# e-Commerce and Sentiment Analysis: Predicting Outcomes of Class Action Lawsuits

**Stacey Taylor**
Dalhousie University
Halifax, Nova Scotia
stacey.taylor@dal.ca

**Vlado Keselj**
Dalhousie University
Halifax, Nova Scotia
vlado@cs.dal.ca

## Abstract

In recent years, the focus of e-Commerce research has been on better understanding the relationship between the internet marketplace, customers, and goods and services. This has been done by examining information that can be gleaned from consumer information, recommender systems, click rates, or the way purchasers go about making buying decisions, for example. This paper takes a very different approach and examines the companies themselves. In the past ten years, e-Commerce giants such as Amazon, Skymall, Wayfair, and Groupon have been embroiled in class action security lawsuits promulgated under Rule 10b(5), which, in short, is one of the Securities and Exchange Commission's main rules surrounding fraud. Lawsuits are extremely expensive to the company and can damage a company's brand extensively, with the shareholders left to suffer the consequences. We examined the Management Discussion and Analysis and the Market Risks for 96 companies using sentiment analysis on selected financial measures and found that we were able to predict the outcome of the lawsuits in our dataset using sentiment (tone) alone to a recall of 0.8207 using the Random Forest classifier. We believe that this is an important contribution as it has cross-domain implications and potential, and opens up new areas of research in e-Commerce, finance, and law, as the settlements from the class action lawsuits in our dataset alone are in excess of $1.6 billion dollars, in aggregate.

## 1 Introduction

Since 1990, over four thousand securities class action lawsuits have been filed alleging violations of Section 10b of the Securities Exchange Act of 1934. While 10b is a very broad section, its main foci are manipulative and deceptive practices in relation to securities. In their communications with stakeholders, companies often refer to financial measures that do not conform with the Generally Accepted Accounting Principles (GAAP). These non-GAAP measures (NGMs) have been shown to positively increase a document's tone, which could be overinflating the company's prospective performance. Consequently, if actual performance falls short, overly favourable wording could be part of the instigation of securities lawsuits.

To evaluate if we could use a NGMs approach to classifying securities class action lawsuits, we use the financial filings submitted to the U.S. Securities and Exchange Commission (SEC) for ninety six random lawsuits, half settled and half dismissed, over the alleged damage period, as our dataset. We propose a novel use of sentiment analysis by examining a key section of the quarterly and annual reports submitted to the SEC in two states: first, the unaltered report as filed with the SEC $(X')$, and second, the report without selected NGMs $(X)$. We then calculated the change in the tone or sentiment (as we use these terms interchangeable) as $(X - X')$ for each report and used it as an input to our prediction model. We found that we are able to predict the outcome of the lawsuits for the aggregate dataset with a recall of 0.8207 using the calculated sentiment (tone) change alone using the Random Forest classifier. When the tone change is used in conjunction with other features, we find that we are able to predict the outcome with a recall of 0.9142, again using the Random Forest classifier.

Securities lawsuits are extremely expensive to companies: the settlements from our sample alone are in excess of $1.6 billion dollars, in aggregate. To our knowledge, this use of Natural Language Processing, in particular the change in Sentiment Analysis of the NGMs in financial reports and approach to potential lawsuit classification has not been done before. We believe that this is an important contribution as it has cross-domain impli-

cations and potential, and opens up new areas of research in e-Commerce, finance, and law.

## 2 Related Work

The recurring theme of research that supports the use of NGMs is altruism — that they provide additional, relevant information that GAAP cannot (Black et al., 2018; Boyer et al., 2016; Bhattacharya et al., 2003; Frankel et al., 2004). These measures are everywhere in the financial ecosphere and have become accepted as part of the fundamental financial narrative. While the use of NGM has its supporters, there are many more detractors who cite evidence that strongly suggests that the motives are opportunistic rather than altruistic. Earnings targets are a fundamental part of measuring corporate goals. Companies set these objectives to help the company grow, but also demonstrably communicate to investors that the company is worth investing in. Researchers have found that there is a higher percentage of companies that are meeting or beating their earnings targets relative to those that do not. This strongly suggests that there is some degree of financial "management" (Burgstahler and Dichev, 1997; Brown and Caylor, 2005; Graham et al., 2005; Roychowdhury, 2006; Lougee and Marquardt, 2004; Bhattacharya et al., 2003; Davis and Tama-Sweet, 2012; Doyle et al., 2013; Black et al., 2018), and one of the tools available to do that are NGMs.

Research has also found NGMs, even as supplementary measures, are misleading given their persuasive nature (Fisher, 2016; Asay et al., 2018) as the company is essentially implying, through the adjustments that they make, that its actual performance is different (and in some cases starkly different) from its audited performance. Alee et al. also raises the concern that non-GAAP earnings, in particular, may confuse and mislead the average investor (Allee et al., 2007) when non-GAAP profits are created through adjustments from what was originally a GAAP loss (Young, 2014). Kang et al. found that when management discloses information to stakeholders, it tends to use "flexibility" in the tone in order to limit the damage by framing the negativity in positive ways (Kang et al., 2018; Li, 2016), which speaks to corporate motivation. Loughran-McDonald (2011) found that this motive entices writers to re-frame negativity into positivity because the impact of negative words on shareholders (or potential shareholders) is inex-

orable (Loughran and McDonald, 2011a). Therefore, careful use of word constructs can help to avoid, or at least, significantly limit the pervasive affect brought on by negative wording. This idea is also echoed by Rogers et al. (2011) who indicate that overly optimistic tones can be catalysts for Securities Class Action Lawsuits (Rogers and Van Buskirk, 2009).

Wongchaisuwat, Klabjan, and McGinnis used clustering classification models to determine the likelihood of patent litigation. If litigation was determined to be likely, SEC financial data was then incorporated into the model to predict the timeline to litigation (Wongchaisuwat et al., 2017). Gruginskie and Vaccaro also researched lawsuit lead time based on data provided by the Tribunal Regional Federal da 4ª Regiã from 2016 (Gruginskie and Vaccaro, 2018). Their model was broken down into four time frames: *Up to 1 Year; From 1 to 3 Years; From 3 to 5 years; and More than 5 Years*. Overall, Support Vector Machines and Random Forest returned the best F1 measure performance of 83.85 and 83.33, respectively, for results *Up to 1 Year*.

Alexander et al. examined features extracted from source documents such as the lawsuit itself, the trial docket, summary judgments, and the magistrate's report to predict the outcomes of a series of lawsuits (Alexander et al., 2018). Using a random forest model, they varied the number of features used in prediction to see which model would provide the most insight. The model that used the full range of features provided the best performance, resulting in 94% accuracy (Alexander et al., 2018).

## 3 Research Design

### 3.1 Methodology

Rogers, Van Buskirk, and Zechman used plantiff complaints to determine which corporate disclosures were most likely to put a firm at risk of litigation (Rogers et al., 2011). Although Rogers et al. did not not disclose which companies were included in their dataset, we based the main idea of our methodology on their work and used lawsuit information and corporate disclosures in conjunction with well-known dictionaries to create our dataset.

We randomly selected 96 lawsuits from the heat map on Stanford's Securities Class Action Clearinghouse (SCAC). 16 lawsuits were gathered from each of the *Top 3* sectors (Technology, Service, and Financial) and 16 lawsuits from each of the *Bottom 3* sectors (Utilities, Transportation, and Conglomer-

ates) during the period beginning in 1990 to 2017. The following criteria were used for a company's inclusion in the dataset:

- the company had to be a public company in order for us to be able to access the company's 10-K and 10-Q reports from the SEC;

- the lawsuits had to be drawn from the *Top 3* and the *Bottom 3* sectors in the SCAC heat map;

- the class action lawsuit had to be promulgated under Rule 10b; and

- the lawsuit's status had to be either "settled" or "dismissed".

Note: *Rule 10b, which is most often addressed under Section 5, addresses deception and making false statements, among other things.* (Congress, 1951).

We then reviewed the information on the information on the SCAC to determine the *alleged* damage period and the length of the lawsuit. Both of these characteristics were then added to the dataset. The 10-K and 10-Q reports were gathered for each company that corresponded to the *alleged* damage period. Our focus was solely on the Management Discussion & Analysis (MD&A) and the Market Risks (following the research of (Loughran and McDonald, 2011a), so we parsed those sections out of the 10-K and 10-Q reports.

We curated a list of NGMs to target by using common NGMs published by Deloitte (Deloitte, 2019) as our starting point. The SEC has very specific rules regarding NGMs. In certain cases, what is normally considered to be a non-GAAP measure is, under SEC regulations, determined to be ***not*** non-GAAP in certain prescribed circumstances (Securities and Commission, 2018). Any NGMs that required contextualization to determine if the measure was actually non-GAAP or *not* non-GAAP under SEC regulation were removed. The following list of NGMs are considered to be *always* non-GAAP under any circumstances:

- Revised Net Income

- Earnings Before Interest and Taxes (EBIT)

- Earnings Before Interest, Taxes, and Depreciation (EBITDA)

- Earnings Before Interest, Taxes, Depreciation, Amortization, and Rent/Restructuring (EBITDAR)

- Adjusted Earnings Per Share

- Free Cash Flow (FCF)

- Core Earnings

- Funds From Operations (FFO)

- Unbilled Revenue

- Return on Capital Employed (ROCE)

- Non-GAAP

- Reconciliation

Note: "Revised" or "Adjusted" variants of measures, such as "Adjusted EBIT" were also included, as were commonly accepted variations of naming of the NGMs such as "debt-free cash flow" and "unlevered free cash flow". Also, we added the word "reconciliation" into our short list.

Using this list, sentences in the MD&A and Market Risks that contained the NGMs were then removed. Our rationale for taking this approach is that the non-GAAP measure is the focus of the sentence, and therefore, the words in that sentence exist only for discussing that measure.

To illustrate that point, we offer the following: "Our EBITDA decreased 2% for the first quarter of fiscal 2012 compared to the first quarter of fiscal 2011, due to a slight decrease in net revenues and a slight increase in operating expenses." (Taken from TD Ameritrade's 10-Q filing made on 2012-02-08.) If we take a Bag-of-Words (BoW) approach to this sentence and only remove the NGM — in this case EBITDA — that leaves the rest of the words in the sentence. Yet, without the NGM, the sentence no longer makes sense: "Our decreased 2% for the first quarter of fiscal 2012 compared to the first quarter of fiscal 2011, due to a slight decrease in net revenues and a slight increase in operating expenses."

Therefore, using the BoW approach, the words from the second non-sensical sentence would be left in when calculating the sentiment (as only the NGM keyword *EBITDA* would be removed). In reality, *all* of the words left in the sentence exist only to discuss and contextualize the NGM and need to be removed. Using both versions of the report — one with the NGMs and one without — we

conducted a sentiment analysis and calculated the change in the sentiment (tone) between the MD&A and Market Risks from the report as filed with the SEC and the report with the NGMs removed *(X - X′)*.

## 3.2 Dictionaries used for Sentiment Analysis in R

The financial lexicon and jargon used by professionals, which subsequently appears in reports, financial statements and filings (such as the 10-K and 10-Q reports we examined for our research), can be quirky and nuanced. As noted by Loughran-McDonald (Loughran and McDonald, 2011b), there are a lot of words which, out of the financial context, elicit emotional responses that may not be warranted. The word "debt" (which is a financial liability) is a good example. When used in a *business* context, the word itself is neutral; it is expected that businesses will have debt and, until that debt has been contextualized by taking into account the rest of the facts, figures, and discussions, it is not appropriate to assign it a tonal label.

We used four dictionaries provided in R to conduct our sentiment analysis, as follows:

- *Harvard-IV:* Psychological dictionary. The implementation of this dictionary in R is strictly a binary classification. There are 1,316 positive words and 1,746 negative words. Words such as *debt*, *interest* and *taxes* are negative words in this dictionary, and are assigned a score of −1 (Feuerriegel and Proellochs, 2019)

- *QDAP:* Collection of dictionaries that include subsets of Harvard-IV, Hu-Liu (Hu and Liu, 2004), Dolch's 220 most common words by reading level (Dolch, 1936), census data collected by the U.S. Government, among others (Feuerriegel and Proellochs, 2019). The R implementation of this dictionary is a binary classification and 1,208 positive words and 2,952 negative words. Words such as *debt*, *interest*, and *taxes* are negative words in this dictionary, and are assigned a score of −1 (Feuerriegel and Proellochs, 2019).

- *Henry:* Financially oriented dictionary. This dictionary has a binary classification with 53 positive words and 44 negative words. Words such as *debt*, *interest*, and *taxes* are, by omission, neutral words in this dictionary, and are assigned a score of 0 (Henry, 2008).

- *Loughran-McDonald:* Financially oriented dictionary. The R implementation of this dictionary is a binary classification only, with 145 positive and 885 negative words. Words such as *debt*, *interest*, and *taxes* are neutral words in this dictionary, and are assigned a score of 0 (Feuerriegel and Proellochs, 2019). The authors have noted, *"Language is dynamic"* and to keep up with that dynamism, they update this dictionary on an annual basis. Since 2012, no words have been deleted from their dictionary, but 343,606 words have been added and 265 words have been reclassified (Loughran and McDonald, 2018).

## 3.3 Dataset Characteristics

Characteristics of our prediciton model are as follows:

1. Date (date that the company filed the report with the SEC). This date is then compared to the *alleged* damage period, to determine which SEC filings are relevant to the lawsuit.

2. Central Index Key ("CIK" which acts as the company number for the SEC). The CIK is used to ensure that the reports and information gathered are for the correct company. It also facilitates calculation of the length of the lawsuit.

3. cgi (change the tone for the General Inquirer dictionary)

4. che (change in tone for the Henry dictionary)

5. clm (change in tone for the Loughran-McDonald dictionary)

6. cqdap (the change in the tone for the QDAP dictionary)

*Notes: Tone changes for each dictionary is calculated as (X - X′). Also, the number of documents included in the dataset for each company was dependent on the length of the alleged damage period.*

The class being predicted was the outcome of the class action lawsuit as either settled or dismissed. Please see Table 1 for the specific composition of the dataset.

| Characteristic | Number/Length/Dollar Value |
|---|---|
| Total Number of Documents | 2, 170 |
| Number of documents per Sector: | |
| (Top 3) Services | 468 |
| (Top 3) Financial | 542 |
| (Top 3) Technology | 536 |
| (Bottom 3) Utilities | 218 |
| (Bottom 3) Transportation | 248 |
| (Bottom 3) Conglomerates | 158 |
| Longest Damage Period | 46 months |
| Shortest Damage Period | 1 month |
| Largest Settlement | $410 million |
| Smallest Settlement | $1.5 million |

Table 1: Dataset Characteristics

### 3.4 Experiments and Evaluation Methods

We performed two different main experiments to test our model, both using 10 fold cross-validation.

The first experiment used aggregated data (all six sectors — *Top 3* and *Bottom 3*) only and leveraged all of the dictionaries. Using Naïve Bayes (NB), Random Forest (RF), and Support Vector Machines (SVM) for our predictive models, we ran a series of tests, varying the number of features used in the class prediction to determine the predictive capacity of each algorithm. In the first run, we used all features in the dataset, as outlined above to predict the outcome. We decreased the number of features used in the second run to only the sentiment and period to predict the outcome. For the third (and final run), we used only the sentiment to predict the outcome. We were particularly interested in the results for the use of sentiment alone given that the change in the sentiment score was driven by the removal of the NGM sentences.

The second experiment used the exact same parameters, reasoning, and interest as the first, with the exception of the data used. Here, we rolled up each individual sector into its major constituent of either *Top 3* (Technology, Service, and Financial) or *Bottom 3* (Utilities, Transportation, and conglomerates).

Class action lawsuits are inherently expensive (regardless of outcome). The settlements from the class action lawsuits in our dataset alone are in excess of $1.6 billion dollars, in aggregate. As indicated in Table 1, the largest individual company settlement was $410 million dollars. Our focus has been on corporate disclosure in the MD&A and Market Risks sections of the 10-K and 10-Q reports filed with the SEC. These disclosures have been meticulously reviewed by company executives, and likely auditors and the company's legal team as well before dissemination to the public. That also means that if a company is to adjust its disclosure to help shield itself from legal action, it has to be done in the drafting and (subsequent) approval stage of the MD&A and Market Risks before release to stakeholders.

From a business point of view, if the cost of **acting** is high (such as making a considerable investment), then precision is the most appropriate measure. But, if the cost of **not acting** is high (such as taking steps to avoid an overly optimistic disclosure tone prior to release), then recall is the most appropriate. Therefore, we chose recall as the most appropriate measure to evaluate our models.

We also make a distinction here between Information Retrieval (IR) and Classification. In IR, a trade-off can be made between precision and recall in that it we can simply return all documents in order to get a high recall, but a very low precision (Manning, Christopher D. and Schütze, Hinrich, 1999). However, in our classification model, we recognize that there is a corporate cost to every action that a company takes — including writing and distributing corporate disclosures. Given this, we see no tangible value for companies and investors alike if all documents are returned in order to trade precision for recall.

## 3.5 Results

The full results from our experiments can be found in Table 2. The *Aggregate* data results used in Experiment 1 are presented first, followed by the major constituents of *Top 3* and *Bottom 3*. The number of features ranges from all to just sentiment alone to predict the outcome (class), as denoted in the table.

Keeping in mind that *recall* is the measure that we are focusing on, we see that Random Forest (RF) is predominantly the best algorithm for this data. The results returned using RF are quite robust, returning a recall of 0.9142 for the Aggregate using all features, and 0.9938 and 0.9407 for the *Top 3* and *Bottom 3*, respectively. When using the tone change alone, the results are robust as well, returning a recall of 0.8207 for the aggregate dataset. At each node, this algorithm is designed to choose the best among randomly chosen predictors to make its decision, and then move on, to prevent overfitting. RF also works well with both numerical and categorical data, which we have. As well, because it employs a boot-strapping method (i.e. that samples are selected and then replaced to be selected again the future), and therefore makes the random tree more robust.

We varied the number of attributes used between tests, determining how far we could strip down the features in the model until the prediction dropped off significantly. The ensemble nature of this algorithm is particularly well suited to this classification task as it uses prediction by committee to overcome the shortcomings of the individual trees.

NB performed the best of all of the algorithms when classifying the Aggregate using the Sentiment Score, the Period, and the Outcome, resulting in a recall of 0.9794. NB also outperformed RF again when classifying both the Top 3 and the Bottom 3 sectors using just the Sentiment and the Outcome. We believe that this is due to the tenet of NB, which is that all of the variables are assumed to be conditionally independent. It also works well with small datasets, which we have.

SVM performed the worst out of the algorithms. The highest recall was 0.6600, was for the *Bottom 3* Sectors using the Sentiment, Period, and Outcome, but was still far off the best performing classifiers. In our dataset, there are a number of filings where the change between the "before" and "after" was zero. This means that the company did not use any of the non-GAAP measures in our extraction list. We believe that due to the fact that this type of paired data cannot be easily separated, that SVM is not well suited to our type of data.

## 3.6 Why This Matters

The arrival of e-Commerce changed the global marketplace forever, giving consumers access to products and services that before they would not have necessarily had access to. e-Commerce also altered the way that businesses *do business* as information that was not readily accessible like shopping habits, tools to infer decision-making, and search history, became available, allowing businesses a keen eye into *who* their customer really is. The economy has also folded in e-Commerce so well, that it is now, more than ever, dependent on it; the global pandemic COVID-19 has made this very clear. Businesses who, before the pandemic, had shied away from, or even made the conscious decision not to engage in e-Commerce have been thrust online, forcing those businesses to pivot quickly for survival.

Investors need to be appropriately protected from adverse financial investment where possible in order for the economy to stay health and strong. e-Commerce is a mainstay in the marketplace, ranging from buying goods and services online to *investing* online. It is, therefore, important that e-Commerce companies are scrutinized, alongside traditional business, to ensure the investment is sound. Sentiment analysis is an excellent tool for such scrutiny as it affords the ability to capture and demonstrate the power of sentiment of both financial professionals and the average financial investor, while allowing research to show the dichotomy that financial language and jargon have on each group's interpretation of company health, risk, and the soundness of an investment. Our research can have an impact on the different understandings of language and how it can help consumers make decisions. It also opens up new avenues of research within the domains of e-Commerce, finance, and law.

## 4 Conclusion

Our research provided the novel approach of performing an extractive sentiment analysis using the tone change between financial reports containing NGMs with those that do not for prediction of the outcome of Securities Class Action lawsuits promulgated under Rule 10b(5). We conducted our

| Algorithm and Features Used | Precision | Recall | F1 | Accuracy |
|---|---|---|---|---|
| *Aggregate (All Features)* | | | | |
| Naïve Bayes | 0.6049 | 0.8822 | 0.7424 | 0.7973 |
| Random Forest | 0.9123 | **0.9142** | 0.9133 | 0.9210 |
| Support Vector Machine | 0.6610 | 0.6390 | 0.6100 | 0.6439 |
| *Aggregate (Sentiment, Period)* | | | | |
| Naïve Bayes | 0.6597 | **0.9794** | 0.7884 | 0.8172 |
| Random Forest | 0.8100 | 0.8879 | 0.8472 | 0.8668 |
| Support Vector Machines | 0.6235 | 0.6280 | 0.6057 | 0.6090 |
| *Aggregate (Sentiment)* | | | | |
| Naïve Bayes | 0.5581 | 0.7152 | 0.6270 | 0.6803 |
| Random Forest | 0.6977 | **0.8207** | 0.7542 | 0.7481 |
| Support Vector Machines | 0.2693 | 0.5194 | 0.3543 | 0.2734 |
| *Top 3 (All Features Used)* | | | | |
| Naïve Bayes | 0.6409 | 0.8822 | 0.7424 | 0.7973 |
| Random Forest | 0.9493 | **0.9938** | 0.9710 | 0.9754 |
| Support Vector Machines | 0.6220 | 0.6240 | 0.6021 | 0.6137 |
| *Top 3 (Sentiment, Period)* | | | | |
| Naïve Bayes | 0.6054 | 0.8761 | 0.7160 | 0.7812 |
| Random Forest | 0.8149 | **0.9446** | 0.8750 | 0.8990 |
| Support Vector Machines | 0.6235 | 0.6280 | 0.6057 | 0.6090 |
| *Top 3 (Sentiment)* | | | | |
| Naïve Bayes | 0.5248 | **0.9610** | 0.6789 | 0.7445 |
| Random Forest | 0.8089 | 0.8468 | 0.8274 | 0.8434 |
| Support Vector Machines | 0.5492 | 0.5492 | 0.4793 | 0.4636 |
| *Bottom 3 (All Features)* | | | | |
| Naïve Bayes | 0.6478 | 0.8067 | 0.7185 | 0.7798 |
| Random Forest | 0.8819 | **0.9407** | 0.9104 | 0.9104 |
| Support Vector Machines | 0.6884 | 0.6568 | 0.6453 | 0.6426 |
| *Bottom 3 (Sentiment, Period)* | | | | |
| Naïve Bayes | 0.6149 | 0.8306 | 0.7067 | 0.7785 |
| Random Forest | 0.8542 | **0.9111** | 0.8817 | 0.8817 |
| Support Vector Machines | 0.6905 | 0.6600 | 0.6491 | 0.6555 |
| *Bottom 3 (SentimentS)* | | | | |
| Naïve Bayes | 0.5248 | **0.9610** | 0.6789 | 0.7445 |
| Random Forest | 0.8156 | 0.7986 | 0.8070 | 0.7993 |
| Support Vector Machines | 0.4965 | 0.5119 | 0.4040 | 0.5002 |

Table 2: Case Study Machine Learning Results

experiments on 96 random lawsuits selected from the Stanford SCAC heat map (organized by sector) from the *Top 3* and *Bottom 3* sectors, equating to 16 lawsuits per sector. We found that using the calculated change in the sentiment $(X - X')$ alone was sufficient to predict the outcome of the securities class action lawsuits to a recall of 0.8207, and when sentiment was combined with other features,

recall rose to 0.9142 - both using RF.

## 4.1 Future Work

Taking an extractive sentiment approach, rather than a classical BoW approach, has provided new avenues of research. In this paper, we only examined the 10-K and 10-Q reports provided to the SEC. It would be valuable to apply this methodol-

ogy on different aspects of e-Commerce, such as buying decisions, based on different user-groups understanding and interpretations of keywords, and how words that characterize and contextualize affect sentiment.

We also suggest that the paradigm be shifted from focusing intently on how customer information can be used to drive bottom-line performance, to incorporating how e-Commerce companies communicate with their stakeholders to determine if there is alignment between what the company says and, ultimately, does, as evidenced in their regulatory and financial filings.

# References

Charlotte Alexander, Khalifeh Al Jadda, and Mohammed Javad Feizollahi. 2018. Using Text Analytics to Predict Litigation Outcomes.

Kristian D. Allee, Nilabhra Bhattacharya, Ervin L. Black, and Theodore E. Christensen. 2007. Pro Forma Disclosure and Investor Sophistication: External Validation of Experimental Evidence Using Archival Data. *Accounting, Organizations and Society*, 32(3):201 – 222.

H. Scott Asay, Robert Libby, and Kristina Rennekamp. 2018. Firm Performance, Reporting Goals, and Language Choices in Narrative Disclosures. *Journal of Accounting and Economics*, 65(2):380 – 398.

Nilabhra Bhattacharya, Ervin L Black, Theodore E Christensen, and Chad R Larson. 2003. Assessing the Relative Informativeness and Permanence of Pro Forma Earnings and GAAP Operating Earnings. *Journal of Accounting and Economics*, 36(1):285 – 319. Conference Issue on.

Dirk E. Black, Theodore E. Christensen, Jack T. Ciesielski, and Benjamin C. Whipple. 2018. Non-GAAP Reporting: Evidence from Academia and Current Practice. *Journal of Business Finance & Accounting*, 45(3-4):259–294.

Benoit Boyer, Ralph Lim, and Lyons Bridget. 2016. A Case Study in the Use and Potential Misuse of Non-GAAP Financial Measures. *Journal of Applied Business and Economics*, 18(3):117–126.

Lawrence D. Brown and Marcus L. Caylor. 2005. A Temporal Analysis of Quarterly Earnings Thresholds: Propensities and Valuation Consequences. *The Accounting Review*, 80(2):423–440.

David Burgstahler and Ilia Dichev. 1997. Earnings Management to Avoid Earnings Decreases and Losses. *Journal of Accounting and Economics*, 24(1):99 – 126. Properties of Accounting Earnings.

United States Congress. 1951. 17 CFR § 240.10b-5 - Employment of Manipulative and Deceptive Devices.

Angela K. Davis and Isho Tama-Sweet. 2012. Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases Versus MD&A. *Contemporary Accounting Research*, 29(3):804–837.

Deloitte. 2019. A Roadmap to Non-GAAP Financial Measures.

Edward W Dolch. 1936. A Basic Sight Vocabulary. *The Elementary School Journal*, 36(6):456–460.

Jeffrey T. Doyle, Jared N. Jennings, and Mark T. Soliman. 2013. Do Managers Define Non-GAAP Earnings To Meet or Beat Analyst Forecasts? *Journal of Accounting and Economics*, 56(1):40 – 56.

Stefan Feuerriegel and Nicolas Proellochs. 2019. *Package 'SentimentAnalysis'*.

Alex Fisher. 2016. Non-GAAP Measures: Academic Literature Review (1996-2016).

Richard M. Frankel, Sarah McVay, and Mark T. Soliman. 2004. Street Earnings and Board Independence.

John R. Graham, Campbell R. Harvey, and Shiva Rajgopal. 2005. The Economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics*, 40(1):3 – 73.

Lúcia Adriana dos Santos Gruginskie and Guilherme Luís Roehe Vaccaro. 2018. Lawsuit Lead Time Prediction: Comparison of Data Mining Techniques Based on Categorical Response Variable. *PLOS ONE*, 13(6):1 – 26.

Elaine Henry. 2008. Are Investors Influenced By How Earnings Press Releases Are Written? *The Journal of Business Communication (1973)*, 45(4):363–407.

Minqing Hu and Bing Liu. 2004. Mining Opinion Features in Customer Reviews. In *AAAI*, volume 4, pages 755–760.

Taeyoung Kang, Do-Hyung Park, and Ingoo Han. 2018. Beyond the Numbers: The Effect of 10-K Tone on Firms' Performance Predictions Using Text Analytics. *Telematics and Informatics*, 35(2):370 – 381.

Feng Li. 2016. Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?

Barbara A. Lougee and Carol A. Marquardt. 2004. Earnings Informativeness and Strategic Disclosure: An Empirical Examination of "Pro Forma" Earnings. *The Accounting Review*, 79(3):769–795.

Tim Loughran and Bill McDonald. 2011a. Barron's Red Flags: Do They Actually Work? *Journal of Behavioral Finance*, 12(2):90–97.

Tim Loughran and Bill McDonald. 2011b. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.

Tim Loughran and Bill McDonald. 2018. Documentation for the LoughranMcDonald Master Dictionary.

Manning, Christopher D. and Schütze, Hinrich. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press.

Jonathan L. Rogers and Andrew Van Buskirk. 2009. Shareholder Litigation and Changes in Disclosure Behavior. *Journal of Accounting and Economics*, 47(1):136 – 156. Accounting Research on Issues of Contemporary Interest.

Jonathan L. Rogers, Andrew Van Buskirk, and Sarah L. C. Zechman. 2011. Disclosure Tone and Shareholder Litigation. *The Accounting Review*, 86(6):2155–2183.

Sugata Roychowdhury. 2006. Earnings Management Through Real Activities Manipulation. *Journal of Accounting and Economics*, 42(3):335 – 370.

U.S. Securities and Exchange Commission. 2018. Non-GAAP Financial Measures.

P. Wongchaisuwat, D. Klabjan, and J. O. McGinnis. 2017. Predicting Litigation Likelihood and Time to Litigation for Patents. In *Proceedings of the 16th Edition of the International Conference on Artical Intelligence and Law*, ICAIL '17, page 257–260, New York, NY, USA. Association for Computing Machinery.

Steven Young. 2014. The Drivers, Consequences and Policy Implications of Non-GAAP Earnings Reporting. *Accounting and Business Research*, 44(4):444–465.