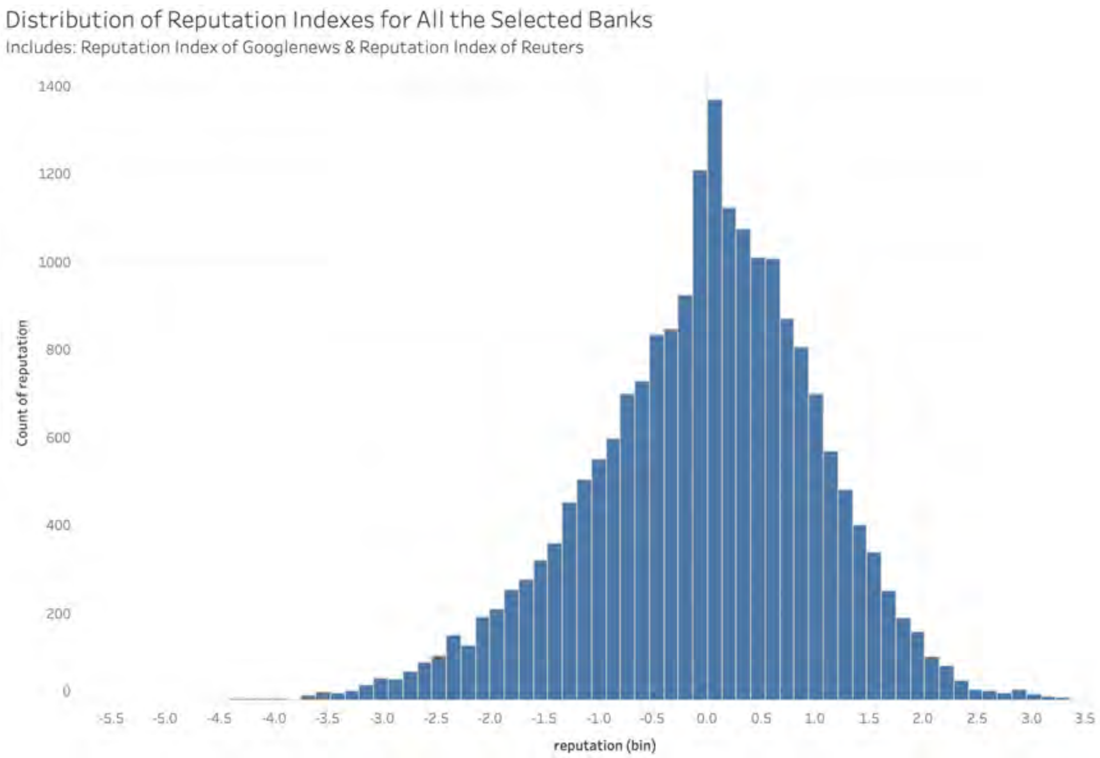

Figure 5.2: Distribution of Reputation Indexes from Googlenews and Reuters for all the selected banks.

since Jun. 2020, the reputation of Credit Suisse increased a lot because few articles were discussing the reputational events nor the covid. Instead, Credit Suisse was highly reported for other good events like Softbank invests in CS funds, CS hires new female executive in China and new CEO for Israel, and so on. So the reputation of Credit Suisse in both news media went up quickly and even reached the highest point in the three years.

- Spying scandal in Sep. 2019.

  > *"A spying scandal that has hit the reputation of one of Europe largest banks and shocked Switzerland financial community." - Reuters*

  We can see from figure 5.4, there is a decrease in reputation of Credit Suisse in the months of Sep. and Oct. 2020. Credit Suisse is famous in this reputational event. However, we can find very few amounts of existed articles talk about it in details.

- Climate controversy on Nov. 2018

  From Wikipedia, in Nov. 2018, about a dozen climate activists played tennis inside Credit Suisse agencies (of Lausanne, Geneva and Basel simultaneously), disrupting operations as a protest against the bank's investments in fossil fuels. The reputational event then pushed Credit Suisse to reviewed and adjusted their ESG investment rules. And we can also found the same decrease of Googlenews and Reuters during Nov. 2018 in figure 5.4. This reputational event has been highly discussed even nowadays as the protesters created a website (called "discreditsuisse") to record the investment activities of Credit Suisse in non environmental-friendly projects.

- Foreign Corrupt Practices Act on 05.July.2018

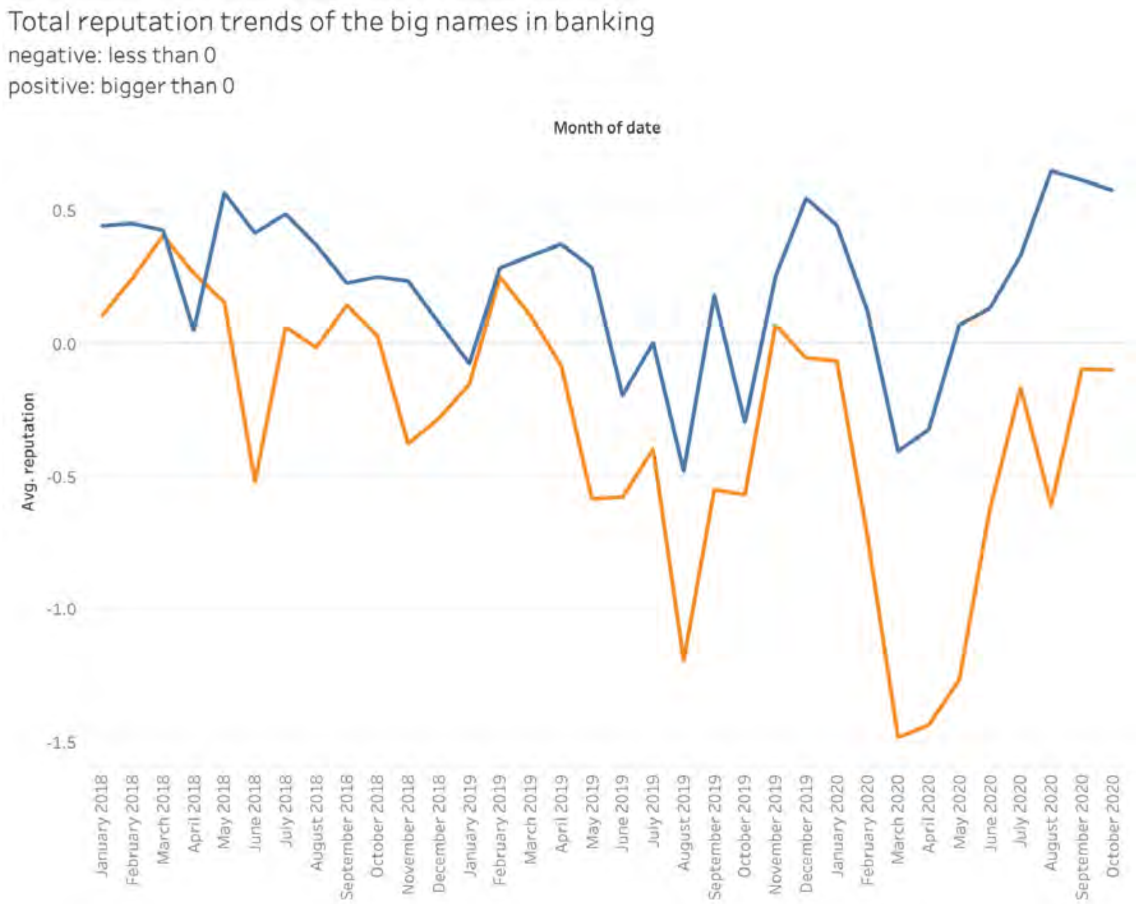

Figure 5.3: Reputation Trends of the Selected Banks to Show the Reputation of Top Level in Banking Industry.

We knew from Wikipedia that on 5 July 2018, Credit Suisse agreed to pay 47 million dollars fine to the US Department of Justice and 30 million dollars to resolve charges of the US Securities and Exchange Commission (SEC). The SEC's investigation said that the banking group sought banking-investment business in the Asia-Pacific region by hiring and promoting more than one hundred Chinese officials and related people in violation of the Foreign Corrupt Practices Act. This event can be also showed in the figure 5.4 as a decrease in July 2018. It is also interesting that Reuters decreased more than Googlenews, because Googlenews contains more general news than Reuters. Thus sometimes, the average sentiment from Reuters is stronger than that from Googlenews.

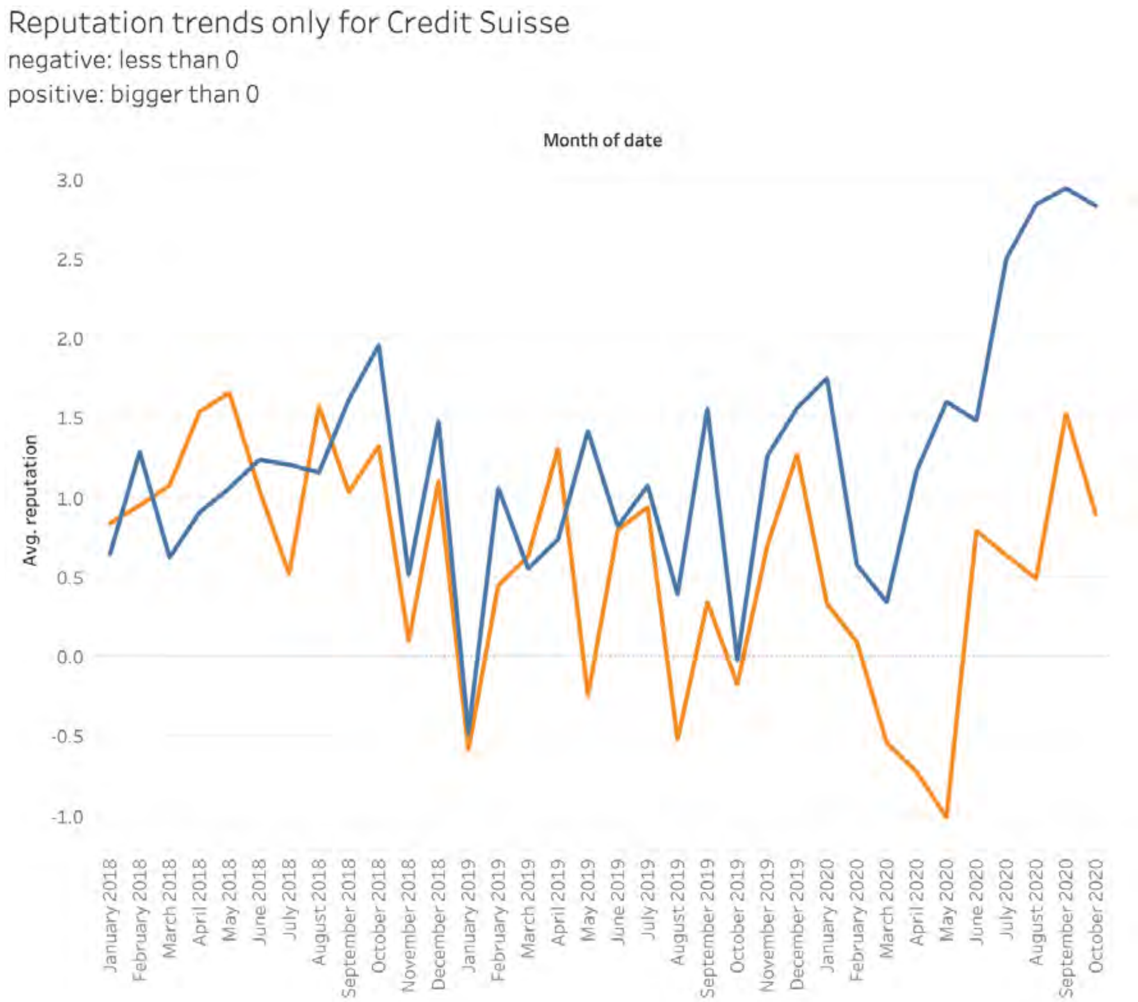

Figure 5.4: Reputation Trends of Credit Suisse Only

Chapter 6

Summary

In this chapter, we will provide the whole conclusion of our work in this project and illustrate the ideas of further improvements. Our work aims to do the research in the topic related to reputational risk and to develop a practical model based on the research for banks to use.

6.1 Conclusion

This study tested currently available methodologies to perform sentiment text analysis and integrated these methodologies into the reputational risk measurement model with implementation on a large self-collected corpus of economic and financial newspaper articles. We tested the machine learning method but finally focused on lexicon-based methods. We introduced the Cumulative Prospective Theory into the sentiment scaling process as reputation and sentiment are psychological terms, which are highly related to the stakeholders' thinking.

We then used our lexical sentiment model to develop a new time-series measurement of reputation for big banks based on text analysis of economics and financial newspaper articles from January 2018 to October 2020. This measure is based on a lexical sentiment analysis model that combines the existing lexicon with another five heuristic rules to consider the effects of context. And we used Autoregressive model to link sentiment and reputation together, showing the memory of reputation.

It is commonly known that developing a reputation risk database and building a reputation risk quantification tool is very important for organizations. Our model can be used to analyze the reputation and reputational risk for organizations, but cannot directly be used to predict reputation trends of the banks because the whole model is based on the published news, which means the reputational events have already happened. But it is helpful for the internal audit team to do the quick action before things go bigger.

More broadly, our event analysis shows that the reputational risk measurement of sentiment extracted from news articles performs well in reputation analysis because of capturing economically meaningful soft information. According to other studies, we know that importantly, this kind of information-based methodology does so at a very low cost and quick answers relative to survey-based measures.

Our data sets are in big size with more than 30k articles for each bank. However, our data sets are still not completed, which means we still don't have all the historical news articles from Googlenews and Reuters because of the limitation of the websites themselves. This will be discussed in the improvement section.

6.2 Discussion of Future Improvement

This section contains the potential for further improvement which focuses on four aspects: identifying more data sources, collecting historical data for a longer period of time, increasing the accuracy of the NLP model, and figuring out other mathematical methods for missing data.

- Identify more data sources

  Currently, the model only contains one type of data - news media, and two data sources - Googlenews and Reuters. We also collected Twitter, a type of social media data, to get a comprehensive view of reputation in multiple aspects even though we ignore Twitter data in our model temporarily due to the high noise of Tweets. In order to analyze the reputation of organizations more comprehensively, it is better to include more live electronic feeds from publicly available media sources supply texts like articles, comments, reports, etc. These feeds could comprise news media (newspapers like CNN, Fox News, etc), new social media (Facebook, blogs), trade reports and, surveys.

- Collect more historical data

  Our historical data sets contain 3 years of articles and Tweets from 1st Jan. 2018 to 30th Nov 2020 for the selected ten banks in the list, in order to prove the efficiency of the methodology. Constructing reputation is always a long topic for a company thus it is better to collect more historical data with a longer time period. And our current historical data sets are not complete with missing days because scraping historical data by auto-crawlers in the front end can be a hard and painful task, especially most of the news media have limited the amount of showing historical articles online. Luckily, there are still some APIs or existed data sets that can be used with calling or downloading fees.

- Add more bank names

  We compare ten top banks from the G-SIBs list to build a high-level view of the reputation of the banking industry. But the ten banks are not enough to get a comprehensive view of the whole industry. So adding more banks' data, especially the banks world wide and in the different levels of the industry.

- Increase the accuracy of the sentiment model by NLP

  We compared different existed lexicons and constructed our model based on the lexical method, which is a relatively static method to analyze the sentiment by comparing the distance of the words. Even though we added heuristic rules to avoid the meaning changing by the context, the method itself is static and requires the updates of the word in the lexicon. We have tried the Machine Learning method to predict the sentiment of each article, but because of lacking a big amount of labeled training set in the financial domain, the accuracy and performance of the Machine Learning method are not better than that of the lexical method.

  Increasing the accuracy of NLP can be a very huge task to complete. There are many aspects that can be researched to help the process of prediction. The most efficient one for the Machine Learning method can be constructing the labeled training data in the specific domain, even though this is time-consuming and high cost. Other algorithm methods may help in special domain, for example dependency parsing is a sufficient way to extract a dependency parse of a sentence that represents its grammatical structure and defines the relationships between headwords and words which modify those heads.

- Figure our other mathematical methods for missing data

  Dealing with missing data can be quite an interesting topic in time series problems. One solution is to re-scraper the missing data by continue to study the website and the crawlers. Another solution is to use mathematical methods to deal with the gaps, e.g. spline interpolation, bootstrap, and Brownian bridge are recommended.

- Find other reputation indexes and signals

  We would like to explore for more possibilities of other reputation signals , for example through leveraging risk data from other risk types and public market data.

- Combine market information in

  We would like to explore the potential relationship between the market and the reputation of banks. For example, the stock price of Credit Suisse can be considered as one factor in our model to measure reputation.

Bibliography

[1] L. J. Ponzi A. Bonime-Blanc. *Understanding Reputation Risk: The Qualitative and Quantitative Imperative.* GEC Risk Advisory LLC ReputationInc, 2016.

[2] L. Huang A. Go, R. Bhayani. *Twitter Sentiment Classification using Distant Supervision.* CS224N project report, Stanford, 2009.

[3] D. Wilson A. H. Shapiro, M. Sudhof. *Measuring News Sentiment.* Federal Reserve Bank of San Francisco Working Paper 2017-01, 2020.

[4] P. Aula. *Reputation Risk and Ambient Publicity management.* Startegy Leadership, 38(6): 43-49, 2010.

[5] Lillian Lee Bo Pang. *Opinion Mining and Sentiment Analysis.* Foundations and Trends in Information Retrieval, 2008.

[6] Jason Brownlee. *A Gentle Introduction to k-folder Cross Validation.* Statistics, May 23 2018.

[7] etc Bruno T. *KNN with TF-IDF based Framework for Text Categorization.* Volumn 69, 2014.

[8] G. van Vuuren R. Lotriet E. Swanepoel, J. Esterhuysen. *Assessing Reputational Risk: A Four Point Matrix.* Journal of Economic and Financial Sciences, 10(2), pp. z313-337, June 2017.

[9] P. Schwizer F. Fiordelisi, M. Soana. *The Determinants of Reputational Risk in the Banking Sector.* Journal of Banking Finance, July 2011.

[10] C. Buckley G. Salton. *Term weighting approaches in automatic text retrieval.* Readings Inf. Retr, 1997.

[11] etc H.Han. *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification.* PAKDD, 2001.

[12] G. Honey. *Corporate reputation: perspectives of measuring and managing a principal risk.* CIMA, 2007.

[13] G. Honey. *A Short Guide to Reputation Risk: Power Publishing Company.* ISBN 978-0-566-008995-4, 2009.

[14] E.E Hutto, C.J. & Gilbert. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.* Eighth International Conference on Weblogs and Social Media (ICWSM-14), 2014.

[15] C. Larrinaga J. Bebbington and J. M Moneva. *Corporate social reporting and reputation risk management.* Accounting, Auditing, and Accountability Journal, 2006.

[16] P. de Fontnouvelle J. Perry. *Measuring Reputational Risk: The Market Reaction to Operational Loss Announcements.* Federal Reserve Bank of Boston, 2005.

[17] S Ramamoorti K. Cravens, F. Goad-Oliver. *The Reputation Index: measuring and managing corporate reputation.* European Journal, 21, 2, 201-212, 2003.

[18] J.Tsujiic K. Masudaa, T.Matsuzakib. *Semantic Search based on the Online Integration of NLP Techniques.* Procedia - Social and Behavioral Sciences 27, 2011.

[19] V. Korde and C. N. Mahender. *Text classification and classifiers: A survey.* Int. J. Artif. Intell. Appl. 3(2) (2012) 85-99, 2012.

[20] J. Larkin. *Strategic Reputation Risk Management.* Palgrave Macmillan, 2003.

[21] B. Liu. *Sentiment Analysis: Mining Opinions, Sentiment, and Emotions.* Sentiments and Emotions, CUP, 2015.

[22] F. Benevenuto K. P. Gummadi M. Cha, H. Haddadi. *Measuring User Influence in Twitter: The Million Follower Fallacy.* Association for the Advancement of Artiï¬*cialIntelligence*, 2010.

[23] G. Cannas M. Micocci, G. Masala. *Reputational Effects of Operational Risk Events for Financial Institutions.* Working Paper, University of Cagliari, 2009.

[24] Peter Mitic. *Reputation Risk: Measured.* Int. J. of Safety and Security Eng., 2018.

[25] H. Lucius N. Mukherjee, S. Zambon. *Do banks manage Reputational Risk? A case study of European Investment Bank.* European Commissions, 2014.

[26] Basel Committee on Banking Supervision. *International Convergence of Capital Measurement and Capital Standards.* Basel Committee on Banking Supervision, Nov.2005.

[27] RepRisk Methodology Overview. *ESG Data Science and quantatitive solutions.* RepRisk, 2020.

[28] A. Ritter S. Rosenthal V. Stoyanov T. Wilson P. Nakov, Z. Kozareva. *SemEval-2013 Task 2: Sentiment Analysis in Twitter.* SemEval 2013, 2013.

[29] M. F. Porter. *An algorithm for suffix stripping.* Emerald Insight 14 (2006) 130-137, 2006.

[30] PricewaterhouseCoopers. *Uncertainty Tamed? The Evolution of Risk Management in the Financial Services Industry.* Economist Intelligence Unit, 2004.

[31] D. Hermansson R. Farha, E. Sekeris. *The hidden cost of reputation risk: an approach to quantifying reputation risk losses.* Oliver Wyman.

[32] RepRisk. *RepRisk Methodology Overview.* RepRisk AG, April.2020.

[33] N. Bt M. Sharef S. Foroozan, M. A. A. Murad. *Sentiment classification of financial news using statistical features.* 2016.

[34] A. Tcersky and D. Kahneman. *Advances in Prospect Theory: Cumulative Representation of Uncertainty.* Journal of Risk and Unvertainty, 1992.

[35] Vargas-Hernandez. *Reputational capital as ethical questioning of technology innovation: the case of Monsanto.* Journal of Applied Business and Economics, 2013.

[36] I. Walter. *Reputational Risk and Conflicts of Interest in Banking and Finance: the evidence so far.* 2006.

[37] An Economist Intelligence Unit white paper. *Reputation: Risk of risks, Economist Intelligence Unit.* sponsored by Ace, Cisco Systems, Deutsche Bank, IBM and KPMG, 2008.

[38] J. Yang and G. Jiang. *Development of an enhanced route choice model based on cumulative prospect theory.* Volume 47, October 2014.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Management, Technology, and Economics
Prof. Dr. D. Sornette

Title of work:

Measuring Reputational Risk by External Data

Thesis type and date:

Master Thesis, Dec. 2020

Supervision:

Prof. Didier Sornette (Chair of Entrepreneurial Risks, D-MTEC, ETHZ)
Makbule Asir (Data Scientist, IA Data Analytics, CREDIT SUISSE)
Wael Malek (Director, IA Data Analytics, CREDIT SUISSE)

Student:

Name:	Siqi Dai
E-mail:	sidai@student.ethz.ch
Legi-Nr.:	17-105-286
Semester:	4

Statement regarding plagiarism:

By signing this statement, I affirm that I have read and signed the Declaration of Originality, independently produced this paper, and adhered to the general practice of source citation in this subject-area.

Declaration of Originality:

`http://www.ethz.ch/faculty/exams/plagiarism/confirmation_en.pdf`

Zurich, 22. 2. 2021: ______________________________