Using Extractive Lexicon-based Sentiment Analysis to Enhance Understanding of the Impact of Non-GAAP Measures in Financial Reporting

Stacey Taylor and Vlado Keselj

Dalhousie University, Halifax, Nova Scotia
stacey.taylor@dal.ca and vlado@cs.dal.ca,

Abstract

For some time, there has been significant disagreement as to whether financial measures that do not conform to the Generally Accepted Accounting Principles (GAAP) should be used in communication to stakeholders, as research points to these measures being used both altruistically and opportunistically. In this paper, we present a novel approach of using Sentiment Analysis to measure the impact that non-GAAP measures have on financial communication. We use an extractive approach in conjunction with the sentiment of four well known and robustly established dictionaries: General Inquirer, QDAP, Henry and Loughran-McDonald. We find that the sentiment declines once the non-GAAP measures are extracted with a statistical significance at the $p = 0.01$ level. We believe that this enhances NLP-based investment management and also has important implications for *Know Your Customer* (KYC) and text-based market provisioning.

1 Introduction

Each year, public companies are required to submit regulatory filings to the U.S. Securities and Exchange Commission (SEC) that provide information on the financial and operational health of their company. While the SEC has provided rules regarding information that must be disclosed under the *Generally Accepted Accounting Principles* (GAAP) [U.S. Securities and Exchange Commission, 2017], anything beyond that is at the discretion of the company. This leaves the company open to discuss financial measures that do not adhere to the Genereally Accepted Accounting Principles (GAAP), which means that these measures are not regulated in how they must be calculated, and are therefore not auditable. These *non-GAAP* measures are ubiquitous in the financial world, have been, and continue to be, a major source of disagreement as prior research has shown these measures to be used in beneficial and predatory ways.

There are two main beneficiaries of this research: investors who are not considered finance professionals (which we term the *average investor*) and companies who prepare the financial filings. For clarity, we define the term *finance profes-sional* in this research to broadly include professional investors, investment and financial analysts, and accountants. We also include those with no formal financial training but who have significant experience in finance and the market, as we recognize that experience and knowledge can be commensurate with training in certain cases.

Given this dispute, we designed an experiment to quantitatively determine the effects that the unregulated non-GAAP measures have on financial reports filed with the SEC. We draw on the well established and robustly proven lexica of General Inquirer, QDAP, Henry, and Loughran-McDonald, using the first two as proxies for average investors, and the latter two as proxies for financially savvy investors. We demonstrate that when non-GAAP measure sentences are extracted, the aggregate sentiment of our sample decreases with statistical significance at the $p = 0.01$ level.

To the best of our knowledge, this *extractive* approach has not been used before in sentiment analysis of financial reports. We see this research as an important step to learning how to better protect the average investor from making poor decision based on measures that can easily obfuscate the information presented. We believe that this will contribute to and enhance NLP-based investment management, as well as have important implications for *Know Your Customer* (KYC) and text-based market provisioning.

The rest of our paper is organized in the following manner: Section 2 provides a discussion of the related work; Section 3 addresses our research design, hypotheses, lexical dictionaries used, as well as the distribution of the data; Section 4 discusses our results and why this research is important; and in Section 5, we provide our conclusion and present some directions for future work.

2 Related Work

Even though non-GAAP measures are not regulated, they have become mainstays of the financial narrative used by companies when communicating to stakeholders. One overarching concern is that by further adjusting audited figures, a company asserts that actual performance differed from audited performance — which in some cases can create an unaudited gain out of an audited loss. Consequently, researchers have determined that non-GAAP measures are influential, deceptive, and can fool the average investor [Young, 2014; Fisher, 2016; Asay *et al.*, 2018]. Written corporate

communication is crafted carefully and purposefully in terms of what information is provided or omitted. It is also designed to evoke emotional responses, and guide decision-making — the effects of which the decision-maker, themselves, may be unaware of. Research has also found that companies employ significant latitude in tone to mitigate bad news by re-framing it in a positive light [Kang *et al.*, 2018; Li, 2006; Loughran and Mcdonald, 2011].

Most Natural Language Processing (NLP) approaches to finance have been sentiment analysis or other forms of text categorization, based on the use of dictionaries (lexica) — developed word lists of positive, negative, and neutral words, as well as other categories such as uncertainty, litigious, and modal [Henry, 2008; Loughran and Mcdonald, 2011].

Kang *et al.* studied the relationship between firm performance and the tone of the 10-K (SEC filing). One of the foci of their research was determining if there was an "over-tone" (inflated positivity) or an "undertone" (less robust positivity) in the text. They used the ordinary least squares regression model and a firm cluster-robust regression model. Their results showed a correlation between sentiment and performance, and that companies that overstate positivity in their financial narratives are less able to deliver on the company's expected future performance. The study also found that investors either do not understand or struggle to fully comprehend the underlying overtone and its true meaning [Kang *et al.*, 2018].

Butler and Keselj evaluated how generating readability indices for corporate annual reports can be used to make class predictions. Using Perl, they created three well known readability indicies for each annual report: Flesch, Flesch-Kinkaid, and Gunning-Fog. Five features were used in the classification — the three readability indices and two financial performance measures from the preceding year. This data was scaled and transformed in order to be used with Support Vector Machines. Results show that their model outperformed previous n-gram techniques and portfolio benchmarks (S&P index), thereby creating more efficient trades. Their approach also offered textual insight related to a company's forecasted performance [Butler and Keselj, 2009].

Jegadeesh *et al.* identified that previous research has considered positive and negative words equal in weighting. To address this, they used the market's reaction to corporate annual reports to determine the weighting that was assigned to each word in an effort to provide a more realistic weighting for each word. They believed this approach provided a much more accurate sentiment evaluation [Jegadeesh and Wu, 2013]. Finally, Sarderlich *et al.'s* work focused on building a novel financial lexcion based on Yahoo Message Stock Boards to determine new weightings for financial terms. They found a strong bias towards positive words — either due to wishful thinking or overconfidence on the part of participants on the message board. Using a sparse vector space model which considers each term in a separate dimension, they developed a "bag of semantic orientation" model that is specific to market terminology (long, short, put call, etc.). In taking this approach, they were able to extend existing lexica and capture both the formal and informal language used in stock trading to better classify documents [Sarderlich and Kazakov,

2018].

3 Research Design

The purpose of this research is to quantitatively measure the effect that non-GAAP measures have on the tone of the Management Discussion and Analysis and Market Risks sections of the annual 10-K and quarterly 10-Q reports filed with the SEC. Our sample dataset comprised 100 randomly selected 10-K and 10-Q reports from each quarter from 1998 to 2019. We drew these samples from the dataset maintained by Bill McDonald [McDonald, 2019]. This gave us a dataset of 10,000 SEC filings.

We followed the main ideas for text pre-processing, extraction, and sentiment analysis from a related *forthcoming* paper [Taylor and Keselj, 2020]. Although the reporting to the SEC is standardized, the format and naming conventions are not. Therefore, parsing out the Management Discussion and Analysis proved to be a significant challenge as it is listed in the Table of Contents, could be listed at the top of each page of the report that contained parts of that section, or could go by other names such as *Financial Review and Analysis*, *Business Outlook*, or *Management's Financial Discusison*, for example. To address this challenge, we used Python's *file read backwards* package to ensure that the first time Python encountered Management's Discussion and the Market Risks would be the actual section itself, rather than a page header or listing in the Table of Contents.
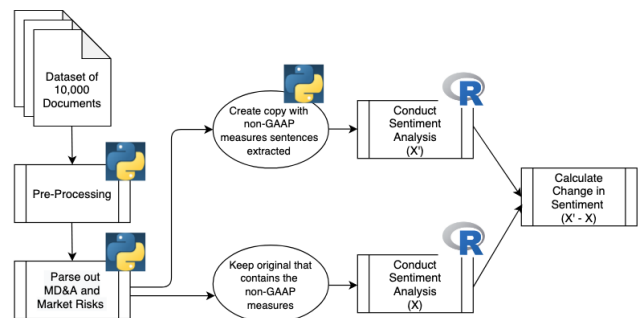

Figure 1: Sentiment Experiment Setup

Sentences that contained any of the following non-GAAP measures were then extracted, using a Python script:

- 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: Commonly accepted name variations of these measures, as well as "Adjusted" or "Revised" variations were also included, along with the term "Reconciliation", which is required when companies use non-GAAP measures.

3.1 Lexical Dictionaries and Sentiment Analysis

We used four dictionaries for our sentiment analysis, each providing scores that range from -1 to 1. The first two dictionaries used, the General Inquirer and the QDAP, are all-purpose dictionaries that are not targeted towards any specific domain. As such, we believe that these act as good proxies for the average investor. The remaining two dictionaries, Henry and Loughran-McDonald, are specifically targeted to the domain of finance, and as such, are good proxies for the financially savvy. The change in sentiment between the reports containing the non-GAAP measures and those without was then calculated as [X′ - X].

The Harvard-IV General Inquirer is a general psychological dictionary. Financial words such as loans and taxes are considered negative [Zimmerman, 1987] We believe that this is a reasonable proxy for the average investor as they, too, would interpret words such as loans and taxes as negative terms. The other general purpose dictionary that, for similar reasons, we believe is a reasonable proxy for the average investor is QDAP. This dictionary has some degree of overlap with the Harvard dictionary as it includes a subset of the Harvard-IV, but also includes words that target opinion mining, government data, and words by reading level [Rinker, 2018].

We selected two well known financial dictionaries, Henry and Loughran-McDonald, as representative of the financially savvy, who have either significant experience in the market/finance, or who have financial training. The Henry dictionary is very small in comparison to Loughran-McDonald, however, but focuses on descriptive words such as "deteriorate" (negative) or "improved" (positive) to characterize the financial terms [Henry, 2008]. This approach provides a robust bridge between a highly financially oriented dictionary and one that is general purpose. Conversely, Loughran-McDonald's dictionary is quite large and is continually being adjusted to keep up with the evolution and dynamism of language. Words such as loans and taxes are assigned a sentiment of 0 in this dictionary [Loughran and Mcdonald, 2011], as contextualization is needed to determine if these words are negative or positive. As such, we believe that this fairly represents the analytical approach that finance professionals would take. As Henry and Loughran-McDonald each take different approaches to their financial lexica, there is no overlap between the dictionaries.

3.2 Data Distribution

Before conducting any statistical tests, we looked at the histograms to determine if our dataset was parametric or non-

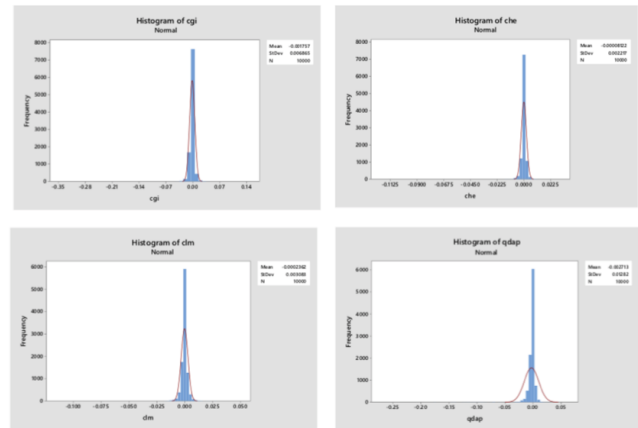parametric, as the result would dictate the statistical testing that could be used.

Figure 2: Dataset Histograms for Each Dictionary

As the histograms in Figure 2 showed normal distribution of the data for each dictionary, we were able to use a paired t-test to evaluate the statistical significance for each dictionary.

3.3 Hypotheses

We developed two research hypotheses to examine the effect of the non-GAAP measures:

Hypothesis 1 *Overall Aggregated Tone*

There is a significant body of existing research that supports the perspective that non-GAAP measures are used opportunistically in order to positively impact the tone of the Management Discussion and Analysis and Market Risks in order to influence investor decision-making. If the non-GAAP measures were not as positively influential as researchers have found, we would expect the tone change, calculated as [X′ - X], to either be zero (or close to it) or to increase once the sentences containing the non-GAAP measures were removed. We examine this hypothesis on a dictionary-by-dictionary basis, using Harvard-IV and QDAP as proxies for the average investor (without financial training) and Henry and Loughran-McDonald as proxies for those with financial training or significant investment training. Taking this approach allows us to capture how the two different groups of investors will interpret the sentiment of the non-GAAP measures, giving us quantitative insight on how influential (or not) these measures are.

Therefore, when the tone changes for each dictionary have been aggregated for all 10,000 reports, we postulate that the aggregated tone will decrease:

Null Hypothesis: The aggregate tone of the dictionary under evaluation is ≥ 0.
Alternative Hypothesis: The aggregate tone of the dictionary under evaluation is < 0.

Hypothesis 2 *Statistical Significance*

Another aspect to our main research question of quantifying the effects of non-GAAP measures is to determine if the

results in our *Aggregated Tone* hypothesis are explainable by chance alone. As we are using two related samples, one with non-GAAP measures and one without, we use a paired t-test to examine the paired observations. If the probability results from the paired t-test are greater than $\alpha = 0.05$, then any differences observed could be explained by chance. If they are equal to or less than $\alpha = 0.05$, then the differences are not from chance alone, and we can, therefore, infer with statistical significance that the difference is a result of removing the non-GAAP measure sentences.

Therefore, we postulate that the aggregated changes in the mean for each dictionary will be less than 0, and will be statistically significant at $\alpha = 0.05$.

Null Hypothesis: After extraction, the mean (μ) of the tone change for the dictionary under evaluation $= 0$.

Alternative Hypothesis: After extraction, the mean (μ) of the tone change for the dictionary under evaluation < 0.

Note: In conducting these tests, we used a 95% confidence interval to evaluate our hypothesis.

4 Results and Why This Is Important

Hypothesis 1: Figure 3 below provides the aggregate totals of the tone change for each dictionary for 10,000 documents over 100 experiments. As can be seen, the *aggregate* tone change for each dictionary is negative, meaning that the sentiment decreased in tone once the non-GAAP measures (and the supporting words) were extracted. The most pronounced negative results are for the two dictionaries that were used as proxies for the average investor. The results show that once the non-GAAP measure sentences have been extracted from the text, the aggregate sentiment score for General Inquirer and QDAP have dropped sharply, as the change in the sentiment scores are -17.57297 and -27.13332 respectively. We can therefore infer that, from the point of view of the average investor, that the text including the non-GAAP measures is much more positive (and therefore influential) than the text that does not include the non-GAAP measures. This also demonstrates that average investors are very sensitive to financially oriented words that are used in conjunction with the discussions of the non-GAAP measures.

It is also notable that change in the sentiment scores for the financially oriented dictionaries of Henry and Loughran-McDonald also show that once the non-GAAP sentences have been extracted, the change in the tone has dropped by -0.81217 and -2.36182 respectively. While this is not as sharp a decrease as for the general purpose dictionaries, it is a decrease nonetheless. These results indicate that even the financially oriented dictionaries recognize that there is inflation of positivity in the text when the non-GAAP measures are included in the text. These results also strongly suggest that savvy investors are not as influenced by non-GAAP measures as average investors.

The results of the Henry dictionary is barely negative which may raise questions as to if the inflationary assertion still holds for the dictionary; we believe it does. The Henry dictionary's focus is on descriptive words that are used in finance such as "growth", "opportunity", "declining", and "deteriorated" [Feuerriegel and Proellochs, 2019], not on the fi-

Experiment Number	Report Quarter	cgi	cqdap	che	clm
100	Q3-18	-0.0101182	-0.0074967	-0.000153	0.002228726
100	Q3-18	-0.00148	0.00104623	-2.52E-05	-0.00411334
100	Q3-18	0.00401739	0.00011235	0.0005106	0.000758538
100	Q3-18	-0.0025215	0.00049252	-0.00172	0.002472894
100	Q3-18	0.00270776	0.00243339	-0.001678	0.000373329
100	Q3-18	0.00246886	-5.053E-05	0.000188	-0.00127438
100	Q3-18	-0.0048782	-0.0050087	-0.000422	0.000396483
100	Q3-18	-0.0080319	-0.0061931	0.000152	0.001665008
AGGREGATE TOTALS		**-17.57297**	**-27.13332**	**-0.81217**	**-2.36182**

change = X' - X, where:
cgi = change in tone for the General Inquirer;
cqdap = change in tone for the QDAP dictionaries;
che = change in tone for Henry; and
clm = change in tone for Loughran-McDonald

Figure 3: Aggregate Sentiment Results

nancial words themselves such as "debt" or "interest". Based on the evidence of the experiments, these descriptive words have been used as supporting words for non-GAAP measures. We can also infer that, based on the results, that sufficient positive descriptive words have been used with the non-GAAP measures that, when removed, have returned an overall decrease in the sentiment, thereby reinforcing that the inflationary assertion still holds.

We also looked at the distribution of the non-GAAP measures over the 100 experiments performed. We first looked at the distribution of the first half of the dataset, up to the 4th quarter of 2005. As can be seen in Figure 4 below, the three most prevalent non-GAAP measures are Earnings Before Interest, Tax, Depreciation, and Amortization (EBITDA), Earnings Before Interest and Tax (EBIT), and Free Cash Flow (FCF).
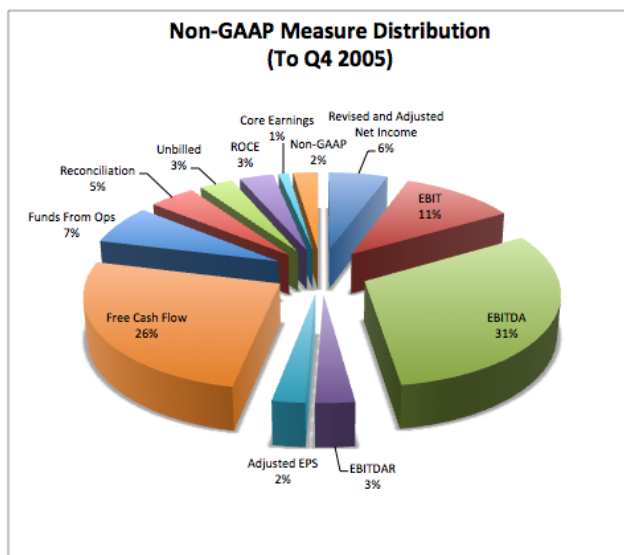

Figure 4: Sentiment Results for the First Half of the Experiments

Over time, however, we see that the use of non-GAAP measures is growing, but the distribution is changing. When

we compare the midway results with the overall results (Figure 5), we find that while Earnings before Interest, Tax, Depreciation and Amortization (EBITDA), Earnings before Interest and Tax (EBIT), and Free Cash Flow (FCF) are still the three main non-GAAP measures used, the percentages for EBITDA and Free Cash Flow have decreased by 6% and 4% respectively, while EBIT has grown by 8%, seen in Figure 5, below.
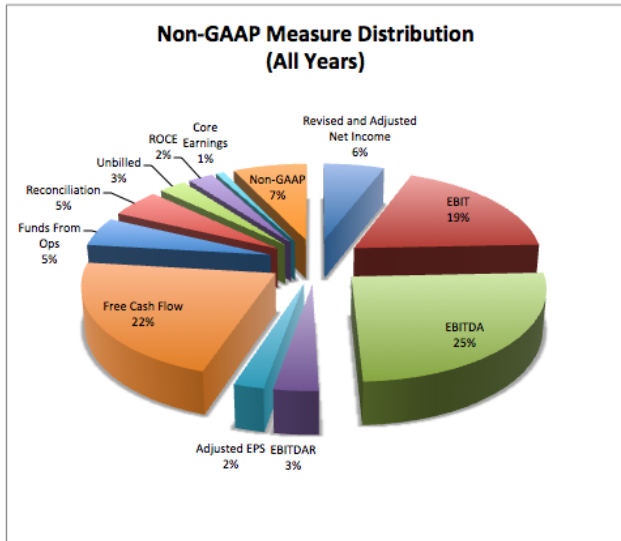

Figure 5: Sentiment Results for All Experiments

The increase in EBIT and simultaneous decrease in EBITDA suggests that companies are changing their communication strategy. In recent years, the SEC has scrutinized the use of EBITDA as companies were including extra adjustments (beyond just interest, tax, depreciation and amortization). Examples include "Further Adjusted EBITDA" or "Structring Adjusted EBITDA" [Scraggs and Powell, 2018]. Therefore, using EBIT instead draws far less attention to the company's reporting than does EBITDA.

With respect to FCF, the SEC has warned companies about using this measure, as it can be very misleading. The word "Free", for example, has a tremendous effect on the average investor, and is seen as a positive word in both General Inquirer and QDAP. The Henry and Loughran-McDonald dictionaries show no effect, as it requires contextualization in order to determine if "Free" is positive or negative. So, we can infer from the drop in the use of FCF that companies are, again, changing their communication strategy in order to draw less regulatory attention.

There are several alternative plausible reasons for the change in the use of these financial measures, however, that should be discussed. The change could be driven by the companies that were included in the random sample. If more capital and intangible intensive companies were included in experiments 50-100, then those types of companies will prefer to use EBIT as it is a better proxy for cash flow. The SEC also requires that companies compare and reconcile the non-GAAP measure with the closest GAAP measure. EBIT is

usually compared to the GAAP measure of Net Income, as the reconciliation is straightforward, only needing to show the difference of interest and taxes. Depending on the adjustments a company makes to EBITDA or FCF, though, it can be harder to find a GAAP measure for comparability. So this too could explain the increase in using EBIT as a non-GAAP measure.

PricewaterhouseCoopers LLP (PWC) (2019) has indicated that there has been a substantial increase in the usage of non-GAAP measures when comparing today's reporting with that of twenty years ago. PWC also indicates that nearly all of the companies listed on the Standard & Poor 500 (better known as the S&P 500) use at least one non-GAAP measure. This is consistent with the change that is seen between Figures 4 and 5.

Hypothesis 2:Using the same four dictionaries, we tested the statistical significance using a paired t-test, given that the distribution of the data for each dictionary was normal. We had hypothesized that the change in the mean of each dictionary, when we considered [X′ - X], that the change would be negative for each dictionary. As seen below in Table 1, the results for each dictionary were determined to be statistically significant at the 0.01 level, meaning that there is a 1% risk that we could incorrectly conclude that there is a difference where none exists.

Dictionary	Number of Samples	Mean	Std Deviation	T-Value	P-value
GI	10,000	-0.001757	0.006865	-25.6	$<0.001*$
QDAP	10,000	-0.00272	0.012801	-21.25	$<0.001*$
HE	10,000	-0.000081	0.002217	-3.66	$<0.001*$
LM	10,000	-0.000236	0.003083	-7.66	$<0.001*$

Table 1: Paired T-Test Results

We see that over 10,000 samples, that all of the dictionaries are statistically significant. This draws attention to the importance of language. As we have extracted both the non-GAAP measures as well as the supporting words in the sentence, we see that the non-GAAP measures are having a pronounced effect for both the non-financial and the financial dictionaries, which act as proxies for the two different types of investors we identified. This is an important finding given that regardless of motive for use, there is a quantifiable effect.

This could have tremndous ramifications on NLP-based investment management, touching on all aspects ranging from due diligence to portfolio selection and maintenance, to client reporting, as prominent companies look to natural language processing (NLP) to aid in these tasks [Xy, 2019; Deloitte, nd]. Training data including non-GAAP measures without proper contextualization or understanding of the sentiment could affect the way that the system functions, which could also affect the way that the system is tested and ultimately evaluated [Bender and Friedman, 2019]

4.1 Inter-Domain Integration

Although our paper was focused on NLP-based investment management, in the finance and business domains, non-GAAP measures are ubiquitous. This creates opportunities for our approach to be applied and integrated into different, but highly related streams of the FinTech. The first stream

that we believe would benefit from our approach is text-based market provisioning. According to the World Economic Forum (WEF), key disruptive trends centre around Artificial Intelligence, Big Data, and Machine Learning [World Economic Forum and Deloitte, 2015]. Using non-GAAP extraction as we have described can help provide better due diligence on companies, which could improve algorithms used to gain insights into the market, as well as those used for processing machine-readable news feeds [World Economic Forum and Deloitte, 2015]. We have also addressed that market participants fall into two main categories — those with financial expertise and those without. Our extractive technique can also be used in for reducing risk as part of a *Know Your Customer (KYC)* approach. One important aspect to the creation and maintenance of an investment portfolio is risk tolerance. Research has shown that risk tolerance is affected by financial literacy [Caratelli and Ricci, 2011; Gentile *et al.*, 2016; Kramer, 2016], which our paper helps to reinforce. Better understanding the influence of non-GAAP measures on investors' perceptions will help investment managers better meet the needs of clients who lack sufficient financial literacy, as well as help to avoid the inclusion of securities with a high chance of facing a class action lawsuit, thereby reducing risk to the client.

5 Conclusion

In this paper, we have presented a novel use of sentiment analysis that extracts non-GAAP measure sentences in order to quantify the effect that non-rule based accounting measures have on financial reporting in the Management Discussion & Analysis and Market Risks section of the 10-K and 10-Q reports filed with the SEC. We found that once the non-GAAP measure sentences have been removed from our sample, the sentiment declines with a statistical significance at the $p = 0.01$ level. We believe that this enhances NLP-based investment management and also has important implications for Know Your Customer (KYC) and text-based market provisioning.

5.1 Future Work

The approach that we have described has opened up new avenues of research, particularly in the areas of Know Your Customer (KYC) and Text-Based Market Provisioning. We see applying our method to those areas a natural *next step* for our research. Also, as we only applied this to the 10-K and 10-Q filings submitted to the U.S. SEC, we believe that extending this approach to financial filings to regulatory bodies (similar to the SEC) in other countries would be valuable.

References

[Asay *et al.*, 2018] H. Scott Asay, Robert Libby, and Kristina Rennekamp. Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, 65(2–3):380–398, 2018.

[Bender and Friedman, 2019] E. Bender and B Friedman. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 2019.

[Butler and Keselj, 2009] Matthew Butler and Vlado Keselj. Financial forecasting using character n-gram analysis and readability scores of annual reports. In *Proceedings of Canadian AI'2009*, Kelowna, BC, Canada, May 2009.

[Caratelli and Ricci, 2011] M. Caratelli and O. Ricci. *The Relationship between Everyday Practices and Financial Literacy: An Empirical Analysis*, 2011.

[Deloitte, nd] Deloitte. *Artificial intelligence — The next frontier for investment management firms*, n.d.

[Feuerriegel and Proellochs, 2019] Stefan Feuerriegel and Nicolas Proellochs. *Package SentimentAnalysis*, 2019.

[Fisher, 2016] Alex Fisher. *Non-GAAP Measures — A 20-Year Echo*, 2016.

[Gentile *et al.*, 2016] M. Gentile, N. Linciano, and P. Soccorso. *Financial advice seeking, financial knowledge and overconfidence*, 2016.

[Henry, 2008] E. Henry. Are investors influenced by how earnings press releases are written? *The Journal of Business Communication*, 45:363–407, 2008.

[Jegadeesh and Wu, 2013] N. Jegadeesh and D. Wu. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110:712–729, 2013.

[Kang *et al.*, 2018] Taeyoung Kang, Do-Hyung Park, and Ingoo Han. Beyond the numbers: The effect of 10-k tone on firms' performance predictions using text analytics. *Telematics and Informatics*, 35(2):370–381, 2018.

[Kramer, 2016] M. Kramer. Financial literacy, confidence and financial advice seeking. *Journal of Economic Behavior & Organization*, 2016.

[Li, 2006] Feng Li. *Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?*, 2006.

[Loughran and Mcdonald, 2011] Tim Loughran and Bill Mcdonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

[McDonald, 2019] Bill McDonald. *Loughran-McDonald 10X File Summaries*, 2019.

[Rinker, 2018] Tyler Rinker. *Package qdapDictionaries*, 2018.

[Sarderlich and Kazakov, 2018] M. Sarderlich and D. Kazakov. Extending the loughran and mcdonald financial sentiment words list from 10-k corporate filings using social media texts. *Proceedings of the 11th LREC Conference on Language Resources and Evaluation in Workshop on Financial Narratives*, 2018.

[Scraggs and Powell, 2018] Alexandra Scraggs and Jamie Powell. *Structuring adjusted ebitda now exists*, 2018.

[Taylor and Keselj, 2020] Stacey Taylor and Vlado Keselj. e-commerce and sentiment analysis: Predicting outcomes

of class action lawsuits. In *Proceedings of the Third EC-NLP Workshop (ACL) on Financial Technology and Natural Language Processing*, 2020.

[U.S. Securities and Exchange Commission, 2017] U.S. Securities and Exchange Commission. *Topic 6 — Foreign Private Issuers & Foreign Businesses*, 2017.

[World Economic Forum and Deloitte, 2015] World Economic Forum and Deloitte. *The Future of Financial Services*, 2015.

[Xy, 2019] Jingya Xy. *How NLP Is Advancing Asset Management*, 2019.

[Young, 2014] Steven Young. The drivers, consequences and policy implications of non-GAAP earnings reporting. *Accounting and Business Research*, 44(4):444–465, 2014.

[Zimmerman, 1987] D. Patrick Zimmerman. Effects of computer conferencing on the language use of emotionally disturbed adolescents. *Behavior Research Methods, Instruments, & Computers*, 19(2):224–230, Mar 1987.