MGMT 590 Web Data Analysis Final Project

**Cutting Edge Technologies’ Effect on**

**S&P500 Companies’ Performance and Mutual Funds**

Executive Summary

**Gagandeep Singh Khanuja Haoyuan Zheng Kalyan Mupparaju Ruthwik Kumar Swapnil Rai**

**1 Introduction**

Covering about 80% of American market by capitalization across industries, S&P500 companies are seamless users of cutting edge technologies in their operations. The purpose of our project is to find out if the application of these technologies in such company operations could affect their future performance. Also, we want to see how appearance of such buzzwords could affect mutual fund manager’s decision on picking holdings for their portfolios.

**2 Data Collecting and Processing**

The company list for S&P500 was scraped from the Wikipedia - “List of S&P 500 companies[[1]](#footnote-1)”. We built a buzzword list of 36 words through secondary research and brainstorming. With the CIKs from Wikipedia, we extracted 10-K links and counted the buzzword appearance using ‘**bag of words’** approach. Using Bloomberg Terminal, we collected the companies’ financial data: profit, ROE, stock price (Year Ending fiscal, 2012-2016). We randomly picked approximately 1,000 mutual funds from EDGAR’s 13-F listings and scraped the holding lists and company holding values through the 13Fs. Finally, we got the change of holdings’ value share for 2014 to 2015 and ’15 to 2016.

**3 Analyses**

Post data -collection, we normalized the individual count of words using log transformation to reduce skewness. Next, we ran a multiple regression with each of the 3 financial variables as the dependent variable (%Stock\_Price\_Change, %Profit\_Change and %ROE\_Change) to test the mention of these words’ effect on company’s financial performance. The result was not statistically significant for the chosen buzzwords. We then split our data by year and ran the regression respectively again. This time, we found only the 2016 stock price change has a statistically significant effect for buzzwords. Additionally, we introduce an interaction term with industry and words to measure the impact on price change, due to these words in different industries

For the next step, we prepared a word cloud to visualize the highest occurring buzzwords used by S&P 500 Companies. ‘**AI**’, ‘**Cloud Computing**’, ‘**Internet of Things**’ appeared with highest frequency overall. Using Tableau, we built a tree map to analyze the importance of different buzzwords across industries.

Next, we ran a **chi-square** test (test of independence), to test if the buzzwords are distributed proportionally across Industries. We observed that industries like IT and Telecommunications see higher prevalence of buzzwords as compared to other industries

To estimate the effect of the word occurrences on mutual fund managers’ decision, we ran a similar regression as earlier case, but with the change in holding values (Between 2014-16) as the dependent variables to test the mention of these words’ effect on mutual fund manager’s decisions.

**4 Results & Conclusion**

* Words like cloud computing, optimization and data center are highly prevalent mentions in 10-Ks of S&P 500 companies. We observe a difference in prevalence of these words across industries.
* Financial performance (in terms of Stock Prices) is positively affected by the occurrence of words like “Intelligent Systems” & “Cloud Computing”
  + The effect is further magnified in the sector of Information Technology.
* However, Mutual Fund managers are not currently being affected by these words in their decisions to drop or buy the stocks of these S&P 500 companies.
* Potential opportunity for the managers with access to our web scraping service to pick the stocks that are going to perform well in future.

1. https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies [↑](#footnote-ref-1)