International Journal of Information Management Data Insights 3 (2023) 100155

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Job satisfaction, management sentiment, and financial performance: Text analysis with job reviews from indeed.com

Sijie Feng

New York University, 44 West 4th St, New York, NY 10012, USA

ARTICLE INFO

Keywords:
Applied machine learning
Data science
Management
Technology
Econometrics
Data analysis

ABSTRACT

Employee job reviews contain information on many firm characteristics that are unobservable to those outside the company. Job review sites such as Glassdoor and Indeed are used by job hunters and current employees to inform each other of prevailing workplace culture. This paper collects over 1.2 million job reviews from Indeed.com. Machine learning and natural language processing methods are used to obtain qualitative data about firms through employees' sentiment regarding company management. This paper seeks to quantify how employees' sentiment affects firms' financial outcomes. By matching review data with firm level financial data, the analyses find that a 1% increase in ratings increases market value by approximately 0.68-0.73%, revenue by 0.62-1.01%. However, increases in management sentiment may have a non-monotonic effect on financial performance, demonstrating that trade-offs are inherent in balancing financial results and employees' satisfaction regarding their roles.

1. Introduction

Job reviews have become a rich source of "inside" information on qualitative factors relevant for employee satisfaction, including workplace culture, management practices, and company strategy. In recent years, large job review sites have collected tens of millions of reviews from dozens of nations internationally. Surveys have indicated that over almost 50 percent of job seekers use job review sites to aid their search. In addition, companies have made impactful changes to their strategies and internal policies in response to reviews that have been made public via these sites (Dube & Zhu, 2021).

While job reviews have proliferated, the open question is now how to extract relevant information from text for analysis. One strand of literature uses machine learning and natural language processing techniques to convert review text into quantitative data. The relevant question is what determines employee satisfaction, based on the aspects of the job discussed in the review (e.g. example work life balance, company culture, relationship with management, benefits). Most studies in this area take the entire review text to be relevant. However, reviews also come with ratings, which already represent the bulk of the employee's sentiment. Thus, the question arises of how to identify further information of value from the review text.

This paper seeks to address the question of how to find valuable information from review text. The approach of this paper is not to use the entire review text, but to focus on portions of the review that discuss *management*. Another goal of this paper is to connect employee sentiment with actual company financial performance, that is, to quantify tangibly how sentiment may have strong real world implications for firms. For this reason, the focus on management is important as management within companies largely control the strategic decision making that affect financial outcomes. Other factors that are discussed in the job review, such as the amenities of the workplace, might not be as relevant and are not highlighted in this study.

From the careers site Indeed.com, approximately 1.2 million job reviews are retrieved dating from 2012 to 2017. Firm level financial data is sourced from Compustat. Because only data from publicly listed companies are available from Compustat, only review text from public companies are mined for this study. To extract the portion of text relevant for management, Doc2Vec is used in order to convert text into embeddings. Then, two supervised learning methods, XGBoost and Deep Neural Networks, are used to fit a sentiment classifier that is trained using review text. These measures capture the dimensions of sentiment that are specific to the domain of job reviews. The use of supervised algorithms takes advantage of the labeled nature of ratings out of five provided by reviews, which makes this study a standard classification problem. Google Sentiments API is also used as a sentiment classifier, which captures broader sentiment within the text in a common language usage context. In addition, this paper highlights the importance of taking into consideration how generic a review is. A measure of genericness is introduced which is based on the term frequency-inverse document frequency (tf-idf) of the text. Further discussion of these techniques, including details on the algorithms used, are in Section 3.

The main contribution of this paper is to quantify how employees' sentiment towards management affects a firm's financial outcomes. OLS

E-mail address: sfeng@stern.nyu.edu

https://doi.org/10.1016/j.jjimei.2023.100155
Received 9 February 2022; Received in revised form 29 November 2022; Accepted 6 January 2023

regression is used to estimate the size of these effects. I find strong evidence of there being a positive relationship between ratings and financial performance in the cross section. A 1% increase in ratings increases market value by approximately 0.68-0.73%, revenue by 0.62-1.01%. There is a significant relationship also between management sentiment, the non-generic index, and financial performance, but their relationships are non-monotonic: positive relationship for certain range of values, and negative relationship in others. This implies that, in certain cases, firms may be able to raise financial performance at the expense of employees in the short term.

These findings are also in line with evidence from literature on the relationship between workplace culture and firm performance. In a real world example, the financial company Wells Fargo implemented a high pressure work environment under the leadership of their management team, which caused short term financial gain but large losses long term, with the widespread reporting of their unlawful and unethical behaviour (Ji, Rozenbaum, & Welch, 2017). While this situation is at one extreme, many companies may find such trade-offs present in their decision making, albeit to a lesser extent. Results from this paper corroborates the existence of such trade-offs, and their importance for corporate strategy.

The paper proceeds as follows. Section 2 discusses the relevant literature informing the analysis. Section 3 specifies the data collection method, machine learning techniques, and estimation models used. Section 4 outlines the results of the analyses and main findings. Section 5 discusses some of this paper's contributions to the literature, as well as practical implications. Finally, 6 summarizes the paper and outlines some of the limitations of this approach as well as suggesting extensions of the study.

2. Literature review

This paper makes novel contributions to several existing literatures. Prior papers utilize the methodology text mining and machine learning methods to find meaningful information in online job reviews. The core rationale behind this literature is that "corporate culture" has an impact on workers' performance. Corporate culture has been defined by Schein (1990) as shared assumptions, values, and beliefs that help employees understand which behaviors are and are not appropriate.

Employee sentiment towards their roles plays both an important role in their on-the-job performance and the success of the business. Employee sentiment is theorized to have a positive correlation to their performance because higher morale would lead to improved productivity (Judge, Thoresen, Bono, & Patton, 2001). While Iaffaldano & Muchinsky (1985) finds weak empirical evidence of this link in a meta-analysis, subsequently Judge et al. (2001) finds a positive correlation of 0.3 between sentiment and performance. Harter, Schmidt, & Hayes (2002) focuses instead on the business as the unit of analysis, and surveys 7939 business units to demonstrate a strong positive correlation between overall employee satisfaction and business outcomes, including profitability. The authors recommend shifting management practice towards bolstering employee satisfaction.

Employee sentiment and satisfaction clearly play a large role in business success. Job review text provide new detail and insight into employee sentiment. Popadak (2013) uses job review data to assess the effect of active shareholder governance on both financial performance and corporate culture. Ji et al. (2017) find that firms with lower levels of job satisfaction and lower levels of "culture and values" are more likely to be subjected to SEC fraud enforcement actions and securities class action lawsuits, although they are also more likely to beat market earning expectations. Luo, Zhou, & Shon (2016) found a positive effect of employee satisfaction on corporate performance using over 250,000 reviews from Glassdoor. Dabirian, Kietzmann, & Diba (2017) look at only the ten best and worst firms to work for on Glassdoor, and find that conventional employee value propositions (such as social element of work, challenging work tasks, and professional development) do not all matter to the same extent and that their relative weightings differ across organizations.

Similarly, a class of literature extends the examination of review text by using machine learning techniques to analyze the text. Originally conceived in Blei, Ng, & Jordan (2003), Latent Dirchlet Allocation (hereby referred to as LDA) uses unlabelled text data as input and outputs "topics" that are probability distribution over words in the vocabulary of the text corpus. Structural Topic Models (Roberts et al., 2014; Roberts, Stewart, Tingley, Airoldi et al., 2013) extend LDA by adding covariates to the model and using domain-specific knowledge. Lee & Kang (2017) used Latent Dirichlet Allocation (LDA) on Glassdoor reviews and found that job satisfaction factors such as "Culture and Values" and "Senior Management" have a positive influence on retention, and a negative influence on turnover. Goldberg & Zaman (2018) develop a text analytics technique specific to the HRM domain to predict employee dissatisfaction using reviews from Indeed. Jung & Suh (2019) extend text-based methods beyond LDA to dominance analysis and correspondence analysis in order to understand the sentiment and importance of each job satisfaction factor at the industry and company level. Sainju, Hartwell, & Edwards (2021) also extends beyond LDA, including the use of Structural Topic Models (STM) and concludes that leadership and management, overwork, and a stressful environment were the dominant factors leading to higher employee turnover.

The novel contribution of this paper is to utilize supervised machine learning approaches in the sentiment analysis of the job review text. This differs from the previous literature, which relies substantially on unsupervised algorithms such as LDA. With methods such as LDA and STM, other methods must be overlaid to gauge the sentiment of text data. For example in Jung & Suh (2019), dominance analysis is used for this purpose. This paper uses more recently popularized supervised learning algorithms, which has the advantage of using the inherent labels in ratings themselves to quantify sentiment in reviews. Details on the algorithms used are discussed in 3.4

Sentiment analysis with machine learning and natural language processing methods also extend beyond the domain of employee satisfaction. Broadly, sentiment analysis seeks to understand the viewpoint or basic human emotion underlying textual data (Pak & Paroubek, 2010; Pang & Lee, 2004). There are a wide range of applications for sentiment analysis. The examination of polarity and subjectivity of positive and negative sentiment has been used to analyze political protest with Twitter data (Neogi, Garg, Mishra, & Dwivedi, 2021); to make recommendations for hotels using tourist reviews (Mishra, Urolagin et al., 2019; Mishra, Urolagin, & Jothi, 2020); to augment drug recommendations for medical patients (Paliwal, Mishra, Mishra, Nawaz, & Senthilkumar). Another approach is to be more agnostic about the domain of application and focus instead on increasing the accuracy of sentiment predictions, particularly using Twitter data (Alharbi & de Doncker, 2019; Srivastava, Singh, & Drall, 2019).

This paper also contributes to a wider strand of literature which links customer and employee performance to firm performance, by using survey data (Bernhardt, Donthu, & Kennett, 2000, Huang, Li, Meschke, & Guthrie, 2015) or "Best Places to Work" lists (Edmans, 2011). These papers argue that there are human capital benefits to healthier company cultures. Cultural effects on employee retention is also the focus of a related literature on firm operations (De Bruecker, Van den Bergh, Beliën, & Demeulemeester, 2015; Corominas, Lusa, & Olivella, 2012; Bordoloi & Matsuo, 2001), which uses mathematical models to derive negative implications of higher employee turnover, particularly for firms that depend more on human-to-human interactions. A clear mechanism in which employee turnover has a negative impact is through its increase of HR costs for companies (Huselid, 1995; Cascio, 1991).

Thus, stakeholders within companies could greatly benefit from investing in information management that would minimize costly and disruptive employee losses. This paper outlines methods by which information from employees can be internalized for key decision making. Review platforms such as Glassdoor and Indeed offer a plethora

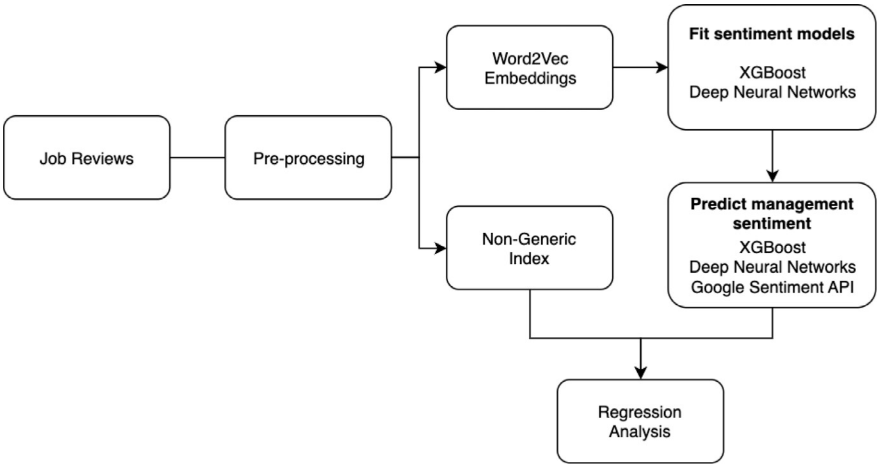

Fig. 1. Structure of methods.

of readily available and valuable information for stakeholders that can improve business decisions and performance. (Sheng, Wang, & Amankwah-Amoah, 2021; Symitsi, Stamolampros, & Daskalakis, 2018; Symitsi, Stamolampros, Daskalakis, & Korfiatis, 2021) Previous literature has argued that firm management and strategy could vastly benefit from the internalization of insights from computational business analytics (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012, Kunc & O'brien, 2019; Xu, Frankwick, & Ramirez, 2016).

The insights provided by this paper are crucial given the effect that high employee turnover and low retention in the wake of "the Great Resignation" (Sheather & Slattery, 2021; Sull, Sull, & Zweig, 2022; Van Dam, 2021), which demonstrates the magnitude of the impact that employee's perception of company culture can have on the wider economy. Poor management practice is being identified as a crucial factor (Robertson, 2021, Meduri & Jindal, 2021) in exacerbating these shifts. With rapidly changing labour market conditions favoring work from home (Hopkins & Figaro, 2021), firms have much higher incentives to understand what keeps employees satisfied.

3. Methodology

The vast quantity of online review data available forms a trove of insights on how employees view their jobs and their companies. In this paper, review data is first collected from Indeed.com (Section 3.1). Then, review text is preprocessed and converted into vector embeddings using Word2Vec (Section 3.2, 3.3). Supervised learning models are fit on the review data, and predictions for management sentiment are made using portions of review text (Section 3.4). A method of dealing with generic reviews is proposed in Section 3.5. Finally, OLS regression model is estimated to quantify effect of sentiment on financial outcomes of companies (3.6). An overview of the methods used in this paper are outlined in Fig. 1.

3.1. Data collection

Job rating and review data is collected from the job site Indeed, the largest job site in the US. The time frame available is 2012–2017. Each review contains the following information: the job title of the position held by the reviewer, the date of the review, a rating from 1 to 5 of the company, the location of the job, as well as the text of the review alongside pros and cons (see Fig. A.2 for an example of a review).

Financial data from firms is sourced from Compustat, a comprehensive aggregated historical dataset for publicly listed companies. From Compustat, I collect data on Earnings Per Share (EPS), Market Value, and Revenue. These variables were selected to be the standard indicators of financial performance of each company over the time frame that reviews are collected.

As the focus of this paper is on explaining the effect of employee sentiment on financial outcomes, only job reviews from companies that are available within the Compustat database are selected. While this approach has drawbacks in generalizability, it is the best available approach without reliable financial data for private businesses or small and medium sized businesses.

Using the list of companies from Compustat, the Python package `beautifulsoup` is used mine review data from Indeed.com.

3.2. Data pre-processing

The texts of reviews are "stemmed" to their base word (e.g. "competitive" to "compet") and stop words that are common in English vocabulary (e.g. "and", "the") are removed. Other words that occur in less than ten reviews are also removed to constrain the dimensionality of the resulting matrix that is fed into the algorithm. Both one and two-grams, that is single words and sequences of two words, are used in the analysis.

3.3. Text mining

In order to measure management sentiment, the Word2Vec algorithm is used as a technique to identify which terms in the review text are actually related to management. Word2Vec (introduced in Goldberg & Levy (2014); Mikolov, Grave, Bojanowski, Puhrsch, & Joulin (2017)) uses a shallow neural network to create vectors (i.e. embeddings) for words so that words which are closer to one another in meaning close to each other in the vector space. This method is suitable for identifying synonyms of specific words, like "management", as it uses the word's surrounding context to identify when two words might be similar.

Word2Vec is performed on a sample of reviews. Then, using the resulting fitted model, a list of words most similar to "management" is retrieved. According to the Word2Vec the top ten words most similar to management are: *manag, manger, supervisor, mgmt, manager, uppermanag, mgt, leader, leadership, mgrs*. While some of these terms are contractions for manager or management, other words, such as *leader* and *supervisor*, are similar only in their contextual use to how manager is used.

For each review that contains a management term, the management phrase consisting of the five words prior and post to the management term is extracted. For example, a entire job review (see Table A.4) is: "I didn't like the work environment at Walmart at all. Some of the employees where rude and very unfriendly. It wasn't hard work if you could keep up. I didn't feel that I could go to the management with issues that I had. I felt like they would judge me and think less of me." The management portion of the review would constitute a management phrase: "I didn't feel that I could go to the management with issues that I had."

3.4. Predicting management sentiment

3.4.1. Supervised training

In order to predict the sentiment in review text towards management, the best performing supervised algorithms are selected and trained on reviews and ratings. The ratings out of five attached to each review will be the label for the supervised classification algorithm, with 4 or 5 stars representing positive sentiment and 1 or 2 stars representing negative sentiment. The features for the model are the Word2Vec embeddings based on review text. This method is selected because it allows for a domain based measure of sentiment that is specific to job reviews, and the semantic associations of sentiment specific to describing areas of work.

More specifically, two supervised algorithms are selected. The first is XGBoost (Chen & Guestrin, 2016; Chen et al., 2015), which is a highly efficient decision tree-based method. The advantages of XGBoost is its ability to not overfit on training data. Since it is a gradient boosting method, this means that the algorithm trains an ensemble of simple models to make predictions. For XGBoost, the component simple models are decision trees. A feature, in this case a word or term, is important when it is used to make an important decision or "split" in the tree. For a single decision tree, the importance of a term is the amount that the split at the term improves performance. Practically, this means that the documents split by whether or not they contain the term have greater accuracy, i.e. matches of the predicted label to the actual training label. Computationally, it is implemented in Python with the `xgboost` package. As a sentiment classifier, it has been shown to perform well in (Giannakas, Troussas, Krouska, Sgouropoulou, & Voyiatzis, 2021; Memon, Patel, & Patel, 2019), largely due to its ability to "combine" the predictive capacity to the decision trees in its ensemble.

The second supervised algorithm is Deep Neural Networks. This class of model is a "latent" algorithm that uses outputs from intermediate layer models as inputs for the following layer of outputs, mimicking the structure of brain neurons (Bishop, 1994; Van Der Smagt, Kröse, Krose, & Smagt, 1993; Anderson, 1995). The implementation was done using `tensorflow` in Python. This model may be particularly suited towards sentiment analysis because of its ability to make nonlinear associations between text embeddings (Severyn & Moschitti, 2015; Alharbi & de Doncker, 2019).

3.4.2. Google cloud sentiment SDK

Supervised methods allow for a domain-specific approach to analyzing sentiment, by restricting itself only to the review texts provided. Management sentiment does not need to be exclusively trained on job reviews data, since other terms unrelated to sentiment (such as industry related terms, for example *sales, transport*, etc.) may affect the predictions. Google Sentiment Analysis is a state of the art API which can analyse the sentiment of any text, and gives a score between $[-1, 1]$. Both a supervised approach and Google Sentiment Analysis are used in order to capture sentiment within the review text that is both domain specific to job review text, and sentiment within the broader context of everyday language.

3.5. Accounting for generic reviews

Not all reviews are equally informative. Some reviews are short and do not discuss the specificity of the job or work environment. For example, a review saying "Management is great" would be a priori less informative than a review that says "Management is great because they are attentive to feedback and give detailed reviews on performance." Without accounting for a measure of genericness, all reviews would have equal weighting.

I introduce a measure called the Non-Generic Index (hereafter known as NGI) to account for the level of genericness in a review. The NGI is based on the term frequency-inverse document frequency (tf-idf) of a particular document, which is a popular method of converting text into vector data (Ramos et al., 2003). Roughly, tf-idf measures how important terms are to a particular document. This is measured by the frequency that the term appears in the document, then offset by how frequently the term appears across all documents. To measure how generic a document is, I calculate the mean tf-idf of all unique terms in a document. As discussed in Section A.1, the more frequent a term across all documents, the lower its inverse document frequency (idf) will be. Thus, if a document contains many words that are commonly used across all reviews, the value of its average tf-idf (and NGI) will be low. Conversely, if a review contains many words that are infrequently used across all documents, its NGI will be high.

For example, a management phrase consisting of "*manag staff great* " is more generic and thus has a low non-generic index of 0.0004, while "*leadership skill includ profession busi ethic* " which includes more uncommon words, has a higher non-generic index of 0.001. Each review and management phrase is given its own non-generic index weighting, so that when ratings are averaged by company, the more generic reviews are weighted less than the less generic reviews. Such a measure would allow firms manage information in a manner that highlights relatively more informative reviews, which could improve their decision making around what manner of feedback to consider more thoroughly.

3.6. Analysis

The goal of the analysis is to determine whether the review text based features derived from machine learning methods discussed above have any effect on the firm's overall financial performance. The purpose of the OLS estimation is to quantify the effect of management sentiment, ascertained through machine learning methods, on the company's financial outcomes. While previous literature (for example Bernhardt et al., 2000) uses survey data to link corporate culture with firm performance, the joining of sentiment analysis to financial data permits a more direct measure of how sentiment towards the company affects performance. With OLS estimation, the coefficient values on sentiment will directly indicate the magnitude of its effect on a range of different financial outcome variables.

Data is aggregated for each variable in the dataset by company and year. The firm's financial performance will be the dependent variable in the regression, and is summarized by variables:

- eps: earnings per share
- $mktval$: market value
- rev: revenue

The firm's review text based features will be the explanatory variables for the regression. They are:

- $rating$: average rating
- $manag\ sent_{NN}$: average management sentiment predicted by the Deep Neural Network algorithm
- NGI_r: average non-generic index for review texts
- NGI_{ms} represents the non-generic index for management phrases within review text

For the regression method, the goal is to ascertain whether or not the estimated beta values for the text based features are (a) statistically significant at the 5% level, and (b) are positive or negative. Positive beta estimates indicate that firm performance is positively correlated with review text management sentiment, while negative estimates indicate that firm performance will increase if management sentiment decreases.

OLS Regression Model

The regression model for the pooled OLS regression is the following equation:

$$\begin{aligned} y_{i,t} =& \beta_0 + \beta_1 rating_{i,t-1} + \beta_2 manag\ sent_{NN,i,t-1} + \beta_3 manag\ sent_{Goog,i,t-1} \\ &+ \beta_4 NGI_{r,i,t-1} + \beta_5 NGI_{ms,i,t-1} + year\ FE + \epsilon_{i,t} \end{aligned} \quad (1)$$

Table 1
Summary statistics of companies with reviews available from Indeed and Compustat.

	2012	2013	2014	2015	2016	2017	All
Mean Rating	3.75	3.77	3.77	3.77	3.74	3.65	3.73
Number of Reviews	115,736	171,864	184,896	203,985	196,718	365,851	1,240,541
Number of Companies	1687	1892	1985	2109	2128	2390	2738
Mean Earnings Per Share	0.4	0.35	0.38	0.2	0.31	0.42	0.23
Mean Market Value	7346.26	8222.08	8937.79	8970.79	9567.04	11579.22	8946.93
Mean Revenue	1909.65	1843.68	1787.14	1632.13	1701.9	1950.19	1851.98

In order to account for the interaction between the non-generic measure of review text and the predicted management sentiment, I also estimate a regression that includes the product of the non-generic indices with their measures:

$$\begin{aligned} y_{i,t} =& \beta_0 + \beta_1 rating_{i,t-1} + \beta_2 manag\ sent_{NN,i,t-1} + \beta_3 manag\ sent_{Goog,i,t-1} \\ &+ \beta_4 NGI_{r,i,t-1} + \beta_5 NGI_{ms,i,t-1} + \beta_6 NGI_r \cdot rating_{i,t-1} \\ &+ \beta_7 NGI_{ms} \cdot manag\ sent_{NN,i,t-1} + \beta_8 NGI_{ms} \cdot manag\ sent_{Goog,i,t-1} \\ &+ year\ FE + \epsilon_{i,t} \end{aligned} \quad (2)$$

Here, $NGI_r \cdot rating_{i,t}$ would represent the average rating for company i in year $t-1$, weighted by the genericness of each review as given by NGI_r. Similarly, $NGI_{ms} \cdot manag\ sent_{NN,i,t-1}$ represents the average deep neural network management sentiment, weighted by the genericness of the management phrase. Thus, the ratings, sentiments, and NGIs will enter non-linearly in the regression. Year fixed effects $year\ FE$ are also included to absorb trends over time. Log transformation is applied to all variables to allow for clear interpretation of coefficients.

4. Findings

4.1. Analyzing review data

Overall, there are 1.24 million reviews for 2738 companies. The average rating across all reviews is 3.73. Most reviewers give their jobs a positive rating of 4 (31%) or 5 (21%), with only 14% of reviewers giving their jobs a negative rating of 1 or 2 out of 5. (See Table A.4 for examples). The number of reviews grew annually, with over 365,000 reviews collected from 2017. The number of companies covered in Compustat also grew annually, to 2390 in 2017. Mean ratings remain roughly stable on average for each year, with 2017 having the lowest mean ratings. This may be due to the influx of reviews in 2017 having greater variation. See Table 1 for a summary of the data.

4.2. Words most related to positive and negative reviews

4.2.1. Naive approach: Word frequency

In order to better understand the difference in words used to describe positive experiences of jobs (ratings 4 or 5) compared to negative ratings (1 or 2), I count the frequency of terms that appear in positive and negative reviews. Table A.5 show the 20 most common words that appear across all reviews; Table A.6 show the common words for positive reviews; Table A.7 for the negative reviews.

While all reviews feature words that describe aspects of the job (*customers, day, company, people, time*), the positive reviews contain words that reflect constructive aspects of work (*learn, enjoy*), while negative reviews mention *pay*, which potentially reflects their dissatisfaction with their compensation.

4.2.2. Supervised learning approach

Another method to uncover words most related to positive and negative reviews is to use the most important features in the fitted XGBoost model. Under this approach, a vivid description arises of mostly what makes certain positions unpopular. Besides emotive language such as *horrible, poor, terrible, worst,* there are mentions of *firings* from employees leaving negative reviews. This aligns with a priori expectations which shows that the algorithm is reasonable in finding discerning terms. More interesting is the inclusion of *turnover*, which clearly represents continuing challenges at the firm leading to many employees leaving unsatisfied with their roles and companies. Descriptions of *unprofessional* and *unorganised* work environments are also important determinants of negative job experiences, as is *micromanagement*. On the positive side, reviews that mentioned that they had *favourite parts* of their work were far more likely to leave high ratings, as well as those that described a work environment *like family*. Firms could harness the information management of their online reviews in order to improve their decision making around investment in human resources, by investing more in reducing the negative aspects of review, for example by providing on-the-job training to reduce unprofessionalism.

Table 2
Most important features for discerning between positive and negative reviews.

Term	Importance
horribl	0.0060
poor	0.0058
terribl	0.0045
worst	0.0042
unprofession	0.0029
told	0.0028
fire	0.0026
dont	0.0022
turnov	0.0022
lie	0.0021
suck	0.0019
let go	0.0018
upper	0.0018
favorit part	0.0017
rude	
like famili	0.0017
dull	0.0017
yell	0.0016
hostil	0.0016
unless	0.0015
micro manag	0.0015
unorgan	0.0015
ensur	0.0015
joke	0.0015
bottom line	0.0015
excel	0.0015
various	0.0015
challeng	0.0015
promis	0.0014
lack	0.0014

4.3. Results from regression

Results from the regression analysis allows for the direct quantification of review based features on financial performance. Job ratings have a positive and significant effect on financial performance measured using all three variables of interest. All results are listed in Table 3. Across firms, a 1% increase in the ratings index increases market value by approximately 0.68-0.73%, revenue by 0.62-1.01%, and earnings per share by 0-0.06%.

Table 3
Regression results explaining financial performance, pooled OLS. Standard errors are reported below the coefficient estimates.

y	eps		$mktval$		rev	
	(1)	(2)	(3)	(4)	(5)	(6)
$const$	3.717	3.6004***	-1.77***	-3.7223***	-1.1609**	-4.086***
	0.0716	0.0996	0.5219	0.7462	0.5339	0.7962
$NGI_r \cdot rating$		-0.0196		0.1443		-0.6499
		0.0482		0.36		0.4037
$rating$	0.0484	0.0579**	0.7306***	0.6764***	0.6192***	1.0132***
	0.0106	0.03	0.0795	0.2313	0.0682	0.264
$manag\ sent_{NN}$	-0.0052	0.0122	0.0343	1.0382***	0.0227	1.0714***
	0.0072	0.0267	0.0515	0.1794	0.0569	0.1969
$NGI_{ms} \cdot manag\ sent_{NN}$		-0.0198		-1.2682***		-1.289***
		0.0348		0.2472		0.2629
$manag\ sent_{Goog}$	0.0039	0.0796	0.0037	0.948***	-0.0323	0.8788***
	0.0087	0.0519	0.0517	0.2871	0.0531	0.2574
$NGI_{ms} \cdot manag\ sent_{Goog}$		-0.0832		-1.0573***		-0.9966***
		0.0587		0.3168		0.2956
NGI_r	-0.0088	0.0071	-0.1046	-0.1749	-0.1173	0.3621
	0.0126	0.0353	0.0924	0.2711	0.0928	0.2925
NGI_{ms}	0.0132	0.0435**	0.161***	0.9503***	0.2022***	0.9611***
	0.0081	0.0177	0.0587	0.1322	0.0548	0.1244

Text based features also have a prominent and significant effect for market value and revenue, but not for earnings per share. The effect of management sentiment related features is also significant, which shows that management sentiment has additional weight on firm outcomes in addition to employee's general sentiment towards their jobs, which is captured by ratings. Interestingly, the NGI for management sentiment is significant, while the NGI for the overall review is not.

The significance of the interaction terms between management sentiment and their NGI shows that the relationship of management sentiment and financial performance appears to be non-linear and non-monotonic. While the effect of Neural Network management sentiment ($manag\ sent_{NN}$) on market value is significant and positive, the non-generic index weighted sentiment is significant and *negative*, for both market value and revenue. A 1% increase in $manag\ sent_{NN}$ increases market value by $(1.04 - 1.27NGI_{ms})$% and revenue by $(1.07 - 1.29NGI_{ms})$%. This means that:

- When $NGI_{ms} \leq 0.8$: an increase in $manag\ sent_{NN}$ would likely increase financial performance
- When $NGI_{ms} > 0.8$ an increase in $manag\ sent_{NN}$ would likely decrease financial performance.

The same holds for the Google management sentiment ($manag\ sent_{Goog}$). A 1% increase in $manag\ sent_{Goog}$ last period increases market value this period by $(0.95 - 1.06NGI_{ms})$% and revenue by $(0.88 - 1NGI_{ms})$%. The significance of Google management sentiment even with the presence of Neural Network management sentiment shows that capturing both domain specific and non-domain specific sentiment is salient for firm outcomes.

Overall, based on OLS, there is likely a positive relationship between last period's job ratings and this period's financial performance. There is likely to be a significant, but non-linear, relationship between last period's management sentiment and non-genericness of management sentiment, and this period's financial performance. The implication for firms is that making decisions that benefit their employees' level of satisfaction are likely to yield positive financial results, at least in the short term. Likewise, lower levels of employee satisfaction in a given quarter could indicate lower performance in the following quarter. The availability of this information would allow firms to make decisions that offset the potential for future losses, for example by cutting costs elsewhere.

5. Discussion

5.1. Contributions to literature

This paper bridges the gap between the strand of literature that measures employee sentiment with machine learning techniques and the literature that explains how employee sentiment might affect firm performance. In the first case, previous papers use methods such as tf-idf, LDA, STM to explain sentiment in job reviews. However, the connection is not made as to why sentiment is important for firm's real world outcomes. The focus is more centered around measuring and explaining employee sentiment itself. In the second case, the focus is on examining ratings data from job reviews (Ji et al., 2017) and their implication for firm performance, but they do not extract text features using machine learning methods. The additional information provided by the text itself is not utilized in their analysis.

The approach of this paper is to do both: it introduces new measures of sentiment analysis of job reviews using more recent supervised learning algorithms, which are shown to have greater accuracy in a range of domains (cite), as well as quantifying how much employee sentiment matters for firms' financial performance. This demonstrates the potential of machine learning techniques in providing new information for the quantification of real-world effects.

In addition, previous papers focused on review text as a whole, and did not consider sentiment towards management in particular. This paper demonstrates the importance of extracting more relevant portions of review text, particularly since ratings may already be representative of the overall review. Thus, it is important to consider what the additional value is of the text. Not all of review text is relevant for firm performance. For example, employees may like the proximity of their workplace to their house and include this in reviews.

Sentiment towards management however, is more insightful, as management controls much of the decision making within a company. Employees are directly observing management, which is not something external stakeholders (such as shareholders) can do. The limitation of this data is that the metrics that the employee uses to form a sentiment of management may be very different from what other interested parties, such as investors, would assess as good or bad management. However, averaged across many employees, the signal is likely to be more informative and thus beneficial for firms to track.

5.2. Practical implications

Firms may find useful insights from job reviews left by their current and former employees. There are strategic implications for decision making from such insights. Firms may find direct feedback for their management strategies that are difficult to gather outside of such a forum; employees may not be as open in their workplace regarding their true opinions. Reviews often are left anonymously, which adds a buffer of protection for the employee. Firms should also focus more specifically on relevant portions of the review, and particularly feedback that leaders at the company may be able to implement changes in. In addition, firms should examine how generic the reviews are in describing their management practices. More generic sentiment may not have as much impact as sentiment that is more specific and detailed.

6. Conclusion

These days, there are a growing number of online data sources on employee sentiment. In this paper, some of the valuable insights acquired from the use of such data is highlighted. When job review data is paired with machine learning techniques, many new avenues for quantification are opened. One such avenue is ascertaining employee sentiment towards management, which is found to have significant effects on a company's financial outcomes. Another avenue is in understanding the importance of how generic review text is in weighting its importance for sentiment. While these measures were previously unobserved characteristics of firms, they become observable through the application of machine learning methods. I find that there is a positive relationship between ratings and financial performance, and a significant but non-linear relationship between management sentiment and financial performance. These findings indicate that there may be trade-offs involved between creating a better workplace culture and short term financial performance. Such effects can have wide ranging implications in an economy greatly affected by high employee turnover and low retention.

This paper also introduces methods that could be applied to job review text in order to implement sentiment analysis. These methods include using Doc2Vec to vectorize review text, supervised algorithms such as XGBoost and Deep Neural Networks to train sentiment analysis models specifically for the domain of job reviews, and a tf-idf based measure of how generic review text is (NGI). These methods could have wide applicability beyond the scope of this domain, particularly when review text is paired with ratings. Another insight from this paper is also to focus analysis on portions of the review text that provides additional information beyond the rating, for example ascertaining sentiment specific to management rather than to the job as a whole. In the analysis, these additional variables were found to have significant effects on the financial outcomes of companies.

This paper has several shortcomings which may also extend into further work. Firstly, the entire distribution of reviews was not used in the analysis. Instead, only the mean sentiment across all reviews were used. This does not take advantage of the variety of sentiments that may exist; for example, using quantiles and standard deviations of sentiment measures may also prove insightful. It's plausible that in certain cases, some minority of reviews may provide more insight over the "average" across all reviews, particularly if these reviews may highlight more serious matters such as management misconduct. Secondly, there are more recent advances in natural language processing techniques, such as Bidirectional Encoder Representations from Transformers (BERT; see Devlin, Chang, Lee, & Toutanova (2018)), which may provide better methods of encoding review text. These methods improve in semantic recognition of context compared to embeddings such as Word2Vec, which may additionally boost the performance of supervised algorithms. Lastly, only companies that are publicly listed are included in this study since financial data is available for public companies within Compustat. However, this means that the estimation of the effect of sentiment on firm performance is biased towards larger companies. For smaller companies and private companies, there may be a different relationship between sentiment and performance.

Another avenue of extension is to examine each industry separately. It is highly likely that the effect of sentiment on firm performance varies greatly by industry. For example, food service employee sentiment is generally low, but may not have a large impact on firm performance which is mostly driven by consumer demand. As a counterpoint, the technology industry may have higher employee sentiment, and may also experience higher impact of sentiment on company performance, since this industry depends largely on highly specialized human capital. This may lead to insights in further areas of research such as the effect of human capital on company performance, which has broad implications for management decision making. Similarly, geographic differences may warrant separate investigation. It may be the case that the effect of sentiment differs geographically, particularly for urban vs. non-urban environments. In urban areas where job markets are saturated, there may be a greater influence of sentiment on performance, since workers have more opportunity to find new opportunities. This may not be true in areas where the job market is thin. Such a study may shed insight on business decision making, especially in terms of how location choices affect employee satisfaction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

1. Ratings: number of stars out of 5
2. Job title
3. Location of review
4. Date of review
5. Text of review

The blue boxes indicate terms that may be important in determining the rating given by the reviewer.

Table A1
Example of reviews for each rating.

Rating	Proportion	Example Review
1	6.24%	just a very bad job overall ... way too much for employees getting paid minimum wage and not getting a substantial amount of hours.'
2	8.07%	The building is a dungeon. The pay is not competitive. They thing if they give you free tv services that makes up for not paying better hourly.'
3	23.73%	I didn't like the work environment at Walmart at all. Some of the employees where rude and very unfriendly. It wasn't hard work if you could keep up. I didn't feel that I could go to the management with issues that I had. I felt like they would judge me and think less of me.
4	30.70%	Its an hustle and bustle kind of job,many duties and responsibilities most of the time its fun to meet and help people with their purchase.Help the customers find exactly what they are looking for,helping them get the best sale advantage,helping them with the on line pick up as fast as possible ans cashing them out as quickly as possible.'
5	21.20%	Wish I could have this job forever! This was such a great experience with great potential for advancement and learning. I appreciate every day what I gained from this company.'

Fig. A1. Example of review.

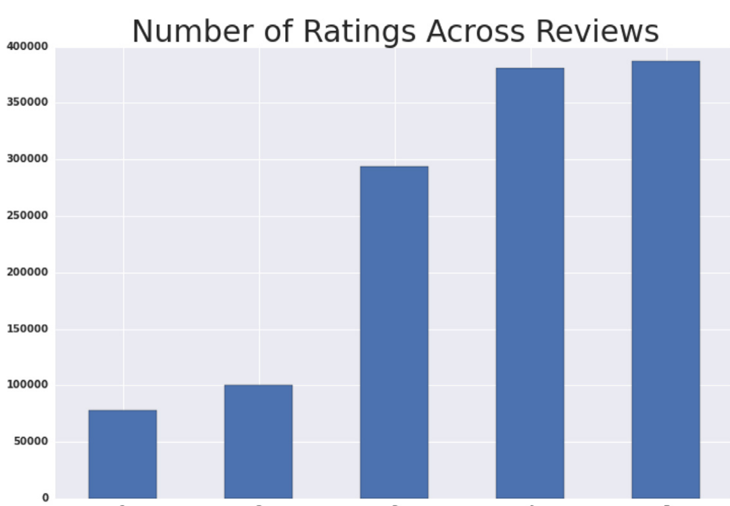

Fig. A2. Number of ratings across different reviews.

Table A2
Most common terms used, all reviews.

Term	Count
great	422,478
custom	405,251
day	348,203
compani	343,606
part	337,977
learn	298,540
good	294,929
time	290,837
peopl	290,505
enjoy	274,470
get	255,953
place	233,976
help	224,524
employe	216,242
would	213,462
part job	193,904
make	187,665
cowork	169,914
lot	163,810
hour	158,269

Table A3
Most common terms used, positive reviews.

Term	Count
custom	263,927
part	221,609
day	213,454
compani	211,669
learn	209,877
enjoy	194,285
good	183,258
peopl	177,892
time	166,793
help	157,438
place	151,836
get	136,694
part job	128,350
would	126,528
employe	113,478
cowork	111,674
make	111,577
team	111,059
alway	105,383
lot	102,723

Table A4
Most common terms used, negative reviews.

Term	Count
compani	57,310
get	52,532
employe	52,051
day	48,343
time	47,999
custom	45,446
peopl	43,750
hour	37,639
would	36,561
good	35,335
part	33,383
pay	32,074
place	32,047
like	31,512
make	30,199
one	26,857
dont	26,160
great	25,584
store	24,105
go	23,751

A1. Tf-idf

Term frequency (tf) of a term t in document d is simply given by the raw count of the number of times a word appears in a document. For the document "the cat in the hat," the word "cat" has a term frequency of 1. The inverse document frequency (idf) is a measure of how frequently a term appears in the entire vocabulary of the corpus. For example, if a second document was "the cat ate the rat," then then idf of the term "cat" would be:

$$\log \frac{\text{Num Documents}}{1+\text{Num Documents that term appears in}} = \log \frac{2}{1+2} = 0.10 \quad (A.1)$$

$tf \cdot idf = 2 \cdot 0.1 = 0.2$

So the tf-idf of the term "cat" in document d is:

The tf-idf grows with (i) the number of times a term appears in the document; (ii) the rarity of the term across all documents.

References

Alharbi, A. S. M., & de Doncker, E. (2019). Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cognitive Systems Research, 54*, 50–61.

Anderson, J. A. (1995). *An introduction to neural networks*. MIT press.

Bernhardt, K. L., Donthu, N., & Kennett, P. A. (2000). A longitudinal analysis of satisfaction and profitability. *Journal of Business Research, 47*(2), 161–171.

Bishop, C. M. (1994). Neural networks and their applications. *Review of Scientific Instruments, 65*(6), 1803–1832.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Bordoloi, S. K., & Matsuo, H. (2001). Human resource planning in knowledge-intensive operations: A model for learning with stochastic turnover. *European Journal of Operational Research, 130*(1), 169–189.

Cascio, W. F. (1991). *Costing human resources: The financial impact of behavior in organizations*. Thomson South-Western.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., et al., (2015). Xgboost: Extreme gradient boosting. *R package version 0.4-2, 1*(4), 1–4.

Corominas, A., Lusa, A., & Olivella, J. (2012). A detailed workforce planning model including non-linear dependence of capacity on the size of the staff and cash management. *European Journal of Operational Research, 216*(2), 445–458.

Dabirian, A., Kietzmann, J., & Diba, H. (2017). A great place to work!? understanding crowdsourced employer branding. *Business Horizons, 60*(2), 197–205.

De Bruecker, P., Van den Bergh, J., Beliën, J., & Demeulemeester, E. (2015). Workforce planning incorporating skills: State of the art. *European Journal of Operational Research, 243*(1), 1–16.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dube, S., & Zhu, C. (2021). The disciplinary effect of social media: Evidence from firms' responses to glassdoor reviews. *Journal of Accounting Research, 59*(5), 1783–1825.

Edmans, A. (2011). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics, 101*(3), 621–640.
Giannakas, F., Troussas, C., Krouska, A., Sgouropoulou, C., & Voyiatzis, I. (2021). Xgboost and deep neural network comparison: The case of teams' performance. In *International conference on intelligent tutoring systems* (pp. 343–349). Springer.
Goldberg, D., & Zaman, N. (2018). Text analytics for employee dissatisfaction in human resources management.
Goldberg, Y., & Levy, O. (2014). Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.
Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*(2), 268.
Hopkins, J. C., & Figaro, K. A. (2021). The great resignation: An argument for hybrid leadership. *International Journal of Business and Management Research, 9*(4), 393–400.
Huang, M., Li, P., Meschke, F., & Guthrie, J. P. (2015). Family firms, employee satisfaction, and corporate performance. *Journal of Corporate Finance, 34*, 108–127.
Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal, 38*(3), 635–672.
Iaffaldano, M. T., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin, 97*(2), 251.
Ji, Y., Rozenbaum, O., & Welch, K. (2017). Corporate culture and financial reporting risk: looking through the glassdoor. *Available at SSRN 2945745*.
Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127*(3), 376.
Jung, Y., & Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decision Support Systems, 123*, 113074.
Kunc, M., & O'brien, F. A. (2019). The role of business analytics in supporting strategy processes: Opportunities and limitations. *Journal of the Operational Research Society, 70*(6), 974–985.
Lee, J., & Kang, J. (2017). A study on job satisfaction factors in retention and turnover groups using dominance analysis and lda topic modeling with employee reviews on glassdoor. com.
Luo, N., Zhou, Y., & Shon, J. (2016). Employee satisfaction and corporate performance: Mining employee reviews on glassdoor. com.
McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68.
Meduri, Y., & Jindal, P. (2021). Manager'S role in employee turnover intentions: A special study during covid-19. *IUP Journal of Organizational Behavior, 20*(4).
Memon, N., Patel, S. B., & Patel, D. P. (2019). Comparative analysis of artificial neural network and xgboost algorithm for polsar image classification. In *International conference on pattern recognition and machine intelligence* (pp. 452–460). Springer.
Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.
Mishra, R. K., Urolagin, S., & Jothi, J. A. A. (2020). Sentiment analysis for poi recommender systems. In *2020 seventh international conference on information technology trends (itt)* (pp. 174–179). IEEE.
Mishra, R. K., Urolagin, S., et al., (2019). A sentiment analysis-based hotel recommendation using tf-idf approach. In *2019 international conference on computational intelligence and knowledge economy (iccike)* (pp. 811–815). IEEE.
Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of indian farmers' protest using twitter data. *International Journal of Information Management Data Insights, 1*(2), 100019.
Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the seventh international conference on language resources and evaluation (lrec'10)*.
Paliwal, S., Mishra, A. K., Mishra, R. K., Nawaz, N., & Senthilkumar, M.. Xgbrs framework integrated with word2vec sentiment analysis for augmented drug recommendation.
Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.
Popadak, J. (2013). A corporate culture channel: How increased shareholder governance reduces firm value. *Available at SSRN, 2345384*.
Ramos, J., et al., (2003). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning: vol. 242* (pp. 29–48). New Jersey, USA.
Roberts, M. E., Stewart, B. M., Tingley, D., Airoldi, E. M., et al., (2013). The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation: vol. 4* (pp. 1–20). Harrahs and Harveys, Lake Tahoe.
Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082.
Robertson, M. B. (2021). Hindsight is 2020: Identifying missed leadership opportunities to reduce employee turnover intention amid the covid-19 shutdown. *Strategic HR Review*.
Sainju, B., Hartwell, C., & Edwards, J. (2021). Job satisfaction and employee turnover determinants in fortune 50 companies: Insights from employee reviews from indeed. com. *Decision Support Systems, 148*, 113582.
Schein, E. H. (1990). *Organizational culture*: vol. 45. American Psychological Association.
Severyn, A., & Moschitti, A. (2015). Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval* (pp. 959–962).
Sheather, J., & Slattery, D. (2021). The great resignation—how do we support and retain staff already stretched to their limit? *BMJ (Clinical Research ed.), 375*.
Sheng, J., Wang, X., & Amankwah-Amoah, J. (2021). The value of firm engagement: How do ratings benefit from managerial responses? *Decision Support Systems, 147*, 113578.
Srivastava, A., Singh, V., & Drall, G. S. (2019). Sentiment analysis of twitter data: A hybrid approach. *International Journal of Healthcare Information Systems and Informatics (IJHISI), 14*(2), 1–16.
Sull, D., Sull, C., & Zweig, B. (2022). Toxic culture is driving the great resignation. *MIT Sloan Management Review, 63*(2), 1–9.
Symitsi, E., Stamolampros, P., & Daskalakis, G. (2018). Employees' Online reviews and equity prices. *Economics Letters, 162*, 53–55.
Symitsi, E., Stamolampros, P., Daskalakis, G., & Korfiatis, N. (2021). The informational value of employee online reviews. *European Journal of Operational Research, 288*(2), 605–619.
Van Dam, A. (2021). The latest twist in the'great resignation': Retiring but delaying social security. *The Washington post*. NA–NA
Van Der Smagt, P. P., Kröse, B. J., Krose, B. J., & Smagt, P. P. (1993). An introduction to neural networks.
Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research, 69*(5), 1562–1566.