Sentiment Analysis

A Vital Tool for Managing the Reputational Risk of Vendors and Clients

- *Ernest D Subah*

Table of Contents

Current State: *A Living Nightmare*

Below are a few words and phrases to describe the current state of sentiment analysis in major corporations

- Non-existent in many organizations
- Inefficient
- Disorganized
- Slow
- Never Current
- Siloed and not integrated
- Waste of Time
- Waste of resources
- Physically and mentally draining
- Extremely costly
- Mis-use and mis-allocation of top skills and abilities

Current State: *Process, Cost and Time*

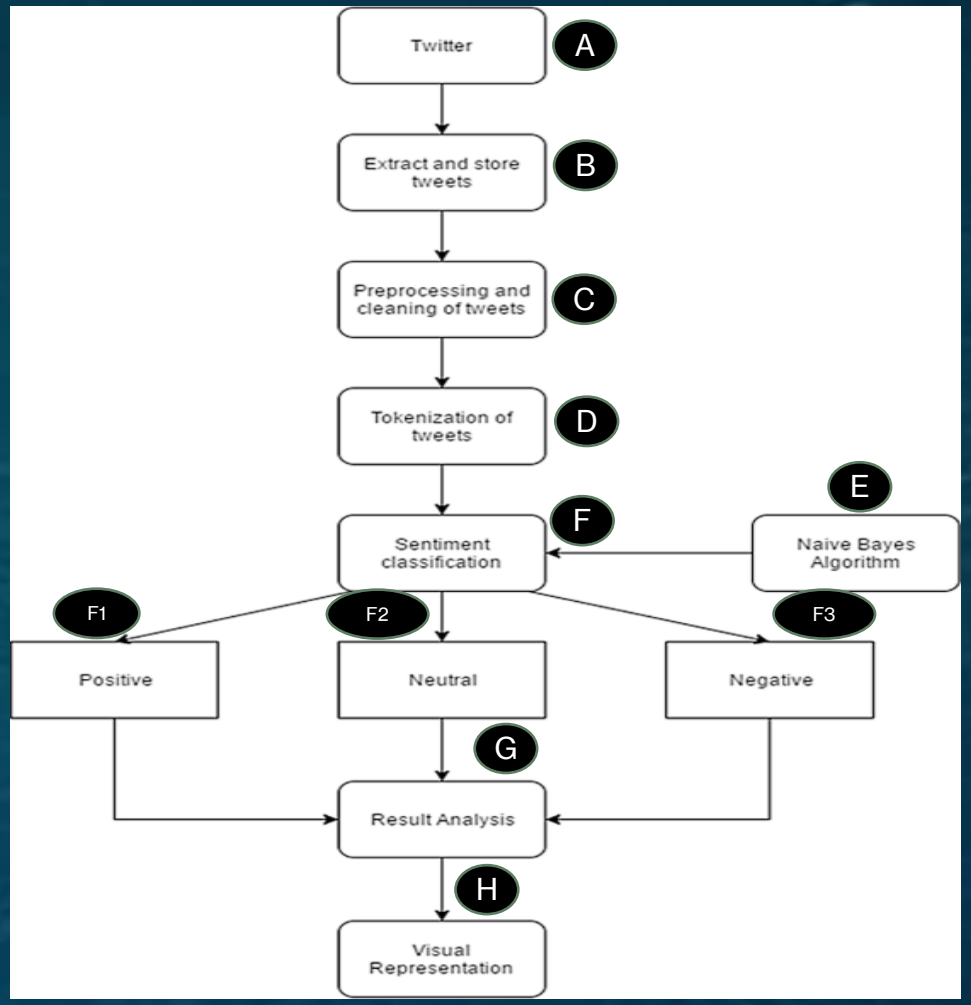
- A team of three employees is tasked with performing sentiment reviews of vendors and clients during onboarding and ongoing monitoring
- The typical process is to search the internet and other social media sites for news stories and other public information on these vendors/clients
- Upon reading, they would have to make a determination on the relevance of the information and write a report that takes into consideration the holistic view of these articles
- At some point, these reports are escalated to a manager for review and decision making
- Assuming the average salary of these three employees is $100,000 per year, and the manager is $150,000 a year, $450,000 combined
- Assuming 100% of reviews team's time is spent performing these reviews and 30% ($45,000) of the manager's time is spent on reviewing reports and making decisions (Total cost: $345,000)
- Total employee hours:
  - Team (3): 8 hrs/day x 5 days/week x 52 weeks/year = 2080 Hours/Year x 3 = 6240
  - Manager (1): 8 hrs/day x 5 days/week x 52 weeks/year = 2080 Hours/Year x 30% = 624
  - Total Hours = 6240 + 624 = 6864
- This is draining, overwhelming and a total mis-management of skilled resources

Regulatory Overview and Best Practices: *OCC 2013-29*

- OCC 2013-29 is arguably the most prescription guidance on Vendor and Third Party Risk Management in the United States
- OCC is an acronym for Office of the Comptroller of the Currency which is part of the US Department of Treasury
- Not every financial service organization is regulated by the OCC, however, other organizations have adopted this guidance as a best practice to help mature their programs
- The guidance identifies several risk domains, including Reputational Risk which is being discussed in this presentation. Here's an excerpt from the guidance on Reputational risk:
    - *"Third-party relationships that do not meet the expectations of the bank's customers expose the bank to reputation risk. Poor service, frequent or prolonged service disruptions, significant or repetitive security lapses, inappropriate sales recommendations, and violations of consumer law and other law can result in litigation, loss of business to the bank, or negative perceptions in the marketplace. Publicity about adverse events surrounding the third parties also may increase the bank's reputation risk."*
- The following slides will provide a rationale why Sentiment Analysis is a vital tool for managing the Reputational Risk of Vendors and Clients

Natural Language Processing: *Sentiment Analysis*

- Natural Language Processing is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how computers process and analyze large amounts of natural language data (Wikipedia)
- Sentiment Analysis refers to the use of Natural Language Processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information (Wikipedia)
- Below is Twitter Sentiment Analysis Using a Modified Naïve Bayes Algorithm, developed by Manav Masrani and G. Poornalatha



An explanation with some slight modifications

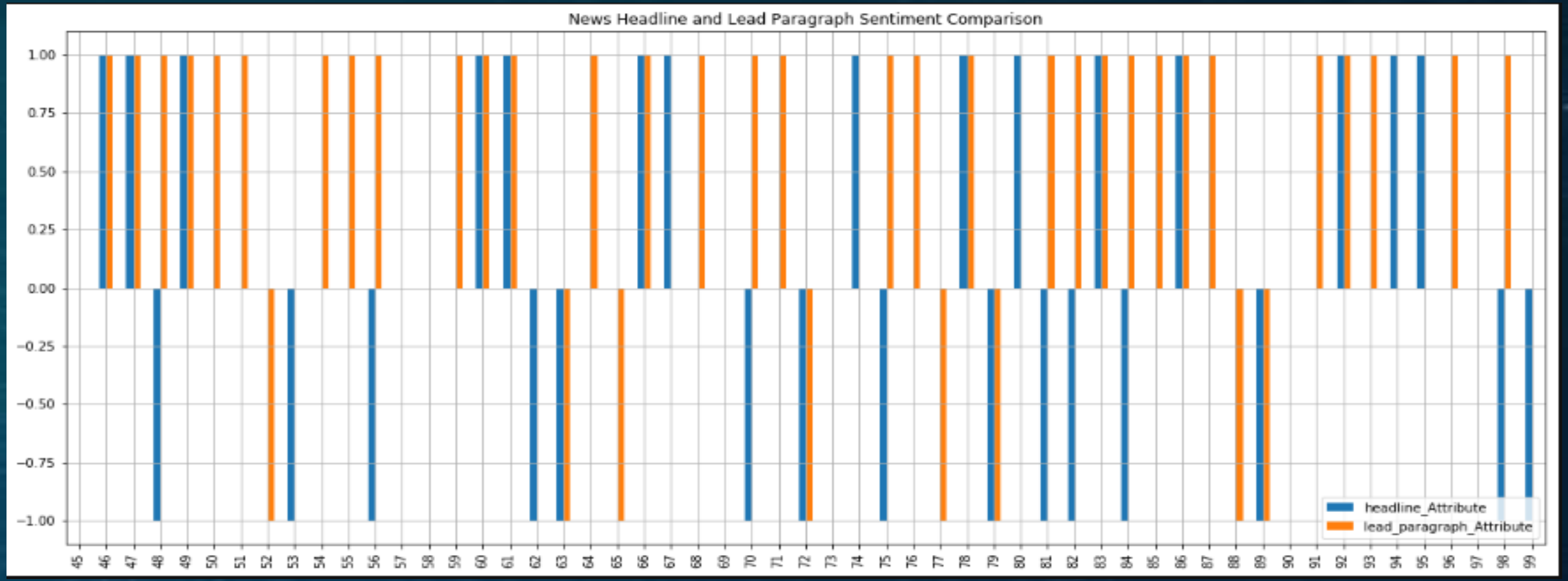
- Data Sources & Platforms (A): Twitter, New York Times, Reddit, Facebook, TripAdvisor...
- Extract & Store Data (B): API and Vendors, Excel & Static Databases, Webscrapping...
- Preprocessing (C + D):
  - Tokenization
  - Lemmatization
  - Stemming
  - Stopwords
  - Concordance
  - Parts of Speech (POS) Tagging
  - Named Entity Reference (NER), etc
- Model (E): Defining the scoring that is applied to sentiment classification
- Sentiment Classification (All Fs): Positive, Neutral, Negative
- Results Analysis (G):
  - Qualitative Analysis
  - Quantitative Analysis
- Refinement (New): loops back one or more of the steps above
- Reporting, Visualization & Decision-Making (H)

NLP Preprocessing: *Some References*

- Preprocessing - *https://towardsdatascience.com/nlp-text-preprocessing-a-practical-guide-and-template-d80874676e79*
- Tokenization - *https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenization-nlp/#:~:text=Tokenization%20is%20a%20common%20task%20in%20Natural%20Language%20Processing%20(NLP).&text=Tokens%20are%20the%20building%20blocks,words%2C%20characters%2C%20or%20subwords*
- *L*emmatization and Stemming - *https://towardsdatascience.com/lemmatization-in-natural-language-processing-nlp-and-machine-learning-a4416f69a7b6*
- Stopwords - *https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dadad47*
- Concordance - *https://avidml.wordpress.com/2017/08/05/natural-language-processing-concordance/*
- Parts of Speech (POS) Tagging - https://nlp.stanford.edu/software/tagger.shtml
- Named Entity Reference (NER) - *https://nlp.stanford.edu/software/CRF-NER.html*
- Much more.....

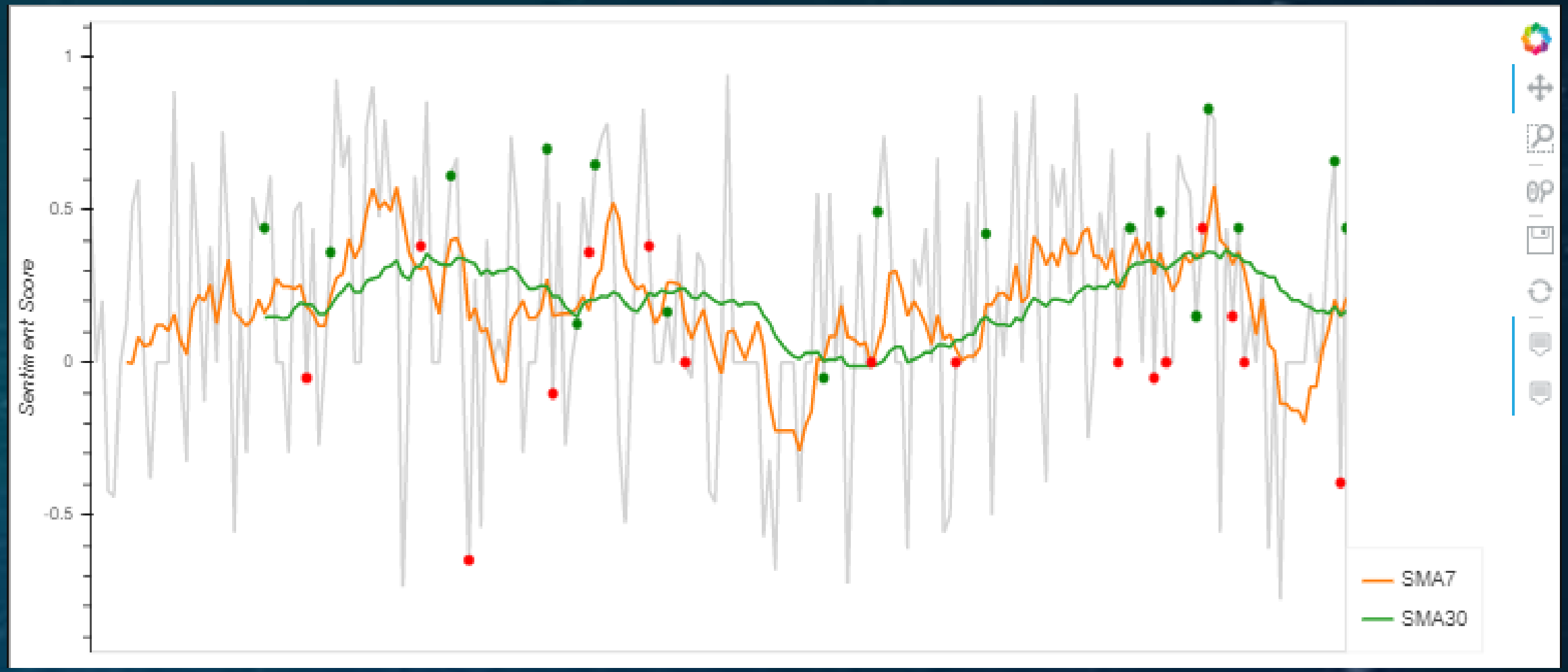
Spotting Click-baits and Positive-spins: *Correlation Analysis*

- News Headline = Orange
- Lead Paragraph/News Story = Blue
- Positive Sentiment = +1
- Negative sentiment = -1
- Click-bait and Positive-spin exist when the news headline and story move in opposite directions



Trend Analysis: *Short-term & Long-term Simple Moving Averages*

- 7-Day Simple Moving Average – Orange
- 30-Day Simple Moving Average – Green
- The Red and Green dots are signals which require action. These signals show the moment when the 7-day SMA is above (Green) or below (Below) the 30-day SMA



Questions and Answers

Questions

pitanja (Bosnian)

Domande (Italian)

Des Questions (French)

Fragen (German)

otázky (Czech)

questão (Portuguese)

vraag (Dutch)

質問 – Shitsumon (Japanese)

ਪ੍ਰਸ਼ਨ Praśana (Punjabi)