Antitrust Evaluation of Mergers and Acquisitions with Machine Learning

Branchie Mbofwana (mbofwana)
branchie.mbofwana@utoronto.ca
Guneek Sahota (sahotag6)
guneek.sahota@mail.utoronto.ca
Zi Wang (wangz654)
thebigzi.wang@mail.utoronto.ca

April 12, 2021

1 Introduction

1.1 Background Significance

Mergers are essentially a legal solidification of two entities into one; whereas, an acquisition takes place when the acquiring company obtains the majority share in the acquired firm. So, given the case a company faces competition, its aim is to simultaneously reduce costs and produce better products/investments. So, the key for a company facing competition is to acquire these competitors in order to discard the threat to the company. Additionally, companies implementing the MA procedure are done in order for them to grow larger (i.e., higher market capitalization). On the other hand, companies may also be interested in working collaboratively (i.e., mergers). With firms engaging in MA, it substantially reduces the risk it can encounter as the main aim of MA is to combat competition and reduce costs. Thus, the main idea is that MA results in non-competitive market behaviour where this type of behaviour can hurt consumers and hinder innovation as it does not optimize social welfare. But, mergers and acquisitions can also result in economies of scale and scope as they increase productivity. Lastly, for the purpose of our research project, we only consider the U.S. antitrust laws which aim for the maximum diffusion of competition. In particular, the U.S. antitrust laws are basically structural in which it anticipates solutions such as dismemberment of monopolies and an imposition of damages in private suits in the pursuit of diminishing anti-competitive behaviour in the market.

1.2 Motivation

The motivation behind our research project is on the antitrust evaluation on MA. Also, MAs are known to be one of the most common behaviours of the firms which eventually lead to anti-competitive markets. So, in basic terms, there are various mergers and acquisitions that have the full potential that can cause a market that is detrimental to consumers as they could be harmful to competition. For example, given the case that two rival companies merge, the merging companies no longer compete to attract customers – therefore, resulting in higher prices but lower marginal costs. Furthermore, there is also a potential for collusion if the merger or acquisition results in fewer competitors but with parallel market shares – resulting in a greater threat to competition. However, there are also some positive mergers and acquisitions which result in socially optimal behaviour through reducing costs leading to enhanced products and valuable investments. Moreover, this is when antitrust laws have an important role in ensuring that mergers acquisitions retain the benefit of competitive prices alongside quality of goods and services. The laws achieve these goals through stimulating and furthering competition in the markets along with precluding anti-competitive mergers and business practices. Moreover, examples of offenses within the U.S. antitrust laws include price fixing, bid rigging, market and/or customer allocations and group boycotts. Also, it is important to note that horizontal mergers (i.e., a merger between direct competitors) could eradicate a competitor in an industry where there are already just a few firms competing. Consequently, with respect to the purpose of our research project, it is imperative that we consider the best methods and practices to implement in order to separate potentially harmful mergers acquisitions from non-harmful mergers acquisitions.

1.3 Research question

Therefore, these best methods and practices would also support the antitrust laws regarding the regulation of mergers acquisitions that possibly pose a threat to competition. So, considering the advanced machine learning algorithms and artificial intelligence, it is critical that we apply them to identify the harmful MAs. So, this leads to our research question which states: "How can machine learning be implemented to identify and separate harmful mergers acquisitions from non-harmful mergers acquisitions?"

1.4 Answer to research question contributions

So, the answer to our research question consists of research that will supplement and contribute to current methods such as merger simulations used in the analysis to accurately identify harmful mergers from non-harmful mergers. Moreover, we will utilize Natural Language Processing (NLP) and Machine Learning (ML) which will provide an additional method that can be used to further assess anti-competitive behaviour in technology firms that traditional methods cannot adequately assess. Therefore, we apply NLP techniques involving text classification that incorporate data retrieval, feature extraction, topic classification, web scraping and other various text analytics which will assist us in converting the text (which were obtained from cases derived from the U.S. Department of Justice database) to numerical data. Then, we will apply supervised and unsupervised ML methods including KNN, decision trees, random forest, Linear Discriminant Analysis (LDA), and Principal Component Analysis (PCA) in order to build the appropriate algorithms which will assist us in identifying the harmful mergers that result in a market with anti-competitive behaviour. Furthermore, with the successful identification of the harmful MA this would result in the reduction of anti-competitive behaviours in the market, increase the overall social welfare of consumers, diminish entry barriers encourage innovation simultaneously, and greatly reduce the cost of litigation and analysis.

1.5 Literature Review

With reference to the motivation behind our research project, we discovered through our literature review process that there was insufficient recent empirical research achieved based upon antitrust laws in combination with machine learning with their empirical analysis. However, there were three key papers that did resonate with the intent of our work. First, Athey (2018) provides a comprehensive introduction of machine learning and its implementation and usability in prediction policy problems in economics. So, the similarity we found between this paper by Athey (2018) and our work is that we both utilize structured and unstructured data to provide parametric estimates and predictions. However, the one dissimilarity is that the contribution we provide through our work is that we utilize algorithms and provide flexibility by comparing different models. Second, Posner (1976) offered a theoretical background and empirical evidence on how antitrust laws can benefit consumers and examined the differentiation between advantageous MAs and harmful MAs. So, the similarity is that we both focus on the field of antitrust law and our research is based on Posner's (1976) theoretical work. But the dissimilarity between our work and the paper by Posner (1976) is mainly through the contribution our research has. We update the date of the data and include richer sets of explanatory variables. Additionally, within the context of our research project, we make predictions rather than causal conclusions.

Third, Datta and Srivastava (2002) discuss how mergers and acquisitions are characterized by high failure rates which are mostly due to the incapability of the acquiring firm to effectively evaluate potential acquisition candidates. So, one similarity between our research project and Datta and Srivastava's (2002) paper is that we both state the high failure rate of corporate mergers and acquisitions which evidently result in anti-competitive markets. In addition, we both provide insights as to how AI assists with the MA evaluation process. However, one dissimilarity between both of our works is through our specific contribution. Specifically, our contribution is that we provide an application of AI on deriving an evidential reasoning approach with respect to the evaluation process regarding potential acquisition candidates.

2 Data and Methodology

In this section we will present the methodology adopted for data retrieval, textual analysis, feature extraction, and the steps taken for this purpose include searching, criteria for inclusion and exclusion, model training, selection and presentation.

2.1 Web Scrapping

Since there is no direct qualitative and quantitative distinction between harmful and non-harmful MAs, we develop our own classification through analyzing the MA transaction case filings, which are publicly available from the US Department of Justice (DOJ) Antitrust case filings databases. Given there are various types of case filings - criminal and civil - and each case often contains different kinds of case filings with complex structures due to factors such as the time of the case, how the DOJ processed the case, and the final outcome of the case. Among all types of case filings, we choose to analyze the Final Judgement document for each case as the availability of such documents indicates the closure of each accusation. Besides, the final judgment is also the most comprehensive document, which contains decisions and recommendations that are provided by the court that are essential for our analysis and classification. Through the collection of final judgement, we applied techniques such as web scraping to obtain the final judgement, which is stored in HTML format. There are in total 2090 cases that are available in the database. We excluded cases that are before 1990 given the unavailability of various predictors for such cases.

We have chosen seven predictors, which are big-mcap, r-mcap, r-profitability, r-revenues, n-sectors , n-competitors, n-acqusitions, and big-mcap which is a dummy variable that takes the value of 1 if all the companies involved in the case have large market caps greater than 1 billion dollars and 0 otherwise .Further, r-mcap represents the ratio of the acquirer's market cap to the acquiree's market cap. In terms of market capitalization, we believe it would mostly cause an antitrust issue if two very large companies are merging, as the merged company would possess market power, thus harming competitors and consumers alike; thus, market capitalization indicators might be a relevant factor to represent such an issue. r-profitability and r-revenues represent the ratio of profitability and ratio of revenues between acquirers and acquirees. We argue that these financial conditions might be relevant due to that acquirers might seek MA targets to increase their competitiveness to improve their financial and operating conditions. n-sectors represents the num of sectors or industries that the acquirers have already been involved in. n-competitors reveals that how many competitors do the acquirers face within the industry of the business is acquired. We believe that this predictor might be highly relevant as there are a few competitors, the merge between these companies can cause great harm to the remaining competitors and consumers. Finally, n-acqusitions represents the number of acquisitions the acquirers had acquired on average for the last two years, we argue that the more companies the acquirers acquired, the more bargaining power or market power the acquirers already contain. All the financial data are obtained from Bloomberg, FactSet, and financial statements. The industry data are collected partially from Bloomberg and most from IBISWorld Industry research, Gartner and Frost Sullivan International Market Research accessed through the Financial Learning Center at the University of Toronto. In order to keep the time of data relevant to the case, we match the year of these with the transaction year. Also, to obtain a complete dataset, we only included observations, which have not only final judgement available but also all the predictors available. We have chosen cases that involve only public companies because both the financial and operational conditions of such companies can be easily and accurately obtained. Overall, we were able to obtain 52 observations in total.

2.2 Textual Normalization and Analysis

After collecting final judgment data from the DOJ, text normalization was implemented to clean, normalize, and standardize the textual data with techniques like removing symbols and characters, HTML tags, removing stop words, lemmatizing the text and finally tokenizing the textual data.

2.2.1 Stop Word and Common Word Removal

In addition to removing stop words, we also removed commonly occurring words from the documents by using a simple algorithm to count the most occurring words. Any words removed were not related to antitrust, and helped reduce the dimensionality of the dataset. For example words like 'judgement','shall'. etc.

2.2.2 Tokenization and Lemmatization of Text

Furthermore, an algorithm was then used to tokenize the text in each document for easier text preprocessing. The tokenized text was then standardized by transforming all words into their base forms for easier feature extraction and dimensionality reduction by using the lemmatizer from the nltk.

2.2.3 Indexing and Labelling

Next, the documents were further cleaned by filtering the documents for nouns and adjectives. This was done to further reduce the dimensionality of the words in the corpus. A vocabulary was then created to count the number of times specific antitrust terminologies from the final judgements to indicate the outcome of the final ruling. That is, did the DOJ deem the mergers as either anti-competitive or not? For example, words like, 'diverstitute', 'divested', and 'remedy' among others were included. These particular words are particularly used when the parties accused are found to be anti-competitive. The vocabulary was then used to create a label for each judgement as either being anti competitive or not by setting a threshold for the number of times identifying words from the vocabulary appeared. Further, a threshold was set at 10. If any identifying words from the vocabulary appeared less than the threshold, then the document was labelled as competitive (0) and anti-competitive (1) if the count was equal or exceeded the threshold.

2.2.4 Feature Matrix and Extraction

As it turns out, many Machine Learning algorithms are incapable of processing strings or plain text in their raw form. Therefore,they require numbers as inputs to perform any sort of processing, for classification or regression. Because of this, we used the Bag of Words model to convert all text documents in our corpus into vectors such that each document is converted into a vector that represents a frequency of all distinct words present in the document vector space for each individual document. We used the countvectorizer to create a count vector matrix M of size D X N. Where D is the list of documents in the corpus, and N is the number of tokens from each document.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
nbsp	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
distributor	15	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0
tampico	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0
compliance	14	5	4	7	6	5	2	1	5	6	0	5	5	5	6	19	5	18	0	5	0	5	5	5	15	6
price	7	4	0	2	2	2	9	0	2	3	0	2	2	2	0	6	1	0	0	2	0	2	0	2	16	0

Now, a column can be understood as a document vector for the corresponding word in the matrix M. For example, the word 'distributor' for document 1 in the above matrix is [2,2]. Here, the rows correspond to the words in the corpus and the columns correspond to the documents in the corpus. Further, an important thing to note for feature extraction is that once we build a feature extractor using some transformations and mathematical operations, we need to make sure we reuse the same process when extracting features from new documents to be predicted, and not rebuild the whole algorithm again based on the new documents. This is done in order to preserve the distribution from which the document matrix is created.

However, in the bag of words model, the created vector is based on the frequency of word occurrences. This is a hindrance because higher frequency words will overshadow less occurring words that might have a greater importance and provide more effective features. To address this, we used the Term Frequency Inverse Document Frequency (TF-IDF) technique. This method transforms the raw occurring frequency count of a token in a given document and scales down the impact of the tokens that occur very frequently in a given corpus and that are hence empirically less informative than features that occur in a small fraction of the training corpus.

The image below shows a TF-IDF document matrix from the final judgment corpus.

	abide	ability	absent	accept	acceptable	acceptance	accepted	accepts	access	accomplish
0	0.032177	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.016227	0.000000
1	0.000000	0.029063	0.017616	0.000000	0.030414	0.0	0.0	0.0	0.056518	0.043594
2	0.027005	0.017508	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.013619	0.000000
3	0.000000	0.015377	0.009321	0.000000	0.016092	0.0	0.0	0.0	0.023923	0.038443
4	0.000000	0.011055	0.006701	0.010035	0.011569	0.0	0.0	0.0	0.021497	0.016582

2.2.5 Model Training

The TF-IDF matrix was then used as a predictor to engineer classification features and the label as the target variable. Further, the train and test set were then randomly assigned using a sklearn package with a weight amounting to 30 percent of the dataset given to the test set. The training data was then fed into five different algorithms. Namely; KNN, Logit, Decision Tree, Naive Bayes and Random Forest. The number of neighbours in the KNN was between 1-25, and the optimal K was 3. In addition, cross validation with 5 K- crossfold utilized for both model and feature selection. Cross validation was implemented in addition to the train and test split because it provided a more accurate out-of-sample estimate and was a more efficient use of the data because every observation was used for both training and testing. Below is a table summarizing the accuracy of the five models.

2.2.6 Financial Classifiers

The collected financial data was imported and a dataset created. The financial data was then used to create several predictor variables, and the label of the final judgment used as the target variable. The data was then again randomly split into train and test datasets, with a weight of 30 percent of the dataset assigned to the test dataset. The training dataset was then used to train five different algorithms. In addition, for each algorithm, an AUROC and Confusion matrix were constructed. Furthermore, cross 5 K - crossfold validation was also used to estimate the accuracy and compare the performance of each model relative to the others. In addition, the ranks for features of the top two performing models were recorded and compared.

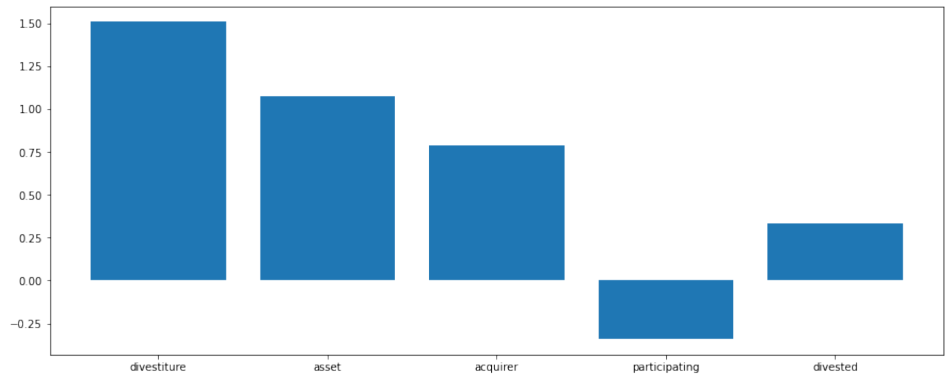
2.3 Results

2.3.1 Engineered NLP Features

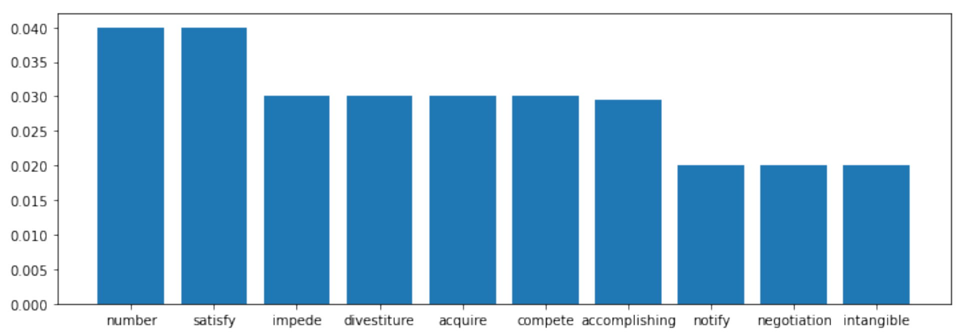
Surprisingly, all five algorithms performed really well. With the Logistic and Random Forest classifiers having a 100 percent test accuracy. However, we believe it is important to note that the test performance, confusion matrices and AUROC scores were significant because of the following reasons:

- Undersampling - This would have led to an imbalance in the distribution of one class with few examples and the other with many classes. In this case, our sample had many Anti-competitive examples.
- During the process of feature engineering, additional features may have been introduced to the model not just from the training set.
- The list of common words to be removed were inadequate for the scope of this analysis.

Additionally, the top five engineered features identified for the top two performing models, based on their test accuracy were the following:



From the image above, 'Divestiture' and 'Divested' are the two most important features pertaining to antitrust that provide the most meaningful features as these features indicate the measures provided by authorities to merging firms in order to ensure markets remain competitive. Further, we also discovered that the list of common words removed during preprocessing needed to be extended, because the documents still contained many words that did not provide any useful information, but the algorithms still considered them as useful despite using the TF-IDF. The image below shows some of the top 10 features identified by the Random Forest Algorithm that had a 100 percent test accuracy.

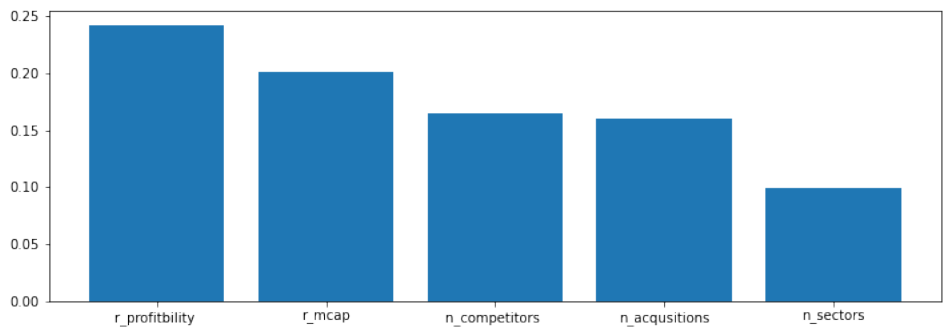


Among these features, only one - 'divestiture' - was a meaningful feature for classifying harmful and non-harmful MAs. This helped us reaffirm our assumption that our dataset was picked from a highly under-sampled distribution with a high number of filings belonging to anti-competitive mergers. This is because many of the documents were similar and had many overlapping words, and therefore, many of these features were used as appropriate classifiers.

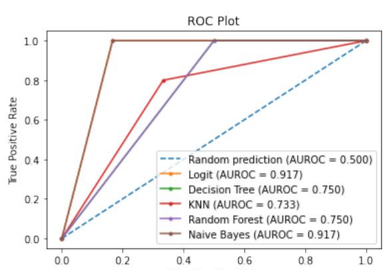
2.3.2 Financial Classifiers

Regarding the financial classifiers, the Logistic classifier had the highest test accuracy at .9 and an AUROC of .91. The AUROC is a better indicator of the performance of the Logit classifier because it measures the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. That is by setting different thresholds for classifying positive classes for data points the thresholds will inadvertently change the Sensitivity and Specificity of the model. And one of these thresholds will probably give us a better result than the others, depending on our aim to lower the number of False Negatives.

In addition, the Logit models features in order of importance can be seen in the image below.



However, we can argue, that the market cap, number of competitors, and number of acquisitions made are better features than the ratio of profitability of merging firms. But the financial classifiers provide a better and more reliable model for classifying harmful and non-harmful MAs.



	Training Accuracy	Test Accuracy
Naive Bayes	0.725	0.818182
Logistic	0.700	0.909091
Decision Tree	1.000	0.727273
Random Forest	1.000	0.818182
KNN	0.725	0.727273

Figure 1: AUROC and Model Accuracy

Additionally, the Random Forest and Naive Bayes, also performed considerably well, relative to the Logit classifier. In addition, their AUROC were also comparatively high. These results show that the financial classifiers were stronger than the NLP engineered features.

3 Conclusion

Our research study pertaining to its use of ML methods and AI in areas of antitrust case filings (i.e., evaluation of MA deals) should be implemented given its ability to produce detailed and precise results regarding the classification of harmful vs. non-harmful MAs. Moreover, it was found that all five algorithms performed extremely well in classifying harmful MAs. Therefore,the financial classifiers are vital to this study as they contain high test performance and AUROC scores. Lastly, the major limitation our study had was that the dataset we used derived from a highly under-sampled distribution with an extensive amount of filings associated with anti-competitive mergers.

So, what we obtained via our study can be extended to applications in MA given real-life settings. Furthermore, NLP techniques and machine learning algorithms which were used in this project have the potential to be further utilized in the way MA deals are conducted. In particular, AI has the ability to further data analytics to a more advanced level which could enable more intelligent and efficient decisions throughout the entire MA evaluation process. Therefore, it can be stated as to how data analytics and AI can be leveraged extensively through the MA cycle. So, as we used NLP to convert the text into numerical data (e.g., web scraping and textual analysis) and then formed our own financial classifiers via ML methods, this same approach can be furthered in MA processes by providing high-speed and at the same time, in-depth analysis of a target company's value drivers and linked value formation opportunities and risks. Thus, through utilizing ML algorithms and AI, one can efficiently convey meaningful results behind the numbers and supply valuable perceptions in MA cases. Additionally, as observed in this project, with the MA evaluation process being advanced by AI, data analytics has the potential to generate sensing tools that could assist in accelerating strategic identification of arising risks and value formation opportunities. Thus, this could make it easier to generate and consolidate deal insights with the application of ML methods and AI in MA. In addition, having MA combined with data analytics and AI will also enhance the value of the data as utilizing the relevant ML methods and NLP techniques will enable for MA-focused analytics which could be used for highly focused moments or decisions in the MA deal execution. Lastly, combining the MA evaluation procedure with ML methods, this will also have the capability to analyze micro-level details and then associate them to macro-level decisions that occur.

4 References

- Antitrust case filings. (2021, March 31). Retrieved April 7, 2021, from https://www.justice.gov/atr/antitrust-case-filings-alpha

- Are Antitrust laws being Violated? (n.d.). Retrieved from https://mnacritique.mergersindia.com/antitrus laws-a-study

- Athey, S. (2018). The impact of machine learning on economics. In The economics of artificial intelligence: An agenda (pp. 507-547). University of Chicago Press.

- Hayes, A. (2021, March 18). Mergers and acquisitions (ma). Retrieved April 5, 2021, from

- Key contacts Dina Marques Director. (2020, October 08). Not using analytics in ma? You may be falling behind. Retrieved April 8, 2021, from https://www2.deloitte.com/ca/en/pages/finance/articles/analytics-m-and-a-ia.html

- Philippine Competition Commission on Mergers Acquisitions. (n.d.). Retrieved from https://www.phcc.gov.ph/wp-content/uploads/2017/04/PCC-MODULE-3-1.pdf

- Posner, R.A 1976. Antitrust Law: An Economic Perspective. Chicago and London: University of Chicago Press

- Srivastava R.P., Datta D.K. (2002) Evaluating Mergers and Acquisitions: A Belief Function Approach. In: Srivastava R.P., Mock T.J. (eds) Belief Functions in Business Decisions. Studies in Fuzziness and Soft Computing, vol 88. Physica, Heidelberg.