Master's thesis

Submitted in fulfilment of requirements for completion of Double Master's Degree

Programme Grande Ecole - 3rd year (PGE 3) / Master of Science International Finance (MSc IF)

How Artificial Intelligence can shorten the pre-M&A process: on the interest of using a Supervised Machine Learning approach to predict firm's acquisition.

-

Adrien BONHOMMEAU

ETU20170447

30/09/2021

Received Grade: 91 / 100

16/05/2022 Supervisor: Dr. Amirhossein SADOGHI

Final Version

Oath of personal work

I undersigned Adrien Bonhommeau, declare that the following graduating project is my own work. No part of this research has been submitted in the past for publication or for degree purposes. I am fully responsible for the truthfulness of this declaration.

Date: 24th September, 2021

Signature:

Foreword

The COVID-19 crisis has allowed us all to reflect on certain subjects, and in particular the contribution of technological solutions. Prediction has always been one of human's most fantasized subjects. Invoking divinities, the interpretation of star constellations, cartomancy, or statistics and probabilities for the most rational, predicting future events is a delicate thing, especially in the long term. In modern finance, predicting what will happen in a few minutes, or even seconds, is already a feat and guarantees the investor potentially huge gains while minimising risks. This is what we are trying to understand in this thesis. Advances in predictive systems, combining computing power, large amounts of data and sophisticated mathematical models, may lead us to believe that we will be able to predict the outcome of major events in the same way that we predict the weather tomorrow. Artificial intelligence is no stranger to this, and in particular its predictive branch: machine learning. Applied to the financial sector, some might hope to reap great benefits while minimising risks, especially in the exciting and complex business of mergers and acquisitions. Having an algorithm that can predict whether a deal is likely to succeed or fail would be a Holy Grail for any acquirer or investment bank. It would drastically reduce the time and energy spent by the teams involved in setting up the project.

I hope you will enjoy reading it.

If you want to know more about this work, please feel free to reach me on my personal email: adrien.bonhommeau@gmail.com.

Acknowledgments

This work concludes my academic journey at Rennes School of Business.

I would like to thank the most important architects of my journey: my teachers and professors. Their advice, their knowledge and their wisdom sharing is definitely what I will keep in mind for a long time.

I would like to thank my parents, that have always supported me, and their advice have been the foundations of who I am today: someone curious, trying to learn each day more and more.

Thanks to my tutor, Dr. Amirhossein Sadoghi. He captivated me from the very first class in data sciences. He opened my eyes to the almost infinite possibilities offered to us today by predictive analysis technologies and tools, which are already structuring the world we live in without us even knowing it. I sincerely thank you.

I would like to thank Rennes School of Business for having allowed me to evolve in an environment as rich as possible, for having been able to be part of its Junior-Enterprise Bretagne Conseil. I had a high-quality education and was confronted to the diversity of other cultures that it offers in its teaching.

Key words

Merger & Acquisition, Firms Screening, Due Diligence, Value Prediction, Machine Learning, Supervised Learning, Logistic Regression, Support Vector Machine, Decision Trees, Lower Risk, Predictive Analytics.

Abstract

In this paper, we investigate the topic of M&A, and more specifically its predictability using supervised machine learning techniques. M&A transactions are known to be particularly complex, and to mobilise significant human and financial resources, and involve many different actors. It is interesting to study the factors influencing the outcome of these transactions in order to reduce the risk of bad targets for companies wishing to use such solutions to grow. In order to understand the problem as well as possible, we first explain what an M&A deal is and why companies resort to it. The currently achieved work have often used Logit and Probit techniques, which are often inaccurate when dealing with non-linear analyses, such as M&A deals. In order to analyse our dataset, coming from deals in the industrial sector (aerospace, manufacturing...), we have created an algorithm in Python language using the main supervised learning techniques. The results obtained confirm the inadequacies of linear models, and highlight the relative efficiency of decision trees to predict deals successfulness.

Contents

1 Introduction

"Some people call this artificial intelligence, but the reality is this technology will enhance us. So instead of artificial intelligence, I think we'll augment our intelligence." – Virginia Rometty, former IBM CEO from 2012 to 2020.

It is common belief to think that Merger & Acquisition (M&A) deals are key strategic options and are mostly the best way to take positions into a new market. However, M&A are rarely creating value for acquirers (Martin, 2020). Indeed, the 2014 Microsoft $7.9 bn acquisition on Nokia handsets have perfectly shown the issue: after writing off 96% of the firm's value, the acquisition has been recognised to be one of the greatest failures in the M&A world. According to Martin (2020), the M&A deals can actually reveal failures at 70 to 90%. The misinterpretation of the firm's value and their current situation results thereafter in a fatal bidding war. Therefore, the M&A prediction researches number has been growing for previous year, enhanced by the AI-driven analysis democratisation.

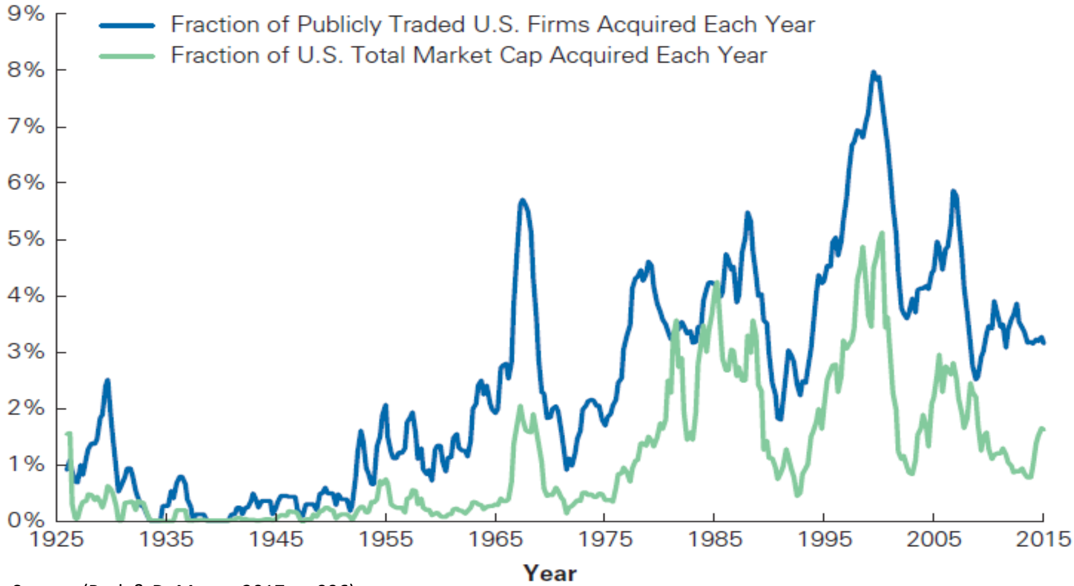
This study has been motivated by the increasing amount of data analysts and financial institutions must compute and analyse to check and run before considering a firm as a potential candidate for a merger and acquisition operation. For the last decades, a large number of M&A deals have been reported to be under of overvalued, especially because the acquiring candidates are artificially increasing the price to carry off the targets (Zhang, et al., 2020).

Since 1990, Berk & De Marzo (2017) have identified six majors M&A trends that have been linked to the real economy. From 1992 to 1999, with a rapid expansion of companies in the tech and internet sectors. From 2000 to 2002, with a contraction in the number of deals due to the bursting of the internet bubble. From 2003 to 2007, with a recovery in M&A, driven by consumer spending and the increase in consumer credit. From 2008 to 2009, with the subprime crisis and the sudden stop of the economy, aggravated by the collapse of the banks and the freezing of loans. The period 2010 to 2014, with an uneven economic recovery and size differences between the various M&A deals, marked by low inflation and historically low interest rates. Finally, from 2015 to 2020, with record deal sizes and then a sudden halt due to the global covid-19 pandemic. All those periodic trends show that after each crisis, M&A strategic option is translated as a compelling come back.

1.1 What is a Merger and Acquisition Deal?

In 1985, 347 merger and acquisition deals have been realised. In 2019, more than 3,700 firms or business units have been acquired by another one, which is 10.66-time take-off in 35 years.

Figure 1: Fraction of U.S. Public Companies Acquired each year, 1926-2015



Source: (Berk & DeMarzo, 2017, p. 996)

It is now accepted that a company can achieve sustainable growth with a well-thought-out strategy. Of course, there are many parameters to consider when defining a good strategy. In The Strategic Vacuum (2012), Baumard presents a rational view of strategy by comparing it to a vector, an object that would move an organization from state A to state B. Baumard nevertheless points out that great leaders such as Lou Gerstner, then CEO of IBM, said that "the last thing IBM needs is a mission statement", while Colin Marshall, Chairman of British Airways, said that strategic plans are not only indispensable, but also a sign of the company's robustness: "the mission of the organization is more than just good intentions and nice ideas. It represents the overall framework of the company's business, the values that drive the company, and the belief in what the company has in it and what it can achieve". Consequently, there are many definitions of strategy, and this concept could only be partial, universalism paradoxically removing all its substance.

From this perspective, it should be noted that strategy is nevertheless one of the milestones of a company's competitiveness. Corporate strategy is the result of a meeting between the current state of affairs and the necessary destiny in which management wishes to orient the entire organization. Whether it is defensive, structuring or conquest, a company's strategy must allow it to consolidate its activity and to engage its transformation at the same time as its environment. In this context, horizontal diversification or vertical integration developments can be achieved through disposals, acquisitions or even mergers with or between similar companies.

These transactions are compared to a puzzle, in which the main stakeholders must assemble vast and some smaller pieces, trying to mortar them together. Even a small mistake within the assembly might run to a failure in the best case, in a financial disaster in the worst. The definition plan as well as the execution must be perfectly managed, thorough fast-paced processes.

Today, investment banks and other financial or consulting firms may now engage powerful tools in a “24/7” mode to watch the market, dynamically adjusting investment strategy, risks, stock price, synergies… That implies that processing to business intelligence is becoming easier and easier thanks to Machine Learning democratisation. Nowadays, such tools must be a main concern for the corporate leaders if they want to keep up with competitive advantage. For instance, Accenture thinks that using Machine Learning can reduce time spent on researching and evaluating firms from 50 to 60%, with Natural Language Processing (NLP). Nevertheless, one of the toughest challenges is to convince the C-levels that an opportunity must be taken. Therefore, providing creative insights about an industry and cleaning the data room to make it clearer for wiser decisions is the main challenge.

1.2 Why do firms resort to M&A?

There are many reasons for management to embark on an M&A deal. The first is the reduction of the consolidated production costs, the search for synergies within the new entity and the acquisition of new market shares. This also includes the acquisition of patents, know-how and knowledge. The second is the desire to invest money in a project. Indeed, latent money in a business can be a weakness and an opportunity cost for the company, since it does not generate cash flow. Finally, other reasons may be in the buyer's desire to benefit from an optimal capital structure and to leverage its tax shield. Nevertheless, one of the most important characteristics belonging to the acquired targets must be their business model potential. Either needing adjustments or any correction, acquired firms must present some prosperities acquiring firms can benefit from. Otherwise, the acquired firm shouldn't be presented as a good opportunity for the acquirer. When it comes to the effectiveness of a good M&A deal, there are several possible combinations. The typical case is a company that is growing well and whose market value is not too high. In this way, the buyer can continue to benefit from the growth effect and the maturity of the business model (Brealey, et al., 2017). Nevertheless, it is tempting for the acquiring firms' management to believe a takeover will result in an absolute target's strengths acquisition, which is far to be verified when deal occurs (Larkin & Lyandres, 2019).

1.3 How are M&A deals organised?

A lot of time and energy is spent on completing the M&A operation and then embedding it successfully into the buyer's organisation. But for a serial acquirer, the initial phase of building a pool of potential targets is the most critical phase of the acquisition process. As a component of the strategic planning process, a company should attempt to identify its current business position and its future direction. Then, in order to meet these new business objectives, a gap analysis will be compiled to determine the requirements necessary to take advantage of the new business opportunities. The following step is to determine how best to meet these requirements: organically using existing capabilities, forming a joint venture or strategic alliance with a friendly partner, or seeking a merger or an acquisition. Therefore, the business strategy is the first part of the M&A process. The M&A strategy aims to align the strategic vision with the business objectives management must meet in order to achieve the desired results. The M&A strategy aims to align the strategic vision with the business objectives. It basically answers the question "Why and for what?". Based on the buyer's business goals and objectives, the business strategy is the first step in the M&A process.

Based on the buyer's business objectives and desired benefits, a first and extensive list of companies that can offer the geographic, technology, market or patent products identified in the gap analysis is established. The main difficult is to ensure the quality of the preliminary list, as it will be used as a basis for the next steps. The information provided by the buyer and the desired benefits are then compiled. Other external databases and information sources may include industry contacts, trade unions, and web searches for targets and their sectors.

Financial data is found in specialised databases such as Thomson Reuters, Orbis or Bloomberg for the Internet, or directly in the company's annual report.
Afterwards, two main different types of stakeholders are clearly identified: the seller and the buyers. A third type of player comes, especially when the acquisition process takes shape and the due diligence stage looms: the underwriters. In general, the process of cross-checking information is long and tedious, and teams of external mediators are mostly called in as back-up. These support teams are composed of lawyers for the legal part, auditors and finance banks for the financial part, and consulting firms specialised in strategy for the commercial part (Berk & DeMarzo, 2017).
In a nutshell, the overall process is articulated as follow: at first, the management board define the growth strategy, related to the corporate strategy. Then the right markets and targets must be assessed, which means some data are available through business intelligence. After this step comes the fast-paced due diligence process during which stakeholders meet in the data room. The planification and its execution is absolutely key at the moment since they will structure the rest of the deal. When parties agree, the integration phase begins and the communication starts. It requires well established resources (financial, operational and management).

1.4 Why do some M&A deals fail?

Due to uncertainty, deliberate management behaviour on cheating and hiding information, or due to human judgemental errors, as well as hubris, it is getting harder and harder to obtain a fair value for the other firm. Those management behaviours altered reported quality and might directly affect the negotiation quality of the deal, while putting pressure on firm’s liquidities. Although M&A deals are extremely complex and involve a multitude of players, ranging from audit to compliance teams, they are nonetheless highly biased. The outcome of the negotiations is based on the quality of the due diligence process, upstream of the negotiations, and the quality of the financial modelling and growth prospects resulting from it. Moreover, one important and non-negligeable factor in such negotiations is the overconfidence of bidders. Indeed, many studies (Bing-Xuan, et al. (2008); Aktas, et al. (2009); Renneboog & Vansteenkiste (2019)), have clearly reported the existence of judgmental biases when completing a bidding offer and when starting negotiations in such

M&A deals. Overconfidence of CEOs is explained in previous outstanding financial results or in their ambitious corporate strategic program. Since CEOs are claiming the good performance of their ruling business, they are more likely to take more risks to increase their global return, and then their bonus and prestige.

A classic M&A scheme groups different processing and negotiation phases within the due diligence process. There are for instance the phases of Financial Due Diligence, Legal Due Diligence and Commercial Due Diligence. Each of them includes their own experts and gather various and large amount of data, often unstructured ones. In that perspective, research teams whose main tasks are to find information and carry out data crunching, spend most of their time carrying out operations with low intrinsic added value. In fact, more than half of a Merger & Acquisition analyst's time is spent monitoring and reporting on databases and importing data into Excel for modelling purposes. In the past, the value of companies was often studied through the latest financial reports, issued by the companies themselves and audited by the world's leading audit firms (PwC, Deloitte, EY, KPMG). Even today, analysts still use those reports, but put them in perspective with the latest "breaking news", or the latest announcements by Twitter influencers, Bloomberg or Refinitv's Analysts, on the prospects for growth or speculation on a sector of activity.

There are two main types of M&A transactions: vertical integrations and horizontal integrations. Vertical deals are the result of an integration strategy on the part of the originating company. In general, these deals are motivated by the desire to gain control over the value chain, such as the acquisition of knowledge or specific technology transfer. These vertical operations often make it possible to reduce costs and optimise operations for greater efficiency. On the other hand, horizontal deals are more motivated by the desire to capture more value and are mainly driven by the acquisition of a player over one of its often-direct competitors (Renneboog & Vansteenkiste, 2019). These horizontal deals are sometimes suspected by antitrust authorities of fostering market imbalances. This concentration of power may even lead to imperfections both for consumers and for the employees of the company itself.

In a nutshell, M&A activities can be considered as a consolidation of assets for an activity development at a given moment. Strategically important for any developing business, those non-organic choices have a key priority and require granular analyses. Any study on M&A has demonstrated how complex were the financial arrangement of such deals, as well as the difficulty to leave the *statu quo* to find the *consensus*.

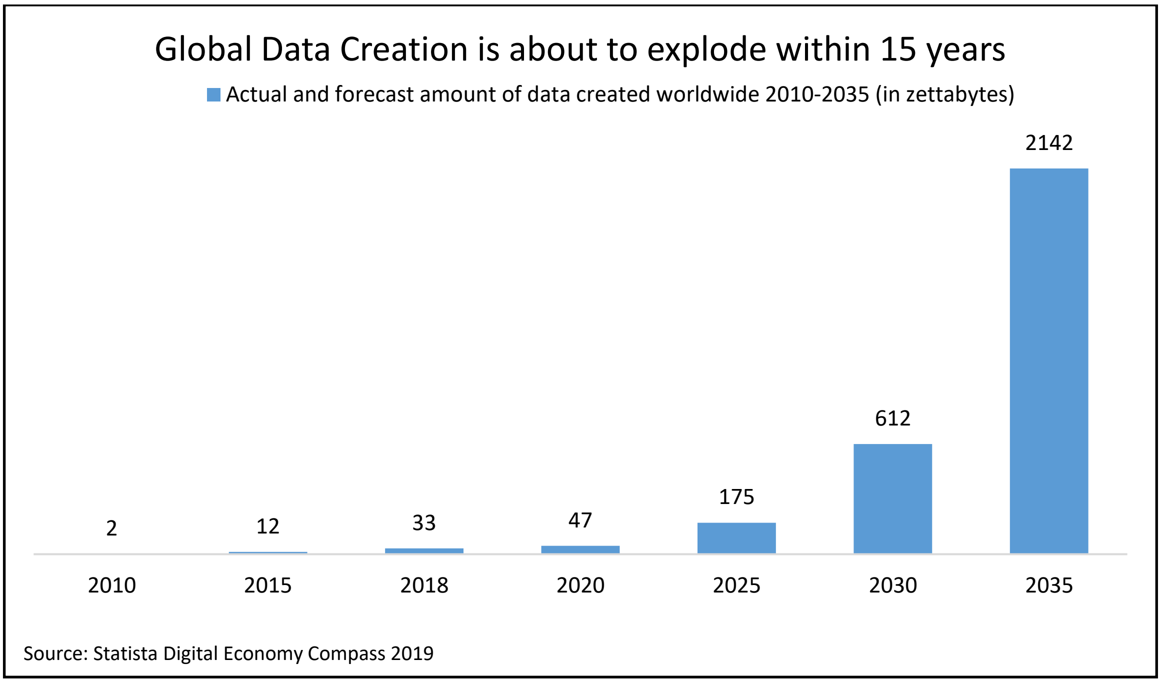
It should be noted that among the many variables to be taken into account when considering a merger and acquisition, the approach of entity A (the buyer) to entity B (the target) is critical to the progress of the negotiations, and therefore to the conclusion of the merger. In September 2020 the management of Veolia, the world's largest water and waste management company, indicated its intention to make a takeover bid for its direct competitor, Suez. The price per Suez share was set at €15.50, then €18, but this was considered too low by the majority of shareholders, which provoked an angry response from Suez's management and prompted the French Minister of the Economy, Bruno Lemaire, to rule (Financial Times, 2020). Only a few days after the official launch of the takeover bid for Suez, Veolia decided to stop the process, in the face of the mistrust of the financial markets and the European competition authorities. A few months later, in May 2021, the situation eased and the parties financially found common ground to increase the offer price to €20.50 per share. Immediately, the managements jointly declared that the combination of their entities was a marriage of convenience in order to strengthen each other and to align their strategy. The shares of the two companies then rose by 30% for Veolia and 20% for Suez (Bloomberg, 2021).

This example alone sums up the decisive role played by the approach strategies chosen by buyers. If the financial markets are confident, success is only partially guaranteed, since the share price might be overlapping the anticipated benefit from the merger (Betton, et al., 2014). If the markets are suspicious of the viability of the offer or doubtful about the takeover project, the penalty is immediate and the takeover costs may explode. On the other hand, a hostile takeover would lead to increase mistrust on the part of employees, for fear of the synergy search that management would carry out.

1.5 Why does Machine Learning represent a more and more prevalent option to assess business success?

Nowadays, statistical learning, through Machine Learning, offers the possibility to compute massive amount of data in order to find and model hidden patterns within the datasets (Mueller & Massaron, 2019). Many choices could orient the research study, such as the dataset nature: structured or unstructured. Today, humanity produces large amount of data at an astonishing speed. According to Statista, the world will produce 46 times the current data production within the next 15 years (Statista, 2019).

Figure 2: Global Data Creation is about to explode within 15 years



According to estimates published in Statista's Digital Economy Compass 2019, the annual volume of digital data created globally has increased more than twenty-fold over the past decade and was expected to approach 50 zettabytes in 2020. This amount of data pales in comparison to what is expected over the next 15 years. Forecasts predict a three- to four-fold increase in the annual volume of data created every five years. With this exponential rate of growth, the astronomical threshold of 2,000 zettabytes should be crossed by 2035. The increasing democratisation of connected objects and the advent of 5G and IoT (Internet of Things) technologies will be the main drivers of this "big bang" of digital data.

Machine Learning is actually not new trend in business analytics. Previously known as Statistical Learning, it gathers sophisticated technics for statistical treatments using most of the time a supervised, unsupervised or an ensemble approach. The main aim is to manage vast amount of data and training an algorithm to find alone the hidden patterns and draw the inductive links between the variables of the dataset. Such tools are now totally available and easy to use for anyone who wants to report market trends or any deep financial analysis. A written press article appeared in Forbes Technology Council and written by Eric Braun (2020), exposed how machine learning is already revolutionizing the financial sector and pretends it is only the beginning. The increasing power improvement of computing and the massive reserve of data provide proficient bases for AI-driven processes, especially for finance and strategy, as well as for their common progeniture: Mergers and Acquisitions. Another recent press article appeared on Techcrunch and written by Rust Wiley (2021) explains that Artificial Intelligence and Machine Learning are already playing, and will continue to play, an increasing role in reducing pre-deal process time. Due diligence processes are often critical to the success of an M&A deal, and their duration is often inversely proportional to the chances of a successful engagement. AI techniques now allow the various parties involved - commercial, legal and financial - to reduce the time required to prepare, evaluate and make available matters by a factor of six. The author also notices that firms that do not invest in such tools are already late and endanger themselves their own business.

Finally, the growing acceptance of AI and ML processes is highlighted in a publication by McKinsey (2021). According to the Partners Shenai and Krishnakanthan, CEOs used to let only financiers run the business units, but now they will be forced to choose financiers who understand the challenges of IT and the contribution of data science technologies.

To date, few articles have been published on the impact of Machine Learning in M&A processes. Since the democratisation of techniques, made possible by computing power and the improvement of analyst training is still in its infancy, it remains difficult to obtain a reliable and objective view of the use of the processes we will study in this paper.

1.6 What is a meaningful M&A target?

As we have said, M&A deals are particularly complex to implement, requiring a lot of work from the strategy, finance, legal, IT and HR teams, as well as a massive investment from the management. It is therefore a question of maximising the return on the time and resources devoted to the whole operation. But how should we define the perfect target for an acquirer? Is there a method to ensure that the deal does not fall through? Obviously, the answer is no. But then, are there any dimensions that are difficult to take into account, to accurately assess? Renneboog & Vansteenkiste (2019) have found that despite the huge variety of M&A deals, it is pretty hard to assess long-run performance based on the short-term performances given the VUCA (volatile, uncertain, complex, ambiguous) context. Renneboog & Vansteenkiste also assess that post-takeover serial acquisitions, CEOs' hubris, business models complementarity and shareholders' baheviour have a major impact of the success of the M&A integration. Such factors are extremely hard to determine since they are mostly related to human behaviour and judgment, so uncertain and complex by nature (Ferris, et al., 2013). De Bodt et al. (2018) bring evidence of such impacts when they studied the existence of overbidding in M&A contests. According to their research, CEOs' narcissism improves misjudgment in M&A deal, and they suggest it could even lead to strategic targeting misleading. Nevertheless, other factors must be mentionned such as the market momentum: is the chess move the most efficient to do given the context? Does the company have the choice? A lot of uncontrolled and unassessable risks surround strategic moves, and as for any investment: there is no free lunch.

The paper is structured as follow. Section 2 reviews the literature, with a focus on how M&A can bring value to firms and why they represent a competitive advantage for the acquirer. Section 3 is dedicated to the methodology and explains the different steps we will check to pursue with the analysis (both qualitative and quantitative parts). Section 4 tackles with the data collection and the mainstream source we get figures from. Section 5 contains the application detailed. Section 6 presents the model analysis and the results we get from them. Finally, Section 7 ends paper results and concludes the work. Section 8 and Section 9 respectively presents the works cited and the table of figures.

2 Literature Review

2.1 Does M&A deal bring any value?

Many studies have already sought to know the impact of mergers and acquisitions on the wealth of shareholders, the growth of companies, as well as their *a posteriori* performance. The literature on mergers and acquisitions has already focused on defining what a successful M&A deal is. Is it a deal in which the buyer has acquired the target for a lower price than analysts thought? Is it a deal after which the acquired entity was profitable after all, supporting the synergy thesis? Is it a deal that enriched the buyer's shareholders?

A very recent meta-analysis presents mergers and acquisitions as a major factor in improving the financial, commercial and legal performance of companies (Hossain, 2021). This is due to the synergies achieved by the acquired company, the market power in which it operates, and the profitability of the new entity. Of course, risk diversification and management strategy usually play a major role in the success of the overall operation and the future of the entity. It is said that companies that share a common cultural mindset are more likely to merge and could generate a more positive impact. Indeed, acquiring companies with boards and directors that are more internationally minded perform better in cross-border acquisitions, not least because they have cultural reading skills, especially if CEOs have already lived in those countries. In addition, the personality of the CEO also plays an important role in the conduct of such deals, especially when he or she is dominant and aggressive (Ding, et al., 2021). When it comes to determining profitability across different industries, M&A has been shown to have the most profitable effect on the IT industry sector, not least because it adds more value in the post-integration phase than for other sectors. A guideline for companies involved in M&A is to build on their business strengths and eliminate their weaknesses. The study also shows that the outcome of an M&A deal is strongly correlated with other characteristics, such as a high level of integration. Therefore, companies on both sides need to understand each other's objectives, culture, capital structure (asset-light, liability and equity allocation), business environments, business models, and more.

Powell & Stark (2005) have identified that an M&A operation is successful when either the new entity succeeds in increasing its share price (which happens quite rarely in the M&A world), or when the combination succeeds in generating a significant increase in market share. A second branch of research focuses on the different implications a bidding offer has on all the stakeholders, and their key determinant capacity to understand whether the process will go further or not. A last branch of research defines success in the simplest of ways: a successful negotiation (Branch & Yang, 2003). We propose to keep this definition to tackle our study.

A recent study assesses different data to determine whether or not mergers and acquisitions play a major role in shareholders’ wealth and in the growth and performance of acquiring companies. Gupta & Banerjee (2017) found that using an M&A strategy positively influences operational efficiency and economies of scale by compressing costs. The authors also found that these business buyouts reduced competitiveness and confirmed the positive performance of the company and the market as a whole. On addition, existing literature shows that firms valued more than $500 Mn and with a significant experience in M&A strategy succeed better into acquiring further firms and also creating profitability and then value for shareholders (Hu, et al., 2020). Nevertheless, de Bodt, et al. (2018) have provided evidence that most of M&A offers are over evaluated, which conduct to overbid the whole operation price. According to their results, such mispricing is mostly related to conflict of interest as well as a strong tendency for winners to bid in an auction that exceeds the target's intrinsic value, due to asymmetrical pieces of information and management hubris. Such mispricing is called "winner's curse". These results have been confirmed by Duchin & Schmidt (2013), and Bo & Anand (2021). Duchin & Schmidt (2013) have actually been showing that mispricing occurs much more often during merger waves, when acquisition operations massively increase. They show that for that special periods, analysts' are forecasting poorer quality reports, making more mistakes, especially due to uncertainty and high volatility of stocks and income results. As a result, Duchin & Schmidt obtain worse results for M&A deals in the long-run than they could be in growing periods. Bo & Anand (2021), using an M&A activity variable to measure the returns of the 12 Fama-French industries, suggested to the researchers that companies with the least M&A activity tended to undervalue enterprise value by 10.3%, while the most M&A active sectors tended to overvalue targets by 13.46%. Evaluating the work of the business analysts and investment banks working on a target-side in the takeover, Eaton, et al. (2021) have found that there is a significant bias on the price target fix to accept the bidder's offer. According to the authors, banks choose top class firms for peers and multiples analyses, and advice a higher takeover price, not in touch with the market, to extend their own bonus. This confirms the thesis that the market is out of touch with reality.

Although complementarity between the targets' and acquirers' products, techniques, know-how and technologies is often a key driver of M&A operation. Associations between firms with similar products and offerings generate better return, bidders offer are more likely to be accepted by the targets and a better business model integration within the new entity's brand portfolio (Bena & Li, 2014). Bena & Li (2014) also mention the reversal pattern between R&D expenses: it is most common to find an acquirer with low R&D investments and a target with high R&D expenses. Indeed, biggest firms have no interests being R&D pioneers but focus the commercial and marketing area whereas smallest entities are more able to disrupt the market with lower means and costs, but they can also their R&D pattern registrations to be pointed out by bigger firms (Phillips & Zhdanov, 2013). As a consequence, innovation has been mentioned to be an important acquisition factor. Larkin & Lyandres (2019) have demonstrated that most of the mergers are inefficient when it comes to find relevant synergies, due to the information friction and lack of relevant information for the bidders. According to their research, the authors suggest that the belief that an acquiring company would exploit the strengths of the target company is far from true due to search frictions. Firstly, because there will be synergies between the business models of the two entities, but also a possibly important part of cannibalism. This part is often neglected by acquirers, because it is difficult to estimate its true value. Secondly, because acquirers are blind to the real risks of lack of information or hidden information during the due diligence phase concerning the real weaknesses of the target. In the best case, the buyer is aware of or has little control over the weaknesses, in the worst case it realises it when it is too late and the deal will be his undoing. As an illustration, this is exactly what happened when Microsoft acquired the Nokia's telephony entity in 2014. Finally, the authors believe that the M&A market is not optimised because it does not most of the time enable complementarity gains between the stakeholder firms. According to Larkin & Lyandres, very few bidders find the optimal targets within the market, these that will optimise the post-integration value of the new formed entity. Investigating this path, Humphery-Jenner & Powell (2014) have demonstrated that the commonly accepted belief that biggest firms perform worse after acquisition is internationally true, but a non-negligeable nuance exists when the market governance is weak, typically in developing countries where regulation might be less restrictive than in developed economies. Complementary to that last study, Shen, et al. (2021) have shown the higher the geopolitical risk, typically in emerging and developing countries, the more likely it is that companies will adopt an aggressive M&A strategy, particularly in the energy production sector that the authors have targeted. According to Shen, et al, acquirers will primarily target other stakeholders in the production chain for the simple reason that this will help them to better control the risks of sourcing and selling their production, and to obtain better synergies. The propensity to acquire companies is positively correlated to local corruption of such areas, where the corruption and geopolitical risk are higher. In contrast, firms targeting is negatively correlated with it. Indeed, when seeking to make acquisitions, buyers tend to pay in shares, certainly because all players prefer to keep low cash reserves. Their debt ratios are also higher, again for the same reasons. On the other hand, when these companies have more liquidity, they will use it for the payment of the transaction in order to protect the liquid assets from expropriation, and making it more difficult to collect (Nguyen, et al., 2020). As a consequence, M&A have also a significant relative value due to the economic and political context in which they belong.

One other area to explore is the fiscal and financial regulation area. Actually, Blouin, et al. (2020), use a difference-in-differences approach to understand the Domestic Production Activities Deduction (DPAD), related to IRC Section 199 of the US regulations. Despite the low tax rate cuts, which never exceed 3.01%, their effect on mergers is significant. The researchers' tests seem to prove that companies benefiting from Section 199 deductions are more prone to acquisitions. In addition, a one standard deviation increase in the DPAD tax rate reduction results in a 1.39% increase in the amount of US dollars that firms spend on acquisitions. The authors are almost certain that DPAD has a major impact on M&A transactions in the US, given the benign tax regulations.

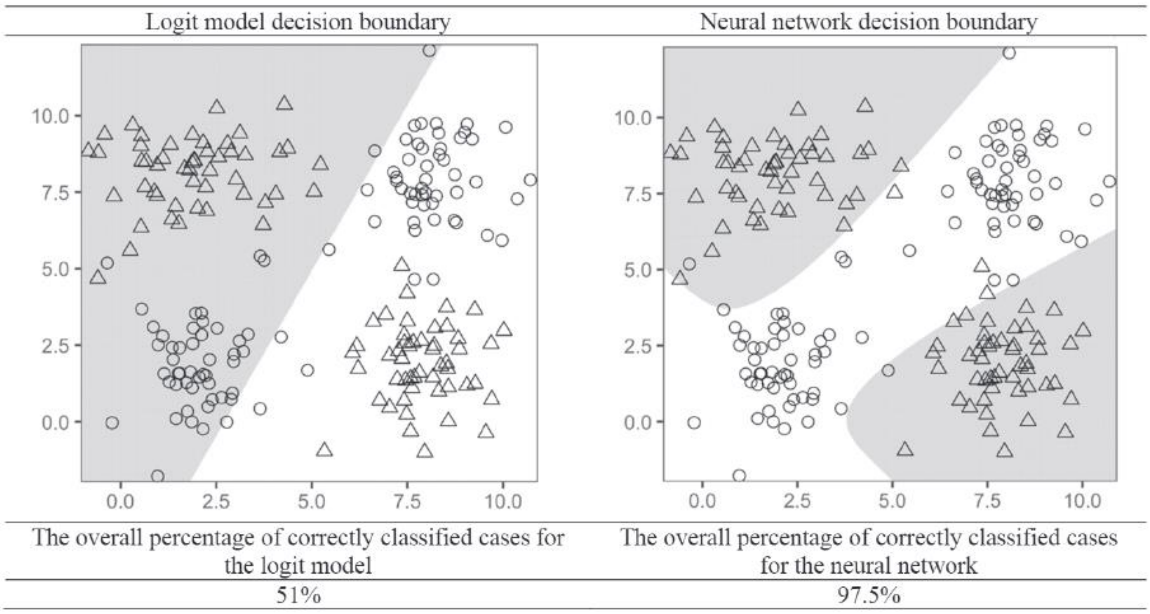
2.2 To what extend statistics could provide evidence on a deal quality?

In order to understand the subtle difference between Statistical Learning and Machine Learning, it is necessary to explain how they work. Statistical Learning is first based on a statistical and empirical analysis of a sample of observations. The input, combined with the marginal laws governing the whole dataset, allows to provide an output and projections, predictions. On the other hand, Machine Learning resonates with the opposite logic: it compares the input and output data to determine which laws govern the set of variables in the sample.

Computer takes larger and larger place in companies and in business in general as almost every transaction must be recorded in a centralised system. The increasing size of the data requires more sophisticated and powerful computational power to get the best from their analysis. Moreover, it is now common to discover or qualify until now unknown parameters into predictive models. The same thing happens with variables. Therefore, keeping regression analysis to tackle wider and wider dataset seems odd. Machine Learning technics have therefore a lot to offer as algorithm are able to manage amount of data even the best analyst will never success to. Support Vector Machine, Decision Trees or Neural Networks offer vast statistical possibilities to interpret data and should be considered as solid and insightful models to run statistical analysis using machine learning technics (Varian, 2014).

Lee, et al, (2019) have shown that a certain number of researches have been achieved using logit and probit approaches, which led them thinking, just like Varian (2014), that economists should extend their analytics toolbox to analyse large amount of data. The linearity of the most used models conducts to relatively inaccurate results. Figure 3 illustrates the linear classifiers imperfection and shows the better quality of non-linear ones, applied to M&A classifier models.

Figure 3: Graphical difference between linear and non-linear methods to classify M&A deals (sucess or failure)[1]



The linear and nonlinear decision boundaries, logit model and neural network, respectively, are displayed, dividing M&A success and failure classes.

Various topics can be cover using statistics, such as asset pricing, real estate pricing, the impact of COVID-19 on the stock market or even the return of the automotive industry, for instance. However, while-decision-makers try to identify hidden patterns between to outstand the market, many well used statistical applications might seem odd regarding recent researches and new models. For example, different authors have proved that ML technics are giving better results than well-established statistical models (Philip, 2020). Philip (2020) has shown that Vector auto-regression models (VAR) to estimate and analyse the permanent price impact, are not efficient to capture the correct inferences when the price functions are non-linear, because they are leading to incorrect inferences and conflicting results due to their

[1] (Lee, et al., 2019, p. 273)

linearity. However, the author proposes to tackle the problem using machine learning technics through an iterative process and obtained significant results. The most interesting aspect here is that Machine Learning technics can deal with an iterative approach and find better results that more traditional ones. The author therefore brings into light an alternative method using Reinforcement Learning (RL), assuming basic conditions for the data (stationarity and Markov). To another extent, Perboli & Arabnezhad (2021) have got pretty significant results when predicting economic difficulties for firms on mid and long-term, compared to traditional statistical methods, such as Altman or Ohlson methods. For similar reasons, recent studies (de Bodt, et al., 2018; Charkrabarti & Mitchell, 2016; Betton, et al., 2014) have demonstrated reliable results using respectively probit and logit models, but indicates that linear methods could not fully represents reality and should suffer some misspecifications, especially because their logit and probit methods cannot solve non-linear problems. De Bodt et al. (2018) clearly specify that no theory specifically support the use of linear models more than non-linear ones on studying empirical evidence of overbidding in M&A deals. According to the authors, limiting oneself to a linear analysis and approach to research topics leads to forgetting about inferences, and thus possibly missing relevant points due to potential misspecifications and measurement errors. Investigating another topic, Giordani, et al. (2014) have found that bankruptcy event could not be successfully predicted using traditionnal linear apporach on financial ratios and brankruptcy risks, but they suggest that other non-included factors influence the poor accuracy of their model, and using a non-linear model could be of great help on catching better granularities between variables relationship, and therefore being more accurate. This paper confirms, in a way, that our world is drastically not linear, and that linear model, perhaps being more pratical sometimes, are far from being the best predicators.

2.3 Current and achieved work with Machine Learning methods

Recent work on Machine Learning and its adaptive technics using Neural Networks can measure the impact of strategic impact on firm performance. Works provide an interesting and comprehensive point of view on the determinants of the market value of Small and Middle size firms (Lee & Kwon, 2016). On another qualitative work, Gökhan and Nihat (2016) show that even if the top management is still cautious about such technology while business operations teams already use those technics in the Enterprise Resource Planning (ERP), more and more managers rely on AI-driven decisions based to decide, which prove, according to the authors, that a total democratisation is ineluctable.

Alternative measurement studies use supervised Machine Learning technics to predict companies to be acquired. With Natural Language Processing, Guang, et al. (2012) have analysed people and public firms’ profile on TechCrunch and Crunchbase to figure out whether there would be hidden patterns about a potential within written-word articles, news or comments from internet users. TechCrunch provides a large variety of data about Start-up and tech-oriented firms. The authors' model was able to identify M&A deals using keywords with a True Positive accuracy of between 60% and 79.8%, with a False Positive margin of error of between 0% and 8.3% depending on the categorisation of the used keywords (Guang, et al., 2012).

On other behalf, Lee et al. (2019) have shown interesting results on predicting M&A failure, using machine learning technics. Their study clearly highlights the difficulty that conventional linear analysis models have in predicting the outcomes of M&A deals. The diversity of deals, in terms of size, value and deal context, makes the datasets complex to analyse using logistic regression alone, leading to a large discrepancy between the predicted results and the reality of the facts. Type I errors (i.e., false positive: predicting a failure instead of a success in the facts) and type II errors (i.e., false negative: predicting a success instead of a failure of the deal in the facts), pollute the results and lead to a contextual misinterpretation. Although Type II errors are more damaging than Type I errors in the conduct of transactions and business, the feverishness of the linear models used in the study does not allow for any conclusion as to the effectiveness of such tools.

3 Methodology

The purpose of this study is to figure out whether Machine Learning can go through the M&A deals prediction. In regards to that question, we propose to use a supervised learning approach. Due to the large amount of data from previous deals, supervised technics seem to be profitable as they already benefit from labels that qualify the observation. As a consequence, supervised learning uses past experiences to predict the future ones.

To analyse the dataset and evaluate the situation, and then understand whether the firm is a good target or not. We will apply supervised machine learning technics to our dataset. Testing and finding the right split between the training set and the test set, the algorithm should be able to identify hidden patterns amongst the dataset.

In order to constitute the dataset, the study should consider different variables explaining the companies' profiles. These different variables should express the size of the company, its financial performance, its profitability and its capacity to generate value. For those reasons, the following data should be collected on Bloomberg on an annual basis, from 1990 to 2020. The sample is made of publicly traded companies.

3.1 Index and key financial ratios

Different ratios are to be studied to measure the significance between companies that are targeted and those which acquired them through the takeovers versus the companies which do not. A study by the M&A Research Centre at Cass Business School[2] and the American public consulting firm Intralinks (2016), has identified key factors that acquirers are focusing on when they are screening another company to figure out whether that last is a golden nugget or not. According to their common study, public companies whose sales growth increases and whose profitability, leverage, size, liquidity and valuation decrease, constitute the best suited targets. The M&A Research Centre at Cass Business School, City University of London and Intralinks also identified that small size and low profitability are the two most statistically signifianct predictors for a publicly traded company that could become a target.

For the purpose of the study, we will focus on the six following key ratios:

a. **Valuation index** (ratio of the Enterprise Value over the Earnings Before Interest, Tax, Depreciation & Amortisation) year prior the takeover initial offer:

$$Enterprise\ Value\ =\ Number\ of\ Share\ Outstanding\ \times Stock\ Price$$

$$Valuation\ index = \frac{Enterprise\ Value}{EBIDTA}$$

This ratio enabled us to measure the valuation of the targeted business on its past and current performances. Targeting the right price when making an offer is extremely important as this

[2] Cass Business School is part of City University London.

estimate will be the basis for further negotiations. The valuation determines whether the buyer makes a fair offer or whether its bid is overvalued, which is key in the financial strategy and budget allocation for the buyer. Included in the Enterprise Value is also the measurement of the target potential. Usually, target companies have a lower valuation multiple than non-targets (M&A Research Centre at Cass Business School, City University of London and Intralinks, 2016), so than acquirers.

b. **Liquidity index** (ratio of the Current Assets over Current Liabilities) prior the takeover initial offer:

$$Liquidity\ index = Current\ ratio = \frac{Current\ Assets}{Current\ Liabilities}$$

The liquidity index reflects the ability of a company to meet its short-term obligations. It informs investors and analysts about the ability of a company to be a proficient payer and to optimise its balance sheet. A company in difficulty will have a liquidity index below 1 and will naturally be a target company during M&A screening.

c. **Sales index** prior the takeover initial offer:

$$Sales\ index = \frac{Sales}{Average\ Sales\ of\ the\ Targeted\ Firms\ of\ the\ same\ Industry}$$

The Sales Index tells us how a company's business performance compared to that of its peers at the time of their involvement in the due diligence process. A company that underperforms in its market could more easily be a target for takeovers, as it has relatively fewer resources at its disposal. The reverse could also be true: a company that is outperforming its peers could also be the target of a larger company wishing to benefit from the target's dynamism.

d. **Leverage index** (ratio of the Debt over Earnings Before Interest, Tax, Depreciation & Amortisation) prior the takeover initial offer:

$$Leverage\ index = \frac{Debt}{EBIDTA}$$

The leverage index is a ratio that will be useful to measure the amount of debt a company has over the wealth it produces after removing production costs. A company with controlled costs will be able to pay back its debts better, and thus accelerate its future investments. Conversely, a company in difficulty will have an increasing debt and its EBITDA to stagnate or decline. In both cases, such firms might be adequate targets following possible acquirers.

e. **Profitability index** (ratio of the average earnings before interest, tax, depreciation & amortisation over the sales) prior the takeover initial offer:

$$Profitability\ index = \frac{EBITDA}{Sales}$$

The profitability index is an indicator of the company's performance in managing industrial costs. A company that is able to hold down its productioncosts is naturally more likely to make a profit, and will therefore be identified as a safe bet for an investor.

f. **Number of employees index** (ratio of the number of employees of the target divided by these of the acquirer):

$$No.of\ employees\ index = \frac{No.of\ target's\ employees}{No.of\ acquirer's\ employees}$$

The Number of employees index can be seen as a human resources size ratio. The headcounts of the publicly traded companies might massively influence a bidding offer since it is highly correlated with the current firm's position on the economic landscape. Comparing the targets number of employees to the acquirers' will provide an efficient metrics to size the possible new created entity.

g. **Acquisition Index** (dummy variable equals to 0 for "no deal" or 1 for "deal")

The Acquisition Index is the "Class" category within our dataset. It has been used to label our data so as to then use supervised techniques.

So as to be able to compare and qualify the ratios between the target and the acquirer, we will then get the ratios of the target to those of the acquirers.

After having compiled the ratios for each firm, we will compute the ratios for the deal, to characterize each one deal:

$$\frac{Target's\ ratio}{Acquirer's\ ratio}$$

3.2 The Machine Learning technics

After compiling the ratios for all the actors in the takeovers, we will obtain our final dataset, which we can then exploit with a Machine Learning algorithm. The chosen method to analyse the dataset is Supervised Learning. Python is nowadays one of the most used platforms for coding, and for Machine Learning because it offers a large pallet for its available libraries such as *numpy*, *pandas, seaborn* or even *sklearn*. Therefore, Python provides very useful tools to adjust tailor-made algorithm for Machine Learning and should be considered as a proficient solution to conduct such study.

Amongst supervised Machine Learning technics, different analysis families must be identified: regression on one hand, and classification on the other one. Regression analysis is used to predict the value of the dependent variables using one or more dependent variables. On the other hand, classification consists into classify cases as similar or not. Our study is based on three main known algorithms: Logistic Regression (LR) as the traditional method, Support Vector Machine (SVM) and Decision Tree (DT) as main techniques.

3.3 What is classification?

Before diving into the details, what is classification? As one of the main family technics of supervised machine learning, classification aims to predict a categorical outcome from input data (Sadoghi, 2020). Most the time, the input are set quantitative, but images, spatial, temporal or text can be used. However, data must be encoded in a quantitative form to be exploited by the algorithm. The most common way to proceed is to designate a dummy variable (dv) and mark each record (with 0 or 1) to enable the supervised learning approach. At the end, the dataset will be characterized by a binary classification. For instance, well known anti-spam filters use "spam(dv=1) or no spam (dv=0)?", or even in fraud detection for firms like Mastercard: "fraud (dv=1) or no fraud (dv=0)?", and so on. In a nutshell, classification technics are very useful to find hidden patterns amongst the dataset, and then classify the data answering the question: "Yes or no?".

The traditional way to proceed is to firstly provide a training set to feed and breed the algorithm, and then let it figure out which input belongs to which category for the testing phase.

The main problem classification models rise are multiple. At first, overfitting issues are well-known phenomenon that occurs a model has to idealistic results. It means it is perfectly finding the solution to a test set because the model has been overfed. As a consequence, the algorithm will no perform well for an out-sample dataset. This problem happens with trees or regression with *n* variables and *n* observations, which may provide a too good fit in-sample but a very poor out-sample fit. Other difficulties might occur, due to our dataset structure: unbalancing, not enough data.

3.4 The classification technics

Classification is efficient when the output takes limited and discrete values, 0 or 1, yes or no, true of false. In general, classification is used to predict a class from the dataset and split them into two categories: binomial or multi-class. Within classification, three technics are denoted as Logistic Regression, Support Vector Machine, and Decision Trees.

The purpose of this paper is to study whether M&A deals outcome could be reasonably predicted using supervised machine learning technics. Therefore, it seems rational to rely on classification technics to run our study.

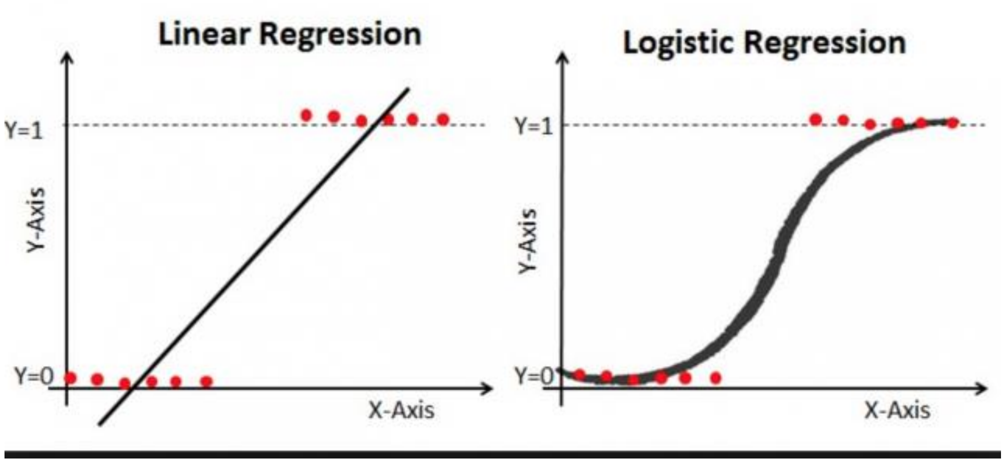
Machine Learning gathers many statistical methods that allow various prediction based on exploring the data. In order to make the most of the dataset and maximise its understanding, we need to follow several key steps. Firstly, the dataset must be split into two distinct parts which will be used respectively for the training phase of the chosen model, and for the test phase, which will serve both to validate the model and to measure its performance. In general, several tests are necessary to find the best settings. Indeed, giving a too important part of the dataset to the algorithm will lead to an overfitting phenomenon. Overfitting occurs when a model has too idealistic results. It means it is perfectly finding the solution to a test set because it has been overfed. Therefore, the algorithm will perfectly match with the dataset, but not with any set extension. On the other hand, not feeding the algorithm enough will cause it to perform poorly. In both cases, the performance of the algorithm will not be optimised. Algorithms will perform better when being fed by balanced dataset meaning a reasonable similar proportion of "0" and "1". In general, an algorithm that is right 70% of the time will be considered efficient

3.5 Logistic Regression

Logistic Regression (LR) is one of the basic techniques of machine learning. In fact, Logistic Regression is a mutation of linear regression. These are simple algorithms that aim to determine which variables might have the largest impact on the dependant variable (Provost & Fawcett, 2013). These algorithms are helpful for a wide range of topic, such as real estate pricing. In the example of real estate price prediction, Logistic Regression can be used as a factor predictor: what is the most important factor influencing the price of real estate?

The main concept in this model is to find an optimised linear equation that best describes the dataset behaviour by reducing as much as possible the error margin. The following graphs[3] shows the difference as well the basic interests of using a Logistic Regression instead of a Linear Regression when it comes to describe the behaviour of binary outcomes: 0 or 1.

Figure 4: Linear Regression vs Logistic Regression graphs



We might say that Logistic Regression (on the right) seems to better describe the dataset's behaviour than Linear Regression (on the left).

Usually, Ordinary Linear Square (OLS) regression are written under this form:

$$Y = a_0 + a_1 X_1 + \cdots + a_n X_n + \varepsilon$$

[3] Open Data Science. Available at https://medium.com/@ODSC/logistic-regression-with-python-ede39f8573c7

Concretely, it is about minimizing the distance between any dots and the trend line. The result is given by the R^2, from 0 for a poor result, to 1 for the perfect one (which is never reached in practice). This R^2 determinant is the quadratic error sum.

$$R^2 = 1 - \frac{Squared\ Error\ Model}{Squared\ Error\ Mean}$$

A $R^2 > 0.9$ may indicates an overfit or a snooping to the dataset. Such technics have been really popular when analysing real estate price evolution in function for different parameters: surface, localisation, period, age of the building, …

More sophisticated models include lags, which conduct to consider temporality in the model. Such models, like well-known ARDL model, are restricted with time series.

Derived from OLS, LR was introduced in order to account for the probability that a new data item belongs to one of the identified classes. The logistic regression function measures the probability of occurrence of an event: the odds. The notion of odds helps to understand the distance the labels have with the s-shaped curve. Basically, the odds is the ratio of the probability of the occurring event to the probability of the non-occurring event. Actually, the result of the ratio is included between 0 and $+\infty$. The natural logarithm (log) is then used to range the distanced from $-\infty$ to $+\infty$ from the boundary. The log-odds linear function is given by:

$$f(x) = \log\left(\frac{p(x)}{1 - p(x)}\right) = a_0 + a_1 X_1 + \cdots + a_n X_n + \varepsilon$$

With **x** the dataset's item, p(x) its probability of occurrence.

From the previous equation, we obtain the logistic function by solving p(x):

$$p(x) = \frac{1}{1 + e^{-f(x)}}$$

To better off the outcome of our research and get the best of Machine Learning technics, we can switch to Logistic Regression model. Whilst Linear Regression is continuous, Logistic

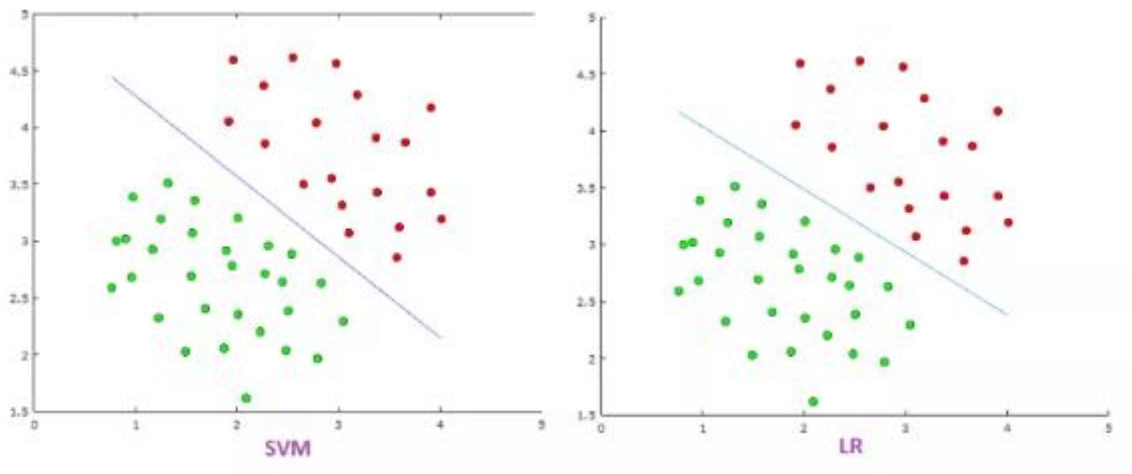
Regression is discrete. In our research, we are using a discrete variable (0 or 1), so Logistic Regression should be used instead of Linear Regression. This model is known as a benchmark for ML multi-model testing. To go into the details, we change the input from real numbers to occurrence probability using a binary system: 0 or 1. Therefore, we start creating classes for the dataset. Logistic Regression is actually about optimising the model so as to better fit the model to the dataset. The challenge of the training phase of the algorithm is to position the curve in s as close as possible to the binary labels, and to maximise the number of points through which the curve runs. The goal is to let the algorithm find the optimal hyperparameters, and to determine the best function describing the dataset.

Actually, the Logistics Regression will be conducted as an iterative procedure that maximizes the likelihood of the prediction: is the M&A deal a success (x = 1), or a failure (x = 0).

3.6 Support Vector Machines

Despite Logistic Regression (LR) remains extremely useful for basic linear analysis, Support Vector Machine (SVM) brings great value to classify a dataset in some different way. While LR has a statistical approach, SVM aims to geometrically classify the dataset as the following picture shows[4]:

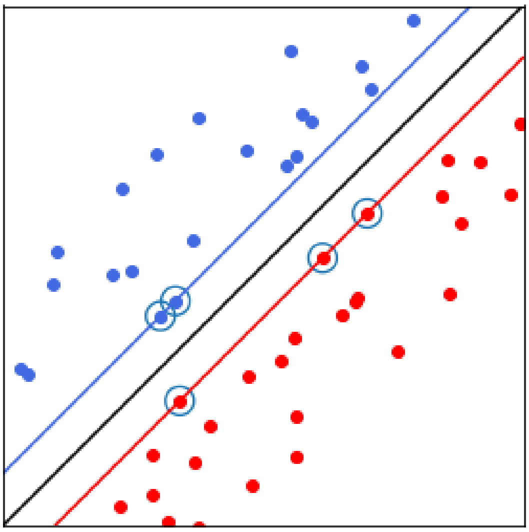
Figure 5: SVR vs Logistic Regression line separation



[4] Picture available at : https://medium.com/axum-labs/logistic-regression-vs-support-vector-machines-svm-c335610a3d16

To return to the start, SVMs were developed by Cortes & Vapnik (1995) in order to present a benchmark study of optical characters recognition by a machine. Their principle is simple: they aim to separate data into classes using as "simple" a boundary as possible, so that the distance between the different groups of data and the boundary separating them is maximum (Sadoghi, 2020). This distance is also called the "margin" and SVMs are thus called "wide margin separators", the "support vectors" being the data closest to the border (the blue and redlines in the picture below).

Figure 6: Support Vector Machine usefulness graph



Technically, the support vectors (those in blue and red in Figure 6), aim to optimise the distance between the dots and the hyperplane (the black line). As an example, this classification technique works perfectly well when it comes to classifying images or recognising handwriting. Using SVMs is of great help when managing high dimensional dataset. However, the bigger the dataset, the less efficient this technic. Moreover, in highly overlapping dataset, SVMs will not perform well as it is unable to fairly separate and then classify the variables.

Unlike Logistic Regression, SVM technics are more robust to overfitting and to noise. Since the support vectors tend to separate the data set by limiting the points within the beam (i.e., the points between the two support vectors), SVMs technics are less sensitive to abnormal datapoints. Moreover, using SVM enable to limit the impact of “massive amount of located datapoints”, whereas Logistics Regression shape will be kind of “attracted” by the local clusters.

The complexity of an SVM model depends on the number of characteristics and the number of samples. The algorithm does not scale well with the number of cases, as the amount of computation is proportional to the number of samples. The computation time increases considerably as well. The question that SVMs answer is: can we linearly separate similar cases in the dataset? To do so, in a given vector space denoted as "E" and for which Dim(E) = 2 (whose basis is composed of two vectors), one must find a linear function like: $ax + b = y$. For instance, let's replace y by x_2, and we get:

$$ax_1 + b = x_2$$
$$\Rightarrow \; ax_1 + b - x_2 = 0$$
$$\Rightarrow \; ax_1 - x_2 + b = 0$$

If $\begin{cases} x = (x_1,\ x_2) \\ w = (a, -1) \end{cases}$, we then get:

(1) $\boldsymbol{x} \,.\, \boldsymbol{w} + \boldsymbol{b} = \boldsymbol{0}$, where $\boldsymbol{w}$ is a scalar in E.

Again, this equation belongs to a hyperplane in vector space E whose dim(E) = 2, and would work with the same logic in a vector space whose dimension is greater than 2.

Therefore, one can use the equation (1) to optimize the model and find the best hyperplane, that satisfies the following conditions:

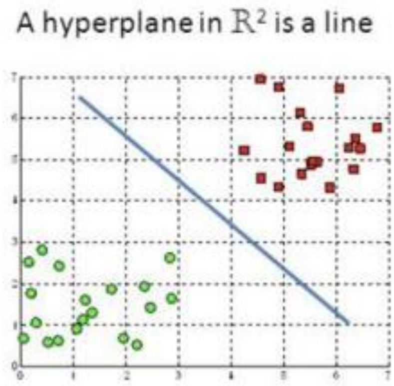
$$h(x_i) = \begin{cases} 1 \; if \; (x_i.w) + b \geq 0 \\ -1 \; if \; (x_i.w) + b < 0 \end{cases} \Rightarrow \; y_i[(w.x_i) + b] \geq 1$$

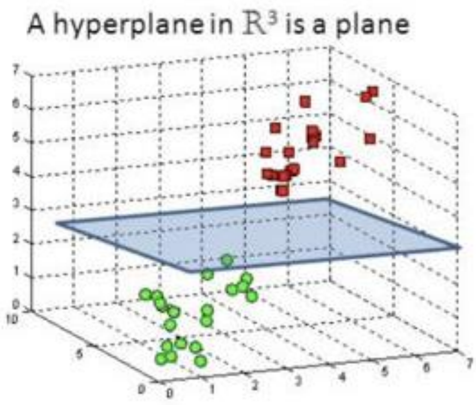
The final condition is the minimization of the hyperplane norm given by:

$$||w||^2 = w.w$$

Figure 7: R^2 line and R^3 hyperplane SVM[5]

[5] The illustration is available at; https://databusiness-ai.com/support-vector-machines-svm/

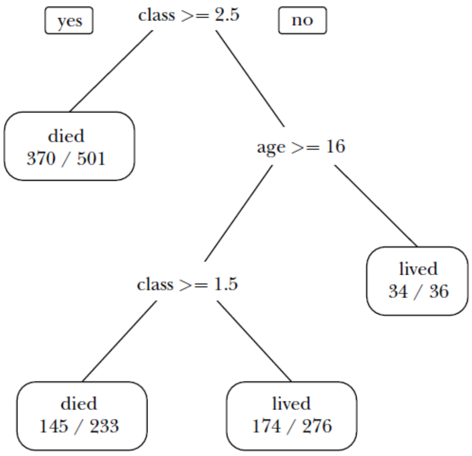




3.7 Decision Trees

Another way to deal with classification problem is using a classification tree (or decision tree). Those are useful and provide graphical representations of a classification method. It tries to make a prediction on categorical variable by dividing the dataset into subsets, conditioned to their distinctive attributes. That technic is very well known today as it has been introduced in the second half of the 20th century by Morgan and Sonquist (1963), but remain unpopular, until Breiman, et al. work in *Classification and Regression Trees* (1984). What is interesting when using such Decision Trees (DT) belongs to their graphical and easy-to-understand attractivity. Everyone can see and read the scheme, which is much appreciated when communicating results: it classifies the data by regions. In practice, DTs are clearly identifying the significant predicators, and generally, distribute the probability using Boolean (yes/no; true/false) on each nod. The leaves give the classification results. The goal is to obtain a decision tree that will be able to make accurate out of the sample predictions, just like the two previous models. On the opposite of Logistics Regression (LR), DTs are convenient models for non-linear dataset's correlation. LR is better for smaller datasets whereas DT handles more efficiently larger ones, and can deal with missing data within datasets (Perlich, et al., 2003). The major benefit from using trees is their ability to interpret problems where interactions and nonlinearities are significant. Lee, et al., (2019), have found that M&A were typically sorted as non-linear cases, which could make us think that Decision Tree can be an efficient predictor model in this case. It is expected that Decision Tree will provide better results than using, for sure Logistic Regression, and in a second time Support Vector Machine technics. Varian (2014) has used Decision Tree technics to investigate Titanic passengers' survival, for which he obtained pretty accurate results, better than using Logistic Regression.

Figure 8: Survival of passengers on the Titanic[6]



Indeed, while using logit model, Varian's (2014) algorithm indicates that age is not an important feature, but Decision Tree suggests that these last is relatively significant into predicting the outcomes. The matter here is not age, but much the youth or the elderly of the passengers, things that did not show LR. This result supports the DT and its interest, in particular the step-by-step dichotomy when classifying the dataset, as well as the details. On the mathematics side, the algorithm must be looking at minimizing the error rate (Sadoghi, 2020). The Gini index (G) is a good example to how to. It provides a synthetic overview of how inequal is the sample. The index stands from 0 (for totally equal) to 1 (totally inequal).

$$G = \sum_{k=1}^{K} p_{mk}(1 - p_{mk})$$

Where p_{mk} is the probability that the item belongs to a given class.

One of the main shortcomings of decision trees is their overfitting of training data, which makes them less able to perform well with out of the sample data. Therefore, one solution is to prune the tree to make it a little "less accurate" during training, but enable it to remain fairly "general" to predict the data in the test phase. Pruning the tree requires back-office mathematics. To handle it, we use cross validation to figure out which tree has the lowest error rate (E):

[6] (Varian, 2014, p. 8)

$$E = \sum_{j=1}^{T} \sum_{i \epsilon R_j} (y_i - \hat{y}_{R_j})^2 + \alpha |T|$$

Where y_i is the testing data and $\hat{y}_{R_j}$ is the predicted data

Where $|T|$ is the number of terminal nods.

At each pair of leaf nodes with a common parent, we evaluate the error on the test data subset. If the testing sum of squares decreases when we remove those two nodes and make their parent a leaf, we should prune, otherwise we shouldn't.

3.8 The classification results and analysis

To allow the supervised approach, our dummy variable will be the effective result of the deal: 0 for “withdrawn” (meaning the negotiations have failed) or 1 for “completed” (meaning negotiations were successful and the offer was accepted / the target was acquired in the takeover).

Thereafter, different classification technics such as Logistic Regression, Support Vector Machine and Decision Tree Model will be applicated to run the supervised learning to the machine. At first, we will test different combinations to identify the best proportions within the training set and the test set (60-40%; 70-30%, 80-20%, 90-10%). This identification should be help us to maximise the accuracy of the Logistic Regression, Support Vector Machine and Tree Decision Model, and thus our results. The Receiver Operating Characteristic (ROC) Curve and the Correlation Matrix will provide more details on the accuracy of the model, in particular the True Positive rate for the Sensitivity of our model, and False Positive rate (Mueller & Massaron, 2019). The main aim of our models is to correctly identify the right solution in any situation, and those ratios are bringing a clear insight for our model.

Figure 9: Confusion Matrix

<u>Confusion Matrix</u>

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	**True Positive**	**False Negative** (Error Type II)	Sensitivity
	Negative	**False Positive** (Error Type I)	**True Negative**	Specificity
		Precision	Negative Predictive Value	Accuracy

About the terminology, *True Positive* (TP) and *True Negative* (TN) means that the model has predicted the right patterns underlying the dataset. For example, in case of disease prediction, *True Positive* means "According to our algorithm, the patient has a disease, and he actually does". Then, *True Negative* means "The patient does not have a disease, and he actually does not". In both cases, the algorithm has correctly predicted the result. On the other hand, *False Positive* (FP) and *False Negative* (FN) are both errors, meaning that our algorithm has made a mistake. In the case of *False Positive*, our algorithm has made an *Error Type I* which can be understood as "The patient has a disease, but he actually does not". This kind of error is polluting the results but is not too serious. The other type of error is *Error Type II*, or *False Negative*. Much serious than Error Type I, *False Negative* is directly impacting the reliability of the outcomes. In this case, the algorithm indicates that "The patient does not have any disease, but he actually does" (Prado, 2018). To measure the algorithm reliability, we use several ratios we detail thereafter. Those ones offer clear indications about the usefulness of the algorithm to provide accurate results, and therefore whether we should reject or not our model. In our study, Error Type I is the worst case as it indicates that the algorithm identified a deal as a success instead of a failure. This case is the most critical in our case. On the other, Error Type II is less critical as it means our algorithm predicted a failure instead of a success.

The *Precision* ratio measures the apparition rate of the correct "yes" prediction. It basically answers to the question: "When the model predicts yes, how often is this correct?", and is given by:

$$Precision = \frac{TP}{TP + FP}$$

The *Sensitivity* ratio (or Recall) provides indications on the apparition rate of correct real "yes". It answers to the question "When it is actually yes, how often does the model say yes?", and is measure by:

$$Sensitivity = \frac{TP}{TP + FN}$$

The *Specificity* ratio corresponds to true negative rate. It answers to the question: "When it is actually no, how much does the model say no?", and is given by:

$$Specificity = \frac{TN}{TN + FP}$$

The *Negative Predictive Value* ratio is a misclassification rate that measures how often the model is wrong. It answers to the question "How often is our model wrong?", and is given by:

$$Negative\ Predictive\ Value = \frac{TN}{TN + FN}$$

Finally, the *Accuracy* ratio measures how performant is our model, and is given by (the higher, the better):

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP}$$

In addition, we will provide a graphic representation on the efficiency of our models with a ROC Curve in each case, and their AUC (Air Under the Curve) score. The following table sums up the models we will use as well as the expected results for each model given what has been presented before.

Figure 10: Summary hypothesis table

Model name	Type	Expected AUC
Logistics Regression	Linear	0.5 to 0.55
Support Vector Machine	Linear	0.5 to 0.55
Decision Tree	Non-linear	Around 0.7

The limitation of our models will be identified and comprehensively explained.

4 Data collection

To run our supervised machine learning models, we will need vast amount of data about previous mergers and acquisitions deals. To do so, we project to use Bloomberg data on that subject. Bloomberg software provides one of the largest qualitative and quantitative set of data relative to finance in the world. Those information are being used by the best and the biggest firms in the world, including worldwide banks.

The sample should include key financial ratios to be analysed as well as labelled variables. Otherwise, we will not be able to take advantage of the full value of supervised learning techniques. Moreover, we should include a large enough dataset to allow algorithms to understand the meaning and the substance of model, and then use supervised machine learning. As a consequence, data from a large period should be preferred to offer a consistent and comprehensive review of the studied question.

To structure the data collection, we will at first focus on having a clean dataset, with no missing values. To better benefit from the dataset will focus on a specific economic sector: industry. The screening procedure will be composed of three parts: at first, a qualitative filtering which consists into cleaning the dataset to make it free from any missing value. The second step consists into a quantitative adjustment to get the key ratios for the supervised treatment. The third step will be the statistical treatment of the models to test our assumption of the possibility to implement a supervised approach to predict firm's acquisition.

5 Application

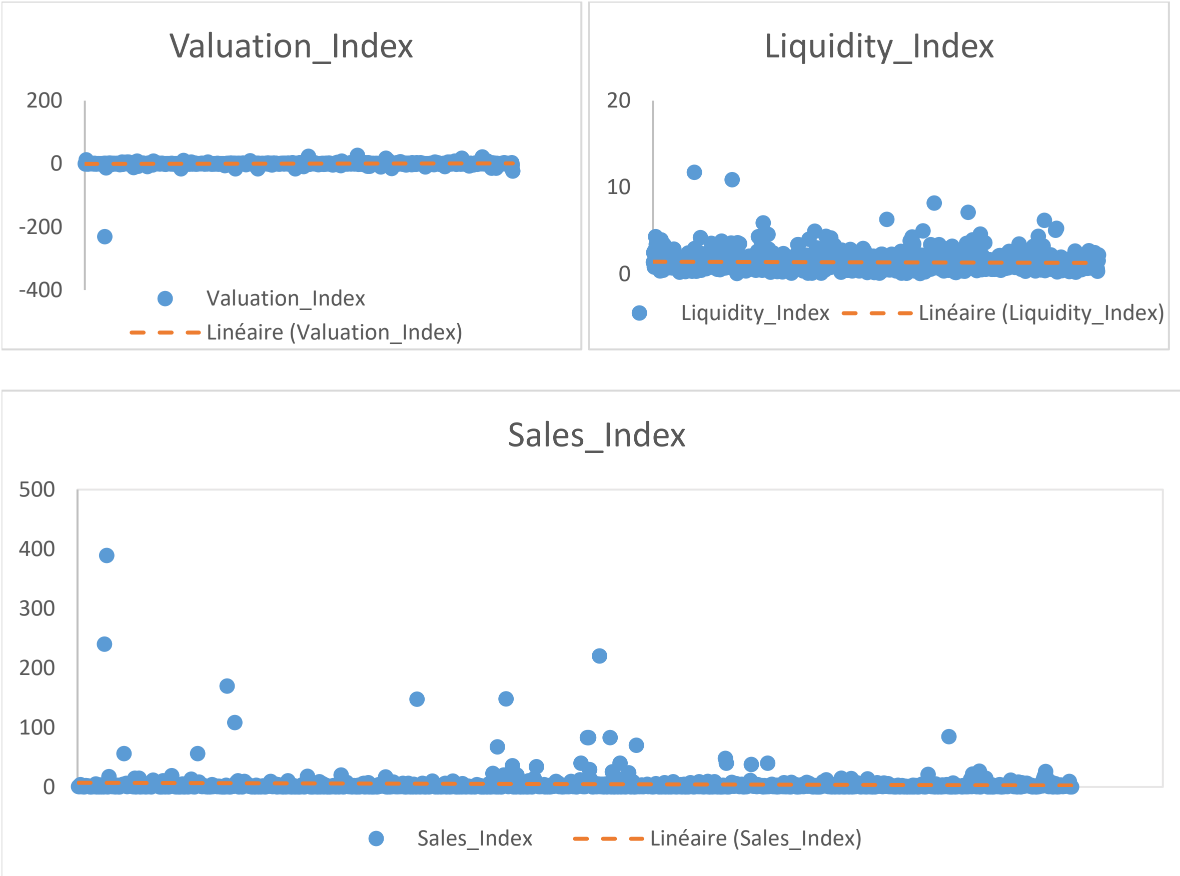
So as to tackle our study, we have got a 22,044 records dataset from Bloomberg and have thereafter applied qualitative filtering to clean the whole dataset, to free it from any missing value. We ensure the new dataset is only made of deals for which we have comprehensive information. After this qualitative filtering, we obtain 921 records that will constitute our study dataset. As a comparison with similar studies, Lee et al. (2019) show that on average, M&A papers’ datasets are composed of 2,099 deal records, for 82% successful deals. In our case, the dataset is composed of 921 deal records and 89% successful deals.

At the end of the filtering phase, we therefore retain only 4% of the data we started with. This loss of 96% of the initial dataset is non-negligible since it will have a major impact on the quality of the results we obtain. This loss is explained by the lack of certain crucial financial data that we need for our methodology. Indeed, some data such as sales, EBITDA or even debt were missing for at least one of the two companies (the target or the acquirer). In order to work with qualitative data that best reflects the context of the deal, these deals were not taken into account.

To process our study, we had to compile the dataset into a csv file "Datasetv2.csv", containing the different ratios of our methodology. In order to compare the results of the companies and to have a single ratio describing a two-company operation, we chose to measure the target's ratios against the acquirer's ratios. This transformation allows us to understand in a single factor the difference between the target and the acquirer. For example, the Sales Index ratio for the target A and the acquirer B is : Sales Index of A / Sales Index of B.

The historical context of the deals is not an issue here since each of the ratios is composed of data from the same deal by comparing the data of the target and the acquirer by means of the ratio of target to acquirer. The ratios should therefore help to qualify, then classify the deals with one unique ratio, and it should be stationary and not cause any obvious bias when applied (due to the large period of the sample). The Figure 11 shows a graphical representation of the data. It seems that the sample is oscillating around the same mean, no matter the period.

Figure 11: Relative graphical stationarity of the data



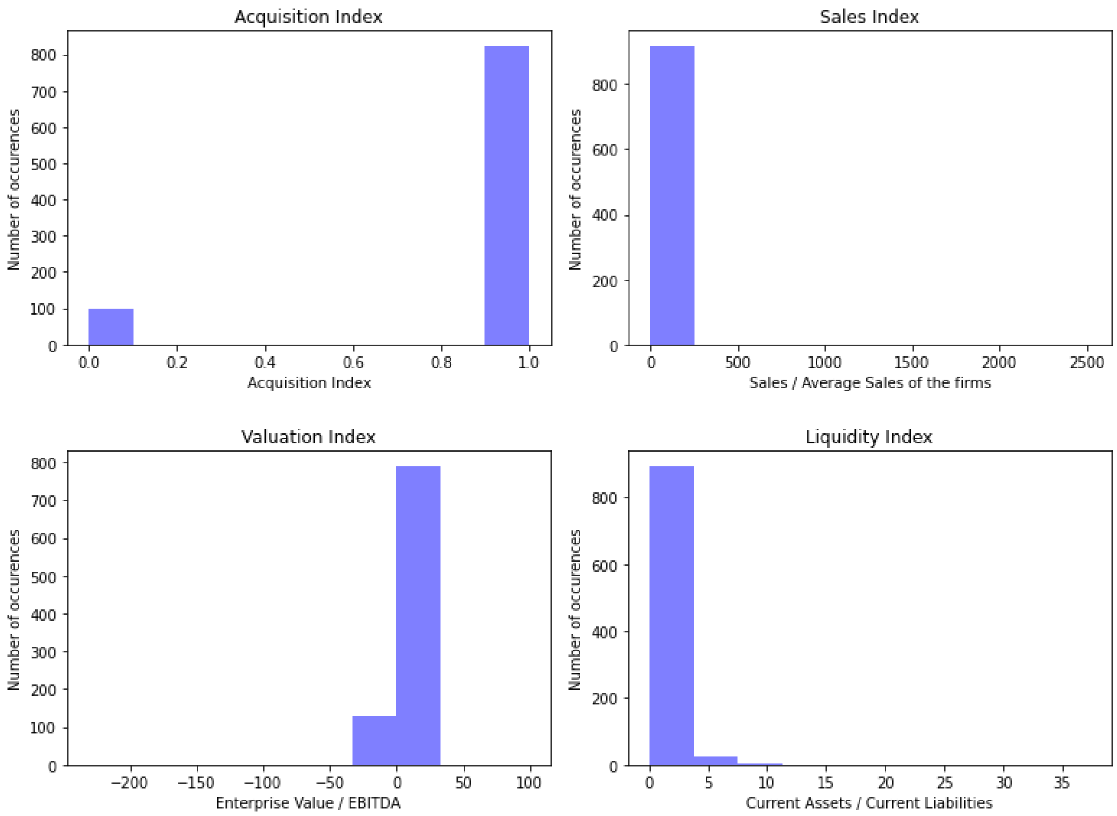
All the records range from the beginning of 1990 to the end of 2020. They all concern publicly traded firms and are sized more than $50 million. The last but the least: all deals have been conducted within the industrial sector firms. In order to remain consistent in the application and analysis of the results, only one sector of economic activity was investigated, so as not to mix together ratios that have no link between them. The latter would have been derived from distant economic contexts because of the heterogeneous inter-sectoral business models. This would have led to a significant decrease in accuracy and an obvious bias when applying the algorithms.

The exploitation of the dataset was carried out using Python. Figure 12 shows the header and the structuring of the dataset used in the application. The dataset is made up of 8 columns, each containing a key financial ratio, and synthesising all 921 deals. Figure 13 shows 4 histograms of the data that will interest us most in our studies, and therefore the distribution of the data.

Figure 12: Dataset header

	Deal Type	Class	Valuation Index (Enterprise Value / EBITDA)	Liquidity Index (Current Assets / Current Liabilities)	Sales Index (Sales / Average Sales of the firms)	Leverage Index (Debt/EBITDA)	Profitability Index (EBITDA/Sales)	Number of Employees Index
0	M&A	1	0.843311	1.380794	0.388009	0.495309	0.681943	0.109091
1	M&A	1	0.268940	2.502929	0.245625	0.601950	0.534041	0.069767
2	M&A	1	13.403385	0.809119	3.748340	2.037274	0.784553	0.526882
3	M&A	1	0.382593	2.725530	0.302446	0.611721	0.715211	0.102545
4	M&A	1	0.772213	1.714339	0.717924	1.211360	0.554317	0.184536

Figure 13: Histograms of three selected variables Sales and Valuation Index

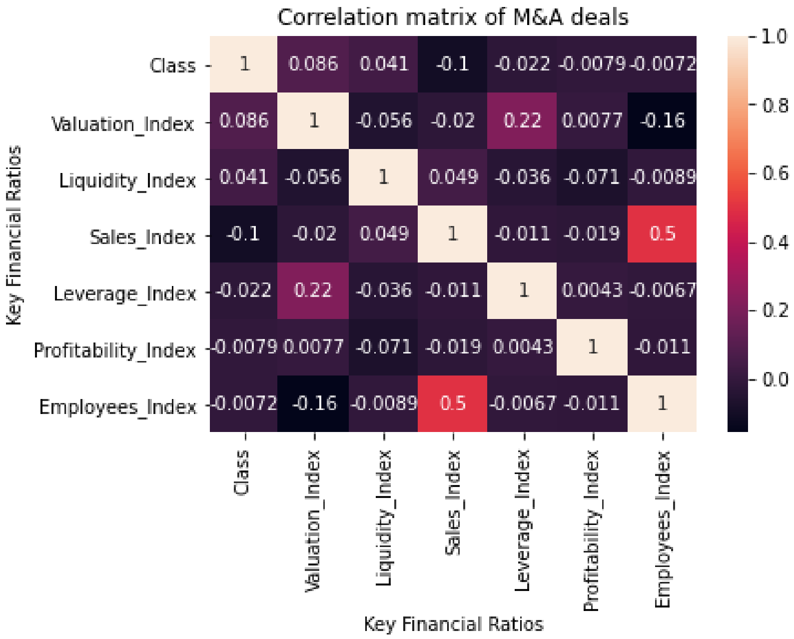


In order to better understand the composition of the dataset and the links that exist within the dataset, we generated a correlation matrix. This matrix uses the Pearson coefficient to measure the mutual importance between two variables. This coefficient is obtained by dividing the covariance of two variables by the product of their standard deviation. This coefficient is included between -1 and 1, indicating respectively a negative or positive exact correlation, and 0 for a non-existent relationship between the two variables.

In the case of our study, presented in Figure 14, we observe that our "Class" label is quite weakly correlated with all the features of the dataset. These results are not very surprising in that what we said before shows the full complexity of M&A deals, so some dimensions are not taken into account. Therefore, and according to those results, it appears that no clear assumption can be made based on the influence of one or more specific variables on the Acquisition Index[7].

Nevertheless, we can identify some correlations that are more important than others, with notably Sales Index (-0.1), Valuation Index (0.086) and Liquidity Index (0.041). We will therefore keep these 3 features to build our models, as they seem to be the most explanatory of our "Class" response.

Figure 14: Correlation Matrix



The previous correlation matrix shows that some variables are more correlated with **Class** than others:

- Sales_Index (corr_coeff = -0.1)
- Valuation_Index (corr_coeff = 0.086)
- Liquidity_Index (corr_coeff = 0.041)

6 Analysis and Results

In this study, we firstly chose to allocate 60% of the data to the training phase and 40% to the testing phase, and then upsizing the training set. In order to measure and report on the overall accuracy of our algorithm, we generated a confusion matrix and the ROC curve for each of the models we planned to use: logistic regression, Support Vector Machine, and decision trees, using the Python programming language. After a first series of tests with the allocation described above (60% - 40%), we will reduce the share of the test set, and move to an increase of the training set by 10%, i.e. 70% - 30%, and so on in order to maximise the

[7] The Acquisition Index has been denoted as *Class* in the algorithm.

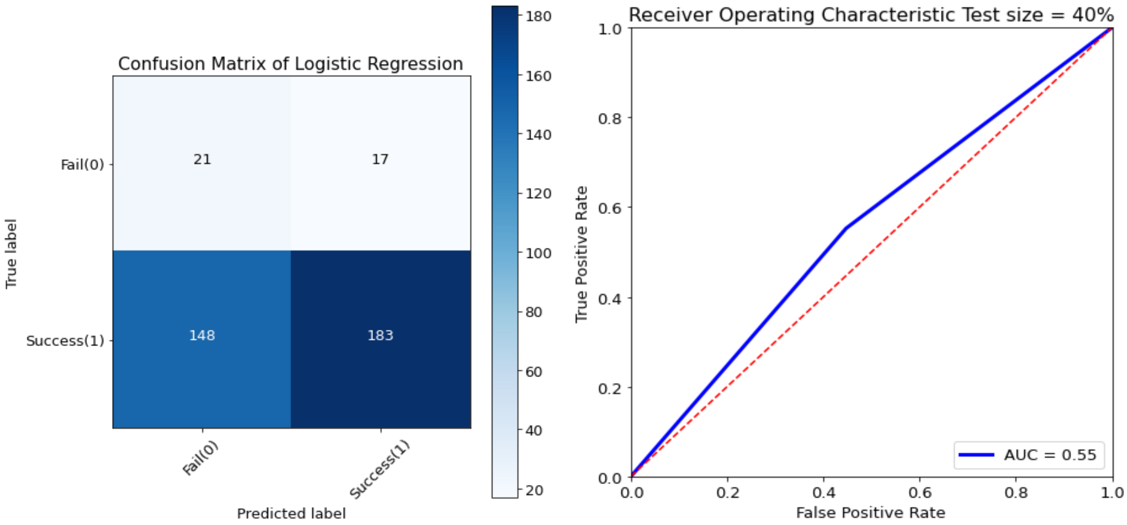
efficiency of our algorithm in an iterative process. The main stake is to figure out the best sweet spot for the best model. The code is available on demand.

6.1 Logistic Regression Model: a benchmark

As we have previously seen, logistic regression is often the first step in supervised learning and therefore serves as a benchmark for studying other models. Here, the algorithm seeks to classify the data by first training on a certain percentage of the dataset, and then testing what it has learned on the rest, using a S-shape modeling.

The logistic regression model gets very imperfect results. The Figure 15 presents the LR's confusion matrix and its ROC curve. That last shows us that the Air Under the Curve is 0.55, which means that our algorithm is actually wrong every almost one out of two times. These results are not surprising and are consistent with the literature review in which we saw that linear models do not capture the complexity of M&A, nor the different inferences that remain. In a second step, we tried to increase the data input in the training phase to check wether the air under the curve score would increase, but we can see in the Figure 16 that this change is not significant. In our study, our algortihme should at worst reduce type I errors as much as possible, i.e. predicting an acquisition when the deal has not taken place. The type II error "the prediction is that the deal failed, when in fact it succeeded" is less problematic, and comes on a second stage. Here, the best prediction is made by the 40% split because its type 1 error rate is very slightly lower. Nevertheless, the two models are equivalent, and remain relatively inefficient for our problem, as we thought earlier. Note that other splits have been used to check whether get better results, but the AUC score decreases if we leave the two presented cases.

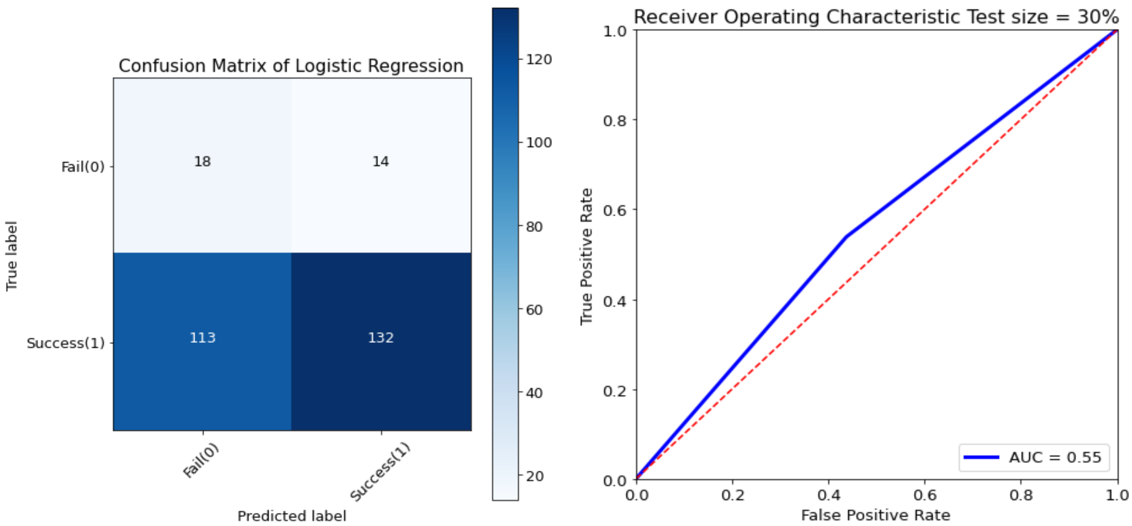
Figure 15: Logistic Regression Confusion Matrix and its ROC curve for a 60% - 40% split



Predicted label

	precision	recall	f1-score	support
0	0.12	0.55	0.20	38
1	0.92	0.55	0.69	331
accuracy			0.55	369
macro avg	0.52	0.55	0.45	369
weighted avg	0.83	0.55	0.64	369

Figure 16: Logistic Regression Confusion Matrix and its ROC curve for a 70% - 30% split



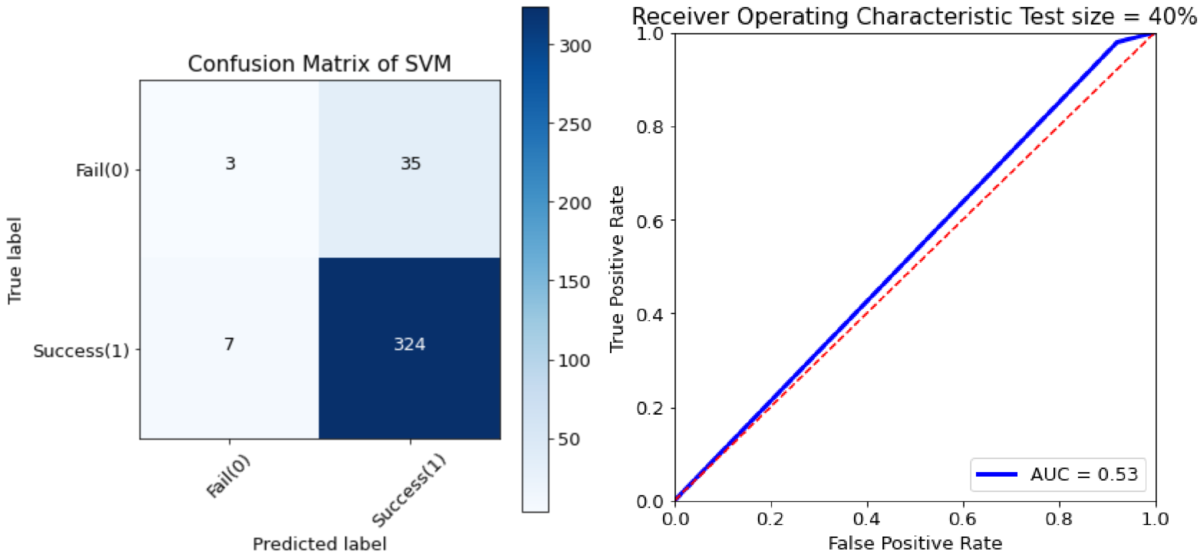
Predicted label

	precision	recall	f1-score	support
0	0.14	0.56	0.22	32
1	0.90	0.54	0.68	245
accuracy			0.54	277
macro avg	0.52	0.55	0.45	277
weighted avg	0.82	0.54	0.62	277

6.2 Support Vector Machine

Our second algorithm is the SVM one. It is used to predict the acquisition of companies by acquirers using a linear discriminant over all the data. As expected, the results for this model are rather disappointing. As with logistic regression, the use of this type of model assumes that we are in a linear framework, which is not the case for M&A. The algorithm is unable to correctly predict which class the deal it is analysing belongs to, regardless the test-train split we allocate to it. Choosing another split could even provoke to get worse results. The ROC curve and the AUC of 0.53 are low, and suggest that the algorithm is a random classifier, which does not allow us to conclude that it is useful for our study. The confusion matrix scores are also quite low, and confirm what the literature has taught us. Another feature is important to highlight: the accuracy is very high: 89%. However, we know that the AUC score is mediocre. That is due to the very unbalanced dataset we have. Indeed, as Lee, et al., (2019) developed in their paper, unbalanced datasets provoke a high bias in the accuracy score, but the precision score is terrible, which make us think that the accuracy score does not tell all the story, and should not be consider as a reliable indicator here.

Figure 17: Support Vector Machine Confusion Matrix and its ROC curve for a 60% - 40% split



Predicted label

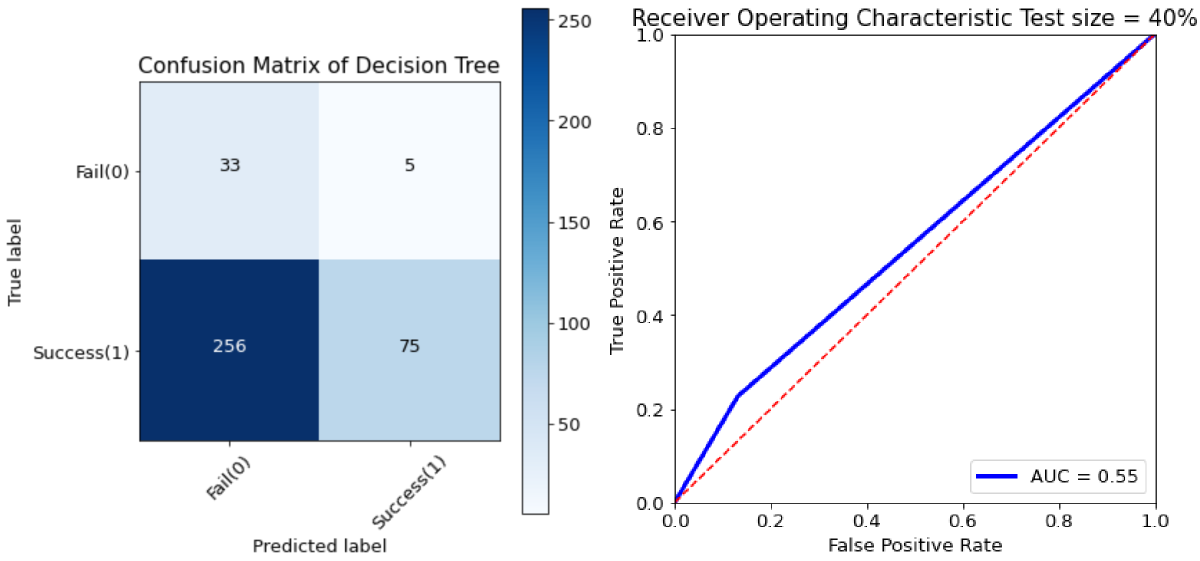
	precision	recall	f1-score	support
0	0.30	0.08	0.12	38
1	0.90	0.98	0.94	331
accuracy			0.89	369
macro avg	0.60	0.53	0.53	369
weighted avg	0.84	0.89	0.86	369

Other splits give similar results and are not presented here.

6.3 Decision Tree Model: the most efficient classifier

Decision Tree is the most interesting model we have obtained in the application. Decision trees are distinguished by their ability to handle non-linear data, and are therefore more appropriate when studying M&A operations than the other two classifiers we have had to study so far. The Figure 18 presents the DT model results for the 60% - 40% split. We can see that the ROC curve sticks very closely to the random classifier red lines, as the AUC score about 55%. It means that with the current parameters, we cannot attest on the efficiency of our model.

Figure 18: Decision Tree Confusion Matrix and its ROC curve for a 60% - 40% split

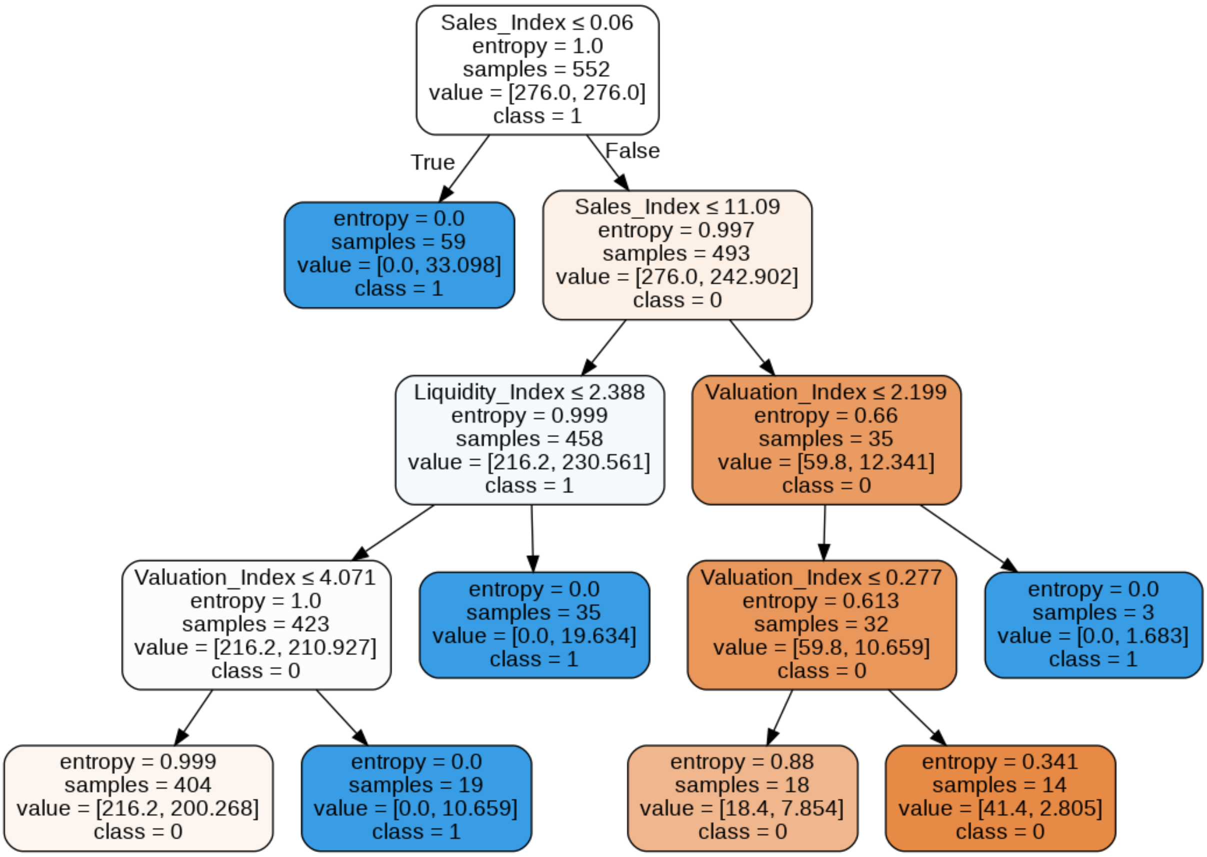


Predicted label

	precision	recall	f1-score	support
0	0.11	0.87	0.20	38
1	0.94	0.23	0.36	331
accuracy			0.29	369
macro avg	0.53	0.55	0.28	369
weighted avg	0.85	0.29	0.35	369

The problem with this model, with the 60% - 40% setting, is that the dataset is too unbalanced, which biases the overall performance of the model. Indeed, the sensitivity which measures the ability of the model to find the true positive boxes. Here, the sensitivity is 23%, which means that out of the total number of successful operations, our model only finds them in 23% of the cases. On the other hand, our model is largely wrong in identifying 77% of successful trades as failures. This error is important, but it is less serious than the opposite error: identifying a deal as a success when it is actually a failure. This misinterpretation could have serious financial consequences for the buyer. Regarding this more serious type of error, the algorithm was not very wrong in the end: it identified 87% of them. But once again, the lack of balance in the distribution of positive and negative transactions may be biasing our vision. Overall, our algorithm has an inclination to conclude that an operation is a failure, and therefore remains globally conservative, despite its overall accuracy score: 29%.

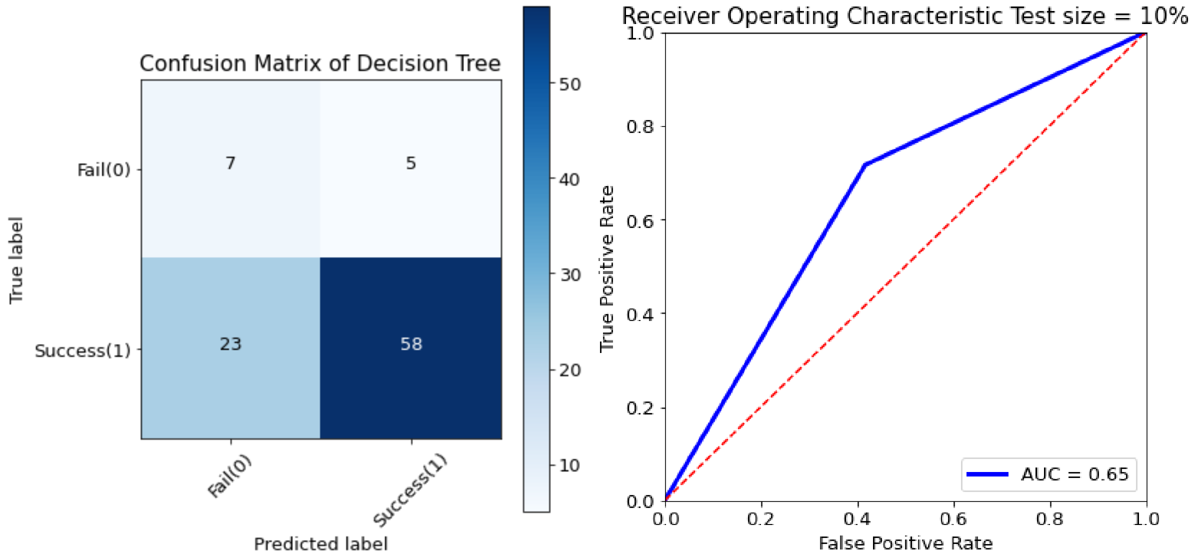
Figure 19: Decision Tree graph for the 60% - 40% split



In a second time, we have tried to increase our training set to feed the Decision Tree algorithm. The obvious danger is to avoid overfeeding, otherwise the algorithm might overfit the dataset, and will show clearly bias results as it will be tailor made for our dataset. The Figure 20 therefore shows the obtained results for a 90% - 10% split between the training and the test sets. We have investigated these parameters to check whether the Decision Tree model would have been slightly outperforming the two previous models, i.e. Logistic Regression and Support Vector Machine. Actually, it seems our current model offers quite satisfying outcomes. However, this split is quite odd as it seems it totally overfed the algorithm given the training set size. Despite this massive amount dedicated to the test set, the ROC curve is still far from the perfect classifier as AUC is about 0.65.

Going deeper in the analysis, the overall accuracy of the model is quite high: 70%. Nevertheless, the other ratios demonstrate the model is not reflecting all the truth, especially when taking into consideration the specificity ratio which is about 42% on only 12 values (13%) of the 93 ones of the test set. On the other hand, the precision ratio seems to be satisfying (92%), showing how efficient is the algorithm to predict a successful transaction when it actually is, avoiding the acquirer to bid a risky deal.

Figure 20: Decision Tree Confusion Matrix and its ROC curve for a 90% - 10% split

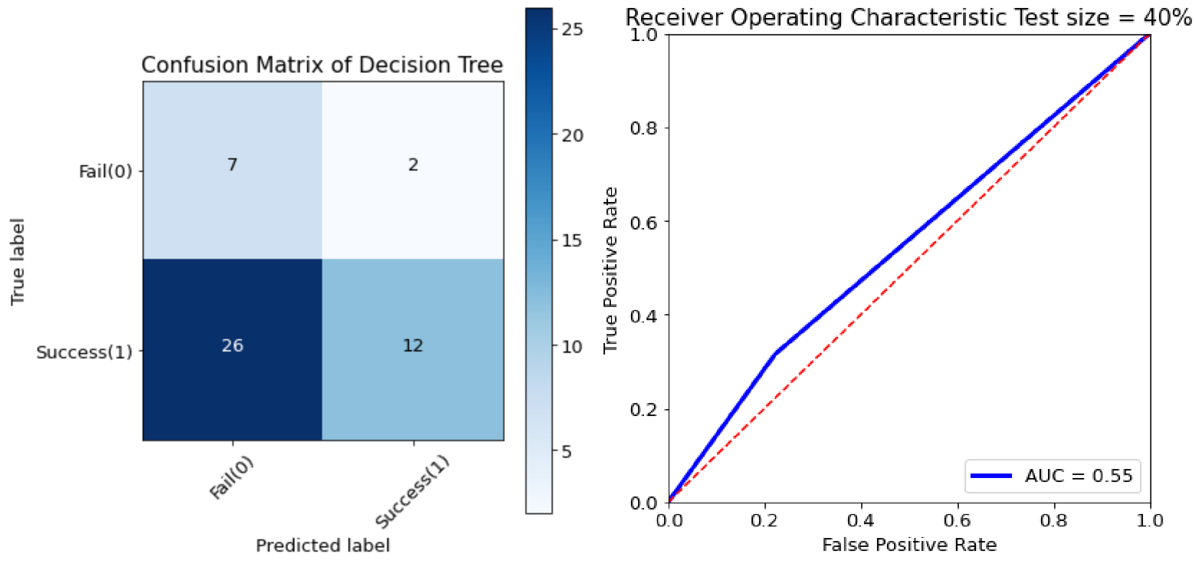


Predicted label

	precision	recall	f1-score	support
0	0.23	0.58	0.33	12
1	0.92	0.72	0.81	81
accuracy			0.70	93
macro avg	0.58	0.65	0.57	93
weighted avg	0.83	0.70	0.74	93

To better understand our model, we have moreover investigated the 95% - 5% case so as to check whether our Decision Tree model would overfit more to our dataset. Actually, Figure 21 presents that results and shows the model has a 55% AUC score, which is actually less than the 90% - 10% split scenario. It illustrates that our previous splitting was probably the sweet spot of our model, meaning we kind of maximized the model complexity and the accuracy of the DT model.

Figure 21: Decision Tree Confusion Matrix and its ROC curve for a 95% - 5% split

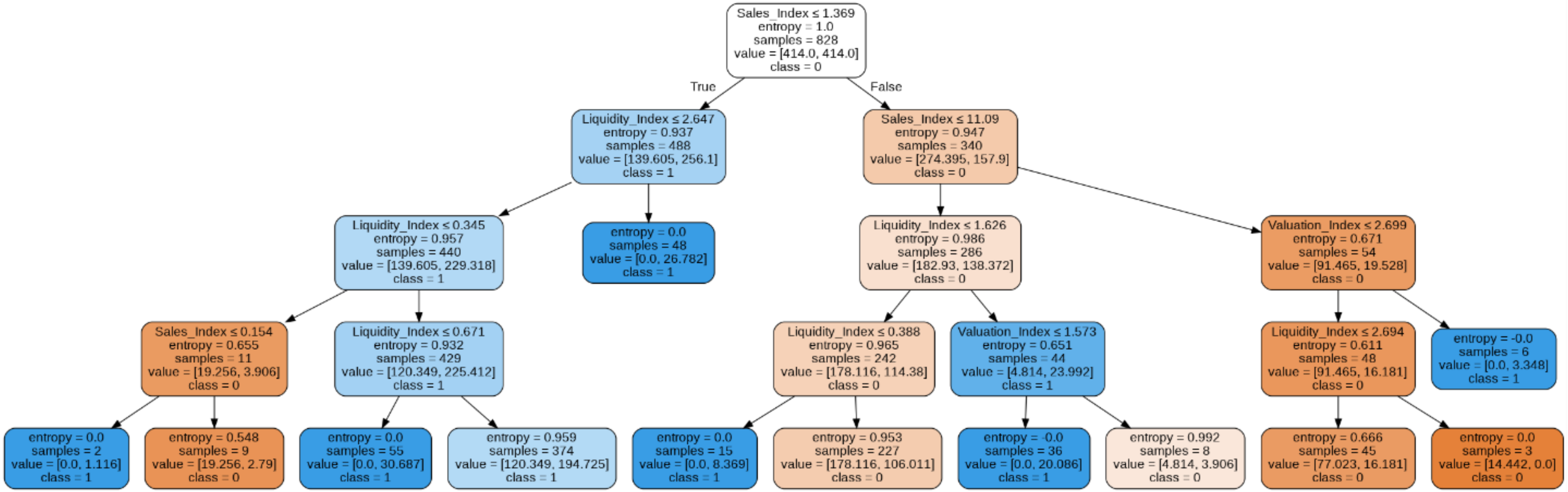


Predicted label

	precision	recall	f1-score	support
0	0.21	0.78	0.33	9
1	0.86	0.32	0.46	38
accuracy			0.40	47
macro avg	0.53	0.55	0.40	47
weighted avg	0.73	0.40	0.44	47

To conclude with this dataset split, we should reject the model or at least investigate it on a wider dataset, but the overfitting risk may be real. Unbalancing is one major reason of this rejection and should be studied to measure its influence on our results.

Figure 22: Decision Tree graph for the 90% - 10% split



The previous decision tree on the Figure 22 provides a visual representation as well as the ratios of the director vectors. For instance, the typical "Success" for the model is given by a Liquidity index higher than 0.671. Reversely, a typical case of failure is identified by a Sales index higher than 11.09 and a Liquidity index higher than 0.388.

The 10% split is rather unrealistic as 90% of the data was used during the training phase of the algorithm. There is therefore a real risk of overfitting the dataset. Consequently, it would be awkward to accept the results our algorithm returned as the split parameters (90% - 10%) are unrealistic. Despite results closer to what we expected, and which would be in line with what Lee et al. (2019) tell us, the 90% - 10% split of the dataset between training and testing should be put into perspective. The split could seem idealistic, as much importance is given to the supervision phase of the algorithm. Indeed, the risk of overfitting the model is real here since a large part of the dataset is dedicated to training. On the other hand, the dataset is also of a size to be relativised, and the model complexity is rather balanced in the way the number of parameters is made of only 3 parameters, when our dataset is made of 921 observations. We should also point out the difficulty to find clean and comprehensive dataset for the M&A transactions. This is because of two things. Firstly, M&A deals are probably one of the most perilous and difficult operations to carry out. The number of dimensions to be taken into account is undoubtedly beyond the scope of this study, which is based solely on financial and commercial ratios. The temporal trends were thus masked by taking into account the financial data of the players at the time of the deals, and the human character, as well as the reception of the markets, public authorities and employees, was not clearly taken into account. Although the ratio of the number of employees in the companies was taken into account, it did not appear to be more important than that in the analysis of the correlation matrix, and is in any case very closely related to the Sales_Index ($\rho = 0.5$). Secondly, M&A transactions are extremely difficult to predict anyway, but our research shows us that, even using a different methodology and different indicators than other studies conducted so far, the algorithm that best juggles with non-linearity does the best job, i.e. Decision Trees. Note that in a context of very high dimensionality, the latter should lose their advantage.

Finally, the methods used in research articles (K-nearest neighbour, ensemble models (Chandler, et al., 2019), neural networks (Lee, et al., 2019), go in the same direction and show better results with non-linear method algorithms.

Another important factor is the choice of the selected ratios. These ratios were selected on the basis of a study conducted by Cass University, as well as on the basis of methods used in practice by analysts in M&A teams, which the author of the present paper was able to experiment during an internship of more than six months.

7 Conclusion

7.1 Main findings and conclusion

Predictive analytics is now a well-advanced branch of artificial intelligence. The use of forecasting techniques requires rigour, reliable, clean data, and if possible, in very large numbers. We have sought to predict the outcome of M&A deals using supervised learning techniques. The results are mixed and need to be put into perspective as there are many limitations. Nevertheless, we have verified that the use of linear classification methods such as logistic regression and SVMs do not provide usable results, regardless of the amount of training we allocate to our model. On the other hand, it seems that with a higher training fraction, our decision tree model performs better than the two previous models, which could support the literature review. Finally, we can say that the linear models fail to predict the outcome of an M&A deal using the ratios we have selected in the methodology. Conversely, the non-linear model provides better results for the studies ratios.

Figure 23: AUC synthesis by test size and by models

Test size	Algorithm Type	Area Under the Curve
40%	**Logistic Regression**	**0.55**
	Support Vector Machine	0.53
	Decision Tree	**0.55**
30%	**Logistic Regression**	**0.55**
	Support Vector Machine	0.53
	Decision Tree	0.53
20%	Logistic Regression	0.54
	Support Vector Machine	0.50
	Decision Tree	**0.60**
10%	Logistic Regression	0.50
	Support Vector Machine	0.47
	Decision Tree (5% test size)	**0.65** (0.55)

7.2 Limitations

An M&A transaction cannot be limited to financial ratios and an analysis of past data alone. Many other variables must be considered in order to conclude that a merger or acquisition project is feasible, such as the managerial dimension, the adherence to the corporate culture, human resources, or simply the question of the legality of such a project with regard to the rules of competition. In any cases, M&A success predictor tool remains a useful asset that should be run at the very first step of the project. It consists into a first-approach tool to which belongs several limits. Management should not make too early conclusion from the results, or they might dramatically regret it in the coming years.

The imbalance in the dataset concerning their Class may have greatly biased the quality of the results, as we have seen above. Indeed, the sample is made of 89% positive cases, and only 11% of negative cases. We could have used technics to balance the dataset, such as weighted extreme learning machine (WELM) or minimum class-specific regulation ELM (MCVCSELM), which has been found to significantly improve the model results when imbalance occurs (Raghuwanshi & Shukla, 2021). MCVCSELM classifier aims to minimize the intra-class variance of the training set, while reducing the output weight norm. In the same time, Chen, et al. (2021) obtained accurate predictions of customer future purchases. They highlight the ability to use maximized marginal category and cost-sensitive ensemble learning to predict travel service purchases, in a clearly imbalanced positive to negative samples of 1:7. In our case, data imbalance might be explained by the fact that companies implied in failed deals do not want to disclose the contract terms, and also because they are not supposed to. Therefore, some key ratios might be missing for the older deals and Bloomberg does not offer the opportunity to catch them. It is therefore logical that the algorithms we have developed in our study cannot be fully trusted. The construction of the code also comes into play insofar as it guides the way in which we have processed the data. In our study, the algorithms remain rather modest and in a supervised Machine Learning framework, which at the end mitigates this bias. The choice of algorithms is also questionable as we have only chosen classification algorithms in line with our methodology. These algorithms have the advantage of being easy to use and the results are transparent and easily transposable, but do not provide the best results.

7.3 Research propositions

Research in predictive analytics is quite advanced, but there are relatively few studies that specifically address M&A transactions and the prediction of negotiation outcomes, as highlighted de Bodt et al. (2018). As we have seen, predicting the outcome of a transaction involves many barriers, including the need for sufficient quality data. However, a screening of potential target companies through clustering has not been encountered. Using unsupervised learning techniques to quickly investigate and search the market to characterize and cluster similar companies, and then classify them according to the acquirer's expectations. Some technics such as K-Mean, or even K-Nearest Neighbors could be assimilated as good starting points regarding an extending study and the discussed topics in this research paper.

7.4 Artificial Intelligence: entering a new spring?

It is clear nowadays that Artificial Intelligence has entered a new spring, after decades of winter. The number of publications on the subject of artificial intelligence continues to grow. The general public is hearing more and more about it as companies invest in it to modernise their services. It is therefore clear that we are only at the beginning of what promises to be a technological upheaval, applied to business and other areas. The small number of scientific studies available on the subject does not allow us to distinguish between fantasy and reality, especially since what seemed unrealistic 10 years ago now seems within reach, as companies such as Palantir, IBM or NVidia demonstrate it every day. As time goes on, companies will deploy increasingly advanced technological resources, leaving computers to find hidden patterns within large data sets. Humans will still have the choice of whether or not to perform the task, but they will now be supplemented by tools that will allow them to gain considerable accuracy and time. Effectiveness will then become efficiency, and no one, be it managers or companies, will be able to ignore the possibilities offered by artificial intelligence when their own competitors are equipped with such tools. Economic warfare is an important issue for any company wishing to position itself on a market and to be sustainable. Technological transformations have been accelerated by the COVID-19 crisis and the hunt for costs is now at the heart of many industrial companies. The consulting firm McKinsey confirms in a recent survey that 2,400 asked firms, more than 50% reported a regular use of AI in their business function (McKinsey Analytics, 2020, p. 2). Industrial companies have suffered greatly from the shutdown of factories, and their automation, whether on the production line or in their administration, is subject to great changes in the current decade. Topics such as Robot Process Automation (RPA) in accounting, E-cobots in warehouses and Chatbots in after-sales services are just the beginnings of what promises to be a future industrial revolution, as the supply chain integrator declared in a recent report (GEFCO, 2019). According to the McKinsey survey, the use of AI has increased over the last few years, enabling significant cost reductions in any business function. Nevertheless, investments have been reduced between 2018 and 2019, resulting from the realisation that AI is not a magic solution to solve all problems, but it may help, just like any other smartly used tool.

8 Bibliography

Aktas, N., de Bodt, E. & Roll, R., 2009. Learning, hubris and corporate serial acquisitions. *Journal of Corporate Finance,* 15(5), pp. 543-561.

Baumard, P., 2012. Tactical and permanent adjustment. In: *Le vide stratégique (The strategic vacuum).* 2nd ed. Paris: CNRS Edition , p. 66.

Bena, J. & Li, K., 2014. Corporate innovations and mergers and acquisitions. *The Journal of Finance,* 69(5), pp. pp 1923 - 1960.

Berk, J. & DeMarzo, P., 2017. Mergers and Acquisitions. In: A. D'Ambrosio, ed. *Corporate Finance, Fourth Edition.* Harlow: Pearson, pp. 994-1024.

Betton, S., Eckbo, E., Thompson, R. & Thorburn, K., 2014. Merger Negotiations with Stock Market Feedback. *The Journal of Finance,* Volume 69(Issue 1), pp. page 1705-1745.

Bing-Xuan, L., Michayluk, D., Oppenheimer, H. R. & Reid, S. F., 2008. Hubris amongst Japanese bidders. *Pacific-Basin Finance Journal,* 16(1-2), pp. 121-159.

Bloomberg, 2021. *Veolia Finalizes Deal to Buy Suez After Long Takeover Saga.* [Online] Available at: https://www.bloomberg.com/news/articles/2021-05-14/veolia-finalizes-deal-to-buy-suez-after-protracted-takeover-saga
[Accessed 31 July 2021].

Blouin, J. L., Fich, E. M., Rice, E. M. & Tran, A. L., 2020. Corporate tax cuts, merger activity, and shareholder wealth. *Journal of Accounting and Economics,* 71(1).

Bo, M. & Anand, M. V., 2021. Stock merger activity and industry performance. *Journal of Banking & Finance,* Volume 129(1).

Branch, B. & Yang, T., 2003. Predicting Successful Takeovers and Risk Arbitrage.. *Quarterly Journal of Business & Economics,* Volume 42(Issue 1/2), pp. pages 3-18.

Braun, E., 2020. *How can AI impact work in M&A?.* [Online] Available at: https://www.forbes.com/sites/forbestechcouncil/2020/03/04/how-can-ai-impact-work-in-ma/?sh=777f957f78bc
[Accessed 11 March 2021].

Brealey, R., Myers, S. C. & Allen, F., 2017. Mergers. In: K. L. Strand, ed. *Principles of Corporate Finance.* 12th ed. New York City: McGraw-Hill, pp. 813-842.

Breiman, L., Olshen, R., Stone, C. J. & Friedman, J., 1984. *Classification and Regression Trees.* 2nd ed. s.l.:Chapman and Hall/CRC.

Chandler, B. et al., 2019. *From Machine Leanring to M&A: Ten M&A Target Predictions through a Machine Learning Model,* Madison: Nicholas Center & Wisconsin School of Business.

Charkrabarti, A. & Mitchell, W., 2016. The role of geographic distance in completing related acquisitions: Evidence from U.S. chemical manufacturers. *Startegic Management Journal,* Volume 37(Issue 4), pp. page 673-694.

Chatt, R., Gustafson, M. & Welker, A., 2021. Firing frictions and the U.S. mergers and acquisitions market. *Journal of Banking & Finance,* 128(1).

Chen, S.-x., Wang, X.-k., Zhang, H.-y. & Wang, J.-q., 2021. Customer purchase prediction from the perspective of imbalanced data: A machine learning framework based on factorization machine. *Expert Systems with Applications,* 173(1).

Cortes, C. & Vapnik, V., 1995. Support-Vector Networks. *Machine Language,* Volume 20(Issue 3), pp. pages 273-297.

de Bodt, E., Cousin, J.-G. & Roll, R., 2018. Empirical Evidence of Overbidding in M&A Contests. *Journal of Financial & Quantitative Analysis,* Volume 53(Issue 4), pp. p1547-1579.

Ding, H., Hu, Y., Li, C. & Lin, S., 2021. CEO country-specific experience and cross-border mergers and acquisitions. *Journal of Corporate Finance,* 69(102039,).

Duchin, R. & Schmidt, B., 2013. Riding the merger wave: Uncertainty, reduced monitoring, and bad acquisitions. *Journal of Financial Economics,* Volume 107(Issue 1), pp. pp 69-88.

Eaton, G. W., Guo, F., Liu, T. & Officer, M. S., 2021. Peer selection and valuation in mergers and acquisitions. *Journal of Financial Economics,* 1(1).

Ferris, S. P., Jayaraman, N. & Sabherwal, S., 2013. CEO Overconfidence and International Merger and Acquisition Activity. *The Journal of Financial and Quantitative,* Volume 48(1), pp. pp 137 - 164.

Financial Times, 2020. *Veolia makes its move for rival Suez after years of talk.* [Online] Available at: https://www.ft.com/content/664ca58d-65c2-4ccd-90d9-fef15fb27fc0 [Accessed 31 July 2021].

GEFCO, 2019. *GEFCO Group: powered by innovation.* La Défense: GEFCO.

Giordani, P., Jacobson, T., von Schedvin, E. & Villani, M., 2014. Taking the twists into account: predicting firm bankruptcy risk with splines of financial ratios. *The Journal of Financial and Quantitative Analysis,* 49(4), pp. 1071-1099.

Gökhan, S. & Nihat, A., 2016. Using or Not Using Business Intelligence and Big Data for Strategic Management: An Empirical Study Based on Interviews with Executives in Various Sectors. *Procedia - Social and Behavioral Science*, 24 November, Volume 235, pp. 208-215.

Guang, X. et al., 2012. *A Supervised Approach to Predict Company Acquisition With Factual and Topic Features Using Profiles and News Articles on TechCrunch.* s.l., Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media.

Gupta, B. & P.Banerjee, 2017. Impact of merger and acquisitions on financial performance: Evidence from selected companies in India. *International Journal of Management and Business Research,* pp. 14-19.

Hossain, M. S., 2021. Journal of Economics and Business. *Merger & Acquisitions (M&As) as an important strategic vehicle in business: Thematic areas, research avenues & possible suggestions,* 116(1).

Humphery-Jenner, M. & Powell, R., 2014. Firm size, sovereign governance, and value creation: Evidence from the acquirer size effect. *Journal of Corporate Finance,* Volume 26(1), pp. Pages 57-77.

Hu, N., Li, L., LI, H. & Wang, X., 2020. Do mega-mergers create value? The acquisition experience and mega-deal outcomes. *Journal of Empirical Finance,* 55(1), pp. 119-142.

Kinateder, H., Fabich, M. & Wagner, N., 2017. Domestic mergers and acquisitions in BRICS countries: Acquirers and targets. *Emerging Markets Review,* Volume 32(Issue 1), pp. Pages 190-199.

Krishnakanthan, K., Shenai, G. & Brown, S., 2021. *In conversation: The CEO's new technology agenda.* [Online]
Available at: https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/in-conversation-the-ceos-new-technology-agenda
[Accessed 16 May 2021].

Larkin, Y. & Lyandres, E., 2019. Inefficient mergers. *Journal of Banking & Finance,* 108(1).

Lee, J. & Kwon, H.-B., 2016. Progressive Performance Modelling for the Strategic Determinants of Market Value in the High-Tech Oriented SMEs. *International Journal of Production Economics*, 25 October, pp. 91-102.

Lee, K. et al., 2019. Unbalanced data, type II error, and nonlinearity in predicting M&A failure. *Journal of Business Research*, 19 March, Issue 109, pp. 271-287.

M&A Research Centre at Cass Business School, City University of London and Intralinks, 2016. *Attractive M&A Targets: Part 1. What do buyers look for?,* London: City University of London.

Martin, R. L., 2020. M&A: The one thing you need to get right. *Harvard Business Review,* June, 94(6), p. 8.

McKinsey Analytics, 2020. *Global survey: The state of AI in 2020,* Silicon Valley: McKinsey Global Publishing.

Mhlanga, D., 2020. Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion. *International Journal of Financial Studies,* 31 July.8(45).

Morgan, J. N. & Sonquist, J. A., 1963. Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association,* 58(302), pp. 415-34.

Mueller, J. P. & Massaron, L., 2019. *Le Machine Learning pour les Nuls (Machine Learning for Dummies).* 2nd ed. Paris: First Intercative.

Mueller, J. P. & Massaron, L., 2019. *L'Intelligence Artificielle pour les Nuls (Artificial Intelligence for Dummies).* 2nd ed. Paris: First Interactive.

Mullainathan, S. & Spiess, J., 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, May, pp. 87-106.

Nguyen, N. H., Phan, H. V. & Simpson, T., 2020. Political corruption and mergers and acquisitions. *Journal of Corporate Finance,* 65(1).

Perboli, G. & Arabnezhad, E., 2021. A Machine Learning-based DSS for mid and long-term company crisis prediction. *Expert Systems with Applications,* Volume 174(1).

Perlich, C., Provost, F. & Simonoff, J. S., 2003. Tree Induction versus Logistic Regression: A Learning-Curve Analysis. *Journal of Machine Learning Research,* 4(1), pp. 211-255.

Philip, R., 2020. Estimating permanent price impact via machine learning. *Journal of Econometrics,* April, Volume 215(Issue 2), pp. 414-449.

Phillips, G. M. & Zhdanov, A., 2013. The Review of Financial Studies. *https://www.jstor.org/stable/23355403,* Volume 26(Issue 1), pp. pp 34 - 78.

Powell, R. G. & Stark, A. W., 2005. Does operating performance increase post-takeover for UK takeovers? A comparison of performance measures and benchmarks. *Journal of Corporate Finance,* Volume 11(Issue 1-2), pp. page 293-317.

Prado, M. L. d., 2018. *Advances in Financial Machine Learning.* 1st ed. Hoboken: Wiley.

Protiviti, 2016. *Guide to Mergers and Acquisitions,* Paris: Protiviti.

Provost, F. & Fawcett, T., 2013. Class Probability Estimation and Logistic "Regression". In: M. Loukides & M. Blanchette, eds. *Data Science for Business.* Sebastopol: O'Reilly, pp. 99-102.

Raghuwanshi, B. S. & Shukla, S., 2021. Minimum class variance class-specific extreme learning machine for imbalanced classification. *Expert Systems with Applications,* 178(1).

Renneboog, L. & Vansteenkiste, C., 2019. Failure and success in mergers and acquisitions. *Journal of Corporate Finance,* 58(1), pp. 650-699.

Rust Wiley, 2021. *As M&A accelerates, deal-makers are leveraging AI and ML to keep pace.* [Online]
Available at: https://techcrunch.com/2021/05/14/as-ma-accelerates-deal-makers-are-leveraging-ai-and-ml-to-keep-pace/
[Accessed 16 May 2021].

Sadoghi, A., 2020. *FI505E - Coding and Data Science for Accounting and Finance.* Rennes, Rennes School of Business.

Shen, H. et al., 2021. Does geopolitical risk promote mergers and acquisitions of listed companies in energy and electric power industries. *Energy Economics,* 95(1).

Statista, 2019. *Global Data Creation is About to Explode.* [Online] Available at: https://www.statista.com/chart/17727/global-data-creation-forecasts/ [Accessed 18 March 2020].

Varian, H. R., 2014. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives,* 28(2), pp. 3-28.

Zhang, C., Zhang, H. & Liu, D., 2020. A Contrastive Study of Machine Learning on Energy Firm Value Prediction. *IEEE Access*, 17 January, p. 99.

9 Table of figures