Deal or No Deal: Predicting Mergers and Acquisitions at Scale

Ryan Moriarty
Courant Institute of Mathematical Sciences
New York University
New York, New York
rpm295@nyu.edu

Howard Ly
Courant Institute of Mathematical Sciences
New York University
New York, New York
howard.ly@stern.nyu.edu

Ellie Lan
Bloomberg LP
New York, New York
zhaolan2011@gmail.com

Suzanne K. McIntosh
Courant Institute of Mathematical Sciences, and NYU Center for Data Science
New York University
New York, New York
mcintosh@cs.nyu.edu

***Abstract*— While research on merger and acquisition (M&A) has been extensive in the finance literature, in the realm of data science, little work has been done in deploying a successful Big Data informed M&A prediction model. In this paper, we explore what can be learned about M&A activity from a firm's annual Form 10-K SEC filings. We utilize natural language processing (NLP) techniques to vectorize each filing's textual data. Next, we cluster firms by industry and identify keywords suggestive of upcoming M&A activity. We then train a classifier to predict acquirers and targets, which we use to forecast the most likely M&As of 2019. Lastly, we deploy an application which enables users to query our forecasts and visualize our data.**

Keywords— Natural language processing, Data analysis, Analytical models, Big data applications, Data visualization, Mergers and Acquisitions, Apache Spark

I. Introduction

A. Overview

A merger or acquisition occurs when one company takes over another, establishing itself as the new owner. Colloquially, the purchasing company is referred to as the acquirer and the seller is referred to as the target. These transactions are an important part of the financial ecosystem as they allow companies to grow, downsize, or shift their competitive position. M&As affect all parties interested in the management decisions of a company, including employees, stockholders, investment banks, and hedge funds. Due to the effect they can have on the stock price of the companies involved, much work has been done on trying to forecast when these deals are likely to take place.

B. Workflow

We attempt to use textual analysis to analyze and predict M&A activity. Specifically, we use a corpus of historical Business Description and Management Discussion and Analysis (MD&A) sections of 10-K filings from 1994 to 2018. Using an NLP pipeline, we vectorize these text files and, based on a dataset of historical M&As, create two Boolean labels for each vector signaling whether they were an acquirer or target in the year following the report.

We then implement several machine learning algorithms to extract actionable data from the resulting datasets. We use K-means clustering to identify commonalities among positively labeled companies. Next, we train two logistic regression classifiers to identify reports suggestive of upcoming M&A activity. Using the classifiers, we then segment 2019 filings into companies likely to be acquirers, targets, or neither. By joining acquirers and targets by industry we formulate our predictions for this year's most likely M&A deals.

To gain some interpretability into our classifier, we use latent Dirichlet allocation (LDA) [5] on positively labeled data and individual clusters to obtain words and phrases associated with M&A activity.

C. Project Design

We follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) in the development of this analytics application. The design stages of our project are shown in Fig. 1 and we provide here a high level description of the stages:

1) Predicting Mergers and Acquisitions: Our project aims to predict mergers and acquisitions before they are announced to the public. This is useful because this enables trading strategies where firms can buy stocks of the target prior to the acquisition and profit from the price jump following acquisition.

2) Data Sets: We obtained a dataset which contains the records of M&A transactions that occurred from 1994 to 2018. We then scraped the Management and Discussions Outlook (MD&A) and the Business Descriptions sections of 10-k reports for all the companies we could find.

3) Vectorizing and Labeling Data: We proceed to vectorize the MD&A and Business Descriptions corpus. For each company we assigned labels of target, non-target, acquirer, non-acquirer.

4) Clustering and Classification: We trained a classifier to enable us to distinguish between firms that are most likely be targets or acquirers. In addition, we performed clustering to identify subgroups within the categories of targets, acquirers, non-targets and non-acquirers. This allows us to single out which topics and terms are most indicative of each category and identify relations across subgroups. This approach supplements our classifier to enable greater interpretability.

5) Cross-validation: We use cross-validation to split our data intro training and test sets. We evaluated our results through the precision levels we were able to obtain across a range of thresholds.

6) Application Development: We created a website that enables users to input a company or industry and receive useful information such as the probability of a company being a target or acquirer and the likely targets or acquirers of that company. We also provide a variety of different ways for investors to visualize our results.

D. Contributions

Our main contributions are as follows:

- We perform a clustering analysis of targets and acquirers, finding that the 10-Ks of targeted companies not only appear more homogenous but also contain more instances of negative and risk-related terms than those of acquiring companies.
- Our classification models achieve an area under the receiver operating characteristic curve (AUC) of .72 and .77 for targets and acquirers respectively. They produced precision rates of 8% for targets and 79% for acquirers.
- We use our classifiers to make predictions for 2019 and develop a novel application for querying and viewing our results.

The paper is organized as follows: in Section II we describe the motivation for developing this analytic application. In Section III we describe related work, Section IV describes the datasets used, Section V describes the backend analytic, Section VI provides an overview of the design of the application, Section VII provides an analysis of our results, and Section VIII describes actuation and remediation.

II. Motivation

We wanted to give the stakeholders of a company a means of inquiring about the likelihood of an upcoming merger or acquisition. Our analysis will provide an estimate of the chances a given company will be an acquirer or a target. To those not interested in specific firms, but the M&A landscape generally, we provide forecasts of the most probable deals by matching acquirers and targets by industry.

While knowledge of upcoming M&A activity can aid employees and other firms within related industries, its primary benefit is to investors. Investors can use insights derived from our application to inform their trading decisions. Historically, a target company's stock price tends to increase during a takeover and the acquirer's stock price tends to decline. Therefore, a viable strategy would be to buy shares of companies that are most likely to be targets of an acquisition, while selling shares of companies most likely to be acquirers.

III. Related Work

In the literature, Routledge, Sacchetto and Smith [1] examine whether the MD&A section of a firm's annual 10-K filing can be used to predict whether that firm will be involved in a merger or acquisition. The authors trained a regularized logistic regression model and evaluated it using a test set. They used the pseudo R^2 measure as a barometer for performance and compared their results to a baseline model which used only financial variables.

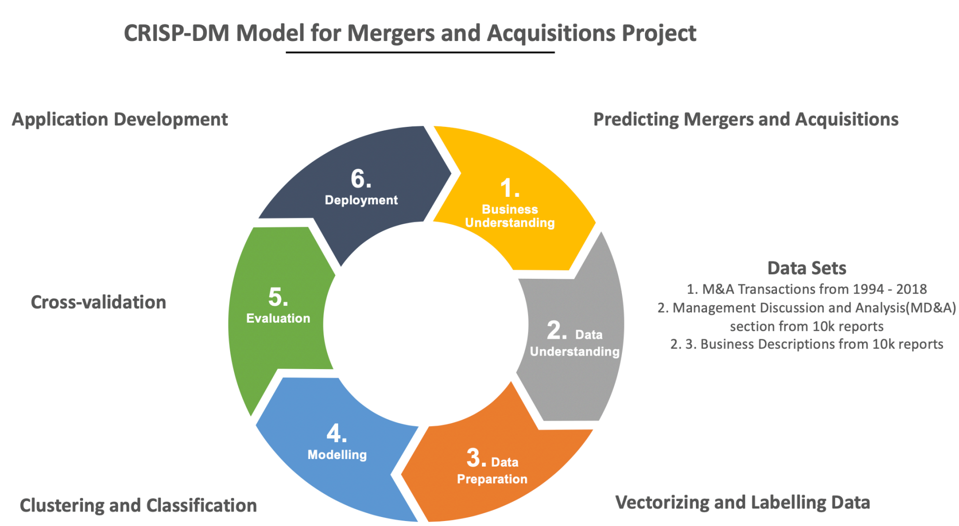

Fig. 1. Analytics development process

They found that the textual model substantially outperformed the baseline model on the task of predicting acquirers and was comparable to that of predicting targets. Our model builds upon this by improving the precision in which one can predict acquirers.

Barnes [2] took a different approach to see if methods used to predict bankruptcy with accounting factors can also be used to predict mergers and acquisitions. While Barnes' model did not see an improvement from the baseline, we were curious to see if bankruptcy itself and overall poor financial performance can be indicative of merger activity.

Xiang et. al [4], also used accounting factors such as financial and managerial rations along with basic features and topic features produced by latent Dirichlet allocation to predict mergers and acquisitions for startups on the Crunchbase website. The authors used Bayesian networks for their prediction and achieved a 60% to 80% true positive rate for most categories of startups. The single most important predictor for almost all of the categories were revisions on profiles in Crunchbase. The other topic features that were predictive towards M&As were features related to founders and funding. This is unsurprising since the network and experience of founders have undue weight on the success of startups. The features far outweighed other features related to the company such as the number of products.

More recently, Morgan [3] found that the more optimistic the language used by management in SEC filings, especially those in the technology and telecommunications industry, the more the returns across 60 and 90 days intervals increased.

Looking exclusively at the food industry, Adelaja, Nayga and Farooq [8] built two logistic models to explain mergers and acquisitions activity in US food manufacturing firms. The food industry is interesting because in contrast to other sectors where M&A activity can be driven by aspects unrelated to the characteristics of the firm [7], M&A activity in the food industry is mostly driven by moves that align with the strategic interests of the company [11]. The authors' models found that for targets the most important features were firm liquidity, debt/leverage and profitability. For acquirers, the important features were degree of control, attitude surrounding the transaction and number of prior bids. These models yielded a predictive accuracy of 74.5% for targets and 62.9% for acquirers. These figures are high because the authors were looking exclusively at the food industry. For perspective, Palepu [6] showed that after correcting for methodological errors for previous studies, the best precision achieved for targets was only 5%. Our model aims to improve upon the predictive accuracy for both targets and acquirers as well as expanding the ability to predict M&A activity in all industries, not just food.

To understand the context and landscape of M&A Activity Martynova and Renneboog [7] analyzed the trends that can be observed through a century's worth of M&A transactions. In the first section, the authors highlight how M&A activity tends to go through periods of boom and quiet that closely mirrors the state of the overall stock market. They concluded that M&A activity is very much a function of stock market health, if the economy is doing better than M&A activity is more likely to occur. Furthermore, the authors suggest that M&A activity generally occurred in waves with similar underlying themes. The authors highlighted five main merger waves which occurred in the 1900s, 1920s, 1960s, 1980s and 1990s. The authors attributed the cause of these waves to the industrial revolution, monopolistic competition, political stability, economic rebound, and globalization respectively.

Martynova and Renneboog's paper also corroborates and gives background to the Adelaja et. al paper. Martynova and Renneboog suggests that while M&A activity occurring prior to the 1970s occurred due to varying factors such as diversification, recent M&A transactions were more unified in that strategic alignment with the firm was a bigger priority. Given that the paper by Adelaja, Nayga and Farooq on food M&A was written about a decade ago, we wanted to see if shifts in modern M&A motives can result in building a successful model in predicting M&As across all industries.

IV. Datasets

Our MD&A dataset consisted of roughly 150,000 reports for filing years 1994 to 2018, totaling roughly 6 GB. These reports were downloaded as a one-time collection from the Electronic Data Gathering, Analysis, and Retrieval system (EDGAR) package in R. The dataset can be updated on a yearly basis as 10-K filings are due 90 days after fiscal year end.

Our second dataset consists of the business descriptions dataset which is roughly 8.5 GB in size with around 150,000 reports from 1994 to 2018. These reports were also gathered through the same EDGAR package in R and as such can be updated at the same time as the MD&A data sets.

We augmented the aforementioned datasets with an M&A dataset gathered from Bloomberg terminal [12]. It is a csv file of all successfully completed transactions dating from 1994 to 2018. The dataset contains the date in which the transaction was announced as well as the target, seller, and acquirer companies. This dataset was used to generate our ground truth labels.

V. Description of Analytic

A. Preprocessing

We processed the MD&A texts using an array of NLP techniques. Each text file was put through a pipeline which included conversion to lowercase, stop word and whitespace removal, lemmatization, tokenization, addition of 2-grams and 3-grams, and tf-idf. We discarded all phrases that appeared in fewer than 100 documents and capped the size of the vocabulary at 100,000 phrases. The resulting data frame was then joined by Central Index Key (CIK), which is a unique corporation identifier, onto our dataset of historical M&As. We then created two labels corresponding to whether the company was an acquirer or acquiree within 365 days of publishing the report.

B. Keyword Extraction

Two new data frames were generated corresponding to positively labeled vectors, one for acquirers and one for targets.

For the MD&A texts LDA was run directly over these dataframes. We generated fifty topics of five terms each.

The business description texts were put through an identical NLP pipeline identified by their ground truth labels. To extract commonalities among the texts, we ran K-Means clustering on acquirers and targets respectively. We set K to twelve corresponding to the major categories companies tend to fall into. Finally we ran LDA on each cluster generating ten topics of five terms each. Both the document concentration and topic concentration parameters were set to 0.001 in order to capture a wide variety of topics.

C. *Classification*

We used the vectorized and labeled MD&A dataset as the input to our classification model. An 80/20 split was used for segmenting data into training and test sets. Before using logistic regression to separately model whether a company is a target vs. non-target and acquirer vs non acquirer, we used oversampling on the minority classes, truetargets and true acquirers, to account for data imbalance. The ratio was set so that the overall number of targets and non-targets were identical after oversampling.

D. *Prediction*

Given these two classifiers we then made predictions on 2019 data. Acquirers and targets were then joined on their Standard Industrial Classification (SIC) codes to generate predictions for specific mergers.

VI. Application Design

We wanted to give users a convenient way to query our 2019 predictions and visualize our data. To do so we built a webpage which connects to a database populated with our predictions. The application can be found at https://m-a-prediction.herokuapp.com/.

Users can query our predictions by company or industry using the screen shown in Fig. 2. Upon searching a ticker symbol, our prediction will be displayed along with our confidence level. If the given company is classified as an acquirer or target, potential partners in an M&A deal will be displayed in order of likelihood. Similarly, when searching via SIC code, the user will be shown all predicted deals within the queried industry sorted by likelihood.

The user can also visualize the results in two fashions. Firstly, as shown in Fig. 3 they can view the output from our LDA model in the form of a word cloud where the size of the word indicates the frequency of occurrence. Our webpage displays the LDA outputs from MD&A texts classified as acquirers or targets respectively. It also displays the LDA output from the clusters generated from business description texts.

We also use Scattertext, a popular interactive scatter plot tool that enables users to perform exploratory data analysis. Shown in Fig. 4 and Fig. 5, the tool allows users to visualize terms and phrases that are more predictive of one category than others. Each point in the plot corresponds to a word or phrase in the corpus. The closer to the top-left or bottom-right of the plot that a word appears, the more disproportionately it appeared in one class vs. the other.

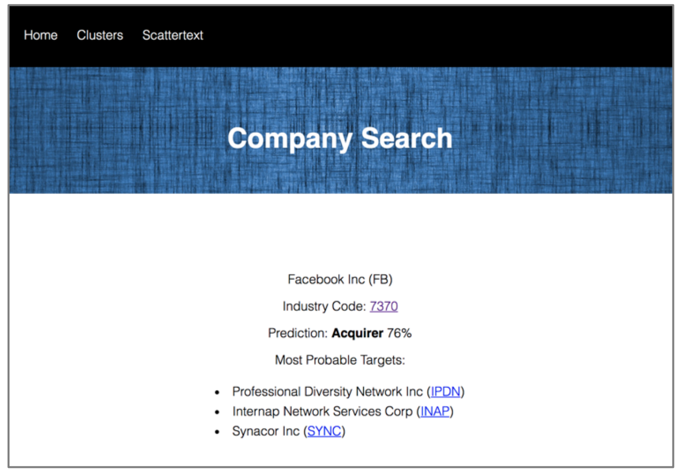

Fig. 2. Query by company ticker symbol

VII. Analysis

A. *Clustering*

Clustering acquirers and targets by their Business description texts revealed that targets tend to fall into fewer clusters than acquirers. After removing outlying clusters, we were left with 6 clusters for acquirers compared to 3 for targets. This suggests there is a certain degree of homogeneity among targets that does not exist among acquirers.

Performing LDA on clusters of targets, we found that the topics produced by targets contained many more negative and risk- related terms that those of acquirers. This makes intuitive sense as poorly performing firms are more likely to take part in a merger as a means of improving their financial position. As seen in Fig. 6 negative and risk-related terms include “undercapitalized”, “unsound”, “risk”, “failure” and “unsafe”.

On the other hand, performing LDA on clusters of acquirers did not yield much meaningful results. The words yielded by the LDA neither had a particularly positive nor negative slant. This discovery coincides with our previous claim that there seems to be more homogeneity in targets than acquirers.

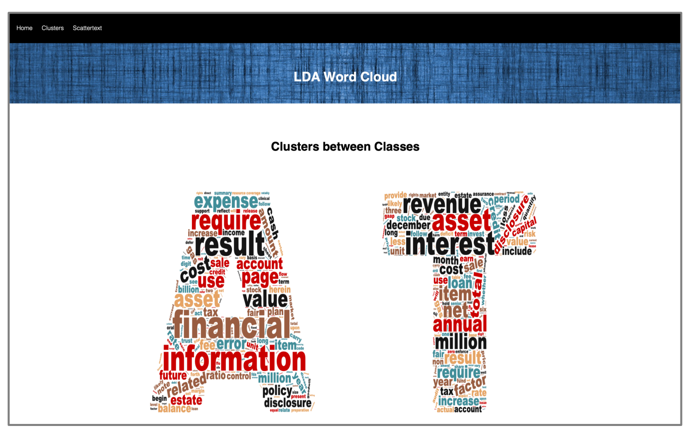

Fig. 3. Word cloud generated using LDA mode

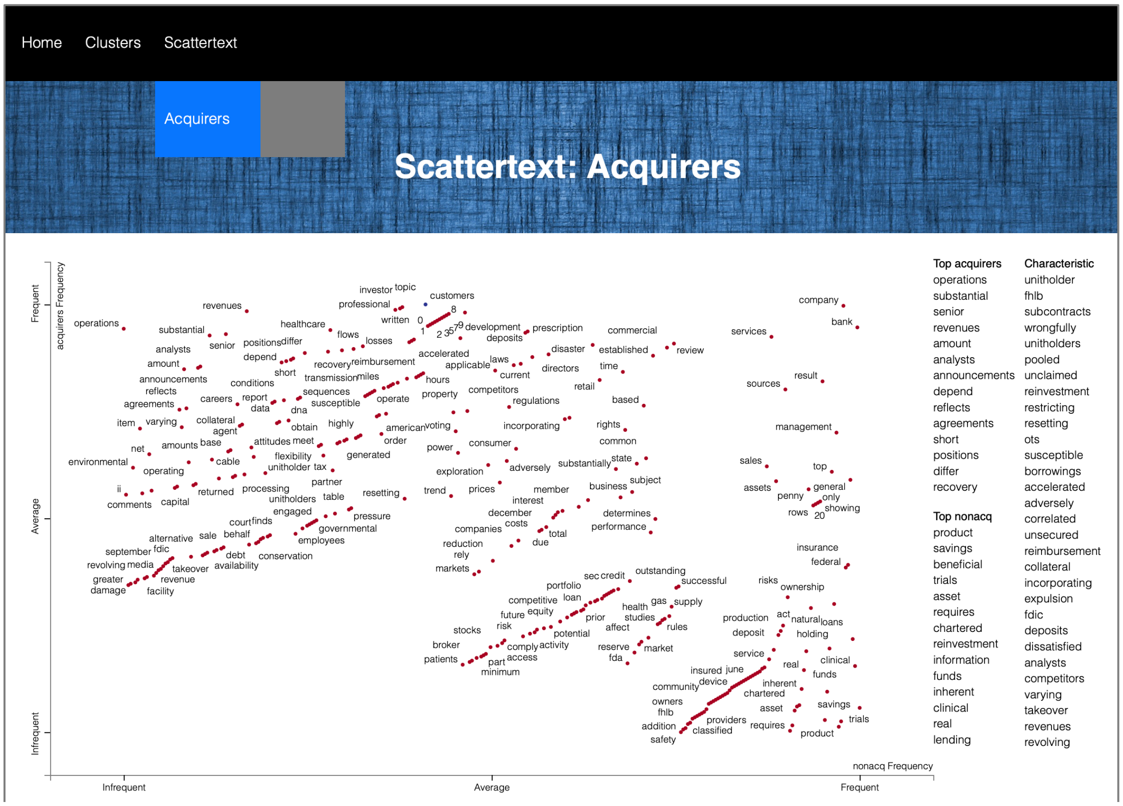

Fig. 4. Acquirers Scattertext

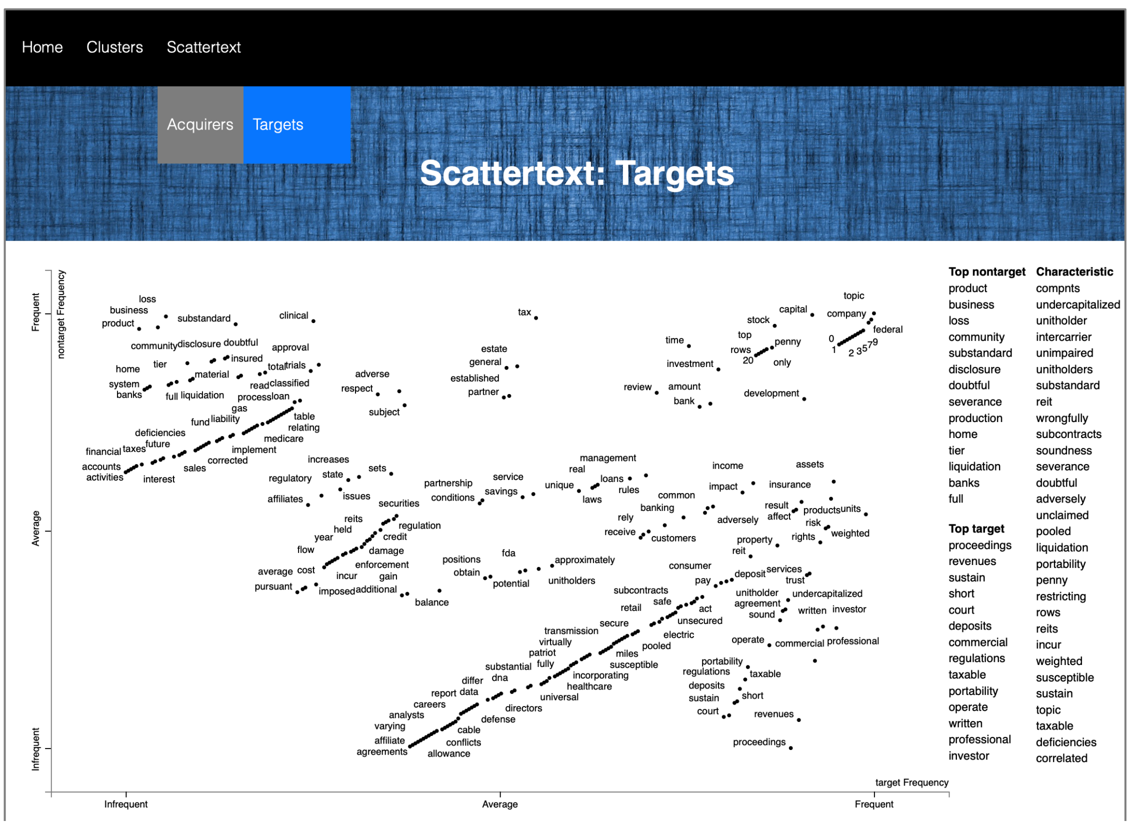

Fig. 5. Targets Scattertext

While target companies generally have similar reasons for being targets of acquisitions, acquiring companies motives for acquisitions are more diverse and unpredictable.

B. *Classification*

Given the vectorized MD&A texts, we then trained a logistic regression classifier on the labeled targets and acquirers dataframes respectively. Support vector machine and naïve bayes models were also tested but ultimately discarded due to poor performance.

We tuned the hyperparameters of each model using a grid search and 10-fold cross validation. Ultimately this led us to choose elastic net parameters and regularization parameters of .5 and .03 respectively for targets and .25 and .01 respectively for acquirers. The performance of the models are shown in TABLE I.

The models resulted in an AUC of around 0.72 and .77 for targets and acquirers respectively. The target audience of our application, investors, care significantly more about precision than recall. The precision rates of both models as a function of the threshold of the logistic regression model are shown in TABLE II. and TABLE III.

While the precision of predicting targets may appear low on an absolute scale, one must take into account the severe data imbalances. Only 1.7% of texts were positively labeled for targets and 15% for acquirers. As such, both models significantly outperform a baseline model of random guessing as shown in Fig. 7. The target model also improves upon the precision of the financial ratio based model reported in [6].

Our results suggest that it is easier to predict acquirers with a high degree of accuracy than it is to predict targets. With this in mind one can structure their investment strategy to focus primarily on acquirers. Of course other factors must be taken into account like the confidence of our prediction and the potential for stock price appreciation post-takeover.

Fig. 6. Word cloud generated on a cluster of targets

TABLE I. LOGISTIC REGRESSION PERFORMANCE

	Targets	Acquirers
Area under ROC	.72	.77
Max. Precision	7.6%	79%

TABLE II. TARGET PRECISION VS. THRESHOLD

Threshold	Precision	# Predictions
.8	7.4%	54
.75	7.6%	225
.7	6.4%	560
.65	5.3%	1257
.6	4.8%	2461

TABLE III. ACQUIRER PRECISION VS. THRESHOLD

Threshold	Precision	# Predictions
.95	79%	62
.9	73%	283
.85	71%	620
.8	65%	1057
.75	57%	1616

VIII. ACTUATION AND REMEDIATION

A. *Optimization Metrics*

When the user queries a particular company, our application returns one of three predictions: Acquirer, Target or No M&A deals this year. We optimized for precision while taking into account sensitivity because, to the typical user of our application, the downside of incorrectly recommending a target is far higher than incorrectly failing to recommend one. That is, to an investor trading on our predictions a false positive may inflict a material loss, whereas a false negative is simply a missed opportunity conferring no financial loss.

B. *Use Case*

Increasing our classification threshold, and thus our precision, produces accurate results. Take, for example, the company Empire Resorts. Empire Resorts is a good use case because it illustrates the importance and efficacy of clustering and classification approaches to predicting M&A.

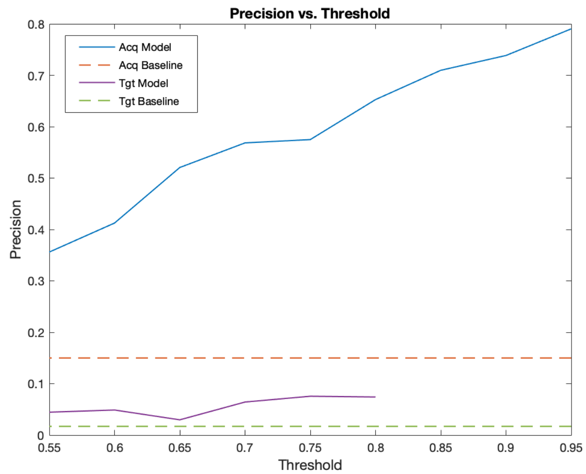

Fig. 7. Precision vs. threshold

Our program predicts that Empire Resorts would be a target company in 2019 with high probability. Recently, on August 7th, 2019, news broke that Empire Resorts would be acquired by Genting Group [9]. Looking more into the context of this acquisition, we can see that a big reason for this acquisition is that Empire Resorts is considering filing Chapter 11 bankruptcy [10]. Recall that we previously revealed that clusters of target companies tend to contain much more negative words than that of acquirers. In fact, in our application, the words "undercapitalized", "banking" and "risk" were among the most indicative words of a target company. Thus, this example illustrates how our application can successfully inform investors of upcoming M&A activity.

C. Limitations

That said, we would still recommend that investors perform their due diligence when it comes to creating an investment strategy around M&A activity. Finance trends are fickle, and just because a pattern exists now does not mean that it will continue to exist in the future. Hence, we do not explicitly state an action that the investor should take. Our application is likely best used to inform areas of further inquiry to a domain expert, who then can make an informed financial decision.

IX. Conclusion

Mergers and acquisitions are an integral part of the financial ecosystem and they effect nearly all stakeholders of participating companies. As such, a lot of effort has gone into trying to predict when these deals are likely to transpire. Our results show that a good amount of actionable information about these deals can be extracted from a firm's annual Form 10-K SEC filings.

We found that the reports of targeted companies tend to cluster into a smaller number of groups than acquirers, indicating that target companies tend to share similar features.

Additionally, an LDA analysis found that the texts of targets tend to have a higher rate of negative terms, implying a certain level of distress that is indicative of a forthcoming acquisition.

Finally, our classification model produced precision rates of 8% for targets and 79% for acquirers. This suggests M&A deals can be forecasted with some degree of certainty and presents an opportunity for investors to profit off this information. In the aim of making our results accessible, we developed an application that allows users to easily visualize and query our predictions.

X. Future Work

1) Factoring in sentiment into our classifier: As it stands, our program relies on two distinct approaches with the classifier and clustering models. While this has been perfectly functional in producing our desired results, we believe it would strengthen our model if we can somehow combine the two models such as through factoring in sentiment from the clusters as a variable in our classifier model. Doing so would enable us to better quantify the impact of negative terms and potentially add to the predictability of our model.

2) Deep Learning: It would be interesting to test the performance of a transfer learning approach. We could take a neural network pre-trained on a large corpus of English text, say Wikipedia. Then we could train it further on 10-K reports and see if it outperforms our logistic regression model.

Acknowledgment

We would like to thank Wensheng Deng and the NYU High Performance Computing team for their assistance with questions regarding the cluster. We also thank Cloudera for providing the Apache Hadoop and Apache Spark distribution (CDH) through the Cloudera Academic Partnership.

References

[1] B. Routledge, S. Sacchetto, and N. Smith, "Predicting Merger Targets and Acquirers from Text," 2013.

[2] P. Barnes, "Can Takeover Targets be Identified by Statistical Techniques?: Some UK Evidence," *Journal of the Royal Statistical Society,* 1998.

[3] P. Morgan, "Predictive Power? Textual Analysis in Mergers & Acquisitions," 2018.

[4] G. Xiang, Z. Zheng, M. Wen, J. Hong, C. Rose, and C. Liu, "A Supervised Approach to Predict Company Acquisition with Factual and Topic Features Using Profiles and News Articles on TechCrunch," ICWSM, 2012.

[5] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, 3 (2003) 993-1022, 2003.

[6] K. Palepu, "Predicting takeover targets: A methodological and empirical analysis," Journal of Accounting and Economics, Volume 8, Issue 1, Pages 3-35, March 1986.

[7] M. Martnova and L. Renneboog, "A Century of Corporate Takeovers: What Have We Learned and Where Do We Stand?" Journal of Banking and Finance, Volume 32, Issue 10, Pages 2148-2177, October 2008.

[8] A. Adelaja, R. Nayga Jr., and Z. Farooq, "Predicting Mergers and Acquisitions in the Food Industry," Wiley, March 1999.

[9] Newsdesk, "Genting Malaysia to form part of new merged entity taking full ownership of New York's Empire Resorts," August 2019.

[10] Mid Hudson News, "Empire Resorts considers filing Chapter 11 bankruptcy," August 2019.

[11] Duff & Phelps, “Food and Beverage M&A Landscape,” 2018.

[12] Bloomberg L.P, Mergers and acquisitions data. Retrieved from Bloomberg M&A database, 2019.