TEXT ANALYTICS FOR CORPORATE SOCIAL RESPONSIBILITY

Analysis of Apple, Google, Microsoft and Facebook

Hult Business School
Text Analytics and NLP

Jason Lee
Masters in Business Analytics

Overview

In this report, I will utilize text analytics on the annual 2019 Corporate Social Responsibility (CSR) reports submitted by Apple, Google, and Microsoft. The CSR report represents information that explains how companies are giving back to the community, pushing towards sustainability and other related practices or goals they look to achieve in the direction of being a better company for the people, society, and environment. Each of these companies dominates the technology industry, with their ability to collect data on their customers. However, the tech industry is under scrutiny with data breaches (Google & Facebook) and battery performance scandals (Apple) in recent years.

Tokenization

Tokenizing each companies report gives us a better understanding of the main topic of focus. It is important to note that Microsoft's CSR report is half the length compared to Apple and Google. Below are the top 3 most frequently used words in their respective CSR reports.

Apple	Google	Microsoft
Energy (341)	Energy (332)	Learn (94)
Renewable (197)	Data (165)	AI (63)
Data (176)	Carbon (146)	Human (52)

Apple and Google have very similar top occurring words – Energy and Data. Apple and Google both are large enterprises that have to store and collect user data every minute. The focus topics of *energy* and *data* signals that they are trying to reassure that they are taking strong energy-saving procedures to manage their data intake. With databases and data centers scattered all over the world, it is reassuring to see that they are taking the necessary steps to improve energy consumption. In comparison, we look at Microsoft's top 3 words, and they are more focused on the actual usability of their technology as it relates to *learn, AI ,and human*. These points position Microsoft's report in finding ways to utilize their technology advancements for ethical practices.

Sentiment Analysis

Understanding the tone of the language is critical in determining what direction the company is trying to relate to their audience. Below is a breakdown of the sentiment used in the reports using R's nrc library.

| NRC Library	Apple	Google	Microsoft
Anger	86	65	54
Anticipation	321	367	245
Disgust	129	76	24
Fear	256	103	94
Joy	411	263	226
Negative	295	182	113
Positive	**1730**	**1197**	**945**
Sadness	87	46	43
Surprise	64	76	67
Trust	**960**	**707**	**481**

Historically CSR reports, have communicated positivity and optimism since they are focused on sustainability practices. However, it is interesting to see that the second most popular sentiment

is trust. All three of these companies are leaders in the technology industry, so it seems like they are trying to convince their audience they are authentic and can be reliable. In 2019, the California Data Privacy Law act was passed, which enhanced the privacy rights and consumer protection laws for California residents - this could relate to past data breaches from Google in 2019, and when Apple had to release a statement informing customers of malpractices for battery performance issues in 2018. These companies are using their CSR report to regain trust and stabilize any doubt investors and customers may have felt.

Combining this sentiment analysis with the tokenization, Apple and Google are ensuring they are building a strong foundation around data and trust policies, whereas Microsoft is aiming to build trust around the ethical use of AI and technology opportunities.

Correlogram and Correlation

Below we can see a correlogram that compares Apple against Google and Microsoft. This information will help us understand if there are any trends in the technology industry.
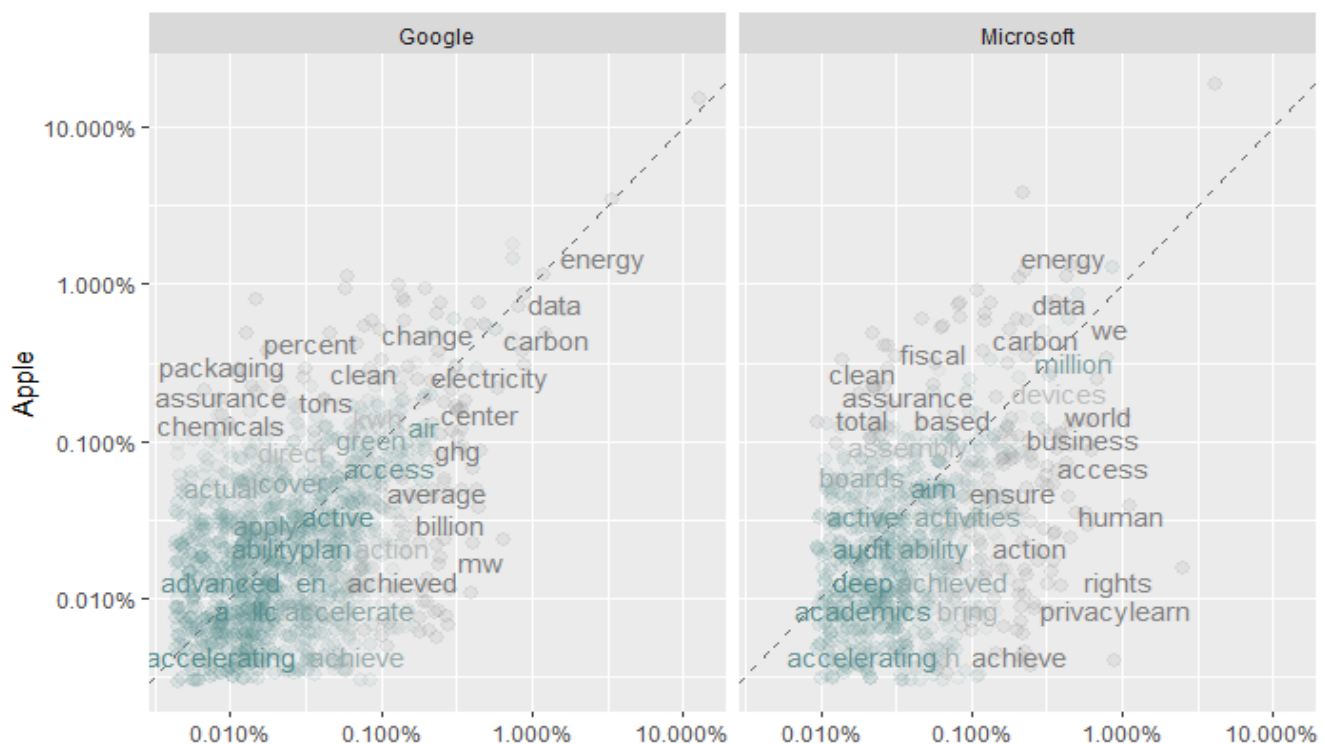
Correlation	Google	Microsoft
Apple	**0.97**	0.66

Microsoft has already made it clear the focus of their report is technology, so it understandable there isn't a strong relationship with Apple. However, looking at this correlogram and the correlation coefficients it is clear that Apple and Google have extremely high correlations 0.97, which means that they are nearly perfectly correlated and have very consistent use of similar language. Since Apple and Google have a high correlation, this provides us with business insights that highlight industry trends. These results mean other high performing tech companies such as Facebook or Twitter could have similar language in their CSR reports, with their focus to also warrant language focused around data privacy and trust amongst their stakeholders.

Word Cloud, NRC Library

Data privacy was a big topic in 2019, and it is important to see if this correlation remains consistent in other leading tech companies. The main company which came under excessive scrutiny for their data privacy laws in 2019 was Facebook. Surprisingly, Facebook does not have a CSR report published for their audience. Since Facebook does not have a CSR report, I extracted text from its sustainability page on their website. Below is a nrc library word cloud analyzing Facebook's website.

Facebook – Word Cloud, NRC Library

Similarly, we can see Facebook's main topics are trust and positivity. Understandably, Facebook is in the process of rebuilding its public relations after its data breach affected 267 million users. They are trying to make amends, and are following Apple and Google in the route of reassuring their customers that data and trust, have a positive direction in the company. Nevertheless, one can be skeptical as to why Facebook did not release an official report since they have felt the bulk of the backlash from government officials last year. Mark Zukerberg could be positioning the company in a way to avoid admitting to the guilt of the data breach, which shows the lack of empathy towards their stakeholders by omitting a CSR report.

Conclusion

From this analysis, it is clear that major companies in the tech industry are looking to rebuild trust with their customers, partners, and investors. With several data management scandals arising in recent years, these giant companies are laying the foundations to make sure they are not met with challenges again by acknowledging that trust and a positive rebuild of the relationship are key to corporate social responsibility. Through text analytics, these results help apply the tokenization, correlogram, and sentiment analysis frameworks, which has given a clear picture of the topic of focus as well as the tone the companies are trying to convey.