**Deriving Stock Market Sentiment with Natural Language Processing:**

**Investigating its Viability in Real World Portfolio Management**

For:

Quantitative Fund Portfolio Manager (name withheld)

Winnipeg, MB

By:

Russel Gerbrandt

University of British Columbia

Vancouver, BC

March 15, 2019

**Table of Contents**

1. **Introduction**
2. Audience
3. Purpose of this report
4. What is natural language processing (NLP)?
5. What is market sentiment?
6. How can NLP be used to derive market sentiment?
7. Description of data sources
8. Scope of this inquiry
9. **Data Section**
   1. Sentiment terms relevant to investing
   2. Term value weightings
   3. Term structure variations
   4. Time frames
   5. Follow-up research requirements
10. **Conclusion**
    1. Summary and overall interpretation of findings
    2. Recommendation
11. **Introduction**
12. *Audience*

This report and its recommendation will be presented to a portfolio manager in charge of quantitative portfolios, who will make investment decisions based on this research. I have worked with him at another firm in the past.

1. *Purpose of this report*

This report will investigate the use of natural language processing in deriving market sentiment. If the process is successful, it would potentially shed light on a company’s managers’ attempts to “sugar coat” the company’s operations being in decline before the majority of investors discover this. As such, stocks which are likely to decline in price can be avoided or sold earlier, thus reducing investment losses.

1. *What is natural language processing (NLP)?*

Natural language processing is the use of computers to process large amounts of text written by humans, creating data from it that can be analysed to find patterns or behaviour that would be difficult or impossible for a single person to find.

1. *What is market sentiment?*

Market sentiment is, in general, the average opinion of investors towards the stock of a company, as implied by its price. Investors pay close attention to quarterly and annual reports from companies, attempting to obtain new information about the company that may influence its stock price. New information that is below previous expectations will result in selling of the stock, which will result in a reduction in its price.

1. *How can NLP be used to derive market sentiment?*

Companies with publicly-traded stock are required to release financial reports both quarterly and annually that reflect the company’s operations. Investors analyse these reports carefully to gain any new information about the company’s operations they can, if they believe it will affect the stock price. NLP could potentially both hasten this process (analysing these reports more quickly than a person could) and find new patterns (comparing new reports to old reports to find changes a person may not see).

1. *Description of data sources*

There has been academic research into how this strategy may be implemented and how successful it could be. Two such academic papers were researched for this report: (need to determine if I can include their names, as the authors are the ones being interviewed, and there may be confidentiality issues)

1. *Scope of this inquiry*

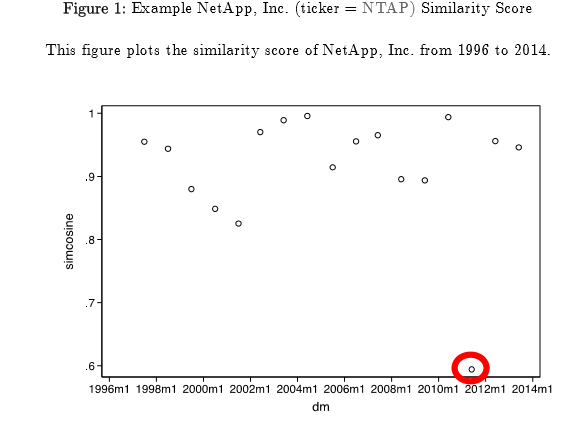
To assess the feasibility of extracting market sentiment through natural language processing, I intend to answer the following questions:

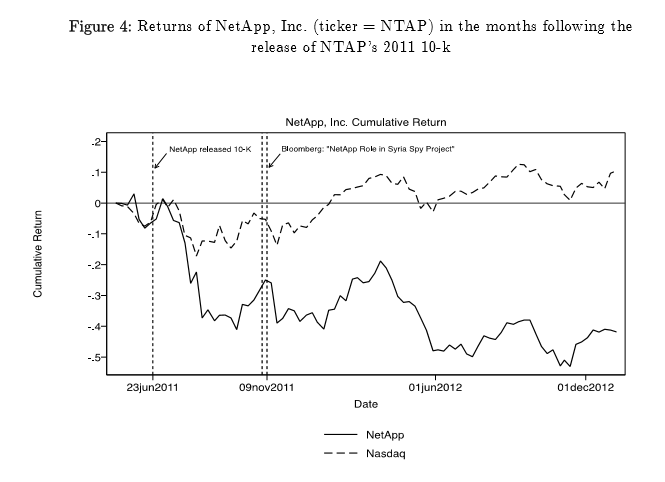
1. What, if any, are the real-world examples of this type of strategy being implemented, and what have their returns been?
2. How were the lists of financially-relevant terms derived? Is it feasible for us to reverse-engineer our own lists, and if so, what is the best method?
3. The research shows positive return over very specific time frames; what about other time frames? Were these time frames chosen before testing, or are they the best examples from a wider swathe of time frames, and if so, what were the results for the omitted time frames?
4. Was any work done on reweighting the value of terms to determine those most impactful on returns, irrespective of frequency (i.e., using machine learning to find the most historically relevant terms)?
5. Were only individual words considered, or phrases as well?
6. Did they look at omitted or reduced sections year-over-year?
7. Did they look at expanded sections (which may be due to a business expansion)?
8. If this strategy becomes widely implemented, what is the projected change to management commentary? Will attempts to conceal or diminish bad news expand?
9. **Data Section**

This section requires feedback from interviewees. The questions have been sent, and once responses are received, this section can be completed.

1. *Sentiment terms relevant to investing*
2. *Term value weightings*
3. *Term structure variations*
4. *Time frames*
5. *Follow-up research requirements*

Note: the following two diagrams are from one of the research papers. I will need to determine any issues with confidentiality before including these, and their sources, in the final report. These diagrams illustrate the effectiveness of the strategy with an example: the first illustrates that NetApp’s November 2011 10-K showed a marked decrease in similarity, and the second illustrates that this was followed by a significant and sustained decrease to the company’s stock price.





1. **Conclusion**

As in the previous section, this one can only be completed after interviewee responses are received.

1. *Summary and overall interpretation of findings*
2. *Recommendation*