

## Research Article

# Anticipating Corporate Financial Performance from CEO Letters Utilizing Sentiment Analysis

Siqi Che <sup>1</sup>, Wenzhong Zhu <sup>2</sup>, and Xuepei Li<sup>3</sup>

<sup>1</sup>School of English for International Business, Guangdong University of Foreign Studies, Guangzhou 510420, China

<sup>2</sup>School of Business, Guangdong University of Foreign Studies, Guangzhou 510420, China

<sup>3</sup>School of Economics and Trade, Guangdong University of Foreign Studies, Guangzhou 510420, China

Correspondence should be addressed to Wenzhong Zhu; 743106832@qq.com

Received 15 January 2020; Accepted 19 February 2020; Published 13 May 2020

Guest Editor: Weilin Xiao

Copyright © 2020 Siqi Che et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the emergence and tremendous growth of text mining, a computer-assisted approach for capturing sentiment viewpoints from textual data is gradually becoming a promising field, particularly when researchers are increasingly facing the problem of filtering bunches of useless information without capturing the essence in the big data era. This study aims at observing and classifying the sentiment orientation in CEO letters, digging the main corporate social responsibility (CSR) themes, and examining the effectiveness of CEO letters' sentiment on forecasting financial performance. A specific sentiment dictionary has been proposed to identify and classify the sentiment orientation in CEO letters by utilizing the appraisal theory. Additionally, the qualitative data analysis software NVivo is applied to explore the CSR topics. Furthermore, a modified Altman's Z-score model and machine-learning approach are employed to predict financial performance. The results of preliminary evaluations validate that approximately 62.14% of the texts represent positive polarity even when companies are not in a promising economic situation. The CSR themes mainly focus on business ethical responsibility, particularly ethical activities. Among various machine-learning approaches, the logistic regression approach is appropriate for predicting financial performance with the state-of-the-art accuracy of 70.46%. The encouraging results indicate that the sentiment information in CEO letters is a vital factor for anticipating financial performance. This work not only offers a new analytic framework for associating linguistic theory with computer science and economic models but will also improve stakeholders' decision-making.

## 1. Introduction

Due to researchers' unrivalled and explosive expansion in data mining, big data, and artificial intelligence, natural language processing (NLP) in handling bunches of textual data becomes an explosive and prevalent field with great future prospects. The high-tech novel technique of sentiment analysis offers a more efficient and accurate way for text processing, and its amazing pace of innovation, low costs, and scalability make it a highly attractive and alternative approach.

Sentiment analysis, also known as "opinion/sentiment mining" or "subjectivity analysis," uncovers a prominent interdisciplinary field of mixing computational linguistics and computer science, which attempts to extract subjective opinions, feelings, and attitudes contained in the text and

analyzes how to use language to deliver subjectivity and viewpoints on a particular topic [1].

Correspondingly, corporate social responsibility report (CSRR), containing abundant sentiments, is crucial for reflecting companies' sustainable standpoints on its operating ideas, strategies, and methods. For this reason, CSRR can be a valuable source for investors' decision-making. For instance, CEO letter contains the ideas that corporations credibly desire to convey some information about themselves to their potential stakeholders, and these ideas may be important in determining investment and lending decisions because they cover valuable background information about interpreting and explaining financial performance, which may not be covered in companies' financial statements [2]. This obviously boosts the requirements of sentiment analysis and turns into extremely precious resources for

stakeholders' decision-making [3]. China, as one of the biggest developing countries in the world, has developed CSR rapidly, especially the Chinese government has paid much more attention to sustainable development. However, fewer researchers have exerted empirical evidence on the sentiment analysis of CSR in China, particularly in the case of information asymmetry.

To our knowledge, prior studies have proposed manifold sentiment analysis techniques towards extracting useful content from massive social media [4, 5]; some scholars have categorized sentences into positive/negative [4], while other researchers have divided sentences into critical/emotional-based on researchers' intuition. The taxonomy results, however, are inconsistent and incomparable as various self-designed frameworks have been applied in diverse studies. Although sentiment analysis of CSR or nonfinancial information appears to be a possible trend for predicting financial performance and assisting investors to make future investment decisions, little research has been done in this field.

As an attempt to make up the deficiency in the above research field, this paper is divided into three dimensions: firstly, in the microlevel analysis, it determines the sentiment polarity and sentiment attributes in letters to shareholders by utilizing appraisal theory; secondly, in the mesolevel analysis, it identifies the prominent CSR themes by employing NVivo 12 plus software; thirdly, in the macrolevel analysis, it explores the sentiment elements so as to anticipate financial performance by using the expression of Z-score.

The following parts of the paper are organized as follows: in the part of literature review, the related CSR, CEO letters, primary sentiment analysis approaches, appraisal theory, and forecasting financial performance through textual information will be reviewed; the part of research methodology will introduce the annotation study and the experimental procedures; in the part of experimental results and analysis, the experimental results will be stated and analyzed; in the conclusion part, the research results, contribution, limitations, and suggestions will be provided.

## 2. Literature Review

**2.1. CEO Letters in CSR Report.** The CEO letter (hereafter, shareholders' letter or letter to shareholders) is the most widely read part of the CSRR and a valuable communication source for stakeholders' decision-making. Generally, letter to shareholders is widely regarded as a promotional genre, which tries to provide companies' subjective sentiment standpoints and aims to portray a positive image [6] such as what stakeholders desired to know about identifying the last year's performance, tracking important events, displaying corporate social responsibility, and predicting the future vision of the companies. Thus, the letter stands in a vital important position to deliver company's competitive advantage [7]. Although a number of scholars recognized the significance of CEO letters, for instance, Kohut and Segars [7] characterize the effectiveness of letter to shareholders and how this information can benefit the company. Patelli and Pedrini [8] indicate that, even under the tough economic

conditions, companies still sincerely disclose nonfinancial information with stakeholders. Surprisingly, little research has focused on how companies try to construct their images and showed to readers through sentiment analysis.

**2.2. Sentiment Analysis.** In artificial intelligence (AI), text mining is an effective and efficient way to process a large number of textual contents through extracting the sentiment polarity based on natural language processing. In particular, sentiment analysis has gradually become a popular technique. A considerable amount of literature has concentrated on analyzing documents' sentiment polarity [9]. Automatic sentiment classification has been extensively applied to reviews of products [10], movies [11], books [12], shopping [13], social networks [14], and students' course evaluation comments [15, 16], which is a common application of classifying positive or negative reviews. Most recent work has involved in extracting the textual information in the financial reports, as the text may contain more information than the numerical part in an annual report [17, 18]. The final results indicate that sentiment analysis is an important technique for forecasting financial performance and, thus, can be used to support the decision-making process of potential stakeholders [19, 20]. Researchers, therefore, have realized the potential value of the textual information analysis of financial reports for predicting financial performance and managing risks.

In fact, the challenge of detecting sentiment from text has been tackled from various perspectives. Nonetheless, previous approaches to spot affect have been categorized into two main approaches: lexicon-based approach and machine-learning approach [21], which are explicitly depicted in Figure 1.

**2.2.1. Lexicon-Based Approach.** Lexicon-based methods are rule-based, requiring a predefined word list and polarity to identify viewpoints and sentiments [4, 22], containing computing orientation for a text from the semantic orientation of words or phrases [23]. In other words, when a new text has been selected, the words inside the text have to match with the words in the sentiment dictionary and then various algorithms have been employed to aggregate values. The total aggregation of positive and negative values of the words assembles the sentiment orientation of the whole text. An idealized operating mechanism is specified in Figure 2.

Table 1 enumerates some related work in the lexicon-based approach for sentiment analysis and illustrates the various types of objectives along with the associated models used and the experimental results produced. We employed precision, recall, and F-measure as evaluation metrics; they are common in information retrieval and document classification research. Precision measures the number of correctly classified items out of the total classified by a classification technique, and recall measures the amount of correctly classified items of those manually classified as the gold standard. The F-measure is the harmonic mean of precision and recall, which offers a better measure than the arithmetic mean of precision and recall.

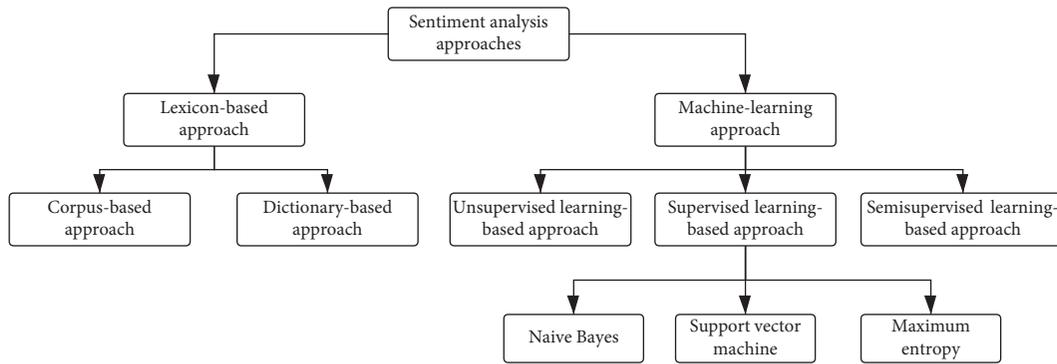


FIGURE 1: Sentiment classification methods.

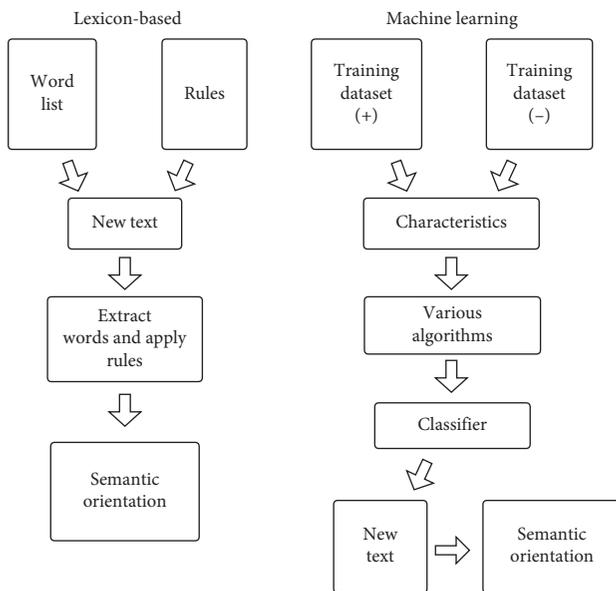


FIGURE 2: Machine-learning and lexicon-based approaches of sentiment analysis.

The merits of employing lexicon-based approach are that, across diverse domains, lexicon-based methods do not require to alter dictionaries [4]. Similarly, Brooke, Tofiloski, and Taboada [29] claimed that, compared with machine-learning, rule-based approach is not a cumbersome task because scholars do not need to take a large amount of time and effort to annotate the training data in a specific domain.

The lexicon-based approach, however, has its own limitations. Sometimes, the individual word extraction may miss important meanings in the text. Since the existing sentiment dictionary cannot meet the specific context of letter to shareholders, for example, in the dictionary, words like *lower* and *decrease* simply do not have a negative meaning in the CSR. Hence, in order to alleviate the dispute, sentiment identification requires a more comprehensive and proprietary sentiment dictionary for the specific CSR domain.

**2.2.2. Machine-Learning Approach.** The machine-learning approaches are presented with training classifiers, such as support vector machine (SVM), Naive-Bayes (NB), logistic

regression (LR), and maximum entropy (ME), to assess the contents' positive or negative orientations based on an initial coding ten test upon the dataset to see if the sentiment is indeed captured [30]. To put it another way, human annotators code a small sample of dataset and then the machine takes over once it "learns" what kinds of words resemble these positive and negative sentiments [31]. A detailed operating mechanism is specified in Figure 2.

Generally speaking, the machine-learning approach can be further divided into supervised, unsupervised, and semisupervised [32]. Normally, the training objective is to be able to classify or distinguish instances. The main difference is that the supervised machine learning-based approach selects labeled instances to construct the model. The unsupervised approach is used for data mining to spot what is inside the unlabeled data. Semisupervised technique is halfway between supervised and unsupervised learning. This technique is about to use unlabeled data to learn a function improving the classification performance.

Table 2 summarizes a great diversity of relevant literature on machine-learning approach for sentiment identification, illustrating that most existing studies analyze text by creating a large dataset to measure subjectivity.

The advantages of the machine-learning approach are that, once the trained dataset is available, the classifier can rapidly determine the text's polarity. For instance, Yang et al. [35] applied a naive Bayes classifier and class association rules to identify the sentiment of consumer reviews and proved to achieve a satisfactory result. Additionally, Troussas et al. [36] classified the people's feelings and attitudes about certain topics and stated how sentiment analysis using naive Bayes can efficiently assist language learning. Specifically, Yuan et al. [39] manifested that the classification accuracy has been improved extremely based on the sentiment analysis of CSR in financial reports through the SVM approach. This provides the inspiration and basis for the further research of this study.

While the machine-learning approach has long been used in topical text classification with good results because of the efficiency and accuracy, it sometimes encounters with shortcomings as a result of high-level human manual intervention. That is to say, the classifiers necessitate sizable human annotated datasets for training and testing, especially amid the big data era, which is extremely costly and time-consuming

TABLE 1: Related work of lexicon-based approach in sentiment analysis.

Authors	Objectives	Models	Data source	Evaluation method	Data set	Accuracy (percent)	Precision (percent)	Recall	F <sub>1</sub>
Hatzivassiloglou and Mckeown [24]	Assign adjective $\pm$	Nonhierarchical clustering	WSJ corpus	N/A	657 adj (+) 679 adj (-)	78.1–92.4	N/A	N/A	N/A
Turney [23]	Assign docs sentiments	PMI-IR	Automobile, bank, movie, travel reviews	N/A	240 (+) 170 (-)	65.8–84	N/A	N/A	N/A
Turney and Littman [25]	Assign docs sentiments	PMI LSA	AV-ENG corpus AV-CA corpus TASA corpus	N/A	1614 (+) 1982 (-)	82.8–95	N/A	N/A	N/A
Taboada et al. [26]	Assign adjectives $\pm$	SO-PMI	Reviews	N/A	521 adj	49.5–56.75	N/A	N/A	N/A
Ding et al. [27]	Assign adjectives, adverbs, verbs, and nouns $\pm$	Opinion Observer	445 customer reviews of products from amazon.com	Macroaveraging (macroaveraging means given a set of confusion tables, a set of values are generated, and each value represents the precision or recall of an automatic classifier for each category)	N/A	N/A	91	90	90
		No context dependency				N/A	92	83	87
		Without using equation				N/A	90	85	87
		FBS				N/A	92	74	82
Taboada et al. [4]	Assign adjectives, verbs, adverbs, nouns $\pm$	SO-CAL	Reviews	Across various domains	N/A	80	N/A	N/A	N/A
Dey et al. [28]	Assign docs sentiments	SO-CAL, Senti-N-Gram	Movie, books, cars, cookware, phones, hotels, music, and computers reviews	3-fold cross validation	Books	68	66	76	71
			Cars		84	76	100	86	
			Computers		80	73	96	83	
			Cookware		74	68	92	78	
			Hotels		74	67	96	79	
			Movies		66	64	72	68	
			Music		64	62	72	67	
			Phones		86	85	88	86	

[31]. Moreover, sometimes, the training datasets are unavailable. Previous attempts at categorizing movie reviews and book reviews used the machine-learning approach with limited success. Because the technique may not suitable for various texts, facing the obstacle of domain specificity, the accuracy of analysis has reduced greatly because the thorough custom-made training dataset may not generalize well to texts from other domains [40, 41]. Consequently, in this study, the particular characteristics should be captured to perfectly satisfy the specific domain of letters to shareholders.

*2.3. Forecasting Financial Performance by Using Text Information.* Recently, sentiment analysis has been widely explored in understanding the relationship between text

information and corporate financial performance. Through evaluating authors' opinions, attitudes, and sentiment polarity, the future financial performance could be predicted. A great number of scholars have recently engaged in analyzing textual information in annual reports, newspapers, and other documents to forecast stock return or financial performance. Table 3 lists some literature to indicate that the sentiment of documents can significantly correlate with financial performance.

From the above literature, our study will try to use letters to shareholders in CSRR to test whether the sentiment in the letter can predict the financial performance or not by employing various kinds of machine-learning methods. Furthermore, in order to have a better comparison of the sentiment attribute categorization and construct a suitable

TABLE 2: Related work of machine-learning approach in sentiment analysis.

Authors	Objectives	Models	Data source	Evaluation method	Data set	Accuracy (percent)	Precision (percent)	Recall	F <sub>1</sub>
Bruce and Wiebe [33]	Assign sentences subjectivity using 1 to 4 scale	LCA	14 articles in Wall Street Journal Treebank Corpus	10-fold cross validation	486 subjective, 515 objective	72.17	N/A	N/A	N/A
Pang et al. [30]	Assign docs sentiments	SVM, NB, ME	Movie reviews	3-fold cross validation	700 (+) 700 (-)	77–82.9	N/A	N/A	N/A
Bo and Lee [1]	Assign docs sentiments	SVM, NB	Movie reviews	10-fold cross validation	1000 (+) 1000 (-)	86.4–87.2	N/A	N/A	N/A
Gamon and Michael [34]	Assign docs sentiments using 4-point scale	SVM	Customer feedback	10-fold cross validation	N/A	77.5	N/A	N/A	N/A
Yang et al. [35]	Assign topic sentiments	Class association rules	Consumer reviews	Macroaveraged Microaveraged (microaveraging means given a set of confusion tables, a new two-by-two contingency table is generated, each cell in the new table represents the sum of the number of documents from within the set of tables)	N/A	N/A	67.64	77.5	72.26
		NB		Macroaveraged Microaveraged		N/A N/A	73.96 68.3	66.79 66.11	70.19 67.19
Troussas et al. [36]	Assign docs sentiments	NB	Facebook 7000 status updates from 90 users	3-fold cross validation	1131 (+) 1131 (-)	N/A	77	68	72
Singh et al. [37]	Assign docs sentiments	NB J-48 BFTree OneR	Product review, movie review	4-fold cross validation	N/A	45.6 96.7 89.2 90	83.1 87.7 88.3 91.3	N/A N/A N/A N/A	81.2 91.7 72.1 97
Rathi [38]	Assign adjectives $\pm$	SVM + KNN, hybrid	Tweet	N/A	267 (+) 264 (-) 168 (0)	76.17	N/A	N/A	N/A

TABLE 3: Related work of forecasting financial performance using textual information.

Authors	Material	Period	Methods	Key findings
Loughran and Mcdonald [42]	Annual report	1994–2008	Sentiment dictionary	Textual analysis can contribute to the ability to understand the impact of information on stock returns. Most importantly, researchers should be cautious when relying on word classification scheme derived outside the domain of business usage.
Schumaker et al. [43]	Financial news articles	2005	Machine learning	The financial news article prediction system can predict price increase or decrease in sentiment analysis with the accuracy of approximately 50%.
Hájek and Olej [44]	520 U.S. companies' annual report	2010	Machine learning	Sentiment information in annual reports can significantly improve the accuracy of the used classifiers.
Xin et al. [45]	U.S. manufacturing industry annual report	1997–2003	Machine learning	Using machine-learning approach employed with the results of text analysis can predict next year's earnings per share.
Hájek et al. [46]	448 U.S. companies' annual report	2008	Machine learning; neural networks	Sentiment information is an important determinant for forecasting financial performance.
Tong et al. [47]	Microblogs in China	2016.2–2016.9	Machine learning	There is a strong correlation between chat room postsentiment and stock price movement.

sentiment classification dictionary for the specific context, appraisal theory will be applied and the seed dictionary will be manually annotated for the domain of CEO letters.

**2.4. Appraisal Theory.** Although machine-learning methods show a better performance, the results of each experiment are not comparable because of various kinds of categorizations. A detailed theoretical framework is highly required to handle this difficulty. Researchers are asked to address more challenging tasks and attempt to perform more sophisticated sentiment analyses based on systematic and well-grounded sentiment theories and frameworks. These frameworks will become common theoretical platforms for comparing results across different studies.

To date, the existing sentiment techniques may not be sufficient for sentiment classification. Linguistic framework, therefore, has been employed as a theoretical platform for sentiment analysis. By far, computational linguists have projected several sentiment frameworks. Wiebe et al. [48] disclosed a sentiment framework based on private states with three types of expressions, that is, explicit mentions of private states, speech events expressing private states, and expressive subjective elements. The framework is designed to expand attitude types to include intention, warning, uncertainty, and evaluation. Asher, Benamara, and Mathieu [49] further denoted an annotation framework involving four categories, that is, reporting expressions, judgement expressions, advice expressions, and sentiment expressions to express various kinds of feelings and opinions. The most comprehensive linguistic theory of sentiment, however, is the appraisal framework [50], also known as interpersonal semantics, developed within the tradition of Systemic Functional Linguistics [51]. Appraisal theory is a framework employed in conveying the language of evaluation in text [52] and also an approach to linguistics that focuses on the semantics of text rather than its grammar [53].

The framework portrayed a taxonomy of the language to convey *attitude*, *engagement*, and *graduation* with respect to the evaluations of other people. The detailed taxonomy is depicted in Figure 3. Attitude considers how people conduct interpersonal interaction to express his or her opinions and emotions; engagement assesses the evaluation with respect to others' opinions; and graduation denotes how to strengthen or weaken the attitude through language functions.

Attitude, engagement, and graduation can be further divided into several subtypes, which are explained more in-depth in the following paragraphs.

*Attitude* can be further divided into three distinct subsystems: affect, appreciation, and judgement. Affect identifies the feelings and emotions of the author (happy and sad), which can be a behavioral process or an internal mental state. Appreciation represents the reaction that a person talks about (beautiful and ugly), such as impact, quality, composition, complexity, or valuation. Judgement considers the author's attitude towards the behavior of somebody (heroic and idiotic) and the evaluation may concern social sanctions or social esteem; social sanctions

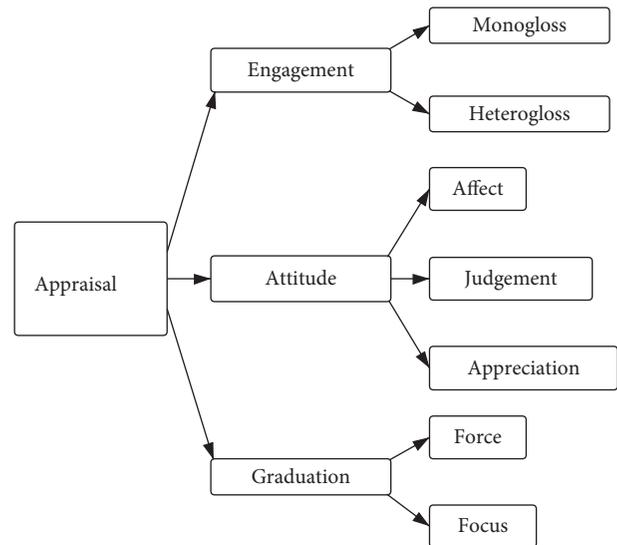


FIGURE 3: Outline of Martin and White's [52] appraisal framework (first two levels).

may involve veracity or propriety; and social esteem may involve the assessment of how normally someone behaves, how capable the person is, or how tenacious the person is.

*Engagement* contains two subcategories: monoglossic and heteroglossic. Monoglossic has no recognition of dialogistic alternatives; that is, the writer/speaker directly expressed the appraisal. Heteroglossic has the recognition of dialogistic alternative positions and voices, which means that the writer/speaker has either attributed to other methods or sources to make it credible.

*Graduation* also include two subclasses: force and focus. Force assesses the degree of intensity; focus covers to sharpen or soften the specification.

The main advantage of utilizing appraisal theory in sentiment identification is that it enables us to take a look deeper inside the thoughts of authors or publishers, revealing their real sensations by employing linguistic and psychological analysis of their texts. In our study, we want to extract the sentiment inside in the shareholders' letters and may possibly draw some statistical conclusions about the types of sentiment that companies of shareholders' letters express via the texts they write. At last, the analysis can help potential investors to make better decision-makings.

So far, scholars have made a great number of useful attempts at the level of sentiment analysis. Table 4 sums up the relevant sentiment analysis literature employing the appraisal framework. From the table, the research integrating the appraisal theory into the analysis of sentiment orientation is relatively scarce. Korenek and Šimko [60] established the method of manually creating an emotional dictionary, utilizing the appraisal theory to evaluate emotional posts on Weibo posts; Taboada and Grieve [54] further made some improvements on how to aggregate the value of adjectives based on appraisal theory. In addition, Khoo et al. [50] extended the appraisal framework for analyzing long news reports, compared with microposts, and

TABLE 4: Related work of appraisal theory in sentiment analysis.

Authors	Models	Data source	Method	Attributes type	Accuracy (percent)
Taboada and Grieve [54]	Lexicon-based	Reviews	Taking into account text structure and adjectives frequency	Attitude	N/A
Whitelaw et al. [55]	Lexicon-based	Movie reviews	Analyzing appraisal adjectives and modifiers	Attitude, orientation, graduation, polarity	90.2
Read and Carroll [56]	Lexicon-based	Book reviews	Measuring interannotator agreement	Attitude, engagement, graduation, polarity	N/A
Argamon et al. [57]	NB, SVM	Documents	Automatic determining complex sentiment-related attributes	Attitude; orientation; force	N/A
Balahur et al. [58]	Robert Plutchik's wheel of emotion; Parrot's tree-structured list of emotions	ISEAR Corpus	EmotiNet (EmotiNet defines to store action chains and their corresponding emotional labels from several situations; in such a way, authors could be able to extract general patterns of appraisal)	Actor, action, object  Emotion	N/A
Bloom [59]	Lexicon-based	MPQA 2.0 Corpus; UIC Review Corpus; Darmstadt Service Review Corpus; JDPa Sentiment Corpus; IIT Sentiment Corpus	Functional local appraisal grammar extractor	Attitude	44.6
Khoo et al. [50]	Lexicon-based	Political news article	Analyzing appraisal groups	Actor, attitude, engagement, polarity	N/A
Korenek and Šimko [60]	SVM	Microblog posts	Structuring sentiment analysis based on appraisal theory	Attitude, graduation, polarity, orientation, engagement	87.57
Cui and Shibamoto-Smith [61]	Lexicon-based	Self-constructed corpus of Chinese newspapers and websites	Discovering the lexical, syntactic, and semantic features of different types of sentiment parameters	Attitude Peripheral sentiment parameters (topic, source, field, process, degree)	N/A

assessed the framework's utility and possible problems, which is definitely a good attempt for our current study.

As indicated in Table 4, the aforementioned work has improved that appraisal framework is a highly useful tool for analyzing sentiment classification [50, 60]. This paper, thus, intends to augment the sentiment classification of president's letter by employing the appraisal theory. The appraisal theory proves to be useful in uncovering various aspects of sentiment that should be valuable to researchers and assists researchers to understand the feelings of shareholders' letters better. By following the doctor dissertation of Bloom [59], who laid the basis for sentiment classification by utilizing appraisal theory, a more systematic and well-grounded approach of sentiment analysis utilizing Martin and White's appraisal theory has been heuristically employed, which can successfully assist potential investors to understand corporations' attitudes accurately and make investment decisions more efficiently.

### 3. Research Methodology

As mentioned earlier, in order to achieve our goals in this study, a three-dimensional research framework has been

established to enable us to systematically test different perspectives of text mining strategies with both quantitative and qualitative methods (see Figure 4).

The specific research questions that we address are the following:

RQ1: what are the prominent sentiment attributes in CEO's letters that can actively promote companies' images and avoid negative views?

RQ2: what are the main sentimental themes in CEO's letters? And which one can best encourage other companies' investment?

RQ3: how can the sentiment attributes of CEO's letters anticipate corporate financial performance that would be useful for stakeholders' decision-making?

The above questions in three dimensions evaluate CEO letters. In the microlevel of identifying sentiment attributes, appraisal theory and lexicon-based sentiment method have been applied to classify various sentiment attributes precisely. In the mesolevel analysis, we utilize NVivo qualitative data analysis software to do the thematic analysis of CEO letters. In the macrolevel, machine-learning approach was

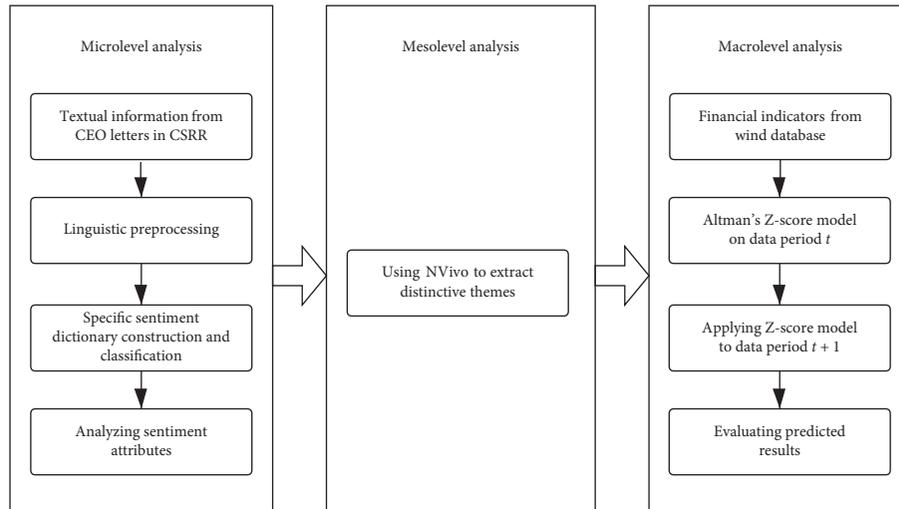


FIGURE 4: An overview of the workflow of the research framework.

selected to assess the power of CEO letters in predicting corporate financial performance in terms of the Z-score model.

**3.1. Data Collection and Description.** To date, various methods have been developed and introduced to measure sentiment. A suitable method to adopt for this study is to propose a specific sentiment dictionary based on the principles of appraisal theory. The advantage of utilizing the appraisal theory is that it allows researchers to categorize and compare sentiment attributes based on a systematic linguistic theory and also enables them to explore authors' thoughts more thoroughly. The major difference from the previous works which use the appraisal theory is that the whole basic appraisal tree has been selected and shareholders' letter specifics have been assessed.

In terms of the variety of genres, such as political news, movie reviews, and product reviews, the analysis results will significantly differ depending on distinct syntactic structure and lexical choices [56]. Therefore, in order to maintain consistency, the same genre has to be examined based on the appraisal framework. In this study, shareholders' letters are good candidates for this study because they are likely to contain similar languages in the same genre of writing; also, examples of appraisal's attributes can be easily detected. Most importantly, identifying valuable information from shareholders' letters is a sufficient way to help companies to improve the quality of released information and facilitate stakeholders' to make investment decisions.

All CSR reports with English version were downloaded from GRI's Sustainability Disclosure Database (<https://database.globalreporting.org>). This database can access to all types of sustainability reports from various industries relating to the reporting organizations. For the sake of preventing problems with both industry-specific attributes and different financial performance evaluation, the financial industries were excluded. Furthermore, in order to ensure accurate comparable information, all selected reports must

have official English version and should be listed in the Hong Kong publicly regulated markets.

Finally, 41 Chinese companies' English version CSR reports from year 2016 had been collected. Then, 41 shareholders' letters were retrieved from the CSR reports in China in 2016. The corpus contains 35,670 tokens in total, named SLC. In addition, all the statistical data for calculating Z-score are gathered from the Wind financial database.

The year 2016 was selected because the United Nations General Assembly unanimously adopted the Resolution 70/1, Transforming our World: the 2030 Agenda for Sustainable Development. This historic document lays out 17 sustainable development goals, which aim to integrate and balance the three dimensions of sustainable development: economic, social, and environmental. The new goals and targets will come into effect on 1 January 2016 and will guide for the next fifteen years.

**3.2. Data Preprocessing.** Most CEO letters were released in PDF format. Firstly, letters were manually transformed.pdf documents to the essential.txt groups and texts were sorted to have only one sentence in each line. Then, the text documents were linguistically preprocessed using tokenization, part of speech tagging, and lemmatization:

Tokenisation—splitting texts into sentences and words

Part of speech tagging—adding a POS tag to each word in a sentence

Lemmatisation—converting a word into its basic lemma form

Deletion of stop words and company name

To control for bias in the analysis, in particular, shareholders' letters have a specific content that differs from other texts, which has to be coped with. For instance, the abbreviation of company's name is Best Buy, including the positive word of "Best", which may significantly convert the

polarity of the text. As a result, the pretreatments were a combination of tokenization, lemmatization, part-of-speech-tagging, and deletion of stop words and named entities, which gave the best result.

*3.3. Specific Dictionary Construction and Classification.* The main issue with the sentiment analysis of textual documents is the right choice of positive terms. Obviously, the categorization of words is not always unambiguous and requires context knowledge. This is due to the various meanings of words and domain specific tone of the words, respectively. Therefore, the appraisal theory has been employed in this study.

To evaluate the appraisal theory, a dictionary tagged with attributes from Martin and White's categorization (2005) has to be constructed. In terms of the work of Korenek and Šimko [60], they built a dictionary based on appraisal theory specializing in microblogs. By following their work, we accumulated approximately five hundred and fifty words from Martin and White's classification. In order to broaden the dictionary, WordNet database, Collins Thesaurus, and Merriam-Webster Thesaurus have been utilized to find synonyms to words identified in the previous step, and this formed the basic sentiment lexicons. The next step is to be aimed at extracting candidate sentiment word lists from the self-constructed SLC corpus, employing the statistical approach based on the amount of pointwise mutual information (PMI) ratio, which can be used to compare candidate words with the existing basic sentiment lexicons. PMI calculation has been defined as follows:

$$\text{PMI}(\text{word}_1, \text{word}_2) = \log \frac{p(\text{word}_1, \text{word}_2)}{p(\text{word}_1)p(\text{word}_2)}. \quad (1)$$

In light of the PMI ratio, whether the word should be identified as a target or not must be decided. On completion of selecting targets, the target candidate words merged with the basic sentiment lexicons to construct a new dictionary specializing in shareholders' letter content, with approximately six hundred words in total.

Due to the specifics of shareholders' letters, the traditional lexicon-based approach cannot identify the complex context. So the statistical method of PMI ratio may not be accurate all the time. Manual screening is highly needed in this step.

To further categorize the new dictionary, each word is manually classified according to the attributes based on appraisal theory and in line with the research of [60], who assigned a 10-point Likert scale, from -5 to +5. The same scaling scope has been considered in this research. For the purpose of manual coding with participants' subjective judgement, eight human annotators, *a* to *h*, were invited to assign polarity independently. The annotators were all well-versed in the appraisal framework. They were asked to specify the type of attitude, engagement, or graduation present and assign a scaling and polarity to candidate

words. In order to address the concern of inconsistent understanding regarding some ambiguous words, the annotation was executed over two rounds, punctuated by an intermediary analysis of agreement and disagreement among all annotators until a consensus has been achieved.

A partial example of entries is presented in Table 5. Each entry contains a word, a part-of-speech tag, a category, and subcategories according to the appraisal theory and appraisal value.

Based on the newly established sentiment dictionary for a specific corpus, we can further precisely analyze the sentiment characteristics of CEO letters from micro-, meso-, and macrolevel analysis.

### 3.4. Data Analysis

*3.4.1. Microlevel Analysis.* Regarding RQ1, we categorized CEO letters based on the specific established sentiment classification dictionary considering the following eleven categories of terms:

- A. Positive attitude affect (e.g., happy, convinced, and satisfied)
- B. Negative attitude affect (e.g., sorry and sad)
- C. Positive attitude judgement (e.g., lucky, fortunate, and famous)
- D. Negative attitude judgement (e.g., imperfect, unknown, and severe)
- E. Positive attitude appreciation (e.g., exciting, dramatic, and excellent)
- F. Negative attitude appreciation (e.g., imbalanced, disharmonious, and conflicting)
- G. Positive engagement (e.g., certainly, obviously, naturally, and evidently)
- H. Negative engagement (e.g., falsely and compellingly)
- I. Positive graduation force (e.g., greatly, slightly, and somewhat)
- J. Negative graduation force (e.g., small and remote)
- K. Positive graduation focus (e.g., true and genuinely)

We assumed that well-performing corporations are being more positive, optimistic tones. Conversely, we expected a more active language in the case of poorly performing companies that need to take positive actions to improve their image and attract more investors.

*3.4.2. Mesolevel Analysis.* With the popularity of computer technology, a range of software packages emerged to assist with the analysis of qualitative data. It is generally accepted that computer-assisted qualitative data analysis software (CAQDAS) can enhance the data handling/analysis process if used appropriately and resolve analysts from complicated data analysis.

TABLE 5: A partial example of appraisal dictionary entries.

Word	Part of speech	Main category	Second level category	Other level categories	Appraisal value
Good	Adj.	Attitude	Judgement	Praise, propriety	3
Harmonious	Adj.	Attitude	Appreciation	Composition, balance	2
Advanced	Adj.	Attitude	Appreciation	Valuation	5
Clear	Adj.	Attitude	Appreciation	Composition, complexity	1
Never	Adv.	Engagement	N/A	N/A	-5
Incredible	Adj.	Attitude	Judgement	Admiration, normality	3
Largest	Adj.	Graduation	Force	Maximization	1

The software package NVivo (now update to version 12) is one of the most distinguished CAQDAS, which can help the analysis to work more efficiently and rigorously back up findings with grounded data [62].

In this study, NVivo has been prompted to do a thematic analysis for identifying the broad CSR topics existing in the shareholders' letters. Thematic analysis is a way of categorizing data from qualitative research through analyzing similar themes and interpreting the research findings [63]. In this study, thematic analysis can be employed to detect the main ideas existing in the CEO letters and through NVivo software to label or code different types of CSR.

NVivo allows nodes to have more than one dimension (tree branch). Therefore, we were able to identify where concepts may have more than one dimension or group them within a more general concept. This is a revolution in finding connections because it prompts the analyst to think about their concepts in more detail, facilitating conceptual clarity, and early discourse analysis [64]. Figure 5 reveals a sample of the tree node structure of CSR.

Coding stripes function is a useful function for researchers to annotate various segments in the whole documents, which may facilitate the comparison of categories. Figure 6 depicts an example of data coded at the ethical responsibility node.

Generally speaking, NVivo offers us a valuable tool to explore the complexities of potential relationships without forcing the data to fit specific categories. In this way, when we identified a possible relationship with CSR, we defined this in NVivo using the free code to represent it first and later created tree node to identify the internal relationship.

**3.4.3. Macrolevel Analysis.** The Altman's model of financial health (Altman's Z-score) was selected for the quantitative evaluation of the assessed companies. The reason of the selection was that this model was created for assessing company financial health in industrial branch with shares tradable on the Hong Kong publicly regulated markets and exactly such companies were selected for the study.

The scope of this so-called bankruptcy model is to predict the probability of survival or bankruptcy of the analyzed company. The nearer to the bankruptcy a company is, the better Altman's index works as a predictor of financial health. It predicts bankruptcy reliably about one to two years in advance. Altman's Z-score model uses the following relation to define the value of a company in industrial branch with shares publicly tradable on the stock market:

$$Z_i = 1.2X_{1,i} + 1.4X_{2,i} + 3.3X_{3,i} + 0.6X_{4,i} + 1.0X_{5,i}, \quad (2)$$

where  $i$  denotes the  $i$ -th company,  $X_1$  is the working capital/total assets,  $X_2$  is the retained earnings/total assets,  $X_3$  is the earnings before interest and tax/total assets,  $X_4$  is the market value of equity/total liabilities, and  $X_5$  is the sales/total assets. Detailed variable definition can be found in Table 6.

The  $Z_i$  value is in range  $-4$  to  $+8$ . The higher the value, the higher the financial health of a company. It holds true that if (1)  $Z_i > 2.99$ , the company is in the "safe zone" (a company with high probability to survive—financially strong company); (2)  $1.80 \leq Z_i \leq 2.99$  "grey zone" (the future of the company cannot be determined clearly—a company with certain financial difficulties); and (3)  $Z_i < 1.80$  "distress zone" (the company has serious financial problems—the company is endangered by bankruptcy).

Due to the special historical conditions of China's stock market formation, the types of stocks formed in China are different from those in foreign countries. In the stocks of listed companies, they can generally be divided into two categories: tradable shares and nontradable shares. In view of the fact that there is no market price for nontradable shares in the Hong Kong stock market, the model has made a slight change:  $X_4 = (\text{share price} * \text{tradable shares} + \text{net asset value per share} * \text{nontradable shares}) / \text{total liabilities}$ , and  $X_5$  is the prime operating revenue/total assets.

The outputs of the forecasting models were represented by the classes of financial performance obtained using the Z-score bankruptcy model, namely, classes "safe zone," "grey zone," and "distress zone." In addition, we obtained additional output classes (increase, no change, and decrease). Given the fact that the classes were imbalanced in the dataset, we use the Synthetic Minority Oversampling Technique (SMOTE) algorithm [65] to modify the training dataset. The algorithm oversamples the minority classes so that all classes are presented equally in the training dataset.

The set of eleven sentiment attributes presented in the previous section was drawn from shareholders' letter. Following previous studies [46, 66], the input attributes were collected for Chinese companies in the year 2016, while the output financial performance (Z-score) was evaluated for the year 2017, and the change in the financial performance was measured as Z-score in 2017 related to its value in 2016. For the sake of preventing problems with both industry-specific attributes and different financial performance evaluation, the financial industry was excluded. As a result, among 41 Chinese companies, 9 companies were classified into "grey zone" and 32 companies were classified into "distress zone;"

Name	References	Sources
Ethical Responsibilities	182	40
Ethical Activity	113	32
Employee	27	19
Client and Customer	17	11
Supplier and Partner	6	5
NGO	4	3
Competitor	1	1
Host Country	1	1
Government	1	1
Ethical Accomplishments	37	20
Employee	4	4
Client and Customer	3	3
Supplier	3	2
Shareholder	1	1
Ethical Ability	34	23
Core Competence	21	17
Organizational Identity	10	8
Technical Responsibilities	58	19

FIGURE 5: Tree node structure of CSR in China.

after making a comparison of financial situation between 2016 and 2017, only 2 companies were improved and the remaining 39 companies were unchanged.

In sum, the Z-score of Chinese companies in 2016 and 2017 could be spotted in Table 7.

Next, a various number of forecasting machine-learning methods have been explored, i.e., logistic regression, support vector machine, and naïve Bayes. Apart from logistic regression, the other two methods can process nonlinear data.

The logistic regression (LR) model has been used with a ridge estimator defined by Cessie and Houwelingen [67]. The classification performance of the logistic regression depends on the number of iterations and ridge factor, respectively.

Support vector machines (SVMs) are a set of related supervised learning methods, which are popular for performing classification and regression analysis using data analysis and pattern recognition.

Naïve Bayes (NB) is a simple multiclass classification algorithm with the assumption of independence between every pair of features. Naive Bayes can be trained very efficiently. Within a single pass on the training data, it computes the conditional probability distribution of each feature given label, and then it applies Bayes' theorem to compute the conditional probability distribution of the label given observation and use it for prediction.

In sum, logistic regression, support vector machine, naïve Bayes are three methods designed to forecast the accuracy of financial performance.

#### 4. Experimental Results and Discussion

In the data filtering, we select 41 CEO letters in Chinese companies CSR report and try to do in-depth analysis at microlevel, mesolevel, and macrolevel, respectively.

4.1. *Sentiment Attributes.* Regarding the eleven sentiment attributes, the detailed statistical data can be found in Table 8. Among all the categories, positive attitude, judgement, appreciation, and positive graduation force are the top three most frequent sentiment attributes.

From the previous data collection part, we know that among 41 Chinese companies, 9 companies were classified into “grey zone” and 32 companies were classified into “distress zone.” None of the companies were classified into the safe zone. Combing the eleven categorizations with our 2016 financial performance, interestingly, we found that poorly performing companies are expected to use a more active language to describe and evaluate their CSR. This phenomenon can be explained further by the impression management effect. Impression management refers to the process by which people try to manage and control the impression others make about themselves [68]. For corporations, the correct impression management can help companies to communicate with stakeholders smoothly. Companies try to use impression management to actively promote their images and avoid negative views. This is probably the reason why companies may encounter with a gloomy economic situation but still concentrate on shaping positive and optimistic images.

4.2. *Sentimental Themes.* The next segment is about to identify the sentimental themes in CEO letters through NVivo software. NVivo’s coding stripes functions enable us to examine all the relevant text and identify the sentences which contributed to that relationship and also to find out which node the sentences belong to. After coding all the relevant text, we gathered comprehensive sentimental themes existing in CEO letters (see Table 9).

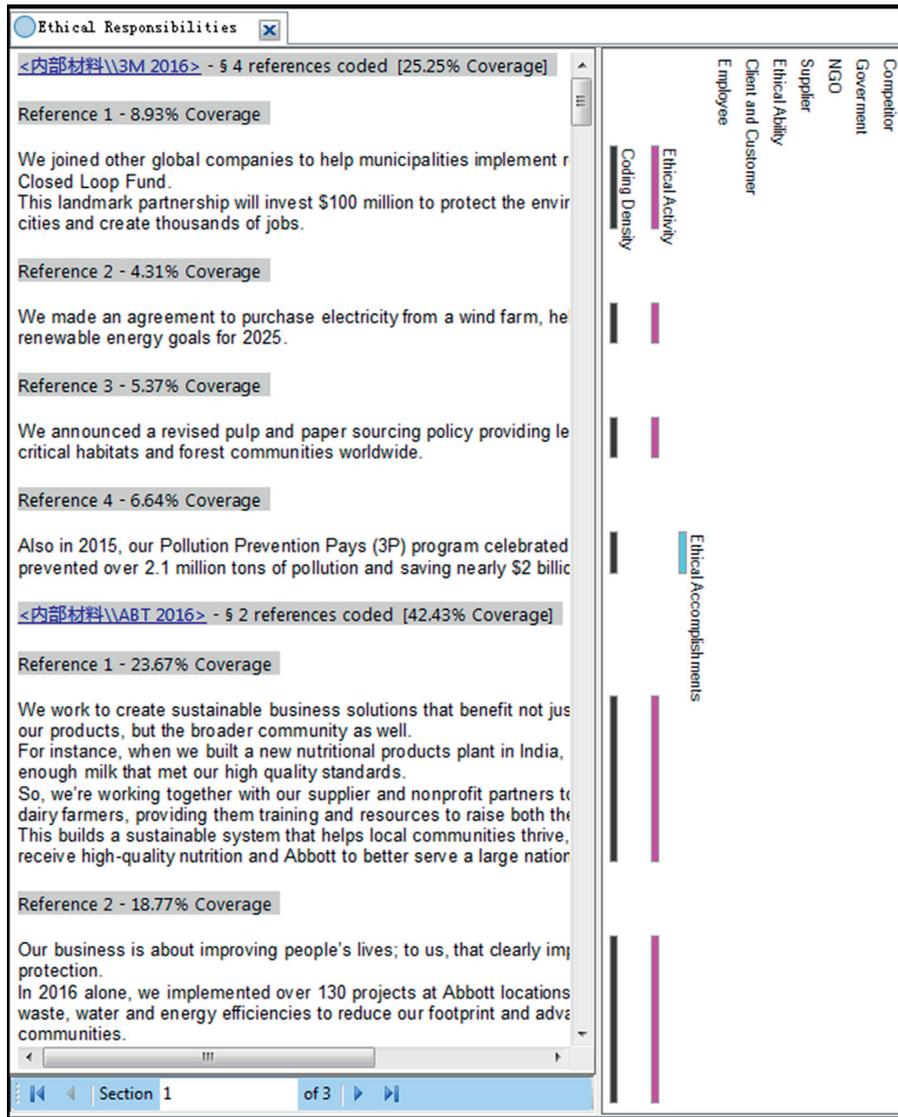


FIGURE 6: Coding stripes on the “ethical responsibilities” node.

TABLE 6: Financial variable definition.

Variable name	Definition
Working capital	Current assets minus current liabilities
Total assets	The final amount of all gross investments, cash and equivalents, receivables, and other assets as they are presented on the balance sheet
Retained earnings	The amount of net income left over for the business after it has paid out dividends to its shareholders
Earnings before interest and tax (EBIT)	Revenue minus expenses, excluding tax and interest
Market value of equity	The total dollar value of a company’s equity
Total liabilities	The combined debts and obligations that an individual or company owes to outside parties
Sales	Total dollar amount collected for goods and services provided

According to Table 9, the most popular sentimental theme in CEO letters is the ethical responsibility. Business ethical responsibility for companies means a system of moral and ethical beliefs that guide companies’ behaviors, values, and decisions and minimizing unjustified harm, suffering, waste, or destruction to people and the environment [69].

This result reveals that a large amount of companies (32 among 41 companies) have put great efforts into business ethics. The concept of business ethics began in the 1960s as corporations became more aware of a rising consumer-based society that showed concerns regarding the environment, social causes, and corporate responsibility. In fact,

TABLE 7: Z-score of Chinese companies in 2016 and 2017.

Stock code	Country	2016 Z-score	2016	2017 Z-score	2017
1958.hk	CN	1.260515439	Distress zone	1.482271118	Distress zone
0606.hk	CN	1.838183342	Grey zone	2.216679937	Grey zone
1800.hk	CN	0.867078825	Distress zone	0.864201511	Distress zone
1829.hk	CN	1.357941592	Distress zone	1.427298576	Distress zone
0941.hk	CN	2.936618457	Grey zone	2.789570355	Grey zone
3323.hk	CN	0.337950565	Distress zone	0.497012524	Distress zone
0390.hk	CN	1.192247784	Distress zone	1.15640714	Distress zone
3320.hk	CN	2.201481901	Grey zone	2.007730795	Grey zone
1055.hk	CN	0.551047016	Distress zone	0.659758361	Distress zone
0728.hk	CN	1.062436514	Distress zone	1.190768229	Distress zone
0688.hk	CN	1.961825278	Grey zone	1.972625832	Grey zone
2196.hk	CN	1.215563686	Distress zone	1.074862508	Distress zone
1618.hk	CN	0.89432166	Distress zone	0.879602037	Distress zone
0857.hk	CN	1.181842765	Distress zone	1.329173357	Distress zone
3377.hk	CN	1.083233034	Distress zone	1.142850174	Distress zone
0347.hk	CN	0.71319999	Distress zone	1.340165363	Distress zone
0992.hk	CN	1.775439241	Distress zone	1.686162129	Distress zone
0753.hk	CN	0.740145558	Distress zone	0.812868908	Distress zone
0883.hk	CN	1.927734935	Grey zone	2.336603263	Grey zone
0232.hk	CN	0.975494042	Distress zone	1.838159662	Grey zone
1186.hk	CN	1.251163858	Distress zone	1.239974657	Distress zone
2607.hk	CN	2.345340814	Grey zone	2.234627565	Grey zone
0763.hk	CN	1.059868156	Distress zone	1.363358965	Distress zone
0670.hk	CN	0.482355283	Distress zone	0.455072698	Distress zone
2202.hk	CN	0.807062775	Distress zone	0.687163013	Distress zone
0489.hk	CN	2.294304149	Grey zone	2.115933926	Grey zone
0836.hk	CN	0.99429478	Distress zone	0.843126034	Distress zone
0392.hk	CN	1.105575201	Distress zone	1.11398028	Distress zone
0604.hk	CN	1.226738962	Distress zone	0.889164039	Distress zone
2380.hk	CN	0.53081341	Distress zone	0.310240584	Distress zone
0257.hk	CN	1.719616264	Distress zone	1.540477349	Distress zone
0103.hk	CN	0.669137242	Distress zone	0.689465751	Distress zone
1211.hk	CN	1.261617869	Distress zone	1.124410212	Distress zone
0916.hk	CN	0.406631641	Distress zone	0.557081816	Distress zone
0175.hk	CN	2.405647254	Grey zone	4.48691365	Safe zone
1088.hk	CN	1.243041938	Distress zone	1.543226046	Distress zone
0123.hk	CN	0.919420683	Distress zone	1.015226322	Distress zone
2128.hk	CN	2.456387186	Grey zone	2.308100735	Grey zone
2866.hk	CN	0.125023125	Distress zone	0.230025232	Distress zone
3360.hk	CN	0.346809818	Distress zone	0.355677289	Distress zone
6166.hk	CN	1.678887209	Distress zone	1.73679922	Distress zone

TABLE 8: The frequency of eleven sentiment attributes.

Positive graduation focus	15
Negative graduation force	15
Positive graduation force	274
Negative engagement	7
Positive engagement	223
Negative attitude appreciation	6
Positive attitude appreciation	345
Negative attitude judgement	11
Positive attitude judgement	378
Negative attitude affect	2
Positive attitude affect	87

the importance of business ethics reaches far beyond the strength of a management tea bond or employee loyalty. As with all business initiatives, the ethical operation of a company is directly related to companies' short-term or

long-term profitability. The reputation of a business in the surrounding community, other businesses, and individual investors is paramount in determining whether a company is a worthwhile investment. If a company is perceived to be unethical, investors are less inclined to support its operation.

In addition, in this study, we have divided the ethical responsibility into three parts: ethical accomplishment, ethical activity, and ethical ability. Ethical accomplishment means some awards that have been achieved by companies. Ethical ability expresses companies' core competence and organizational identity. Ethical activity makes investors to know about the activities and functions taking place in the company. Detailed samples are scheduled below:

- (a) We implemented new energy efficiency projects worldwide, including the implementation of the ISO

TABLE 9: Main sentimental themes in CEO letters.

Responsibility type	Detailed categories	References	Sources
Ethical responsibility	Ethical ability	34	23
	Ethical activity	113	32
	Ethical accomplishments	37	20
Technical responsibility	Technical ability	8	6
	Technical activity	33	13
	Technical accomplishments	18	13
Economic responsibility	Economic ability	5	5
	Economic activity	12	9
	Economic accomplishments	39	23
Philanthropic responsibility	Philanthropic ability	2	2
	Philanthropic activity	47	26
	Philanthropic accomplishments	4	3
Administrative responsibility	Administrative ability	3	3
	Administrative activity	20	15
	Administrative accomplishments	2	2
Political responsibility	Political ability	12	10
	Political activity	11	6
	Political accomplishments	0	0
Legal responsibility	Legal ability	1	1
	Legal activity	2	1
	Legal accomplishments	0	0

50001 Energy Management System in all our European Union locations (Ethical activity, Lenovo 2016).

- (b) In 2016, we have achieved safe flight hours of 2.38 million, transported 115 million passengers, eliminated incidents by human errors, continued to maintain the best safety record in China’s civil aviation (Ethical accomplishments, China Southern Airlines 2016).
- (c) Consequently, China Telecom is among the first batch of the national demonstration bases for entrepreneurship and innovation (Ethical ability, China Telecom 2016).

From the coding results, ethical activity occupies the largest proportion, which means that firms would prefer to demonstrate their actions and practices to the society. Corporations have more incentives to be ethical as the area of socially responsible and ethical investing keeps growing. An increasing number of investors are seeking ethically operating companies to invest, which drives more firms to take this issue seriously. With the consistent ethical behavior, an increasingly positive public image can be established, and to retain a positive image, companies must be committed to operating on an ethical foundation as it relates to the treatment of employees, respecting the surrounding environment and fair market practices in terms of price and consumer treatment.

*4.3. Machine-Learning Approach Forecasting Financial Performance.* We tested many machine-learning approaches and selected only the best outcomes. The measures of classification performance were, in addition to Acc, represented by the averages of standard statistics applied in classification tasks [70]: true-positive rate (TP), false-positive rate (FP), precision (Pre), recall (Re), F-measure (F-m),

TABLE 10: Average accuracy of the analyzed methods and weighted average TP, FP, Pre, Re, F-m, and ROC for classification of financial performance (safe zone, grey zone, and distress zone).

Model	Acc (%)	TP	FP	Precision	Recall	F-m	ROC
NB	0.6925	0.1187	0.3333	0.3097	0.1187	0.0451	0.3358
SVM	0.7022	0.1183	0.3333	0.3130	0.1183	0.0442	0.3298
LR	0.7046	0.1183	0.3333	0.3138	0.1183	0.0442	0.3324

and ROC curve. F-m is the weighted harmonic mean of precision and recall. ROC is a plot of the true-positive rate against the false-positive rate for the different cutpoints of a diagnostic test. In order to validate the accuracy, the experiments were realized using 10-fold cross validation. The best results in terms of the accuracy of correctly classified instances Acc (%) are presented in Table 10 (for the financial performance classes).

The results obtained by modeling point out that the logistic regression is more suitable for forecasting financial performance, reaching the highest accuracy of 70.46%. This evidence suggests that there exists a linear relationship between the sentiment and financial performance.

## 5. Conclusions

CEO letter contains information about corporate social responsibility performance in CSRR, which is designated for stakeholders to make their investment decisions. In this study, sentiment analysis has been applied to the evaluation of CEO letters from three perspectives: sentiment dictionary (microlevel), sentimental themes (mesolevel), and machine learning (macrolevel). In the microlevel analysis, a designated sentiment dictionary has been constructed for classifying sentiment attributes. The results denote that no

matter the companies are in an active or passive economic situation, they are focusing on using a great proportion of positive words to establish a company image and attract investors. In the mesolevel analysis, a comprehensive tree node structure was identified to discover the sentimental topics relevant to CSR. In terms of outcome, the CEO letters contain a large quantity of ethical-related information, especially concentrating on the ethical activities that firms have organized. In the macrolevel, the logistic regression approach achieves the best result in forecasting future financial performance, which proves that there is a linear relationship between the sentiment and economic performance. In other words, sentiment information in CEO letters can be regarded as a vital determinant for forecasting financial performance.

The distinct contribution of this paper is threefold. Firstly, an advanced technique of sentiment analysis utilizing appraisal theory has been conducted, which is potentially useful for detecting the “concealed” information in letters. Secondly, a sentiment dictionary has been constructed successfully and specifically for shareholders’ letters, which can significantly upgrade the accuracy of sentiment classification. Lastly, this research can guide companies to further enhance the technique of releasing nonfinancial information and display fundamental causes for diverse texts.

The current research has its own limitations. Utilizing Z-score, the quantitative assessment to define companies’ economic situation may not be adequate. In the Z-score model, the stock market value is only a static value at some point in time, which cannot reflect a dynamic fluctuation. In fact, the need to interpret the firms’ released information is to predict its future performance, it is insignificant which economic model was selected, and the critical point is the influence of sentiment on the perception of the company by its stakeholders.

In the future research, it is possible to apply other economic models for predicting financial performance. Especially with the popularity and high development of sentiment analysis, a great number of new approaches would be probed by text mining to detect more concealed information for investors’ decision-making and to be applied to other languages as well.

### Data Availability

All data, models, and code generated or used during the study appear in the submitted article.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This study was supported by the 2019 Project of the National Social Science Foundation of China: On the Overseas CSR Driving Forces and Influencing Mechanism for Chinese Enterprises (Grant no. 19BGL116).

### References

- [1] P. Bo and L. Lee, “A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts,” in *Proceedings of the Meeting on Association for Computational Linguistics*, Philadelphia, PA, USA, June 2004.
- [2] E. Abrahamson and E. Amir, “The information content of the president’s letter to shareholders,” *Journal of Business Finance & Accounting*, vol. 23, no. 8, pp. 1157–1182, 1996.
- [3] P. Hájek, V. Olej, and R. Myšková, “Forecasting stock prices using sentiment information in annual reports - a neural network and support vector regression approach,” *Wseas Transactions on Business & Economics*, vol. 10, no. 4, pp. 293–305, 2013.
- [4] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-based methods for sentiment analysis,” *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.
- [5] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis,” *Computational Linguistics*, vol. 35, no. 3, pp. 399–433, 2009.
- [6] C. J. Anderson and G. Imperia, “The corporate annual report: a photo analysis of male and female portrayals,” *Journal of Business Communication*, vol. 29, no. 2, pp. 113–128, 1992.
- [7] G. F. Kohut and A. H. Segars, “The president’s letter to stockholders: an examination of corporate communication strategy,” *Journal of Business Communication*, vol. 29, no. 1, pp. 7–21, 1992.
- [8] L. Patelli and M. Pedrini, “Is the optimism in CEO’s letters to shareholders sincere? Impression management versus communicative action during the economic crisis,” *Journal of Business Ethics*, vol. 124, no. 1, pp. 19–34, 2014.
- [9] P. Bo and L. Lee, *Opinion Mining and Sentiment Analysis*, Now Publishers Inc., Boston, MA, USA, 2008.
- [10] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery: opinion extraction and semantic classification of product reviews,” in *Proceedings of the International Conference on World Wide Web*, Banff, Canada, May 2003.
- [11] A. Kennedy and D. Inkpen, “Sentiment classification of movie reviews using contextual valence shifters,” *Computational Intelligence*, vol. 22, no. 2, pp. 110–125, 2006.
- [12] K. S. Srujan, S. S. Nikhil, H. R. Rao, K. Karthik, B. S. Harish, and H. M. K. Kumar, *Classification of Amazon Book Reviews Based on Sentiment Analysis*, Springer, Singapore, 2018.
- [13] J. Feng, C. Gong, X. Li, and R. Y. K. Lau, “Automatic approach of sentiment lexicon generation for mobile shopping reviews,” *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 9839432, 13 pages, 2018.
- [14] B. Huang, G. Yu, and H. R. Karimi, “The finding and dynamic detection of opinion leaders in social network,” *Mathematical Problems in Engineering*, vol. 2014, no. 1, 7 pages, 2014.
- [15] R. Quratulain, G. Sayeed, Sajjad, and Haider, “Lexicon-based sentiment analysis of teachers’ evaluation,” *Applied Computational Intelligence and Soft Computing*, vol. 2016, Article ID 2385429, 12 pages, 2016.
- [16] M. Stewart, “The language of praise and criticism in a student evaluation survey,” *Studies in Educational Evaluation*, vol. 45, pp. 1–9, 2015.
- [17] C. L. Chen, C. L. Liu, Y. C. Chang, and H. P. Tsai, “Opinion mining for relating subjective expressions and annual earnings in US financial statements,” *Journal of Information Science & Engineering*, vol. 29, no. 4, pp. 743–764, 2012.
- [18] J. L. Hobson, W. J. Mayew, and M. Venkatachalam, “Discussion of analyzing speech to detect financial misreporting,”

- Journal of Accounting Research*, vol. 50, no. 2, pp. 393–400, 2012.
- [19] C. Huang, X. Yang, X. Yang, and H. Sheng, “An empirical study of the effect of investor sentiment on returns of different industries,” *Mathematical Problems in Engineering*, vol. 2014, Article ID 545723, 11 pages, 2014.
- [20] C. Xie and Y. Wang, “Does online investor sentiment affect the asset price movement? Evidence from the Chinese stock market,” *Mathematical Problems in Engineering*, vol. 2017, Article ID 2407086, 11 pages, 2017.
- [21] M. Taboada, “Sentiment analysis: an overview from linguistics,” *Linguistics*, vol. 2, no. 1, 2016.
- [22] C. J. Hutto and E. Gilbert, VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text, in *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, Ann Arbor, MI, USA, 2014.
- [23] P. D. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” in *Proceedings of Annual Meeting of the Association for Computational Linguistics*, pp. 417–424, Philadelphia, PA, USA, July 2002.
- [24] V. Hatzivassiloglou and K. R. Mckeown, “Predicting the semantic orientation of adjectives,” in *Proceedings of the ACL*, pp. 174–181, Autrans, France, 1997.
- [25] P. D. Turney and M. L. Littman, “Measuring praise and criticism,” *Acm Transactions on Information Systems*, vol. 21, no. 4, pp. 315–346, 2003.
- [26] M. Taboada, C. Anthony, and K. Voll, “Methods for creating semantic orientation dictionaries,” in *Proceedings of the Language Resources & Evaluation*, Genoa, Italy, May 2006.
- [27] X. Ding, L. Bing, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in *Proceedings of the International Conference on Web Search & Data Mining*, Cambridge, UK, 2008.
- [28] A. Dey, M. Jenamani, and J. J. Thakkar, “Senti-N-Gram: an n-gram lexicon for sentiment analysis,” *Expert Systems with Applications*, vol. 103, Article ID S095741741830143X, 2018.
- [29] J. Brooke, M. Tofiloski, and M. Taboada, “Cross-linguistic sentiment analysis: from English to Spanish,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Borovets, Bulgaria, September 2009.
- [30] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in *Proceedings of the Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, vol. 10, Philadelphia, PA, USA, July 2002.
- [31] E. Boiy and M.-F. Moens, “A machine learning approach to sentiment analysis in multilingual Web texts,” *Information Retrieval*, vol. 12, no. 5, pp. 526–558, 2009.
- [32] L. I. Jun and M. Sun, “Experimental study on sentiment classification of Chinese review using machine learning techniques,” in *Proceedings of the IEEE International Conference on Natural Language Processing & Knowledge Engineering*, Beijing, China, 2007.
- [33] R. F. Bruce and J. M. Wiebe, *Recognizing Subjectivity: A Case Study in Manual Tagging*, Cambridge University Press, Cambridge, UK, 1999.
- [34] Gamon and Michael, “Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis,” in *Proceedings of the International Conference on Computational Linguistics*, 2004.
- [35] C. Yang, X. Tang, Y. C. Wong, and C. P. Wei, “Understanding online consumer review opinion with sentiment analysis using machine learning,” *Pacific Asia Journal of the Association for Information Systems*, vol. 2, no. 3, 2010.
- [36] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, “Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning,” in *Proceedings of the Fourth International Conference on Information, Mikrolimano, Greece, 2013*.
- [37] J. Singh, G. Singh, and R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” *Human-centric Computing and Information Sciences*, vol. 7, no. 1, p. 32, 2017.
- [38] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, “Sentiment analysis of tweets using machine learning approach,” in *Proceedings of the 2018 Eleventh International Conference on Contemporary Computing (IC3)*, Noida, India, August 2018.
- [39] S. Yuan, H. Wang, and M. Zhu, “Sustainable strategy for corporate governance based on the sentiment analysis of financial reports with CSR,” *Financial Innovation*, vol. 4, no. 1, p. 2, 2018.
- [40] A. Aue and M. Gamon, “Customizing sentiment classifiers to new domains: a case study,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, Varna, Bulgaria, September 2005.
- [41] S. González-Bailón and G. Paltoglou, “Signals of public opinion in online communication,” *The ANNALS of the American Academy of Political and Social Science*, vol. 659, no. 1, pp. 95–107, 2015.
- [42] T. Loughran and B. McDonald, “When is a liability not a liability? Textual analysis, dictionaries, and 10-ks,” *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.
- [43] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, “Evaluating sentiment in financial news articles,” *Decision Support Systems*, vol. 53, no. 3, 2012.
- [44] P. Hájek and V. Olej, “Evaluating sentiment in annual reports for financial distress prediction using neural networks and support vector machines,” *Engineering Applications of Neural Networks*, Springer, Berlin, Germany, 2013.
- [45] Y. Q. Xin, P. Srinivasan, and H. Yong, “Supervised learning models to predict firm performance with annual reports: an empirical study,” *Journal of the American Society for Information Science & Technology*, vol. 65, no. 2, pp. 400–413, 2014.
- [46] P. Hajek, V. Olej, and R. Myskova, “Forecasting corporate financial performance using sentiment in annual reports for stakeholders’ decision-making,” *Technological and Economic Development of Economy*, vol. 20, no. 4, pp. 721–738, 2014.
- [47] S. Tong, W. Jia, P. Zhang, C. Yu, and D. Wang, “Predicting stock price returns using microblog sentiment for Chinese stock market,” in *Proceedings of the 2017 3rd International Conference on Big Data Computing and Communications (BIGCOM)*, Chengdu, China, August 2017.
- [48] J. Wiebe, T. Wilson, and C. Cardie, “Annotating expressions of opinions and emotions in language,” *Language Resources and Evaluation*, vol. 39, no. 2-3, pp. 165–210, 2005.
- [49] N. Asher, F. Benamara, and Y. Y. Mathieu, “Appraisal of opinion expressions in discourse,” *Linguisticae Investigationes Revue Internationale De Linguistique Française Et De Linguistique Générale*, vol. 32, no. 2, pp. 279–292, 2009.
- [50] C. S. G. Khoo, A. Nourbakhsh, and J. C. Na, “Sentiment analysis of online news text: a case study of appraisal theory,” *Online Information Review*, vol. 36, no. 6, pp. 680–686, 2012.
- [51] M. A. K. Halliday, *An Introduction to Functional Grammar*, Oxford University Press Inc., London, UK, 1994.
- [52] J. R. Martin and P. R. R. White, *Language of Evaluation: Appraisal in English*, Palgrave Macmillan, London, UK, 2005.

- [53] S. Eggins, *An Introduction to Systemic Functional Linguistics*, Pinter, London, UK, 1994.
- [54] M. Taboada and J. Grieve, "Analyzing appraisal automatically," in *Proceedings of the Aaai Spring Symposium on Exploring Attitude & Affect in Text Theories & Applications*, pp. 158–161, Stanford, CA, USA, January 2004.
- [55] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the ACM International Conference on Information & Knowledge Management*, Bremen, Germany, October 2005.
- [56] J. Read and J. Carroll, "Annotating expressions of appraisal in english," in *Proceedings of the Linguistic Annotation Workshop*, Prague, Czech Republic, June 2007.
- [57] S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, "Automatically determining attitude type and force for sentiment analysis," in *Proceedings of the Human Language Technology Challenges of the Information Society*, Poznan, Poland, October 2009.
- [58] A. Balahur, J. M. Hermida, A. Montoyo, and R. Muñoz, *EmotiNet: A Knowledge Base for Emotion Detection in Text Built on the Appraisal Theories*, Springer, Berlin, Heidelberg, 2011.
- [59] K. Bloom, "Sentiment analysis based on appraisal theory and functional local grammars," Dissertation thesis, Illinois Institute of Technology, Chicago, IL, USA, 2011.
- [60] P. Korenek and M. Šimko, "Sentiment analysis on microblog utilizing appraisal theory," *World Wide Web*, vol. 17, no. 4, pp. 847–867, 2014.
- [61] X.-L. Cui and J. S. Shibamoto-Smith, "A corpus-based study on Chinese sentiment parameters of Chinese sentiment discourse," *Chinese Language and Discourse*, vol. 5, no. 2, pp. 185–210, 2014.
- [62] A. Edwardsjones, *Qualitative Data Analysis with NVIVO*, Sage Publications, vol. 15, p. 868, Thousand Oaks, CA, USA, 2010.
- [63] R. Boyatzis, *Transforming Qualitative Information: Thematic Analysis and Code Development*, Sage Publications, Thousand Oaks, CA, USA, 1998.
- [64] P. Bazeley, *Qualitative Data Analysis with NVivo*, Sage, London, UK, 2007.
- [65] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.
- [66] P. Hájek, "Credit rating analysis using adaptive fuzzy rule-based systems: an industry-specific approach," *Central European Journal of Operations Research*, vol. 20, no. 3, pp. 421–434, 2012.
- [67] S. L. Cessie and J. C. V. Houwelingen, "Ridge estimators in logistic regression," *Applied Statistics*, vol. 41, no. 1, pp. 191–201, 1992.
- [68] E. Goffman, *Presentation of Self in Every Day Life*, vol. 21, pp. 14–15, Doubleday, New York, NY, USA, 1959.
- [69] A. B. Carroll, "The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders," *Business Horizons*, vol. 34, no. 4, pp. 39–48, 1991.
- [70] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 1, no. 2, pp. 37–63, 2011.