Use of Modern Machine Learning Techniques to Predict the Occurrence and Outcome of Corporate Takeover Events

by

Georges Geha
Master of Finance, Massachusetts Institute of Technology, 2021

Submitted to the MIT Sloan School of Management
at the
Massachusetts Institute of Technology
January 2021

Signature of Author_____ MIT Sloan School of Management

Certified by_____
David Jean Joseph Thesmar
Franco Modigliani Professor of Financial Economics
Thesis Advisor

Accepted by_____
Heidi Pickett
Assistant Dean, Master of Finance Program
MIT Sloan School of Management

This page intentionally left blank.

ABSTRACT

The objective of this project is to use machine learning to predict the occurrence of corporate takeovers. The findings show that random forest yields the best predictions out-of-sample based on the area under the curve (AUC) metric. As such, 8 independent variables are considered statistically significant. A time series machine learning approach is also used at the end of the study to predict these events in 2019 based on each company's data from 2010 to 2018. Random forest is still determined as the model with the best out-of-sample performance. A strategy of investing equal amounts across the companies predicted to be takeover targets in 2019 based on the model yields a profit of 7.4%.

Acknowledgments

I would like to thank the MIT Master of Finance Program and especially professor David Thesmar, my thesis advisor, for his guidance and amazing support on this work.

Thank you to the MIT Sloan Assistant Dean Heidi Pickett and my academic advisor Debra Luchanin for their continuous support and assistance.

Thank you to the MIT Dewey Library, especially Ms. Shikha Shirma for her help in navigating the different databases to collect the data.

This page intentionally left blank.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Project Motivation and Scope

A takeover is when a successful bid is placed to purchase a majority stake to control a company or simply acquire it. It usually falls under merger and acquisition activities. The company making the bid is considered the acquirer and the one that would be controlled is referred to as the target (Cremers et al., 2005).

Takeovers are interesting to individuals or entities looking to control specific companies, but also to investors and funds who want to make profits out of the share price jumps associated with such events. This suggests that this project is useful to different stakeholders who are interested in being able to predict the next corporate takeover in the market for different reasons.

This is done by using statistical and machine learning techniques to predict if a public company is a target for a takeover. There have always been growing concerns on how to merge analytics and corporate finance to be able to predict the occurrence of such events. This project compares the out-of-sample performance of different machine learning techniques and determines the statistically significant variables to look at.

1.2. Literature review

Different studies have used statistical models in corporate finance. Cremers et al. (2005) assessed the impact of takeovers on firm valuation using analytics. Zhang et al. (2012) relied on machine learning to predict the success of takeovers after announcement. Guang et al. (2012) used different models to predict company acquisitions based on data available on TechCrunch.

These studies among others use specific machine learning models to make such predictions. They rely on logistic regressions, LASSO, ridge, artificial neural networks, random forest, CART, kernels, AdaBoost, etc. According to Mullainathan et al. (2017), cross-validation should also be taken into account to choose the parameters with the best average performance. In typical classification problems, certain performance metrics are used such as the true positive rate, false positive rate, accuracy, precision, area under the curve, etc. The area under the curve is usually preferred (Bourne et al., 2019) and random forest is regularly considered to have the best out-of-sample performance (Zhang et al., 2012).

When it comes to the choice of variables, studies rely on very different parameters. For instance, Cremers et al. (2005) relied on a broad set of independent variables from SDC and Compustat including the return on assets, leverage, property, plant and equipment to total assets, cash to total assets, market to book value, etc. They concluded that the likelihood that a takeover impacting the firm valuation increases with leverage, but decreases with return on assets and market capitalization.

Finally, different variables initially used to predict other events can be used actually to predict the occurrence of takeovers. As such, some of the independent variables incorporated by Tunyi (2014) in the model to predict bankruptcies can be used again. The parameters in this case

include retained earnings, earnings before income and taxes (EBIT), working capital to total assets, etc.

1.3. Thesis Organization and Contribution

This thesis is organized into two main parts. Section 2 goes over the data collection process, the different models used and finally the outcome of the analysis focusing on out-of-sample performance. Section 3 details another methodology adopted, a time series approach, to eliminate the look ahead bias in Section 2. Accordingly, the outcome of an investment strategy of $100M is assessed to get a sense of the real impact of such a model. To conclude, the findings on this project are summarized at the end of the paper.

2. Analysis

This analysis is nothing but a classification problem, where the occurrence of corporate takeover events is predicted. For that, data was collected using different sources. Furthermore, several statistical models were used and compared based on their performance on the test sets.

2.1. Statistical and machine learning models

2.1.1. Logistic regression

The first model is a simple logistic regression given that the dependent variable, i.e., the occurrence of a takeover event, is categorical.

A logistic function is increasing and non-linear. It ranges between 0 and 1 with the dependent variable being a Bernoulli random variable. In other words, Y = 1 ("success") if a company is the target of a takeover and Y = 0 ("failure") otherwise. The probability of a success outcome of the dependent variable Y is still predicted without seeking to forecast the value of Y, but rather the likelihood of Y = 1.

The outcome of the model includes p-values for each independent variable showing the level of confidence of how much they are statistically significant. It also comprises coefficients for each variable giving a sense of their incremental effect on the likelihood of the event. These coefficients can be positive or negative (Kotsiantis et al., 2006).

2.1.2. Regularized regression

In a simple logistic model, some independent variables can be removed to reduce the p-values of the remaining ones and enhance the performance of the model. However, this is subject to the modeler's judgement. This is where regularization can come into play and 2 methods were used in this paper: Ridge and Least Absolute Shrinkage and Selection Operator (LASSO).

Put it simply, regularization calculates a lambda to penalize large coefficients. A penalty of lambda equal to 0 leads to the same coefficients as a logistic regression with the best in-sample fit, but probably with an overfitted out-of-sample. In the contrary, a very large lambda leads to coefficients all equal to 0 (similar to a baseline model) with the worst in-sample fit possible and maybe an underfitted out-of-sample fit.

A moderate lambda is the target for both LASSO and ridge. It will provide small coefficients with an intermediate in-sample fit. Therefore, both methods shrink parameters and end up preventing multicollinearity problems in most cases. They also reduce the model complexity by coefficient shrinkage when running a new regression. Nonetheless, there exists some differences between them which are discussed below (Tibshirani, 1996).

2.1.2.1. Ridge

One of the main advantages of this method is that the solution is not as sensitive to noise in the data as the one in logistic regression. However, many variables are kept with annoying small coefficients. In addition, there is not a convenient way to test the significant of coefficients.

2.1.2.2. LASSO

This regularization method is considered more practical and interpretable since a lot of variables' coefficients are taken to 0. Nonetheless, similarly to Ridge, there is no convenient way to test for significant coefficients.

2.1.3. Classification and Regression Trees (CART)

CART is simply a tree-based machine learning model used for regression and classification. It is one of the most interpretable and visual models. It is based on a simple decision rule embedded in a tree following a series of steps similar to human logic. The most "salient" variable is chosen at each node, yielding the highest predictive capacity at this node.

In an attempt to find a trade-off between model complexity and predictive power, k-cross validation is used. It is a resampling technique with the "k" referring to the number of groups that a certain data is going to be split into. The model is then trained and tested multiple times on these sub-groups. This is a powerful and simple method yielding less biased estimates than a typical train versus test splits. In this paper, a 10-fold cross validation is adopted (Biau, 2012).

Other parameters in CART include:

- Complexity parameter (**cp**), which penalizes the model if the tree has too many splits. The higher the cp, the smaller the tree. A very small value leads to overfitting, and the opposite for a very large one. Luckily, the best cp can be found using cross-validation as the one maximizing the accuracy of the latter. In this case, cp is set at 0.002.
- **Minbucket**, which is the minimum number of observations on each terminal node. It is equal to 7 in this model.

2.1.4. Random forest

A random forest is an ensemble of CART trees where a prediction on the dependent variable is made by each tree. Then, all the predictions from individual trees are aggregated and the most common prediction across all trees will be taken.

This is a powerful tool, but the fact that trees are combined make it much less interpretable. One solution for that is to look at the average improvement in the prediction's accuracy across all classification trees using importance metrics. In other words, the independent variables are ranked based on how often each appears in trees, which characterizes the relative importance of these variables (Biau, 2012).

This importance measure is actually considered the average increase in the purity of a node when it splits on that variable. For classification trees, the node purity is measured by the "Gini" index which calculates the probability that a randomly chosen variable is falsely classified. This purity provides valuable information on the main drivers of the outcome. A 10-cross validation is also used for this model, and additional parameters include:

- **ntree**, being the number of trees in the forest. It is set at 500 in this model.
- **mtry,** as the number of variables examined at each node. It is equal to 2 in this case.
- **nodesize,** which is equivalent to minbucket for CART and is equal to 20 in here.

It is worth noting that one of the main advantages of both CART and random forests is their ability to capture non-linearity in the data. Also, random forest usually provides better quality of out-of-sample predictions. However, it is clearly not the most computationally efficient.

2.1.5. Performance measure

For linear regression models, it is common to rely on R^2, root mean square error (RMSE), mean absolute error (MAE), mean square error (MSE), etc. to assess the performance of models. When it comes to classification, different performance measures can be considered. This paper focuses mostly on 3 of them:

- **Accuracy**, representing the proportion of correctly classified observations.
- **Confusion matrix**, including the number of true positives, true negatives, false positives and false negatives among observations. Some rates are calculated using this matrix such as the true positive rate (referred to as the sensitivity), as well as the true negative rate (known as the specificity).
- **Area under the receiver operating characteristic (ROC) curve**, usually referred to as AUC. It is the area under the curve of the true positive rate (y-axis) and the false positive rate (x-axis). The closer the ROC curve is to the top left corner (TPR = 1, FPR = 0), i.e. the closer AUC is to 1, the higher the performance of the model. AUC is used as a single value to compare the performance of models.

2.2. Data collection

2.2.1. Variables

To collect the data, takeovers are considered on a yearly basis because fiscal years are not the same for all companies. Typically, 3 to 4 months should be left at the end of fiscal years to account for this discrepancy and include all the data. For simplicity, end of calendar year is adopted in this case.

As mentioned, the dependent variable is the state of a public company being the target of a takeover.

Different independent variables are also collected. These constitute the inputs for the models. The complete list includes the following:

- **Return on assets** (ROA), as net income over total assets
- **Log of total asset**s
- **Market capitalization** normalized by total assets and book equity

$$\frac{Market\ Capitalization + Total\ Assets - Book\ Equity}{Total\ Assets}$$

- **Leverage ratio**, as total liabilities over total shareholders' equity
- **Long-term leverage ratio**, similar to leverage ratio but with long-term liabilities

- **Total cash to total assets ratio**
- **Basic earnings per share**
- **Total working capital to total assets ratio**
- **Asset structure ratio**, as total property, plant and equipment (PPE) over total assets
- **Book to market value ratio**
- **Goodwill to total assets ratio**
- **Capital expenditures (CAPEX) to property, plant and equipment (PPE) ratio**
- **Total value of common shares issued to market capitalization ratio**
- **Total interest expense to total assets ratio**
- **Total revenues to total assets ratio**
- **Total selling, general and administrative expenses (SG&A) to total revenues ratio**
- **Total earnings to book equity ratio**

It is worth highlighting that different variables from the list above are used to build 2 distinct logistic models. The first one is very simple and includes a limited number of independent variables, precisely the first 6 in the list above. Most of these are used by Cremers et al. (2005). This first model also serves as a benchmark to compare its outcome to other ones. Then, a second logistic model relies on the complete list of these independent variables.

Moreover, the share price jump at the takeover announcement (in %), the deal structure (% of cash versus % of stock), the announcement dates and the % of shares sought are also extracted for analyses outside of these models.

2.2.2. Timeline and databases:

The timeline for the data collected goes from 2010 to 2019. 2 different databases are used to gather this information, as detailed below.

2.2.2.1. Securities Data Company (SDC) Platinum database

This platform is provided by Refinitiv. It is used in this case to extract data on takeover announcements indicating if a company had been a takeover target. Data also includes a target's industry and enterprise value (in $M), the percentage of shares sought, the date of closed transactions, the deal value (in M$) and structure, and the % jump in share price at the takeover announcement.

2.2.2.2. CRSP/Compustat Merged database

This database is included in Wharton Research Data Services (WRDS). All independent variables used in the models are set based on the data from this database. Extracted data points include: net income, total assets, market capitalization, book equity, total debt, long-term debt, total earnings, share price, total cash, total working capital, total property, plant and equipment, sales of property, plant and equipment, total goodwill, total capital expenditures, total value of common shares issued, total interest expenses, and total selling, general and administrative expenses (all in $M). These data points are then used to compute all ratios and logs mentioned

in section 2.2.1. This is an important step to normalize all independent variables between companies of different sizes and characteristics.

2.2.3. Cleaning

Cleaning the data is essential to get sound and precise estimates from the models. A simple method is adopted in this case. All N/A values are removed. Then, for every independent variable, all data points outside 5 times the interquartile range are considered outliers and are therefore removed from the dataset.

Table 1 below shows different characteristics for each variable following the cleaning process.

Variable	Mean	SD	Min.	Max.	p25	p50	p75
Basic earnings per share ($)	1.2	1.5	-3.3	5.1	0.2	1.0	2.2
ROA (%)	4.3%	5.9%	-20.9%	23.3%	1.7%	4.6%	7.7%
Total assets (log)	3.0	0.8	1.1	5.4	2.5	3.1	3.6
Normalized market capitalization	1.7	0.7	0.7	3.9	1.2	1.6	2.1
Leverage (%)	45.3%	41.4%	0%	187.4%	3.3%	31.4%	68.0%
Long-term leverage (%)	43.4%	46.7%	0%	198.7%	1.2%	28.9%	69.0%
Cash to total assets (%)	12.0%	9.9%	0%	43.2%	4.1%	9.4%	17.4%
Working capital to total assets (%)	24.8%	19.2%	-26.3%	84.1%	10.8%	23.3%	36.5%
PPE to total assets (%)	43.8%	35.3%	0.2%	91.4%	17.3%	32.2%	61.9%
Book to market (%)	55.2%	31.9%	4.9%	162.8%	30.6%	46.9%	72.3%
Goodwill to total assets (%)	16.2%	14.8%	0%	51.8%	2.3%	12.5%	27.7%
CAPEX to PPE (%)	9.7%	5.9%	0%	27.6%	5.4%	8.3%	12.6%
Common shares to market cap (%)	8.0%	8.9%	0.2%	50.2%	2.4%	4.5%	9.9%
Revenues to total assets (%)	92.7%	52.1%	3.0%	247.1%	54.7%	81.2%	122.3%
Interest expense to total assets (%)	2.0%	4.4%	0%	7.2%	0.2%	1.1%	2.2%
SG&A to revenues (%)	25.8%	17.3%	0.6%	88.3%	12.8%	21.9%	35.4%
Earnings to book equity (%)	29.2%	81.7%	-306.7%	303.4%	-5.9%	44.7%	80.1%

Table 1: Characteristics of independent variables following the cleaning process

The cleaned dataset includes a total 14,370 companies of which 1760 have been takeover targets, with 50% or more of their shares sought. Figure 1 below indicates the number of takeovers announced and the total corresponding $ value, whereas Figure 2 highlights the total number of deals across different industries.

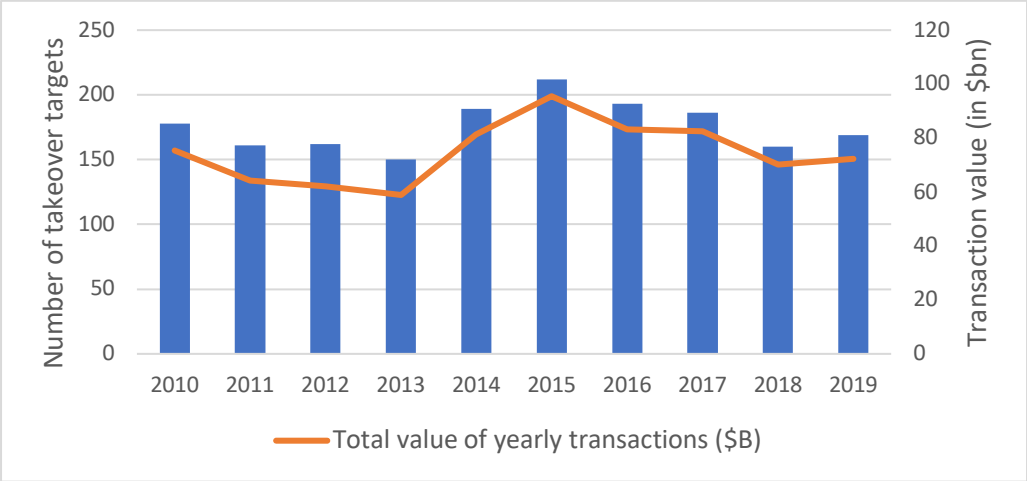

Figure 1: Total number of deals and the associated total transactional value per year from 2010 to 2019

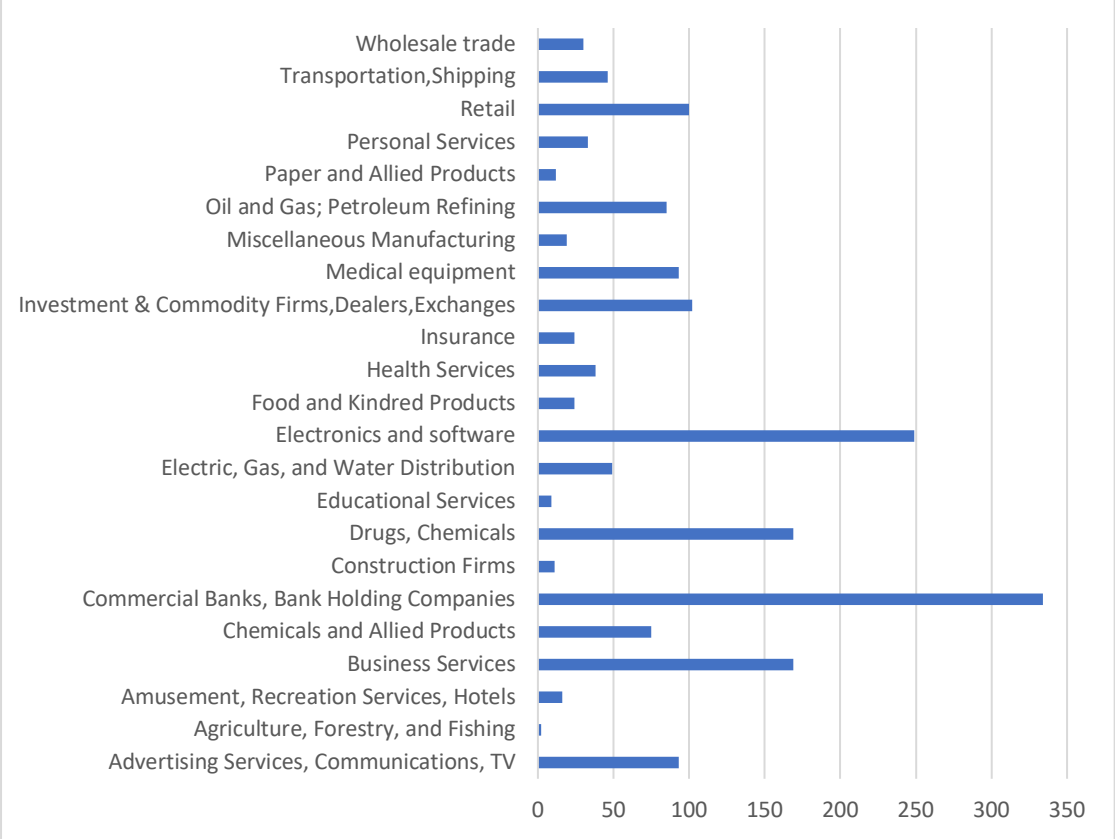

Figure 2: Total number of takeover targets from 2010 to 2019 across industries

Figure 1 shows that 2015 registers the highest number of takeovers announced, but the numbers are overall well distributed throughout 2010 to 2019. It indicates the absence of any bias towards a particular year in the data. The associated total transactional values vary between approximately $65bn (in 2013) and $100bn (in 2015).

Figure 2 designates the most popular industries for corporate takeovers. The top 3 is as follow: commercial banks and bank holding companies in first place, followed by electronics and software in second place then by drugs and chemicals alongside business services in third. Overall, the announcements are also distributed among a broad range of 23 industries in total. This is considered a good and representative sample even if some industries are more popular than others.

Finally, Figure 3 below indicates the number of takeover announcements for different deal structures, i.e., % cash versus % stock.

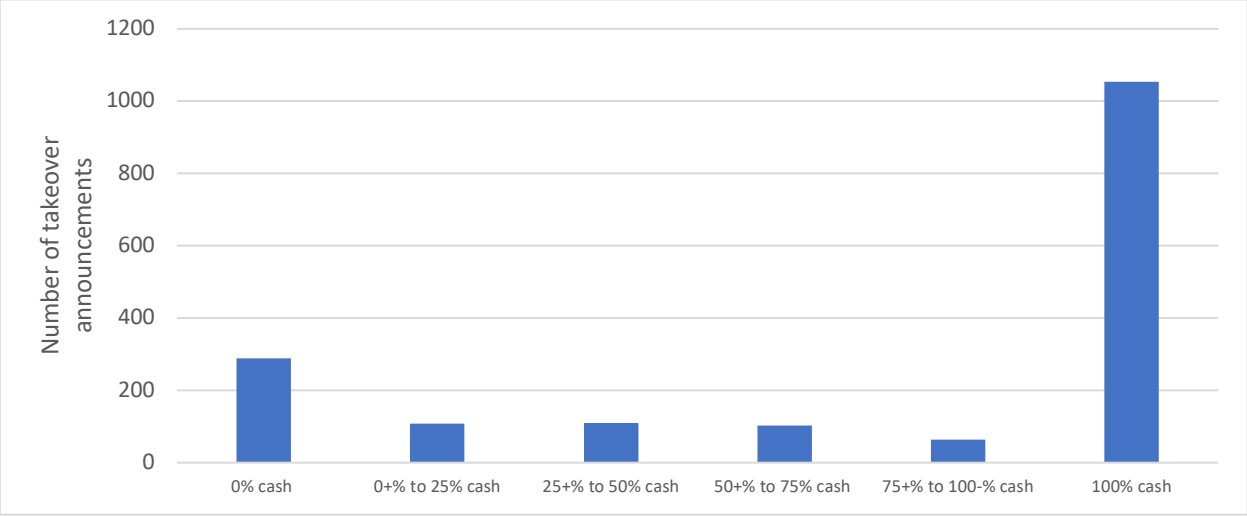

Figure 3: Distribution of deal structures among takeover announcements

Clearly, 100% cash transactions are the most common among deal structures, which is normal within the mergers and acquisition space. Target companies prefer not to bare any risk holding the stock of the acquirer, which will likely be affected by such announcements and transactions.

2.3. Training and testing

The dataset is divided into training and testing cross-sectionally. That said, each entry is treated independently. For instance, a company with data spanning from 2010 to 2019 constitutes 10 independent and distinct data points. The dataset is randomly divided into 70% training set and 30% test set. Note that cross-validation for CART and random forest is conducted within the training set.

In Section 3, the data is divided differently. The models are tested under a time series assumption. The dataset is still divided into 70% and 30% for training and test sets respectively.

However, data points are not treated independently. The models are using all the data provided on a company to make a prediction on this company. In other words, the dataset is divided "vertically". This will be discussed later on.

2.4. Results

2.4.1. Performance

Metrics mentioned in section 2.2.3. are used to assess the performance of all models. The results are highlighted in table 3 below.

Model	Accuracy	Sensitivity	Specificity	FPR	FNR	AUC
Logit with limited variables	82.6%	28.6%	84.9%	15.2%	71.4%	0.56
Logit with all variables	83.0%	31.8%	85.3%	14.7%	68.2%	0.58
LASSO with limited variables	82.9%	32.6%	85.2%	14.8%	67.4%	0.60
Ridge with limited variables	82.7%	31.1%	85.0%	15.0%	68.9%	0.59
LASSO with all variables	82.8%	34.0%	85.1%	14.9%	66.0%	0.62
Ridge with all variables	82.6%	32.7%	85.0%	15.0%	67.3%	0.60
Random forest with all variables	83.1%	35.7%	85.2%	14.8%	64.3%	0.66
CART with all variables	83.0%	34.9%	85.1%	14.9%	65.1%	0.64

Table 2: Predictive performance measures among models used

One can clearly get a sense of the prediction performance of these models based on the values above. All of them confirm that random forest has the highest predictive capacity, having the highest AUC and sensitivity, followed by CART and LASSO with all variables respectively. In fact, the closer AUC is to 1, the better the performance. Accuracy is also important, and it is actually the highest for random forest. However, it can sometimes be misleading. It is calculated by dividing the sum of true positives and true negatives by the total number of observations. This is usually very informative and useful. Yet, imagine the model predicts 80 true negatives and 20 false negatives. The accuracy would be equal to 80%. This seems encouraging, but it does not actually point out that the model has absolutely no predictive power. This is why we mostly rely on other measures including AUC.

For a purely random process, AUC would be equal to 0.5. Obviously, a model with an AUC higher than that is desired. The table shows that the AUC does not decrease considerably for other models compared to random forest. However, it still favors random forest.

2.4.2. Significant variables

Each model determines a set of statistically significant variables A number of them are common between some models. Table 4 below shows the complete list of these variables.

Model	Significant variables
Logit with limited variables	- Log of total assets - Cash to total assets
Logit with all variables	- Log of total assets - Capital expenditure to PPE - Revenues to total assets - Cash to total assets - PPE to total assets - Earnings to book equity
Logit with limited variables (LASSO)	- Log of total assets - Leverage
Logit with limited variables (Ridge)	- Log of total assets - Leverage - Long-term leverage
Logit with all variables (LASSO)	- Leverage - CAPEX to PPE. - Cash to total assets - PPE to total assets - Earnings to book equity
Logit with all variables (Ridge)	- Log of TA - CAPEX to PPE. - Revenues to TA - Cash to total assets - PPE to total assets - Earnings to book equity
RF with all variables	- Return on assets - Log of total assets - CAPEX to PPE - Revenues to total assets - Cash to total assets - PPE to total assets - SG&A to revenues - Earnings to book equity
CART with all variables	- Log of total assets - Leverage - CAPEX to PPE - Cash to total assets - PPE to total assets - SG&A to revenues - Revenues to total assets - Earnings to book equity

Table 3: Statistically significant variables according to each model used

The results from the random forest show 8 significant independent variables. Based on these results, return on assets, log of total assets, revenues to total assets, cash to total assets, PPE to total assets and earnings to book equity have a positive effect on the likelihood of being a takeover target whereas CAPEX to PPE and SG&A to revenues have a negative effect on that. These metrics are therefore considered important to monitor when it comes to takeover activity.

CART shares most of these variables, but considers leverage to be statistically significant instead of return on assets, with a positive coefficient on the prediction of the dependent variable.

3. Alternative exercise

3.1. Methodology overview

The models used above are based on a "horizontal" approach with the data, i.e., all entries are treated independently. In this section, a "vertical" approach is tested where a prediction for a company is made based on the data available for this same company.

All targets in 2019 with available data until 2010 are selected. Independent variables from the first 9 years will be used as inputs for each company to make a prediction for the 10^{th} year (2019). One can think about it as a time series approach using machine learning models. Interestingly, it eliminates the look ahead bias that was included in the previous section. That is considered a more realistic approach. In fact, investors are just provided with timely information that should be used to make predictions.

In this case, 138 targets were identified in 2019 among 1231 companies in total, all of them with data going back 10 years. The models included earlier are used again in this section: logistic regression, LASSO, ridge, random forest and CART.

3.2. Performance and results

The performance of each model is measured according to the same metrics mentioned in Section 2.1.5. Table 4 below provides additional details on that.

Model	Accuracy	Sensitivity	Specificity	FPR	FNR	AUC
Logit with all variables	82.5%	31.3%	85.0%	15.0%	68.7%	0.56
LASSO with all variables	82.2%	34.7%	84.6%	15.4%	65.3%	0.64
Ridge with all variables	81.9%	32.7%	84.4%	15.6%	67.3%	0.59
Random forest with all variables	79.5%	37.7%	84.8%	15.2%	62.3%	0.67
CART with all variables	81.4%	33.3%	84.0%	16.0%	66.7%	0.62

Table 4: Predictive performance measures among models used with the alternative methodology

Even though performances appear to change and improve for some models, AUC and sensitivity indicate once again that random forest is the best alternative when it comes to out-of-sample predictions for this exercise. However, LASSO seems to outperform CART as opposed to the earlier findings in Section 2.4.1.

Note that accuracy ranks models differently. In fact, random forest appears to have the lowest performance, but AUC will be prioritized because of the misleading indications associated with accuracy as mentioned in Section 2.1.5.

3.3. Investment outcome

To test these results in a concrete way, an investment strategy is brought forward. It consists of investing $100M equally across all companies predicted to be taken over in 2019. Then, the financial outcome of this investment is compared to the same $100M equally invested across all companies in the US that have actually been targets of takeover announcements in 2019. The price jumps or drops for each company are retrieved from SDC platinum and used to calculate the financial gains or losses from these investments.

Based on the random forest model and the confusion matrix built, 52 companies are correctly identified as targets in 2019 among 138 in total. At the same time, 166 false positives are indicated. These are wrongly predicted as targets by the model. However, they will still be invested in since the strategy is based on the model's predictions.

Based on the data from SDC, the average jump in daily stock returns for the actual 138 positives is equal to 24.3%. Nonetheless, the strategy does not invest in all these companies because not all of them are identified by the model. Then, the average jump in the model is 19.2% when accounting for the true positives only.

Therefore, this strategy would invest $100M equally in the 218 companies identified but would only benefit from the jumps in 52 of them instead of 138 in reality. In other words, $4.6M would be made from the price jumps in these 52 targets whereas $24.3M could be made if all real targets are identified.

However, one should also account for the remaining 166 companies falsely identified as positives by the model, and in which this strategy invests. When considering the average daily returns for these companies and combining them with the returns of the 52 true positives, the average daily stock return in this period equals 7.4%. Consequently, $7.4M are actually made in total by investing the $100M based on the model developed.

4. **Conclusion**

The goal of this project is to use machine learning techniques to predict the occurrence of takeover events. These models include a simple logistic regression, LASSO, ridge, random forest and CART. They are built using different independent variables and tested by means of several performance metrics. The most important indicator is the area under the ROC curve (AUC), which proves in the first analysis that random forest yields the best predictions out-of-sample.

Another methodology is used relying on time series data to eliminate the look ahead bias for potential investors. Random forest is found once again to have the best performance out-of-sample. This was tested by a strategy investing $100M proportionally across all companies identified as targets by the model. It turns out that $4.6M are made from the price jumps in the companies correctly identified as targets by the model, and $7.4M across all investments.

Bibliography

[1] Biau, G. Analysis of a Random Forest Model. Journal of Machine Learning Research. 2012.

[2] Bourne, M., Dircks, T., Kardatzke, J., Miller, D., Roberts, M., White, C. From ML to M&A. Nicholas Center for Corporate Finance and Investment Banking. 2019.

[3] Cremers, K. J., Nair, V., John, K. Takeovers and the cross-section of returns. Yale International Center for Finance. 2005.

[4] Gu, S., Kelly, B., Xiu, D. Empirical Asset Pricing via Machine Learning. Yale International Center for Finance. 2018.

[5] Kotsiantis, S., Zaharakis, I. D. Machine learning: A review of classification and combining techniques. Artificial Intelligence Review. 2006.

[6] Mullainathan, S., Spiess, J. Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives. 2017.

[7] Tibshirani, R. Regression Shrinkage and Selection via the LASSO. Journal of the Royal Statistical Society. 1996.

[8] Tunyi, A. Takeover likelihood modeling: target profile and portfolio returns. University of Glasgow. 2014.

[9] Xiang, G., Zheng, Z. A Supervised Approach to Predict Company Acquisition with Factual and Topic Features Using Profiles and News Articles on TechCrunch. Carnegie Mellon University. 2012.

[10] Zhang, M., Johnson, G., Wang, J. Predicting Takeover Success Using Machine Learning Techniques. Journal of Business and Economics Research. 2012.