

# Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor.com

*Completed Research Paper*

**Ning Luo**

Gabelli School of Business  
Fordham University, New York, USA  
nluo1@fordham.edu

**Yilu Zhou**

Gabelli School of Business  
Fordham University, New York, USA  
yzhou62@fordham.edu

**John J. Shon**

Gabelli School of Business  
Fordham University, New York, USA  
jshon@fordham.edu

## Abstract

*In recent years, Big Data has created significant opportunities for academic research in a wide range of topics within the social sciences. We contribute to this growing field by exploiting the unique social media data from Glassdoor.com. We extract anonymous employee reviews for textual analysis to reveal the relation between employee satisfaction and company performance. Using categories from corporate value studies, our analysis not only provide a “bird’s eye view,” but also provide specific aspects of employee satisfaction are responsible for driving these correlations. We found that while Innovation is the most important category for technology industry, Quality category drives retailing and financial industry. We confirmed the significant correlation between overall employee satisfaction and corporate performance and discovered categories that are negatively correlated with performance: Safety, Communication and Integrity. We hope that this research encourages other researchers to consider the rich environthat a text analytics methodology makes possible.*

**Keywords:** Social media, text mining, employee satisfaction, Tobin's Q, company value

## Introduction

In recent years, Big Data has created significant opportunities for academic research in a wide range of topics within the social sciences—economics, finance, and political science among others. For instance, data from social media has been used to fine-tune movie box office predictions (e.g., Zhou and Duan 2015), to provide new, insightful analyses for consumer preferences (e.g., Asur and Huberman 2010, Lassen et al. 2014) and voter preferences in political races (e.g., Bermingham and Smeaton 2011, Bae et al. 2013, Jahanbakhsh and Moon 2014). Other studies examine the relation between employee attitudes/satisfaction with the financial performance of the companies they work for. For instance, Edmans (2011) finds that firms that voluntarily apply (and get accepted to) an annual survey called the “100 Best Companies to Work for in America” have higher long-run stock returns. Guiso et al. (2015) finds that an increase in perceived managerial integrity is associated with increases in a firm’s Tobin’s Q (similarly, Koys, 2001; Harter et al., 2002).

In this study, we contribute to this growing field by exploiting the unique social media data from Glassdoor.com to extract anonymous employee reviews for textual analysis. In doing so, ours is the first

study that uses *text mining* to study the relation between employee satisfaction and company performance. Extant studies that examine the relation between employee satisfaction and corporate financial performance (e.g., Edmans, 2011) find that employee satisfaction is positively correlated with equity prices—however, such studies do not delve into granular data about the specific machinations of employee satisfaction; or, such studies suffer from non-trivial selection bias, thus making inferences and generalizability limited. Our study thus showcases the types of analyses that are possible using text mining, and hope that it encourages other researchers to consider the rich environ that a text analytics methodology makes possible. Big Data methodologies can thus be exploited to help researchers and practitioners better understand the relationship between firm performance and unconventional metrics that may drive firm valuation—like an individual employee’s comments about job satisfaction.

We contribute to the extant literature by examining a phenomenon that has previously been examined from a “bird’s eye view,” but has not, as of yet, been analyzed using specific, granular data. Specifically, while higher levels of employee satisfaction are documented to be positively correlated with superior business performance, it remains unclear which specific aspects of employee satisfaction are responsible for driving these correlations. By analyzing company websites, Guiso et al. (2015) find that 80% of S&P 500 companies state that *innovation* is their primary company value; and 70% state that *integrity* and *respect* are core values. However, it is unclear whether such values are simply buzzwords in a public-relations/marketing campaign; or whether such values are embraced by employees. Similarly, Edmans (2011) examines the stock market performance of those companies that are selected to be on the list of “100 Best Companies to Work for in America”; however the cross-sectional variation of these 100 firms remains unexamined—does the company ranked #1 differ significantly from the one that is ranked #100? And if so, *why* are there differences? Such questions have not been examined yet—and indeed, cannot be examined at such granular a level until text mining is performed.

Using our text mining methodology, our study also sheds light on the significant improvements in research design that are possible, relative to prior research methods. Specifically, extant studies often rely on data obtained from survey instruments. While a well-designed survey instrument is certainly valid, such methodologies suffer from potential low response rates, as well as the prohibitive costs of administering and tabulating such surveys on a large-scale, cross-industry basis. Moreover, individual surveys are rigid and closed end, precluding researchers from performing preliminary exploratory analysis, as well as ex post, back-end analysis (unless another survey is administered). Similarly, there exist issues of generalizability with current studies, as the survey cannot be administered on a large scale basis, thus making inferences limited to the industry(s) that is specifically examined. It is in this vein that our approach thus opens up several new potential avenues of research.

## **Literature Review**

### ***Employee Satisfaction and Corporate Performance***

Employee is a valuable company asset. Many studies found that companies with higher employee satisfaction or emphasis more on human resources tend to perform better (Denison, 1990; Ostroff 1992; Harter et al., 2002; Koys, 2001). Shneider et al. (2003) used 35 companies’ survey data over 8 years from a consortium of U.S. corporations to measure employee satisfaction and suggested that overall employee satisfactions and satisfaction with security, pay and work group have a positive impact on return on assets (ROA) and earnings per share (EPS). A study by Edmans (2011) used “100 Best Companies to Work For in America” (1984 to 2009) survey data as a measurement of employee satisfaction and found these listed companies have higher long run stock returns in the overall stock market. Guiso et al. (2014) used the Great Place to Work® Institute (GPTWI) survey data and found that one standard deviation increase in integrity is associated with a 0.19 standard deviation increase in Tobin’s Q. A meta-analysis provided by Rubera and Kirca (2012) suggested that innovation has direct positive effects on firm value. O’Reilly et al. (2014) used the survey data from respondents in 32 high-technology companies and showed that CEO personality affects a firm’s culture and that culture can effect firm’s financial performance (revenue growth and Tobin’s Q). Popadak (2013) conducted a study on employee reviews collected through survey by career intelligence firms and concluded that the corporate culture is positively associated with long-term firm value. However, the data source is possibly biased with the average user of 43 years old and an annual income of \$106,000.

Apparently, existing employee satisfaction measurement relies heavily on survey, and publicly available datasets such as GPTWI and “Best Companies” rank. There are limitations with these approaches. They suffer from non-trivial selection bias, thus making inferences and generalizability limited. More importantly, they do not delve into granular data about the specific machinations of employee satisfaction. A possible direction to alleviate these limitations is to use social media data.

### **Social Media and Glassdoor.com Data**

Social media has produced large impacts on multiple areas. Data on consumers’ opinions and sentiments such as online reviews, tweets and blogs can be used to predict sales (Asur and Huberman 2010; Lassen et al. 2014) and stock price change (Bollen et al. 2011; Oh and Sheng 2011). Twitter has been frequently used as a data source to measure public opinions in order to predict the results of political elections (Birmingham and Smeaton 2011; Bae et al. 2013, Jahanbakhsh and Moon 2014). Social media data can also be used to generalize pros and cons of products that can both help individual customer to make better purchase decision and manufactures to monitor their reputations (Kim and Hovy 2006; Scaffidi et al. 2007; Abrahams et al. 2012). The abundant of social media data made the data gathering easier for employee satisfaction and corporate value study. Among them is the Glassdoor website. *Glassdoor.com* was founded by Robert Hohman, Rich Barton and Tim Besse in 2007 (Glassdoor: About Us 2016). It is a website where employees and former employees anonymously review companies and their management. Glassdoor covers a diverse user population where users’ profiles are fairly distributed across different sectors such as age, income, education (Moniz’s, 2015).

There are limited studies using Glassdoor as their data source. Huang et al. (2015) used 102,888 Glassdoor employee ratings from 993 publicly-traded U.S. companies during the 2008-2012 period. Their analysis focus on six rating dimensions: overall satisfaction, career development, compensation and benefits, work/life balance, senior leadership and CEO approval. They found a statistically significant causal link between employee overall satisfaction and the market value of companies, measured by Tobin’s Q. Although this study relied on Glassdoor as their data source, it neglected the rich information embedded in employees’ textual reviews. The rating is limited in both scope and scale. Moniz (2015) is the only study we are aware of that performed textual analysis on employee reviews from Glassdoor.com. They used topic modeling and extracted keywords with the meaning of “goal-setting” such as “planning”, “goals”, “incentives” and “direction”. The author claims that goals motivate high performance and found that “GOAL” is positively associated with firm value measured in Tobin’s Q. However, we are not aware of studies that use text reviews to examine employee satisfaction on varies corporate value dimensions.

### **Corporate Advertised Value Categories**

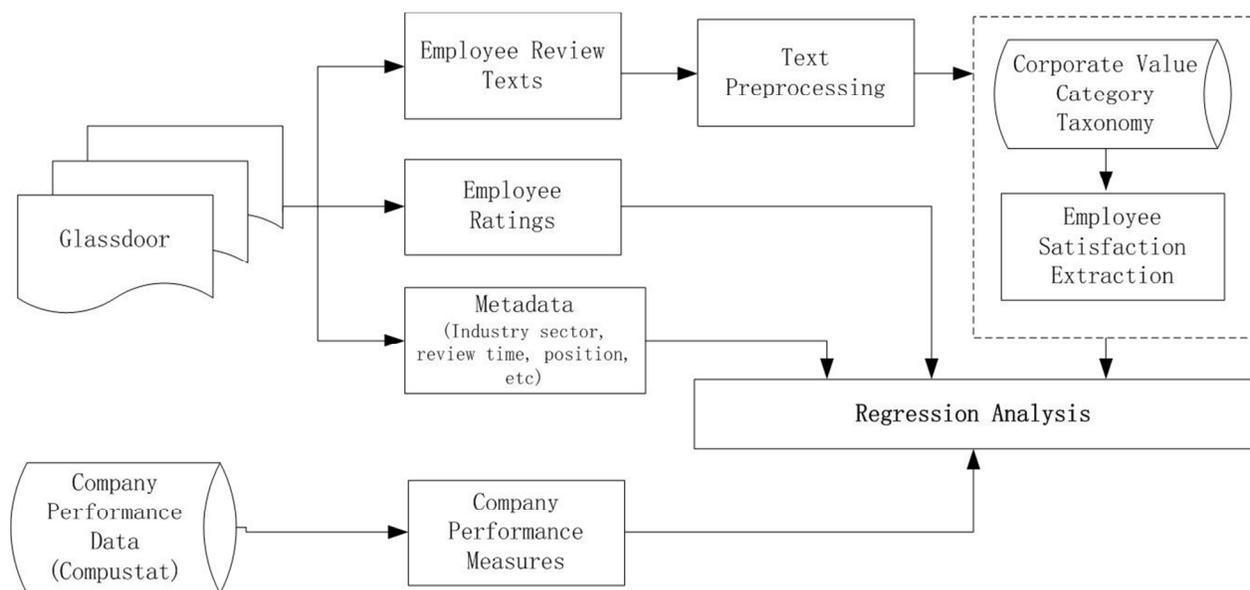
We believe that ratings alone from Glassdoor is too generic and fail to reflect the granularity of employee satisfaction. Thus, we are interested in finding employee satisfaction dimensions from textual reviews. Corporate value categories provide one potential categorization of employee satisfaction dimensions and have been studied in a number of research. Among them Guiso et al. (2015)’s category is considered comprehensive and objective. They collected all the advertised values from S&P 500 companies’ websites and found the nine most recurring categories, *Integrity, Teamwork, Innovation, Respect, Quality, Safety, Community, Communication and Reward*.

Guiso et al. (2014) studied each of the nine categories and how they affect company performance. *Integrity* is found to be positively associated with company performance. Other categories proposed in this study have also been researched. For example, many studies (Büschgens al et, 2013, Rubera and Kirca 2012, Jimenez-Jimenez 2011) showed that *innovation* has direct positive effect on firm value. Another category, *teamwork* cohesion was also found to improve organizational performance in a study by Montes et al. (2005). Past studies on work/life balance cross multiple fields such as economics (e.g., Johnson and Provan, 1996; Whitehouse and Zetlin 1999), information systems (e.g., Baines and Delder 2003; Frolick et al. 1993), and management (e.g., Konrad and Managel 2000; Perry-Smith and Blum 2000). They all confirmed that work/life balance practices enhanced organizational effectiveness. Studies on company *community and environment* (Zahra 1993; Daft et al. 2011) suggested that working environment plays an important role in improving financial performance. *Quality* is studied in marketing fields extensively.

They revealed the positive impact of “customer” on company financial performance (Kamakura et al. 2002; Reinartz and Kumar 2000; Reinartz et al. 2010). The impact of Reward (Kerr and Slocum 2005) Communication (Barrett 2002), Respect (Fuller et al. 2009) were also studied in management, psychology and sociology fields. These nine categories used in Guiso et al. (2014) are representatives of important dimensions in employee satisfaction. Thus, in this research, we will use them as our indicator of employee perceived corporate culture and employee’s satisfaction.

## Research Methodology

The emergence of the social media data and availability of advanced data collection and analytical techniques provide new directions in social science research. The capacity of collecting and processing massive amount of data is catalyzing a paradigm shift to “Computational Social Science”. There are also some computational text analysis researches utilizing text mining, machine learning and NLP techniques to study text data such as archives, press release and speech that addresses research questions in social science area (O’Connor et al. 2011). Text mining refers to the process of discovering knowledge from unstructured text using both Natural Language Processing (NLP) and data mining techniques and has been widely used textual data collected from Web 2.0 applications (Chen et al. 2012). It has the capacity to turn large bulk of unstructured noisy data into structured forms such as records in database from which valuable information can then be discovered using data mining or statistical analysis (Tan 1999). It has the potential to provide detailed information on specific aspects of employee satisfaction. Thus, our research methodology combines techniques from text mining, descriptive data analysis and regression analysis. As shown in Figure 1, it contains several major components: data collection from Glassdoor.com, text pre-processing, employee satisfaction extraction, corporate performance measurement and regression analysis.



**Figure 1. Research Model**

### ***Data Collection from Glassdoor.com***

Glassdoor contains 475,000+ companies with more than 8 million company reviews, company ratings, CEO approval ratings, and salary reports from worldwide locations. The employment positions held by Glassdoor reviewers vary as much as the overall economy; from software engineers in the Technology industry to investment banking analysts in the Financial Services industry. For each individual review, Glassdoor.com provides several metrics to gauge employee satisfaction. Most salient among the metrics is an Overall rating, which is a number ranging from 1 to 5. Each employee is also given the option to rate their employer along five additional criteria: (i) Culture and Values, (ii) Work/Life Balance, (iii) Senior

Management, (iv) Compensation and Benefits, and (v) Career Opportunities. Lastly, there is an open-form section of the review that allows the reviewer to state in his/her own words any other comments about the employer or the job that he/she would like to include. Thus, we are able to analyze employee satisfaction across these five specific ordinal metrics, as well as examining the text of open-ended comments.

We develop a Web crawling algorithm to crawl the review data from Glassdoor.com. We examine the reviews for all Fortune 500 firms (2015) and crawl a total of 274,061 reviews from 2008 to 2014. Companies who have less than 10 reviews are then deleted from the list. Our final dataset is comprised of 257,454 reviews, garnered from 425 companies, representing 21 industry sectors. In Table 1, we provide descriptive statistics on the number of reviews, by industry. (Companies that are not classified by Fortune 500 were included in “Other”.) Technology, Retailing, Financial Services, Telecommunications and Health care are the five industries with the most reviews, respectively.

**Table 1. Glassdoor Dataset Summary Statistics**

| Industry                      | Number of firms | Number of reviews | Mean    | Median | Maximum | Minimum |
|-------------------------------|-----------------|-------------------|---------|--------|---------|---------|
| Aerospace & Defense           | 10              | 7,406             | 740.6   | 353    | 1,923   | 55      |
| Apparel                       | 4               | 1,096             | 274.0   | 216.5  | 626     | 37      |
| Business Services             | 13              | 4,141             | 318.5   | 187    | 1,500   | 39      |
| Chemicals                     | 14              | 3,203             | 228.8   | 148.5  | 681     | 11      |
| Engineering & Construction    | 6               | 2,222             | 370.3   | 399    | 683     | 12      |
| Energy                        | 56              | 5,218             | 93.2    | 49     | 555     | 11      |
| Food, Beverages & Tobacco     | 23              | 4,753             | 206.7   | 128    | 1,500   | 13      |
| Food & Drug Stores            | 7               | 9,521             | 1,360.1 | 1,120  | 2,844   | 398     |
| Financials                    | 45              | 37,654            | 836.8   | 488    | 5,498   | 10      |
| Health Care                   | 37              | 13,321            | 360.0   | 262    | 1,624   | 22      |
| Household Products            | 13              | 4,002             | 307.8   | 201    | 1,762   | 31      |
| Hotels, Restaurants & Leisure | 9               | 7,813             | 868.1   | 176    | 3,648   | 12      |
| Industrials                   | 16              | 4,437             | 277.3   | 125.5  | 1,420   | 13      |
| Material                      | 15              | 918               | 61.2    | 42     | 160     | 11      |
| Media                         | 7               | 1,399             | 199.9   | 125    | 554     | 35      |
| Motor Vehicles & Parts        | 14              | 2,948             | 210.6   | 61     | 772     | 28      |
| Retailing                     | 38              | 50,329            | 1,324.4 | 646.5  | 6,870   | 10      |
| Technology                    | 39              | 70,383            | 1,804.7 | 948    | 10,000  | 27      |
| Telecommunications            | 16              | 16,445            | 1,027.8 | 424    | 4,777   | 24      |
| Transportation                | 17              | 6,986             | 410.9   | 252    | 1,980   | 23      |
| Wholesalers                   | 22              | 2,896             | 131.6   | 75     | 462     | 24      |
| Other                         | 4               | 363               | 90.8    | 56     | 1,675   | 7       |
| Total                         | 425             | 257,454           |         |        |         |         |

### ***Text Processing and Satisfaction Category Extraction***

Once the text from each individual employee is extracted and processed, we must consider (or construct) a systematic methodology for categorizing the text. Because the process of categorization can be a prohibitive cost unto itself, we rely on prior extant literature. Guiso et al. (2015) propose a framework of nine categories of *corporate advertised value*: (i) Integrity, (ii) Teamwork, (iii) Innovation, (iv) Respect, (v) Quality, (vi) Safety, (vii) Community, (viii) Communication, and (ix) Hard work. The authors also provide several relevant keywords for each category. The nine categories thus provide a framework for categorizing aspects of corporate values and employee satisfaction that is derived from such values. In Table 2, we provide the nine categories and their relevant keywords.

Armed with this framework, we perform standard text processing for data cleaning, noise and stop-word removal and stemming, to extract information relevant to these nine categories. Using a bag-of-words approach, we index all reviews and extract term frequencies for each category and its respective relevant keywords.<sup>1</sup> Once these categories are extracted, sentiment analysis will be performed to assign the polarity of each review.

**Table 2. Nine Categories of Advertised Corporate Value (Guiso et al. 2015)**

| <i>Advertised Corporate Value</i> | <i>Keywords</i>  |
|-----------------------------------|--|
| <i>Integrity</i>                  | <i>Integrity + Ethics + Accountability + Trust + Honesty + Responsibility + Fairness + Transparency + Ownership</i>      |
| <i>Teamwork</i>                   | <i>Teamwork + Collaboration/Cooperation</i>  |
| <i>Innovation</i>                 | <i>Innovation + Creativity + Excellence + Improvement + Passion + Pride + Leadership + Growth + Efficiency + Results</i> |
| <i>Respect</i>                    | <i>Respect + Diversity + Inclusion + Development + Talent + Dignity + Empowerment</i>                                    |
| <i>Quality</i>                    | <i>Quality + Customer + Needs + Commitment + Dedication + Value</i>  |
| <i>Safety</i>                     | <i>Safety + Health + Work/Life balance + Flexibility</i>   |
| <i>Community</i>                  | <i>Community + Environment + Caring + Citizenship</i>  |
| <i>Communication</i>              | <i>Communication + Openness</i>  |
| <i>Reward</i>                     | <i>Hard work + Reward + Fun + Energy</i>   |

### **Corporate Performance Measure**

For each company in our sample, we collect financial data (using the firm's unique CUSIP number) from the CRSP/Compustat merged database, a standard financial database in the Accounting and Finance literatures. We use Tobin's Q as our main financial measure (Tobin, 1969). Though several other metrics are available for analysis—indeed, a review of the Accounting and Finance literatures suggest what seems an endless number of potential metrics—we choose Tobin's Q because it reflects the “value added” of intangible factors, ranging from intellectual property like R&D expenditures which are pervasive, to the benefits garnered from superior management and corporate governance (e.g., Hermalin and Weisbach 2001). Because our sample period is 2008-2014, we choose to examine Tobin's Q because of the pervasive and prevalent role that intangibles play in our modern-day economy.<sup>2</sup> Tobin's Q is defined as:

$$\text{Tobin's Q} = \text{Total market value} / \text{Total asset value} \quad (1)$$

In later tests, we use Tobin's Q as our dependent variable in OLS regression tests.

<sup>1</sup> Some keywords were modified or dropped due to their ambiguity in meaning and/or extremely high document frequency.

<sup>2</sup> Nonetheless, in untabulated robustness tests, we also consider several other financial metrics, including ROA (return on assets) and abnormal stock returns,

## Exploratory Data Analysis

### Ranking of Employee Satisfaction Categories by Frequency

From our dataset of 257,454 employee reviews (which include open-form textual comments), we extract relevant words for each of the nine Guiso et al. (2015) categories (see Table 2), and perform word counts to ascertain the frequency word counts. For the overall sample, we find *Quality* is the most mentioned word category among employees, followed by *Innovation*, and then *Respect*. In Table 3, we present, by industry, the three most common categories of words employees mention in their Glassdoor reviews.

The *Quality* category contains many customer-service oriented words, such as *customer*, *quality*, *needs*, and *expectations*. Perhaps not surprisingly, *Quality* is most frequently mentioned for employees of firms in service industries, including Hotels, Restaurants & Leisure, Telecommunications, and Transportation. However, it is also the most frequently mentioned for Food & Drug Stores, Retailing, and Wholesalers, suggesting that employee satisfaction is driven by a pride in the quality of not only the services provided to customers, but also the quality of goods sold, as well.

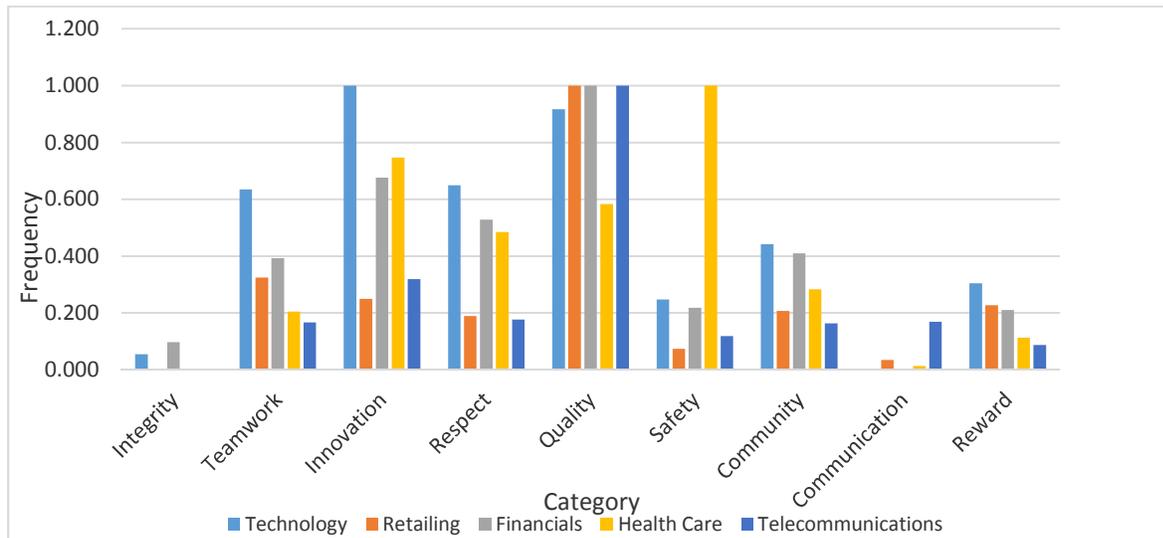
The *Innovation* category consists of words such as *innovation*, *creativity*, *growth* and *improvement*. Not surprisingly, *Innovation* is most frequently mentioned among employees of firms in technology-oriented industries, including Aerospace & Defense, Chemicals, Technology, Engineering & Construction, Industrials, and Motor Vehicles & Parts. Lastly, the *Respect* category is composed of words such as *development*, *dignity*, *empowerment* and *diversity of employees*. *Respect* is most frequently mentioned among employees of firms in the Energy and Material industry. It also has and has a high frequency of mentions as the second and third most referred category in several industries (e.g., Aerospace & Defense, Media, Food & Beverage). Other findings for frequency counts of employee satisfaction are also informative. For instance, the top category for the Health Care industry is *Safety*, suggesting that the safety of the products and services provided within the health care industry rank highest among its employees, as human lives are at stake. In the Retailing industry, the second most common category is Teamwork, conjuring up images of teamwork on the sales floor among sales staff.

**Table 3. Analysis of Employee Satisfaction Focus by Industry**

| Industry                      | Top 1 Category | Top 2 Category | Top 3 Category |
|-------------------------------|----------------|----------------|----------------|
| Aerospace & Defense           | Innovation     | Respect        | Quality        |
| Apparel                       | Innovation     | Quality        | Community      |
| Business Services             | Innovation     | Quality        | Respect        |
| Chemicals                     | Innovation     | Quality        | Respect        |
| Engineering & Construction    | Innovation     | Respect        | Quality        |
| Energy                        | Reward         | Innovation     | Quality        |
| Food, Beverages & Tobacco     | Innovation     | Respect        | Quality        |
| Food & Drug Stores            | Quality        | Innovation     | Teamwork       |
| Financials                    | Quality        | Innovation     | Respect        |
| Health Care                   | Safety         | Innovation     | Quality        |
| Household Products            | Innovation     | Respect        | Quality        |
| Hotels, Restaurants & Leisure | Quality        | Reward         | Innovation     |
| Industrials                   | Innovation     | Quality        | Respect        |
| Material                      | Innovation     | Respect        | Quality        |
| Media                         | Innovation     | Respect        | Reward         |
| Motor Vehicles & Parts        | Innovation     | Quality        | Respect        |
| Retailing                     | Quality        | Teamwork       | Innovation     |
| Technology                    | Innovation     | Quality        | Respect        |
| Telecommunications            | Quality        | Innovation     | Respect        |
| Transportation                | Quality        | Innovation     | Respect        |
| Wholesalers                   | Quality        | Safety         | Innovation     |
| Other                         | Quality        | Respect        | Innovation     |
| Total                         | Quality        | Innovation     | Respect        |

### Top Five Industries: Comment Excerpts

In this subsection, we consider the Technology, Retailing, Financials, Telecommunications and Health Care industries individually because these are the industries that received the most reviews among our 21 industry sectors. The nine category word frequencies are normalized accordingly for cross-sectional comparisons. For each industry sector, we divided each category weight by the maximum weight of all nine categories. We present our results of this analysis in Figure 2.



**Figure 2. Normalized Frequencies of Employee Satisfaction Dimensions for Top Five Industries**

Overall, *Teamwork*, *Innovation*, *Respect*, *Quality*, *Safety*, *Community*, *Reward* are the most mentioned categories in these five industries—though there is of course cross-sectional variation. For example, while *Safety* appears to be most important for the Health Care industry, *Integrity* was hardly mentioned at all. Meanwhile, in the Technology industry, *Innovation* has the highest word frequency.

To give the reader further context into the types of comments that we extract and categorize, we present several representative excerpts below. For instance, an employee from Advanced Micro Devices comments:

*“AMD is constantly looking for new ways to compete with Intel. It is this rivalry that pushes innovation and product quality, pushing everyone to think creatively about how to do more with much less resources than Intel.”*

Another review from employee at Agilent Technologies comments:

*“I have been working at Agilent Technologies full-time for more than 10 years; Great people, innovation at the core.”*

A review by a Microsoft employee states:

*“There is very little innovation and creativity to be found, Microsoft culture rewards competition against coworkers, conformity and process.”*

In the Financial industry, employees frequently mention *Quality*. For instance, one employee comments:

*“Bank of America as a company is out of touch with what customer's value, and their most important priority has become the bottom line.”*

Another comments:

*“So much energy is spent trying to comply with over-demanding leadership that sadly the customers' satisfaction and other associates' growth is lost.”*

From these simple examples, we illustrate how future researchers can exploit our research method to enable deeper analysis of the text, as well as a broader scope of search terms that extend beyond the nine categories we examine. Indeed, even our current dataset provides rich content for exploratory analysis that goes well beyond the simple five-star ratings that Glassdoor explicitly requires. In this manner, we are able to deepen the analysis of employee satisfaction—and its association with corporate financial variables, such as financial valuation and performance.

### **Industry Radar Chart Analysis**

To compare industry differences, we further performed radar chart analysis of three major industries, Retailing, Technology and Financials in Figure 3, 4 and 5 respectively. Each chart shows the distribution the nine categories within an industry and give us an idea of the industry employee satisfaction focus.

Retailing industry is heavily concentrated on the category of Quality. This reflects the fact that they deal with customers. In contrast, there are much less mentioning of Integrity, Communication and Safety. Teamwork, Innovation, Reward and Community is somewhat important in retailing industry. In comparison, technology industry is more balanced between Teamwork, Innovation, Respect and Quality. Among them, Innovation is the more important category. We believed our results captured the common trait in Technology industry. Innovation remained to be the most valued culture among IT companies. Google (HRZONE 2015) has its best well know innovation mechanism “20% time” to empower its employees to be more creative and innovative. Apple’s CEO, Tim Cook (Gadgets360 2016) listed innovation as Apple’s biggest strength. Microsoft (Microsoft Innovation Centers 2016) has its innovation centers to support new ideas. Financial industry is also heavily concentrated on Quality category. Similar to technology industry, it is more balanced between Teamwork, Innovation, Respect and Quality. The focus on Quality is much higher than Innovation in financial industry, while Quality is lower than Innovation in technology industry.



**Figure 3. Rader Chart Analysis of Retailing Industry**



Figure 4. Rader Chart Analysis of Technology Industry



Figure 5. Rader Chart Analysis of Financials Industry

## Regression Analysis

In this section, we estimate a standard OLS regression model to determine the potential impact that employee satisfaction metrics have on a firm's Tobin's Q. We estimate the following model, which controls for both fixed firm effects and fixed quarter effects:

$$\begin{aligned} \text{TQ} = & \beta_0 + \beta_1 \text{OVERALL} + \beta_2 \text{INTEGRITY} + \beta_3 \text{TEAMWORK} \\ & + \beta_4 \text{INNOVATION} + \beta_5 \text{RESPECT} + \beta_6 \text{QUALITY} + \beta_7 \text{SAFETY} \\ & + \beta_8 \text{COMMUNITY} + \beta_9 \text{COMMUNIC} + \beta_{10} \text{REWARD} + \beta_{11} \text{LOGMV} + \varepsilon \end{aligned}$$

where TQ = Tobin's Q for quarter  $t$ , defined as: Total market value / Total asset value,

OVERALL = average quarterly overall employee satisfaction score, ranging from 1 to 5,

INTEGRITY = aggregate normalized frequency of *Integrity*-related keywords during quarter  $t$ ,

TEAMWORK = aggregate normalized frequency of *Teamwork*-related keywords during quarter  $t$ ,

INNOVATION = aggregate normalized frequency of *Innovation*-related keywords during quarter  $t$ ,

RESPECT = aggregate normalized frequency of *Respect*-related keywords during quarter  $t$ ,

QUALITY = aggregate normalized frequency of *Quality*-related keywords during quarter  $t$ ,

SAFETY = aggregate normalized frequency of *Safety*-related keywords during quarter  $t$ ,

COMMUNITY = aggregate normalized frequency of *Community*-related keywords during quarter  $t$ ,

COMMUNIC = aggregate normalized frequency of *Communication*-related keywords during quarter  $t$ ,

REWARD = aggregate normalized frequency of *Reward*-related keywords during quarter  $t$ , and

LOGMV = the natural log of firm size (market value).

Thus, the model examines the potential impact that each of the nine categories of employee satisfaction word frequencies have on a firm's valuation and performance, measured by Tobin's Q. Importantly, this model is estimated while controlling for the overall employee satisfaction measure explicitly provided by Glassdoor (OVERALL). We also control for firm size (LOGMV) because it is a standard control in the finance and accounting literature.

In Table 4, we present our main results. We present two preliminary models and one full model. Model 1 shows the relation between TQ and OVERALL, the average overall rating. The estimated coefficient for OVERALL is positive and statistically significant (0.1717, t-statistic=12.27), suggesting that the overall employee satisfaction rating is positively correlated with Tobin's Q, our measure of firm performance. In Model 2, we present the initial results of our nine-category word frequencies. (Results are qualitatively similar to the full results in Model 3, so we defer discussing economic implications till our examination of Model 3.) We find that three of the employee categories for satisfaction— INNOVATION, RESPECT, and TEAMWORK—have a positive association with firm performance. We also find four categories—SAFETY, REWARD, COMMUNIC, and INTEGRITY—are negatively associated with firm performance.

In our main model, Model 3, we present results that include our nine categories, but controlling for employees' OVERALL rating, as well as LOGMV, firm size. First, consistent with Model 1, we find that the estimated coefficient for OVERALL is positive and statistically significant (0.0521, t-statistic=3.87), suggesting that this overall rating—despite its lack of detailed, granular aspects of employee satisfaction—is related to firm performance/value. More interestingly, we next examine our nine categorical word frequencies, which we garnered from text mining. We find that QUALITY is positively related (0.0574, t-statistic=2.82) to TQ, suggesting that employees' satisfaction with the quality of products and services provided by their employers is positively related to firm performance. Similarly, we find that INNOVATION is positively related (0.0748, t-statistic=2.90) to firm performance, suggesting that employee satisfaction with employers efforts to innovate in their industry is also positively related to firm performance. TEAMWORK is also positively related (0.1769, t-statistic=5.08) to firm performance, suggesting similar notions related to the camaraderie and teamwork that is at the employer workplace and its relation to firm performance.

We find several aspects of employee satisfaction are actually negatively associated with firm performance. Specifically, we find the estimated coefficient for SAFETY is negative and statistically significant (-0.1567, t-statistic=-5.51), suggesting that higher levels of employee satisfaction has a negative impact on firm performance; perhaps because high levels of safety are costly to implement, thus dragging down firm performance/value. Similarly, we find REWARD (-0.1430, t-statistic=-4.80) and COMMUNICATION (-0.1184, t-statistic=-2.84) are negative, suggesting that employee satisfaction for rewards and internal communication have a drag on firm performance. Lastly, INTEGRITY is negative and significant (-0.1347, t-statistic=-2.72), suggesting that satisfaction with the integrity of upper management has a drag on firm performance, perhaps because lack of integrity or unethical, cut-throat is a component in, say, garnering new business or successfully entering competitive markets.

The results of our nine categories of employee satisfaction thus paint a vastly different picture of the components of employee satisfaction as they relate to firm performance. This granularity and depth of

detail thus can help researchers, practitioners, as well as regulators about the aspects of human resource management that are important in creating and maintaining firm value and firm performance. For instance, our results suggest that there is a direct and measurable cost for firms to raise the level of their SAFETY, or even their COMMUNICATION. Similarly, it raises policy implications about how regulators should approach issues of management INTEGRITY, as such important values come at a cost.

Such view is also supported by a number of studies in the past. Bowie (2000) describes that ethics could be seen as a constraint on profitability. Safety has always been a major concern in chemical, energy, and construction industry, a more effective construction of safety and health program may lead to more direct labor costs in short term and lower company performance. McMurrian and Matulich (2006) stated that some managers consider ethics programs in their organizations to be very expensive activities that are only societally rewarding. Donaldson (2003) summarized 52 research projects examining the correlation between ethics and profits. Fourteen of the 52 studies reported no effect and five reported a negative effect on profits. White et al. (2003) analyzed the data from national surveys of British employees from 1992 and 2000 and their results suggested a conflict between high-performance management practice and work/life balance policies. Beauregard and Henry's (2009) study also showed that there is insufficient evidence to support the notion that work/life practices enhance performance by means of reduced work/life conflict.

**Table 4: OLS Regression of Firm Performance on Employee Satisfaction Metrics**

|                   | <b>Model 1</b> | <b>Model 2</b> | <b>Model 3</b> |
|-------------------|----------------|----------------|----------------|
| Intercept         | 0.4474         | 0.4830         | -1.4502        |
| stderr            | 0.0450         | 0.0491         | 0.0703         |
| t-stat            | 9.95           | 9.84           | -20.64         |
| <b>OVERALL</b>    | <b>0.1707</b>  | <b>0.1630</b>  | <b>0.0521</b>  |
| stderr            | <b>0.0139</b>  | <b>0.0141</b>  | <b>0.0135</b>  |
| t-stat            | <b>12.27</b>   | <b>11.59</b>   | <b>3.87</b>    |
| <b>QUALITY</b>    |                | 0.0277         | <b>0.0574</b>  |
|                   |                | 0.0218         | <b>0.0204</b>  |
|                   |                | 1.27           | <b>2.82</b>    |
| <b>INNOVATION</b> |                | <b>0.0667</b>  | <b>0.0748</b>  |
|                   |                | <b>0.0277</b>  | <b>0.0258</b>  |
|                   |                | <b>2.41</b>    | <b>2.90</b>    |
| <b>RESPECT</b>    |                | <b>0.0796</b>  | 0.0048         |
|                   |                | <b>0.0305</b>  | 0.0285         |
|                   |                | <b>2.61</b>    | 0.17           |
| <b>TEAMWORK</b>   |                | <b>0.2073</b>  | <b>0.1769</b>  |
|                   |                | <b>0.0374</b>  | <b>0.0348</b>  |
|                   |                | <b>5.55</b>    | <b>5.08</b>    |
| <b>SAFETY</b>     |                | <b>-0.2222</b> | <b>-0.1567</b> |
|                   |                | <b>0.0305</b>  | <b>0.0285</b>  |
|                   |                | <b>-7.29</b>   | <b>-5.51</b>   |
| <b>REWARD</b>     |                | <b>-0.1407</b> | <b>-0.1430</b> |
|                   |                | <b>0.0319</b>  | <b>0.0298</b>  |
|                   |                | <b>-4.40</b>   | <b>-4.80</b>   |
| <b>COMMUNITY</b>  |                | 0.0068         | 0.0311         |
|                   |                | 0.0384         | 0.0358         |
|                   |                | 0.18           | 0.87           |

|           |              |                |                |
|-----------|--------------|----------------|----------------|
| COMMUNIC  |              | <b>-0.2459</b> | <b>-0.1184</b> |
|           |              | <b>0.0446</b>  | <b>0.0417</b>  |
|           |              | <b>-5.51</b>   | <b>-2.84</b>   |
| INTEGRITY |              | <b>-0.1178</b> | <b>-0.1347</b> |
|           |              | <b>0.0532</b>  | <b>0.0496</b>  |
|           |              | <b>-2.21</b>   | <b>-2.72</b>   |
| LOGMV     |              |                | <b>0.2413</b>  |
|           |              |                | <b>0.0067</b>  |
|           |              |                | <b>36.24</b>   |
| n         | 8664         | 8664           | 8664           |
| Adj-R2    | <b>0.017</b> | <b>0.0324</b>  | 0.1598         |

## Conclusion and Future Direction

This study aims to provide a new direction of analyzing employee satisfaction and its relation to corporate performance. Specifically, we propose to use Web data from Glassdoor.com to extract relevant information to Employee Satisfaction. Our study has three contributions. First, compared to traditional survey, this method exploits the unique social media data Glassdoor.com to extract anonymous employee reviews for textual analysis. Our dataset allows for various exploratory analysis including cross-industry and in-depth category analysis. Second, our study demonstrated the types of analyses that are possible using text mining approach. After the extraction of key categories, both exploratory and regression analysis were performed on the dataset to reveal interesting trends and correlations. We hope that it encourages other researchers to consider the rich environ that a text analytics methodology makes possible. Third, our correlation analysis confirmed the positive correlation between general employee satisfaction and corporate performance. Employee satisfaction category were further explored and negative correlations were found between Safety, Communication and Integrity. This finding aligns with several previous studies.

In the future, we plan to extend our study in several directions. We plan to adopt more advanced regression analysis, control industry sector and study each industry individually. We will also incorporate more sophisticated sentiment analysis in text mining to achieve more accurate result. Besides the nine categories/keywords used in this study, we will also explore other categories/keywords and develop more advanced extraction techniques.

## References

- Abrahams, A.S., Jiao, J., Wang, G.A. and Fan, W. 2012. "Vehicle defect discovery from social media." *Decision Support Systems*, (54:1), pp.87-97.
- Asur, S. and Huberman, B.A. 2010. "Predicting the future with social media." *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on Vol. 1, pp. 492-499. IEEE.
- Bae, J.H., Son, J.E. and Song, M. 2013. "Analysis of twitter for 2012 South Korea presidential election by text mining techniques." *Journal of Intelligence and Information Systems*, (19:3), pp.141-156.
- Baines, S. and Gelder, U. 2003. "What is family friendly about the workplace in the home? The case of self-employed parents and their children." *New Technology, Work and Employment*, (18:3), pp.223-234.
- Barrett, D.J., 2002. "Change communication: using strategic employee communication to facilitate major change." *Corporate Communications: An International Journal*, (7:4), pp.219-231.
- Beauregard, T.A. and Henry, L.C. 2009. "Making the link between work-life balance practices and organizational performance." *Human resource management review*, (19:1), pp.9-22.
- Bermingham, A., & Smeaton, A. 2011. "On using twitter to monitor political sentiment and predict election results." *In Proceeding of IJCNLP conference*, Chiang Mai, Thailand
- Bollen, J., Mao, H. and Zeng, X. 2011. "Twitter mood predicts the stock market." *Journal of Computational Science*, (2:1), pp.1-8.
- Bowie, Norman E. 2000. "Companies Are Discovering the Value of Ethics," *Business Ethics 00/01*, 12th ed., Guilford, Connecticut: McGraw-Hill/Dushkin, pages 150-152.
- Büschgens, T., Bausch, A. and Balkin, D.B. 2013. "Organizational Culture and Innovation: A Meta-Analytic Review." *Journal of product innovation management*, (30:4), pp.763-781.
- Chen, H., Chiang, R.H. and Storey, V.C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS quarterly*, (36:4), pp.1165-1188.
- Chung, J. and Rao, V.R. 2012. "A general consumer preference model for experience products: application to internet recommendation services". *Journal of marketing research*, (49:3), pp.289-305.
- Daft, R.L., Sormunen, J. and Parks, D. 1988. "Chief executive scanning, environmental characteristics, and company performance: An empirical study." *Strategic Management Journal*, (9:2), pp.123-139.
- Denison, D.R. 1990. *Corporate culture and organizational effectiveness*. NY: Wiley.
- Donaldson, Thomas (2003), "Adding Corporate Ethics to the Bottom Line," *Business Ethics 03/04*, 15th ed., Guilford, Connecticut: McGraw-Hill/Dushkin, pages 98-101.
- Edmans, A. 2011. "Does the stock market fully value intangibles? Employee satisfaction and equity prices." *Journal of Financial Economics*, (101:3), pp.621-640.
- Fortune. 2015. "Fortune 500", *Fortune*, available at <http://fortune.com/fortune500/>; retrieved May 9, 2016.
- Frolick, M.N., Wilkes, R.B. and Urwiler, R. 1993. "Telecommuting as a workplace alternative: an identification of significant factors in American firms' determination of work-at-home policies." *The Journal of Strategic Information Systems*, (2:3), pp.206-220.
- Fuller, J.B., Hester, K., Barnett, T., Beu, D., Frey, L. and Relyea, C. 2009. "Extending the group engagement model: An examination of the interactive effects of prestige, respect, and employee role identity." *Journal of Managerial Issues*, pp.119-139.
- Gadgets360. 2016. "Apple's 'magic' and innovation its biggest strength, says CEO Tim Cook", NDTV Gadgets360.com, available at <http://gadgets.ndtv.com/mobiles/news/apples-magic-and-innovation-its-biggest-strength-says-ceo-tim-cook-330109>; retrieved May 9, 2016.
- Glassdoor. 2016. "About Us", *Glassdoor*, available at [https://www.glassdoor.com/about/index\\_input.htm](https://www.glassdoor.com/about/index_input.htm); retrieved May 9, 2016
- Guiso, L., Sapienza, P. and Zingales, L. 2015. "The value of corporate culture." *Journal of Financial Economics*, (117:1), pp.60-76.
- Harter, J.K., Schmidt, F.L. and 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), p.268.
- Hermalin, B.E. and Weisbach, M.S., 2001. "Boards of directors as an endogenously determined institution: A survey of the economic literature (No. w8161)." National Bureau of Economic Research.

- HRZONE. 2015. "Why did Google abandon 20% time for innovation?", HRZone, available at <http://www.hrzone.com/lead/culture/why-did-google-abandon-20-time-for-innovation>; retrieved May 9, 2016.
- Huang, M., Li, P., Meschke, F. and Guthrie, J.P. 2015. "Family firms, employee satisfaction, and corporate performance." *Journal of Corporate Finance*, 34, pp.108-127.
- Jahanbakhsh, K. and Moon, Y., 2014. "The predictive power of social media: On the predictability of US Presidential Elections using Twitter." *arXiv preprint arXiv:1407.0622*.
- Jiménez-Jiménez, D. and Sanz-Valle, R. 2011. "Innovation, organizational learning, and performance." *Journal of business research*, (64:4), pp.408-417.
- Johnson, N.B. and Provan, K.G. 1996. "The relationship between work/family benefits and earnings: A test of competing predictions." *The Journal of socio-economics*, (24:4), pp.571-584.
- Kamakura, W.A., Mittal, V., De Rosa, F. and Mazzon, J.A. 2002. "Assessing the service-profit chain." *Marketing science*, (21:3), pp.294-317.
- Kerr, J. and Slocum, J.W. 2005. "Managing corporate culture through reward systems." *The Academy of Management Executive*, (19:4), pp.130-138.
- Kim, S.M. and Hovy, E. 2006, "July. Automatic identification of pro and con reasons in online reviews." In *Proceedings of the COLING/ACL on Main conference poster sessions*, pp. 483-490. Association for Computational Linguistics.
- Konrad, A.M. and Mangel, R. 2000. "Research notes and commentaries the impact of work-life programs on firm productivity." *Strategic Management Journal*, (21:12), pp.1225-1237.
- Koys, D.J. 2001. "The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study." *Personnel psychology*, (54:1), pp.101-114.
- Lassen, N.B., Madsen, R. and Vatrapu, R. 2014, September. "Predicting iphone sales from iphone tweets." In *Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International* pp. 81-90. IEEE.
- McMurrian, R.C. and Matulich, E., 2006. "Building Customer Value And Profitability With Business Ethics." *Journal of Business & Economics Research (JBER)*, (4:11).
- Microsoft Innovation Centers. 2016. "Microsoft Innovation Centers - About Microsoft Innovation Centers", [Microsoftinnovationcenters.com](https://www.microsoftinnovationcenters.com), available at <https://www.microsoftinnovationcenters.com/aboutus>; retrieved May 9, 2016.
- Moniz, A. 2015. "Inferring Employees' Social Media Perceptions of Goal-Setting Corporate Cultures and the Link to Firm Value."
- Montes, F.J.L., Moreno, A.R. and Morales, V.G. 2005. "Influence of support leadership and teamwork cohesion on organizational learning, innovation and performance: an empirical examination." *Technovation*, (25:10), pp.1159-1172.
- O'Connor, B., Bamman, D. and Smith, N.A. 2011. "Computational Text Analysis for Social Science: Model Assumptions and Complexity." *public health*, (41:42), p.43.
- Oh, C. and Sheng, O. 2011. "Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement." In *Proceedings of the International Conference on Information Systems*.
- O'Reilly, C.A., Caldwell, D.F., Chatman, J.A. and Doerr, B. 2014. "The Promise and Problems of Organizational Culture CEO Personality, Culture, and Firm Performance." *Group & Organization Management*, (39:6), pp.595-625.
- Ostroff, C. 1992. "The relationship between satisfaction, attitudes, and performance: An organizational level analysis." *Journal of applied psychology*, (77:6), p.963.
- Perry-Smith, J.E. and Blum, T.C. 2000. "Work-family human resource bundles and perceived organizational performance." *Academy of management Journal*, (43:6), pp.1107-1117.
- Popadak, J.A. 2013. "A corporate culture channel: How increased shareholder governance reduces firm value." *Available at SSRN 2345384*.
- Rubera, G. and Kirca, A.H. 2012. "Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration." *Journal of Marketing*, (76:3), pp.130-147.
- Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H. and Jin, C. 2007, June. "Red Opal: product-feature scoring from reviews." In *Proceedings of the 8th ACM conference on Electronic commerce*, pp. 182-191. ACM.

- Schneider, B., Hanges, P.J., Smith, D.B. and Salvaggio, A.N. 2003. "Which comes first: employee attitudes or organizational financial and market performance?" *Journal of Applied Psychology*, (88:5), p.836.
- Reinartz, W., Krafft, M. and Hoyer, W.D. 2004. "The customer relationship management process: Its measurement and impact on performance." *Journal of marketing research*, (41:3), pp.293-305.
- Reinartz, W.J. and Kumar, V., 2000. "On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing." *Journal of marketing*, (64:4), pp.17-35.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. 2010. "Predicting elections with twitter: What 140 characters reveal about political sentiment." *ICWSM*, 10, pp.178-185.
- Tan, A.H. 1999. "Text mining: The state of the art and the challenges." In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases* Vol. 8, pp. 65-70.
- White, M., Hill, S., McGovern, P., Mills, C. and Smeaton, D. 2003. "High-performance Management Practices, Working Hours and Work-Life Balance." *British Journal of Industrial Relations*, (41:2), pp.175-195.
- Whitehouse, G. and Zetlin, D. 1999. "Family friendly policies: distribution and implementation in Australian workplaces." *The Economic and Labour Relations Review*, (10:2), pp.221-239.
- Zahra, S.A. 1993. "Environment, corporate entrepreneurship, and financial performance: A taxonomic approach." *Journal of business venturing*, (8:4), pp.319-340.
- Zhou, W., and Duan, W., 2015. "An empirical study of how third-party websites influence the feedback mechanism between online Word-of-Mouth and retail sales." *Decision Support Systems* 76: 14-23.