Master's Thesis

Double Degree in Management: Major in International Management

Department of Business and Management
Chair of Quantitative Methods for Management

Machine Learning in Financial Asset Management: a challenge between perception and reality

Prof. Marco Pirra

SUPERVISOR

Prof. Francesco Cappa

CO-SUPERVISOR

Simone Maria Caranci
Luiss 730151 – King's College London AC36468

CANDIDATE

Academic Year 2021/2022

Table of Contents

List of Figures and Tables

Acknowledgements

Prima di procedere con la trattazione, vorrei dedicare qualche riga a tutti coloro che mi sono stati vicini in questo percorso di crescita personale e professionale.

Vorrei cogliere l'occasione per ringraziare il mio relatore, il professor Marco Pirra, per la sua guida e supporto durante il processo di completamento di questa tesi e nei due anni che l'hanno preceduta.

Vorrei anche esprimere il mio apprezzamento a tutto il personale accademico del Dipartimento di Management.

La LUISS e il King's College London, per avermi permesso di vivere questa meravigliosa esperienza.

Infine, un immenso grazie alla mia famiglia, a Claudia e ai miei amici per il loro sostegno morale e incoraggiamento. Senza di voi questo percorso non sarebbe stato lo stesso.

Abstract

This paper provides an empirical examination of the effect of four variables on the use of Machine Learning (ML) within the decision-making process of managers operating in the Asset Management sector. The four variables are: Company Size, Knowledge of ML, Manager's Perception of the impact of the ML on Company Performance and Investor concerns. Analysing a sample of 53 managers from different countries operating in the Wealth Management sector, this study finds a non-significant impact on the Company Size. Furthermore, there is a positive effect of the Knowledge of ML on its deployment in the investment process. Similarly, the perception that its use improves the Company's Performance positively impacts its use frequency. Finally, there is a negative impact on ML use provided by the extent to which managers meet investor concerns on this technology. The results, controlled for the geographic variable of where the company's headquarters are located, show a greater likelihood of using ML if the company is in America and a lower probability if it is in Europe. The results on the other continents are not significant, and it is therefore impossible to estimate the effects. By applying the OLS regression model to the collected data, this work extends the existing literature in the organisational and managerial theory field, as well as in the use of new technologies.

1. Introduction

The use of Machine Learning (ML) within the world of Asset Management has become increasingly popular (Dhall *et al*., 2020): the main reason for this is that the best investment strategies are based on economic and financial theories. ML is a powerful tool to implement them and discover new ones (Lopez de Prado, 2020).

Machine Learning Definition:

ML, a branch of Artificial Intelligence (AI), is the "experiential learning" that a machine carries out using computational algorithms (Helm *et al*., 2020).

These algorithms utilise data input and output systems to recognise patterns and models that the machine "learns" autonomously and effectively to formulate recommendations and decisions without human intervention (Blum, 2007).

Through several iterative cycles, the algorithm can predict the outputs following the reception of an input (Agakov *et al*., 2006).

Asset Management Definition:

According to ISO 55000, International Standard issued by the International Organisation for Standardisation (ISO), Asset Management can be defined as the "coordinated activity of an organisation to realise value from assets" (ISO, 2014).

Companies operating in this sector make decisions, plans and activities to balance risks, costs, and opportunities in order to achieve the desired asset performance (IAM, 2022).

Some examples of figures operating in the Asset Management sector are investment managers who manage the assets of equity funds, pension funds and hedge funds.

The main activities of which Asset Management is composed and within which in recent years there has been increasing use of ML are portfolio construction, risk management, capital management, infrastructure and distribution, and sales and marketing (Snow, 2020). Both the use of these techniques and their studies have significantly increased in recent years.

Recent studies in the literature tried to analyse the phenomenon of ML use within the Asset Management sector, and two interesting points of reflection have emerged from them:

First, there has been a growing concern in the literature about the use of ML in many aspects of life, including investors' perception of its use in finance (Ipsos Moris, 2017). Other recent studies conducted on hedge funds have also shown a positive correlation between the use of ML and Performance, but it is little used, mainly in small businesses (Grobys, 2022). However, so far, the literature does not present an answer about the reasons behind these phenomena and an overview of the state of the art of ML use and its perception at a more enlarged level.

Therefore, this study aims at providing some new insights into this field.

First, I will investigate further if there is indeed a negative correlation between company size and the use of ML (Grobys *et al.*, 2022); Second, I will extend prior research to find out what are the reasons behind the use or not of ML within companies operating in this sector, analysing its use in relation to three other variables, which are: Knowledge of the ML, Perception of the impact of the ML on the Performance and perception of managers about how much the ML appeals to investors (Ipsos Moris, 2017); Third, I will contribute to a more holistic understanding of the phenomenon analysing not only Hedge Funds but to broaden the research to the larger category of Asset Management companies (Grobys, 2022).

The rest of the work is organised as follows: After presenting the analytical framework, I will verify the four hypotheses related to ML use in the Asset Management sector. The model tests the interdependencies between the use of this technology and the four variables mentioned above: Company size, Knowledge of the technology, perception of effectiveness on Performance and perception regarding the level of Trust placed by investors in this technology. This work will conclude with the analyses of the obtained results, a conclusion on the overall topic, and a suggestion for further research on the subject.

2. Literature Review

This literature review aims to give an overview of the topics analysed within the research work, examining the existing literature on ML use within the Asset Management sector in order to lay the foundations for the completion of the purpose of this study. First, I will explore the existing theories on the relationship between company size and the use of ML. Subsequently, following the definition of Supervised and Unsupervised ML, I will analyse the current literature on the managers' Knowledge of ML, their Perception of Performance and their Perception of the existence or not of Bias by investors on the technologies taken for examination, and if it influences the decision-making process.

2.1. Machine Learning and Company Size

The first part of the Literature Review aims to analyse the information on the relationship between the use of ML by companies operating in the Asset Management sector and their size. In their research, Grobys *et al.* (2022) performed a study of 826 Hedge Funds. Their results show a predominant use of ML in small companies operating in the Asset Management sector compared to larger ones. It should be noted that this study was conducted exclusively on hedge funds: a particular type of fund in which relatively liquid assets are traded, using more sophisticated trading, portfolio construction and risk management strategies in an attempt to increase performance (Lemke *et al.*, 2014). A possible explanation Chen *et al*. (2004) provided for this "size factor" is that using ML leads to creating unique strategies that are difficult to scale. Dietvorst, Simmons and Massey (2015) found a reason for the size of companies using ML in investors' aversion towards new technologies that replace human judgment. This scepticism leads to lower investments in these types of funds with consequent downsizing of the AUM of the companies that use these technologies. The article by Grobys *et al.* (2022)

concludes that small Hedge Funds use ML more than large ones and, because of this, they perform better. However, it should be emphasised that there may be other reasons behind the better performance of small-sized funds: in fact, Ammann and Moerth (2005) state that the reasons for the outperformance of small funds compared to large ones are a higher risk appetite that leads to higher returns, in addition to being more liquid, as a consequence of their more minor financial positions. Thus, considering what was said above, it is necessary to conduct an analysis focusing on different variables to study phenomena.

Hence, the first contribution that this work wants to provide is to understand the actual relationships between the use of ML and the size of companies, shifting the focus from hedge funds only, to the larger category of Asset Management (in which hedge funds are included).

2.2. Machine Learning and Management Perception

In this second part, I will investigate the present literature on the perception of ML use by managers operating in the Asset Management sector through several filters: namely, the manager’s Knowledge of the subject, their perception of Performance and, finally, their perception of the Trust placed by investors in these technologies.

But, before proceeding with this analysis, it is necessary to explain two fundamental concepts of ML that will be very important to fully understand the study of the manager’s perception of the ML itself: Supervised ML and Unsupervised ML.

Supervised ML

Supervised Learning is an ML technique that makes use of labelled datasets. A “training dataset” is created at first by manually tagging the data. (Mohri *et al.*, 2012). These datasets are intended to “supervise” algorithms as they categorise data or forecast outcomes (Norvig and Russell, 2010).

Unsupervised ML

Unsupervised Learning (or Deep Learning) uses ML algorithms to analyse and group unlabeled datasets. These algorithms discover hidden patterns in data without the need for human intervention (“unsupervised”) (Hinton and Sejnowski, 1999).

It can therefore be seen that the main difference between these two approaches to ML is the degree and level of human intervention within the processes and algorithms. This difference is visible in the initial phase, the affixing or not of tags, but it detects an essential factor to consider in analysing managers’ perception of ML: human intervention in the decision-making process. Therefore, it will be of interest for the present study to investigate how this variable impacts the research results.

2.2.1. Managers and Machine Learning Knowledge

Once the first phenomenon has been stated, it is possible to deepen the existing literature on the relationship between the use of ML and managers’ Knowledge of this topic.

In an article published in 2021, Ferreiraa, Ruivob and Reis (2021) analysed the relationship between data analysts and managers in companies that use ML, trying to understand how necessary the knowledge of ML is in managerial decisions and whether it impacts the extent to which this technology is used. According to this paper, at the various stages of ML adoption, the experience of data scientists and managers in this area can significantly influence results.

Too often, investment decisions are based on the opinion of the highest-paid person rather than on objective data structured and analysed through processes and technologies, as in the case of ML (McAfee and Brynjolfsson, 2017).

Managers familiar with ML approaches can ensure that the projects in their portfolio add the most value to their organisations. With a thorough understanding of ML, they can recognise good opportunities and potential benefits, effectively utilising their potential and giving their businesses a competitive advantage (Lee and Shin, 2020).

Their involvement in implementing innovation within one company area can give rise to a process that promotes innovation in all business areas (Rai *et al.*, 2009; Ramamurthy *et al.*, 2008). Conversely, a lack of their participation and understanding of potential technologies will produce little results. Therefore, for a project to be successfully implemented and increase the value of the business, there is a need for those who prepare the software (Ruane, 2019) and who will analyse the output (i.e. data scientist and manager) (Beech, 2000) to have a shared understanding (Sarker *et al.*, 2006). Therefore, it can be said that the relationship between the knowledge of ML management and the best performance in business terms, as well as the importance of communication and knowledge shared between managers and data analysts, are topics already debated and present in the literature. However, what is not present is an analysis of how the knowledge of this technology impacts managerial choices and the frequency of use of ML in decision-making processes.

There is, therefore, no answer to the presence or absence of a correlation between the knowledge of ML and its greater use to understand if it is one of the reasons for its adoption.

Starting from this question, this analysis will be the second contribution this work aims to give to research in this field.

2.2.2. Managers and Machine Learning Performance

The third point to be taken into consideration for the development of this work is the relationship between the frequency of ML use and the manager's perception of its impact on business performance, that is, whether performance increases as ML use increases.

In their article on hedge funds, Grobys *et al.* (2022) divide funds into four categories depending on how much ML is integrated into business processes:

- Traditional: they do not use ML.
- Discretionary: they are based on both ML and managerial choices with a greater focus on managerial decisions.
- Combined: they are based on both ML and managerial choices in equal measure
- Systematic: they are based on both ML and managerial choices with a greater focus on ML.

They conclude that funds that use ML as a tool in investment decision-making (Systematic Funds) outperform those that do not.

However, the study also underlines that funds that use an average level of ML and human decision together (Combined Funds) perform worse than those guided only by the instinct of managers (Traditional Funds) and those driven by the calculation of ML algorithms (which, as mentioned, show the best results).

The article by Ferreiraa, Ruivob and Reis (2021) highlights how using platforms that exploit ML can improve company performance, both from a financial and organisational point of view, and that many factors influence the result.

These, in fact, are the maturity of the platform, the feasibility of the project, the quality of the data, the intensity of the information and the compatibility between the existing systems within the organisation. As for the first factor, the maturity of the platform, it is easy to understand that the most complex and best-performing platforms are those used first in the processes and therefore are the result of continuous improvement over time (Lismont *et al*., 2017). The best performance exists because they allow a deeper understanding of data and its value creation potential (Anand *et al*., 2016).

Second, the project's feasibility is crucial when the goal is to produce value and obtain financial benefits (LaValle *et al*., 2011). ML projects could reinforce this effect (Kietzmann and Pitt, 2020).

The third factor is the quality of the data, which must be reliable, extensive in number and continuously updated (Liu *et al*., 2020). Since ML uses a massive amount of data, its quality is vital for the project to increase the value generated by improving business performance (Lee, 2003). Poor data quality can negatively impact company performance (Du and Zhou, 2012).

Fourth, the amount of information a company uses in its processes is called information intensity. Projects using ML as technology and involving a lot of information could produce advantageous commercial value (and, therefore, performance) because ML handles complexity and large amounts of data well (Jarrahi, 2018; Kuzey *et al*., 2014).

Finally, increasing the compatibility between existing systems increases the likelihood of achieving positive effects (Ruivo *et al.*, 2013; DeLone and McLean, 2003; Bradford and Florin, 2003).

Once the main factors affecting the performance of companies that use ML have been defined, it is necessary to highlight a sphere that is not covered by the existing literature: in fact, whether the impact of performance is studied, what is not examined is the perception that managers have on the practical implications for the performance, in particular of companies operating in the Asset Management sector, and on how this affects the choices of use of these platforms and technologies. Furthermore, as anticipated in the first part of the chapter, the difference between the use of Supervised and Unsupervised Learning technologies is significant for deepening the topic and digging into the reasons behind managerial choices. Indeed, the research aims to be able to study how managers perceive the benefits of using ML on performance, compared to the risks that these methodologies may entail (Castagno and Khalifa, 2020; Ipsos Moris, 2017). It also tries to figure out if the degree of participation of the manager in the decision-making process in investment choices (Supervised/Unsupervised Learning) plays a role in the frequency of use of the ML in business activities.

2.2.3. Managers' and Investors' Perception of Machine Learning

The last part of the literature review will analyse the existing literature on investors' perceptions of using ML in investments. It will also be investigated how and if this impacts the choices of use of this technology by managers operating in the Asset Management field.

In 2017, the English company specialised in market research Ipsos Mori carried out a study on behalf of the Royal Society about the public's view on ML (Ipsos Mori, 2017).

This study focused on various fields of application, including the financial one.

In this research, great emphasis was placed on the risk/benefit trade-off perceived by the interviewees regarding the technology in the various application areas.

While the survey participants favoured using ML to monitor their transactions to safeguard them from possible fraud, few agreed with the idea that ML provided a consultancy service. In fact, they did not want technology to decide on investment choices and that a machine could daily check their portfolio.

Other concerns are related to the vision of human beings as unique entities. Therefore, it is difficult to generalise one's actions and ideas through algorithms, in addition to the fact that investors perceive the use of machines as a danger to their freedom of decision in investments.

Given a sample of 978 adults, when asked if they were in favour of a computer, adapting to the financial market, and investing money on their behalf, only 18% replied in the affirmative, 41%

responded negatively, and 30% replied they remained indifferent. The remaining 11% did not know how to answer (Ipsos Mori, 2017).

Once we understand investors' perception regarding ML in private investments, it is interesting to understand the current state of the relationship between the asset manager and investor from a regulatory point of view.

Wealth managers are one of the longest-running targets for regulation (Clark, 1981). In fact, their work and their relationship with investors are regulated by MiFID I and II in order to avoid Liability of Asset Managers problems (European Parliament and Council, 2004; European Parliament and Council, 2014; Moloney, 2012). Therefore, managers are obliged by law to pursue the interests of investors without exposing them to big significant risks and listening to their needs.

To the author's best knowledge, a study of the relationship between asset managers and investors' concerns on ML is missing in the literature. Specifically, regarding how investor concerns about the use of ML in investment choices are considered by managers and affect the use of this technology. This analysis will again consider how and if the difference between Supervised and Unsupervised Learning affects perception.

2.3. Analytical Framework

As highlighted in the first part of this chapter, the purpose of this work is to fill some of the gaps in the literature regarding the relationship between the use of ML technologies and the four variables previously explained: Company size, Knowledge, Managers' Perception on Performance, and relation with Investors' Perception.

In the Literature Review, the contents of the texts by Chen *et al*. (2004), Dietvorst, Simmons and Massey (2015) and Grobys *et al*. (2022) were analysed.

These findings highlighted a relationship between the size of companies operating in the financial branch of hedge funds and the use of Artificial Intelligence (AI) technologies (in particular ML). As previously stated, according to the author's best knowledge, however, a study on the broader sector of Asset Management is not present in the literature.

Therefore, taking into consideration all of the above, the following hypothesis can be developed:

Hypothesis 1*: Small companies are more likely to use ML than larger companies.*

Subsequently, the importance of the knowledge of ML by managers for the improvement of company performance was analysed, together with how good knowledge shared between managers and data analysts can influence the success of investment projects in the operations carried out by asset management companies and more (Ferreiraa, Ruivob and Reis, 2021).

Once again, however, this thesis has the ambition to study the relationship between the variable taken into consideration (in this case, the knowledge of ML) and the use of ML within the decision-making process implemented by the manager.

Hence, in conclusion to this review of the existing documents on the relationship between managers' knowledge of ML and its use, it is possible to formulate the second hypothesis of this study:

Hypothesis 2*: Managers who have more knowledge of ML are more likely to use it than managers who have less knowledge of it.*

The third point analysed in the Literature Review is the relationship between the use of ML and company performance.

In their work, Grobys *et al.* (2022) analysed how the use of ML impacted the performance of hedge funds. As seen above, a positive correlation was found between the use of technology and the results obtained. However, the presence of machine-manager hybrids (Combined Fund) was found to be less performing than "pure" forms (Traditional and Systematic funds).

As already mentioned, this study focuses on the hedge funds segment and studies its current performance without focusing on what the perception that managers have of ML is and, therefore, on how and if this affects the frequency of use of this technology in the decision-making process.

So, returning to the central theme focused on Asset Management businesses, once the significance of data and these aspects affecting corporate performance have been grasped, it is possible to start researching the manager-performance relationship from the formulation of the third hypothesis:

Hypothesis 3*: Managers who believe that using ML leads to better performance are likelier to use it than Managers who think it leads to lower performance.*

This hypothesis will be studied by analysing both Supervised and Unsupervised ML presences to understand if these lead to differences in the manager's perception.

Finally, the last point considered in the literature review is the manager-investor relationship and how the latter perceives ML's Usefulness.

The report drawn up by the company Ipsos Mori (2017) highlighted how investors are reluctant and averse to the extensive use of Artificial Intelligence technologies in the financial field for the purpose of advice and choice in investment, while they proved to be favourable to other services, such as detecting any fraud.

We have also seen how international banking regulations place great emphasis on the relationship between managers of a wealth management company and investors (European Parliament and Council, 2004; European Parliament and Council, 2014).

However, it has been seen that this need for ad hoc regulation was necessary as fund managers are often encouraged to pursue their own interests rather than those of investors by creating Liabilities of Asset Managers (Moloney, 2012).

Therefore, considering the statements mentioned above, the latest contribution that this work aspires to provide is a study on the impact that investor concerns have regarding the use of ML in the financial field on the investment choices of managers.

Once again, the study will consider the presence of Supervised and Unsupervised ML and their impact on investors' perception first and then on managers.

Hence, it is possible to formulate the fourth and final hypothesis of the paper:

Hypothesis 4*: Managers who believe investors like ML are more likely to use it than Managers who believe investors don't like it*.

This paper's analytic framework (Figure 1) is a Fixed Analytical Frame, generally used for quantitative studies and tests a series of hypotheses against the data (Ragin and Amoroso, 2011).

The relationships between the use of ML and the four variables just illustrated will, in fact, be realised through a regression model on the data collected through a survey sent to companies operating in the asset management sector (Appendix 2).

This model will be better explained in the next chapter of the paper.

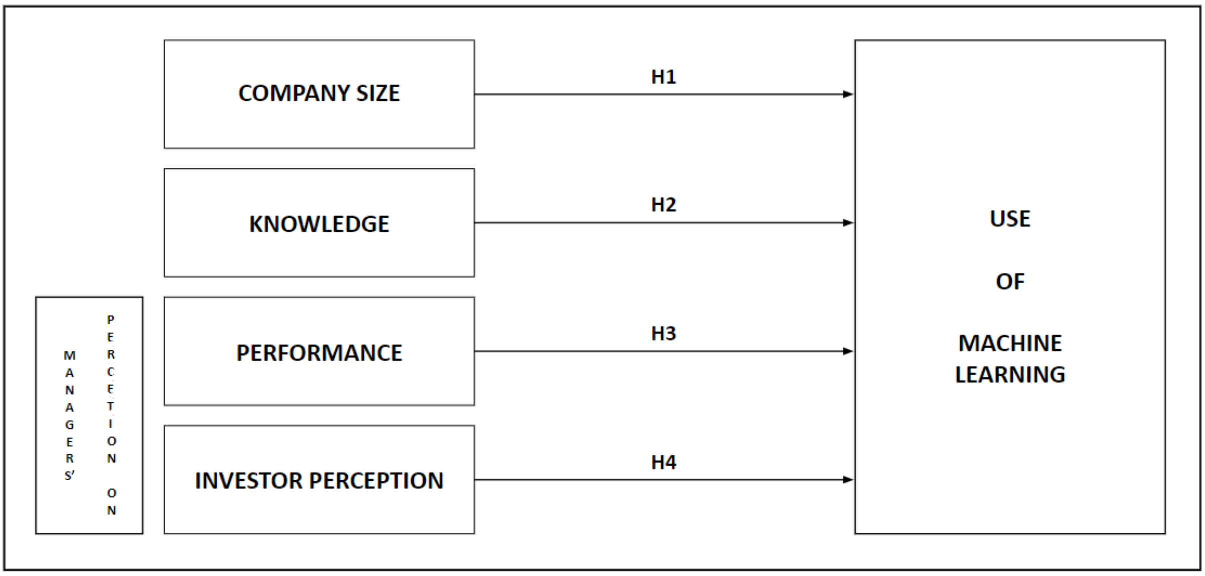

Figure 1: Analytical Framework

3. Research Methodology

3.1. Sample

The study and analysis were conducted through the administration of an anonymous survey (Appendix 2). Respondents were given the opportunity to reply only once. This survey's target sample is composed of managers of companies operating in the Asset Management sector.

The survey was sent via email and LinkedIn to an initial sample of managers who work in 150 companies operating in the world of wealth management. These companies had a different dimension regarding Asset Under Management (AUM), headquarters distributed evenly on all continents and a foundation year in a time horizon ranging from the eighteenth century to 2021. Only 53 companies responded to the survey out of 150.

As shown in Table 1, out of the 53 responses received, 35 came from managers whose companies have their headquarters in Europe, 15 in America, 2 in Africa and one in Asia.

The time range of the foundation is extensive. In fact, a company was founded in the eighteenth century, 8 in the nineteenth century, 24 in the twentieth century and 20 after 2000.

Headquarter		Foundation Year	
Africa	2	1700-1799	1
America	15	1800-1899	8
Asia	1	1900-1999	24
Europe	35	2000+	20
Oceania	0		
Tot	53	Tot	53

Table 1: Sample Composition - Headquarters and Foundation Year

Table 2 shows how the sample is composed in terms of AUM and Integration of the ML in business processes.

Using the division that Deloitte uses for the size of the funds according to the AUM, it is possible to divide the companies into three groups (Deloitte, 2020): Small, with an AUM of less than 100 billion; Medium, with an AUM between 100 and 500 billion; and Big, with AUM exceeding 500 billion.

The sample includes 32 small firms, 11 medium-sized and 10 big firms.

Finally, according to the categories Grobys *et al.* (2022) determined on the extent to which ML is integrated within the decision-making processes of the fund, there are 9 Traditional companies, 13 Discretionary companies, 15 Combined companies and 16 Systematic companies.

AUM		ML Integration	
Small (< 100 bln)	32	Traditional	9
Medium (100 bln - 500 bln)	11	Discretionary	13
Big (500+ bln)	10	Combined	15
		Systematic	16
Tot	53	Tot	53

Table 2: Sample Composition - AUM and ML Integration

3.2. Measures

The model used in this analysis is a multiple linear regression that uses the least squares method (OLS) to study the relationship between the dependent variable "Use of the ML" and the independent variables.

This method, discovered and published by Adrien-Marie Legendre (1805) (Mansfield, 1877) and co-credited to Carl Friedrich Gauss (1795) (Stigler, 1981), is a standard approach in regression analysis that minimises the sum of squares of the residuals (difference between an observed value and the fitted value provided by a model) of the results of each individual equation in order to study the fit of the data.

The model is constructed as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, i = 1, \ldots, n$$

Where,

Y is the dependent variable.

X_1, X_2 and X_k are the independent variables (regressors).

(Y_i, X_{1i}, X_{2i}, X_{ki}) denote the i-th observation on Y, X_1, X_2, X_k .

β_0 is the intercept of the unknown population.

$\beta_m, m = 1, \ldots, k$ is the effect on Y of a change in X_m, holding all the other variables constant.

u_i is the regression error (omitted factors) (Berry and Feldman, 1985).

Therefore, I will study the effect that independent variables have on the dependent variable.

3.2.1. Dependent Variable

The dependent variable is the use that managers make of ML within the decision-making process in companies operating in the Asset Management sector.

This variable is of interest in the understanding state of the art in using this technology. Studying the interactions with independent variables allows us to understand the reasons behind its use.

In order to collect information about this variable, in the first part of the survey, three questions were asked relating to the frequency of use by the respondent manager, the frequency of use by the company where the manager works and the category in which the company falls according to the description of Grobys *et al*. (2022).

Of these three variables, the only one that will be considered as the dependent variable is the first, since I am interested in studying the behaviour of the individual manager.

The other variables will be used as control variables, as discussed later.

In order to measure the frequency of use in the survey, a 5-point Likert scale was adopted with a value of 1 "Never" and 5 "On a daily basis".

The questionnaire questions result from an adaptation of papers and surveys already validated in the literature.

Specifically, the questions related to the frequency of use of the ML have been adapted from the aforementioned paper by Ipsos Mori (2017).

3.2.2. Independent Variables

In the previous chapters, four areas have been identified that can be considered as the reason and, therefore, the object of the study for the use of ML: (1) the Company Size, (2) the Knowledge, (3) the Manager's Perception of Performance and (4) the Manager's Perception of investor concerns.

In order to study these phenomena, within the survey, more questions were asked by the research area in order to find the most significant ones, refine the model and go deeper into the question (Siemsen, 2010).

This study will use the answers to these questions as independent variables.

(1) The *Company Size* is calculated through three measures: the amount of AUM, the number of employees and the number of countries in which the asset management company operates (Appendix 2).

These values were modified and cleaned for analysis by eliminating the outliers and using a logarithmic transformation in order to "normalise" them, reduce or eliminate skewness and thus make them more manageable and valid in the application of the statistical model (Ives, 2015).

(2) *Knowledge of ML* is calculated by answering the question, "How much, if anything, would you say you know about Machine Learning?". The variable is calculated as a 5-point Likert scale which takes the value of 1 when the answer is "Not at all" and 5 when the answer is "A great deal" (Appendix 2).

The formulation of these questions is the adaptation of two surveys present in the literature on the subject, respectively, by Ipsos Moris (2017) and Castagno and Khalifa (2020).

(3) The *Manager's Performance Perception is calculated using* four questions (Appendix 2). A 5-point Likert scale was adopted for all four:

The first variable tests how much the manager believes ML is valuable and therefore impacts performance positively. The scale range has values of 1 for "Of no use at all" and 5 for "Extremely useful"; The second variable is the manager's perceived level of risk/benefit with respect to the application of ML in managerial choices. The range of the scale has values of 1 for "The risks are much bigger than the benefits", and 5 for "The benefits are much bigger than the risks"; The third and fourth variables test how comfortable the manager feels in using a technology of Unsupervised Learning (Q3) and Supervised Learning (Q4). The scale range has values of 1 for "Not at all comfortable" and 5 for "Very comfortable". These questions were adaptations of the papers by Ipsos Moris (2017) and Castagno and Khalifa (2020).

(4) The *Managers' Perception of Investor Concerns* is calculated through three questions (Appendix 2). A 5-point Likert scale was adopted for all three:

The first variable is the manager's level of importance to the investors' concerns. The range of the scale has values of 1 for "Not important at all" and 5 for "Extremely important"; The second and third variables test how comfortable the manager thinks the investor feels about using Unsupervised Learning (Q3) and Supervised Learning (Q4) technology. The scale range has values of 1 for "Not comfortable at all" and 5 for "Very comfortable". These questions were adaptations of the articles by Ipsos Moris (2017) and Castagno e Khalifa (2020).

3.2.3. Control Variables

Various control variables have been implemented to control for other influencing factors (Bryman and Cramer, 2005).

To this end, a series of questions were asked in the questionnaire whose validity is the result of adaptation to the aforementioned papers by Ipsos Mori (2017) and Grobys *et al*. (2022).

1) *Use of the ML by the company where the manager works*. This variable is calculated as a 5-point Likert scale which takes a value of 1 when the answer is "Never" and 5 when the answer is "On a daily basis".
2) *Company category*. A dummy variable that identifies the four categories framed by Grobys *et al*. (2022) - Traditional, Discretionary, Combined, and Systematic - and, therefore, the general level of integration of the ML in the processes. It is composed as follows:

Company Category (Traditional = 1, Discretionary = 2, Combined = 3, Systematic = 4)

3) *Foundation Year*. The variable indicates the year the company was founded and its longevity.

4) *Headquarters*. Set of dummy variables indicating the continent in which the company's headquarters is located.

   It is composed as follows:

   Africa (Africa = 1, America = 0, Asia = 0, Europe = 0)

   America (Africa = 0, America = 1, Asia = 0, Europe = 0)

   Asia (Africa = 0, America = 0, Asia = 1, Europe = 0)

   Europe (Africa = 0, America = 0, Asia = 0, Europe = 1).

4. Results and Analysis

In this section, I will present and discuss the main findings of my analysis. I will start with the statistical description of the variables and the presentation of the correlation with the dependent variable. I will then present the multiple regression analysis of the use of ML by managers against independent variables.

4.1. Descriptive Statistics and Correlation

Table 3 shows the Descriptive Statistics of the variables indicating the number of cases, the minimum, the maximum, the mean and the standard deviation of the variables.

Variables	n	Minimum	Maximum	Mean	Std. Deviation
N. Application Manager	53	0	4	2.42	1.447
Log_AUM	53	13.82	28.71	23.1301	3.90302
Log_N.Employees	53	1.10	18.13	6.3991	3.34841
Log_N.Countries	53	0.69	4.71	2.2986	1.10008
Knowledge of ML	53	1	5	3.75	1.343
Usefulness of ML	53	1	5	3.74	1.195
Risks/Benefits	53	2	5	3.83	0.914
Unsupervised ML - Manager	53	1	5	3.26	1.347
Supervised ML - Manager	53	2	5	4.11	0.913
Investor Importance	53	3	5	3.47	0.639
Unsupervised ML - Investor	53	1	5	2.91	1.290
Supervised ML - Investor	53	2	5	3.75	0.979
N. Application Firm	53	1	5	3.34	1.556
Dummy_Categories	53	0.00	3.00	1.7170	1.08091
Foundation Year	53	1700	2021	1960.77	81.373
Dummy_Africa	53	0.00	1.00	0.0377	0.19238
Dummy_Asia	53	0.00	1.00	0.0189	0.13736
Dummy_America	53	0.00	1.00	0.2830	0.45478
Dummy_Europe	53	0.00	1.00	0.6604	0.47811
Valid n (listwise)	53				

Table 3: Descriptive Statistics

Table 4 is a correlation matrix showing the interaction between the individual variables (Pearson, 1895).

The first column, which shows the correlation between the dependent, independent and control variables, is interesting for the analysis. The AUM variable is negative, albeit with a coefficient close to zero (-0.03). This negativity goes against what is present in the literature (Grobys, 2022). It must be said, however, that the variable is not significant. This value can be given by a small sample size related to the variability of the data (Sandelowski, 1995). The other two variables relating to the size of the company (no. Employees and no. Countries in which the company operates) are positive (0.098 and 0.158), but they are also not significant.

Except for the foundation year and the dummy variables Africa and Asia, all other variables are significant at the 0.01 level. The Knowledge variable is positive with a high correlation (0.716). This element implies that the use of ML tends to increase as the manager's knowledge of the technology increases. All the variables that measure the manager's perception of the ML impact on performance have a positive correlation coefficient. In fact, the use of ML tends to grow as the manager's perception of its Usefulness in increasing financial performance increases (0.832), as much as the manager believes that the benefits of its use outweigh the risks (0.752) and the manager's trust of the efficiency of Unsupervised Learning (0.683) and Supervised Learning (0.604) technologies. The variables on the relationship manager-investor also have a significant correlation coefficient. The use of ML decreases as the manager supports the thoughts and concerns of investors (-0.278). It is in line with the literature, as investors have an aversion to using ML in financial practices (Ipsos Mori, 2017; Dietvorst, Simmons and Massey, 2015). Therefore, the greater the manager's importance to investors' judgments, the lower the propensity to use ML.

On the other hand, the relationship between investors' confidence in using the methods of Unsupervised (6.19) and Supervised (6.30) Learning is positive. The reason is probably that the more managers feel that investors are comfortable using these technologies, the more they will be pushed to use them. Regarding the control variables, the ML frequency of use by the company and the category in which the firm falls (Grobys *et al.*, 2022) are both significant and positive. These have very high coefficients (0.876 and 0.900). It could create multicollinearity problems when multiple regression is performed, such as erratic coefficient signs and inconstancy in significance (Greene, 2000). The foundation year has a positive but not significant correlation coefficient (0.165). Finally, out of the dummy variables of the continent in which the headquarters is located, only America and Europe are significant. The result may depend on the number of survey replies for the other two continents, which are only one for Asia and two for Africa. Variable America has a positive coefficient (0.461), while Europe has a negative coefficient (-0.376). This data is interesting because it shows that the company's presence within the American territory increases the probability that the manager uses ML. Conversely, this probability drops if a manager works in a European company.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
N. Application Manager	1																		
Log_AUM	-0.03	1																	
Log_N.Employees	0.098	0.595**	1																
Log_N.Countries	0.158	0.727**	0.736**	1															
Knowledge of ML	0.716**	0.177	0.204	0.337*	1														
Usefulness of ML	0.832**	0.029	0.031	0.177	0.594**	1													
Risks/Benefits	0.752**	-0.089	-0.098	0.107	0.717**	0.785**	1												
Unsupervised ML – Manager	0.683**	-0.225	-0.014	0.092	0.706**	0.677**	0.818**	1											
Supervised ML – Manager	0.604**	-0.114	-0.176	0.05	0.447**	0.645**	0.761**	0.695**	1										
Investor Importance	-0.278**	0.171	-0.022	0.023	-0.064	-0.186	-0.222	-0.304*	-0.324*	1									
Unsupervised ML – Investor	0.619**	-0.374**	-0.109	-0.06	0.630**	0.570**	0.736**	0.889**	0.597**	-0.225	1								
Supervised ML – Investor	0.630**	-0.228	-0.206	-0.065	0.539**	0.650**	0.726**	0.750**	0.807**	-0.181	0.850**	1							
N. Application Firm	0.876**	0.134	0.06	0.268	0.777**	0.835**	0.825**	0.727**	0.677**	-0.222	0.639**	0.687**	1						
Dummy_Categories	0.900**	-0.016	0.01	0.124	0.746**	0.849**	0.787**	0.753**	0.637**	-0.249	0.684**	0.715**	0.927**	1					
Foundation Year	0.165	-0.514**	-0.661**	-0.670**	-0.133	0.064	0.072	0.093	0.107	-0.262	0.12	0.127	0.042	0.139	1				
Dummy_Africa	-0.265	0.15	-0.027	0.046	0.037	-0.207	-0.072	-0.188	-0.134	0.322*	-0.218	-0.154	-0.044	-0.133	-0.395**	1			
Dummy_Asia	0.153	-0.229	-0.167	-0.153	0.13	0.148	0.179	0.18	0.136	-0.103	0.227	0.178	0.149	0.166	0.086	-0.027	1		
Dummy_America	0.461**	-0.082	-0.059	0.002	0.399**	0.459**	0.488**	0.504**	0.292*	-0.071	0.538**	0.461**	0.487**	0.596**	0.114	-0.124	-0.087	1	
Dummy_Europe	-0.376**	0.084	0.115	0.023	-0.432**	-0.396**	-0.486**	-0.455**	-0.263	-0.032	-0.489**	-0.428**	-0.488**	-0.562**	0.026	-0.276*	-0.193	-0.876**	1
n	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 4: Correlation Matrix

4.2. Test of Hypotheses

Table 5 presents the results of the multiple regression analysis for the sample using all the variables. To test the various effects of Company Size, ML Knowledge, Perception of ML usefulness on Performance and the relationship between managers and investors, I run sixteen different multiple regression models using a different number of independent variables at the same time. Specifically, the models are constructed by analysing all the independent variables individually, in the groups relating to the study areas connected to the hypotheses and, finally, all together. All models are controlled for the factors previously presented. The dependent variable is how often managers use ML. As can be seen from Models 1 to 16, none of the regression coefficients of the independent variables is significant, except for the Log_N.Employees in Models 5 and 16.

Variables	Model 1 Controls Only	Model 2 AUM	Model 3 N. Employees	Model 4 N. Countries	Model 5 Company Size	Model 6 Knowdlege	Model 7 Perceived Usefulness	Model 8 Risk/Benefit Perception	Model 9 Manager Unsupevised	Model 10 Manager Supevised	Model 11 Manager Perception on Performance	Model 12 Investor Importance	Model 13 Investor Unsupervised	Model 14 Investor Supervised	Model 15 Manager Perception on Investors	Model 16 All
Log_AUM		-0.019			-0.042											-0.076
Log_N.Employees			0.064		**0.091***											**0.119***
Log_N.Countries				0.000	-0.031											-0.009
Knowledge of ML						0.111										0.230
Usefulness of ML							0.150				0.126					0.213
Risks/Benefits								0.097			0.182					0.121
Unsupervised ML - Manager									-0.038		-0.068					-0.314
Supervised ML - Manager										-0.050	-0.103					0.229
Investor Importance												0.007			0.017	0.061
Unsupervised ML - Investor													-0.016		0.066	0.040
Supervised ML - Investor														-0.090	-0.151	-0.120
N. Application Firm	**0.325***	0.317	**0.326***	0.326	**0.389***	0.280	0.284	0.282	**0.336***	**0.345***	0.268	**0.325***	**0.329***	**0.341***	**0.340***	0.252
Dummy_Categories	**0.820****	**0.827****	**0.785****	**0.820****	**0.702****	**0.774****	**0.732****	**0.821****	**0.837****	**0.823****	**0.782****	**0.821****	**0.825****	**0.851****	**0.852****	0.562
Foundation Year	0.00015	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
Dummy_Africa	-1.01931	-0.961	-0.708		-0.602	**-1.042***	-0.888	-0.987	-1.068	-1.033	-0.964	-1.024	-1.048	-1.071	-1.003	-0.669
Dummy_America				-1.020												
Dummy_Asia	0.157	0.052	0.328	0.157	0.140	0.134	0.134	0.133	0.174	0.170	0.148	0.160	0.171	0.189	0.161	0.049
Dummy_Europe	0.315	0.306	0.255	0.316	0.244	0.314	0.299	0.340	0.300	0.323	0.336	0.317	0.303	0.296	0.336	0.095
n	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53	53

a. Dependent Variable: N.

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 5: Multiple Regression n.1

As said before, the reason behind this can be given by the variables' multicollinearity present in the model (Greene, 2000). Furthermore, the large number of variables inserted in the model could lead to an over-parameterised model (Harrell, 2001). Therefore, it is good to eliminate the variables that create problems. To do this, I first excluded from the model the variables that were not significant from the correlation study and then I carried out a study on the multicollinearity of the remaining variables. The Variance Inflation Factor (VIF) of the variables inserted in the model (James *et al.,* 2017) is used to study multicollinearity. The variables with a multicollinearity threshold higher than 5 (Shrestha, 2020) are excluded. Thus, I exclude the independent variables on the size of the company (AUM, No. of Employees and No. of countries in which the company operates) and the control variable Foundation Year, as they have insignificant correlation coefficients. In Table 6, it is possible to see the VIFs of the remaining variables. Considering these data, I exclude from the model the variables with VIFs greater than 5. The only variable that does not respect this condition but I still maintain is the Utility of the ML. It is to have a variable that measures managers' perception concerning the impact of the ML on performance and because it slightly exceeds the threshold (5.731): data that with a reduced variable regression could collapse.

In addition to the non-significant variables in the correlation, the variables excluded with the related VIFs are: Risk/Benefit (7.121), Unsupervised ML - Manager (11.937), Supervised ML - Manager (8.645), Unsupervised ML - Investor (18.933), Supervised ML - Investor (14.259), N. Application Firm (11.203) and Dummy_Category (12.536).

Variables	Beta	VIF
Knowledge of ML	0.173	3.983
Usefulness of ML	0.256	5.731
Risks/Benefits	0.003	7.121
Unsupervised ML - Manager	-0.348	11.937
Supervised ML - Manager	0.256	8.645
Investor Importance	-0.058	1.474
Unsupervised ML - Investor	0.375	18.933
Supervised ML - Investor	-0.427	14.259
N. Application Firm	0.163	11.203
Dummy_Categories	**0.786****	12.536
Dummy_Africa	-0.883	1.771
Dummy_Asia	0.017	1.090
Dummy_Europe	0.288	2.321

a Dependent Variable: N. Application Manager

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 6: Multicollinearity

Once the variables presenting multicollinearity have been excluded, it is possible to perform new regressions and study the results. To analyse the remaining set of variables, I again run several regressions. Again, I examined the control variables alone, the three variables Knowledge of ML, Usefulness of ML and Investor Importance individually, but controlled by the control variables, and finally all together. Thus, I built five new models presented in Table 7.

Variables	Model 1 Controls Only	Model 2 Knowdlege	Model 3 Manager Perception on Performance	Model 4 Manager Perception on Investors	Model 5 All	VIF
Knowledge of ML		**0.695****			**0.402****	1.683
Usefulness of ML			**0.916****		**0.658****	1.863
Investor Importance				-0.413	**-0.246***	1.143
Dummy_Africa	**-2.967****	**-2.550****	-1.044	**-2.512***	-1.074	1.358
Dummy_Asia	0.533	0.255	0.167	0.368	0.011	1.057
Dummy_Europe	**-1.438****	-0.564	-0.339	**-1.414****	-0.129	1.507

a. Dependent Variable: N. Application Manager

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7: Multiple Regression n.2

4.2.1. Results for Control Variables

Looking at the control variables, as seen previously, the only ones included in the model are the four dummy variables representing the four continents to which the companies' headquarters, where the managers who responded to the survey work, belong.

Looking at the results of the models, it is immediately apparent that the variable representing America has been excluded from them by the software. On the contrary, the other three dummy variables (Africa, Asia and Europe) are included and have different impacts and significance in each model used.

The variable Dummy_Africa has a negative beta coefficient in all models. However, it has significance only in models 1, 2 and 4. We can deduce from this that the condition in which the respondent manager lives and works on the African continent is a deterrent to the use of ML in the decision-making process. Its non-significance in models 3 and 5 can probably be linked to the small number of observations that present this positive variable (with value=1).

The second control variable, Dummy_Asia, on the contrary, has positive beta coefficients. In all five models, this variable is not significant. Therefore, it is impossible to determine a relationship between it and the dependent variable. In this case, the motivation is probably linked to the number of observations with a value of 1. In fact, the number of responding managers from Asia equals one. Finally, the third variable representing managers operating in Europe has a negative beta and a p-value lower than 0.01 (and therefore is significant) only in models 2 and 4. In significant surveys, the interpretation that can be given is, as in the case of Africa, that the presence of the company in Europe is a factor that determines a lower use of ML compared to other continents.

The other control variables are the frequency of use of the ML by the company in which the manager operates, the Company Category (that is, the degree of ML integration in the business processes) and the Foundation Year. As previously mentioned, these variables were excluded from the definitive regression models. In the first set of regressions, the first variable has a positive beta in all 16 cases of the first set of models and is significant in the first, third, fifth, ninth, tenth, twelfth, thirteenth, fourteenth and fifteenth. This result indicates that, as this variable increases, the utilisation of the ML also increases. Although the variable has significance in many models, I decided to exclude it because it has a high VIF (11.203) and, therefore, high multicollinearity. This information is not unexpected as the variable has a coefficient close to 90% (0.876) in the correlation matrix (Perkins, 2014). The same goes for the variable Company Category, which is significant at the 0.01 level in all models from 1 to 15 but has a correlation of 90%. This level of correlation resulted in a VIF of 12.536. Lastly, the Foundation Year presents beta close to, if not equal, 0 in all models from 1 to 16. In fact, the coefficient in the correlation matrix is very low (0.165) and not significant.

4.2.2. Results for Machine Learning and Company Size

The company's size was calculated using three variables: the AUM, the number of employees and the number of countries in which the company is present. The amount of AUM was present in models 2, 5 and 16 of the first group of regressions and had a negative beta. This sign would have run counter to previous literature. Indeed, as seen in the literature review, Grobys *et al*. (2022) found an inversely proportional relationship between company size and the use of ML. However, it must be said that they had carried out the study only on companies in the hedge fund category. The data is not significant in any of the three models and was not significant in the correlation analysis.

For this reason, I decided to exclude it from the variables used in the second and final round of regressions. The number of employees is the second variable used to identify the size of companies. Its beta was positive in all models from the first round, confirming the literature (Grobys *et al*., 2022). Although it was the only independent variable that presented significance in the first set of regressions (in models 5 and 16), not presenting significance in correlation with the dependent variable, it was excluded from the final models considered valid. Finally, as in the case of the AUM, the variable representing the number of countries in which the company operates presented a negative beta in all the models of the first set in which it was present (4, 5, 16) but, having no significance nor in them or in the correlation matrix, it was excluded from the definitive model set.

4.2.3. Results for Managers and Machine Learning Knowledge

The Knowledge of the ML variable is significant in the models in which it is present (Table 7: models 2 and 5) and has a positive beta coefficient. It is possible to deduce that as a manager's knowledge of ML increases, its use also tends to increase.

4.2.4. Results for Managers and Machine Learning Performance

The variables that measure the manager's perception concerning the impact on financial performance are the level of perceived utility for this purpose, the degree of the perceived risk/benefit ratio and the confidence that the manager has in using Unsupervised and Supervised Learning technologies. Only the first of these was included in the second set of models, as the other three presented multicollinearity that displaced the model. The Usefulness variable is present in models 3 and 5 and is positive and significant at the 0.01 level. It means that as the utility perceived by the manager of ML use (and therefore the better performance obtained from its use) increases, its use also increases.

4.2.5. Results for Managers' and Investors' Perception of Machine Learning

Finally, the independent variables that represent the relationship between the manager and the investor are the level of importance that the manager gives to the investor's opinion on ML use and his perception of the confidence that the investor relies on Unsupervised and Supervised Learning technologies. Also, in this case, the last two variables show a high degree of multicollinearity, which is why they have been excluded from the models. On the other hand, the first variable is present in models 4 and 5 (Table 7), it is significant, and its betas assume negative values. It means that as the importance given by the manager to investor opinion on ML matters decreases, its use in decision-making processes increases. It confirms what is present in the literature (Dietvorst, Simmons and Massey, 2015). In fact, the report prepared by Ipsos Mori (2017) highlights an aversion towards ML on the part of investors, and therefore, if managers listen to their concerns, they will be led to use this technology less.

4.3. Robustness

I have conducted several robustness tests to verify the results. I used General Linear Models to ensure the model does not exhibit heteroskedasticity. First, I calculated the values of the Unstandardised Residuals squared and substituted them for the managers' ML frequency of use as a dependent variable (Hayes and Cai, 2007). In Table 8, it is possible to see that the p-value is always greater than 0.05 in all models; therefore, the hypothesis of heteroskedasticity is rejected. Second, I performed the White Test (Table 9) (White, 1980), which, with a p-value again above 0.05 in all models, rejects the hypothesis of heteroskedasticity. Third, I performed the Breusch-Pagan test (Table 10) (Breusch and Pagan, 1979) and the modified Breusch-Pagan test (Table 11) (Bolakale and Oyeyemi, 2021). Once again, the heteroskedasticity hypothesis was rejected, presenting non-significant values in all models. Model 3 of the unmodified Breusch-Pagan test is the only model with a significant value, which is borderline (0.047). It must also be said that the most precise estimate is found in the table of the modified Breusch-Pagan test. Fourth, I carried out the F-test (Table 12), which also rejects the hypothesis of heteroskedasticity with a p-value higher than 0.05. Finally, I compared the results of the analysis carried out using the regression models with the Parameters Estimates values resulting from the use of Robust Standard Errors (Table 13) (Hayes and Cai, 2007). To calculate the Parameters Estimates with Robust Standard Errors, I used the HC3 method (Long and Ervin, 2000). As can be seen from the comparison of the two tables in Models 1, 2 and 4, the standard errors are almost equal to the robust standard errors, which again implies that the models are robust. Instead, in Models 3 and 5, in which the variable Usefulness is present, the robust standard errors are equal to zero since the model considers the variable redundant. This data suggests that the models containing this variable present heteroskedasticity and are not robust.

Regression with Unstandardised Residuals Squared: Significance				
Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
0.129	0.201	0.136	0.123	0.114

a. Dependent Variable: Square of Unstandardised Residuals

Table 8: Robustness Test n.1

White Test for Heteroskedasticity: Significance				
Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
0.108	0.521	0.165	0.308	0.492

a Dependent Variable: N. Application Manager

Table 9: Robustness Test n.2

Breusch-Pagan Test for Heteroskedasticity: Significance				
Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
0.236	0.559	0.047*	0.379	0.053

a Dependent Variable: N. Application Manager

Table 10: Robustness Test n.3

Modified Breusch-Pagan Test for Heteroskedasticity: Significance				
Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
0.150	0.686	0.096	0.286	0.080

a Dependent Variable: N. Application Manager

Table 11: Robustness Test n.4

F Test for Heteroskedasticity: Significance				
Model 1	Model 2	Model 3	Model 4	Model 5
0.156	0.693	0.099	0.295	0.083

a Dependent Variable: N. Application Manager

Table 12: Robustness Test n.5

Standard Errors vs. Robust Standard Errors - HC3 Method									
Model 1		Model 2		Model 3		Model 4		Model 5	
SE	RSE	SE	RSE	SE	RSE	SE	RSE	SE	RSE
0.913	0.746	0.106	0.126	0.110	0.000	0.286	0.279	0.091	0.000
1.273	0.237	0.673	0.740	0.597	0.000	0.951	0.550	0.108	0.000
0.387	0.341	0.951	0.307	0.840	0.000	1.266	0.285	0.158	0.000
0.000	0.000	0.314	0.301	0.283	0.000	0.384	0.347	0.537	0.000
		0.000	0.000	0.000	0.000	0.000	0.000	0.712	0.000
								0.243	0.000

Table 13: Robustness Test n.6

5. Conclusions, Limitations, and Further Research

This paper provides an empirical analysis of the use of ML in the managerial choices of companies operating in the asset management sector and the reasons that influence this phenomenon. The findings, controlled for the territorial variable in which the company has its headquarters, reveal a series of important implications for using ML. The previous literature analyses the relationship between the size of companies and the use of ML by hedge funds (Grobys, 2022) without expanding the study to the broader asset management category and studying the possible reasons behind this phenomenon. This study does not reveal a significant correlation between these two factors, thus finding a massive use of this technology within small boutiques and medium-sized firms and giants on the market. The studies in the literature have also analysed the impact of AI on performance (Ferreiraa, Ruivob and Reis, 2021) without considering the managers' perception of its use and how it affects it. In fact, it has been studied how more excellent ML knowledge by managers, working together with data analysts, improves the company's financial and organisational performance (Lee and Shin, 2020). Instead, this work investigates how managers' understanding of ML technologies and platforms increases their use. A significant correlation between this effect was found in the present research.

Furthermore, as seen above, some papers define ML as unsupervised and supervised learning (Mohri *et al.*, 2012; Hinton and Sejnowski, 1999) and explain its impact on financial performance. Despite this, the managers' perception of their effect has not been studied, as well as how the trust placed in these methodologies affects the more frequent application of ML. From the analysis carried out in this thesis, it emerges that the greater the manager's perception that ML positively impacts financial results, the greater they will use it. However, regarding the difference in the degree of human participation in the decision-making process (Unsupervised and Supervised Learning), the data collected are not statistically significant.

Therefore, it is impossible to perform statistical inferences on the subject. Another area covered by the literature is how these technologies are perceived by the public and by investors (Ipsos Mori, 2017). It emerges that they are, on average, opposed to its use in investment decision-making, considering it dangerous and thinking it reduces their freedom of choice (Dietvorst, Simmons and Massey, 2015). Other literature studies the relationship between the manager and the investor by analysing the relevant international regulations (European Parliament and Council, 2004 and 2014). What is missing again is a study of how the investors' concerns about the use of ML affect the decisions on its use by managers. The analysis conducted here revealed an inverse relationship between managers' perception of investor concerns and the use of ML. It means that the more important the manager thinks the investor's thinking on the ML topic is, the less he will use it. This finding is consistent with the literature (Ipsos Mori, 2017) as, as mentioned, investors are averse to its use (Dietvorst, Simmons and Massey, 2015). So, if the manager listens to their ideas, he will be led not to use ML.

Once again, the difference between the presence or absence of human intervention in the decision-making process is not statistically significant.

Using multiple regression (Mansfield, 1877), the study extends the existing literature on using new technologies in the financial field.

Therefore, combining the literature with my results, it is possible to state that ML should be more studied and used as knowledge increases its application: the greater the knowledge, the greater the application; the greater the application, the better the financial results obtained.

This thesis has a number of limitations. First of all, it presents a relatively small sample. A more extensive data collection could lead to expanding the number of significant correlations within the model. It is linked to the second limitation: a large part of the initial set of variables was excluded as not statistically significant or because they presented multicollinearity

(Greene, 2000). Increasing the number of cases would have been possible to correct these problems and raise the number of factors studied.

Furthermore, the model explains only some of the phenomena that can influence the use of ML by managers in the decision-making process. Further research could be performed from this starting point, expanding the sample size and collecting data from a higher number of companies in a higher number of countries. The analysis could therefore be extended to more areas searching for a broader and more complete set of reasons behind the phenomenon. Some possible interesting variables to analyse could be the relationship between the use of ML and the countries of origin of the managers and, above all, the age of the managers, which could profoundly influence the choice of using ML in the financial field.

Reference List

Agakov, F. *et al*. (2006). “Using machine learning to focus iterative optimisation” *International Symposium on Code Generation and Optimization* (CGO’06). pp. 11-305.

Ammann, M., and Moerth, P. (2005). “Impact of Fund Size on Hedge Fund Performance”. *Journal of Asset Management* 6 (3). pp. 219–238.

Anand, A., Coltman, T., Sharma, R. (2016). “Four steps to realising business value from digital data streams”. *MIS Quarterly Executive*. 15. pp. 259–277.

Beech, N., (2000). “Narrative Styles of Managers and Workers: A Tale of Star-Crossed Lovers”. *The Journal of Applied Behavioral Science*. 36. pp. 210–228.

Berry, W. D., Feldman, S. (1985). “Multiple Regression in Practice”. Sage 50.

Blum, A. (2007). “Machine learning theory”. *Carnegie Melon Universit, School of Computer Science.*

Bolakale, A. H., Oyeyemi, G. M. (2021) “A Modified Breusch–Pagan Test for Detecting Heteroscedasticity in the Presence of Outliers”. *Pure and Applied Mathematics Journal*. 10 (6). pp. 139-149.

Bradford, M., Florin, J. (2003). “Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning systems”. *International Journal of Accounting Information Systems*. 4. pp. 205–225.

Breusch,T. S., Pagan, A. R. (1979). “A Simple Test for Heteroskedasticity and Random Coefficient Variation”. Econometrica. 47 (5). pp. 1287–1294.

Bryman A., Cramer D. (2005). “Quantitative data analysis with SPSS 12 and 13: a guide for social scientists”. *New York, NY: Psychology Press.*

Castagno, S., Khalifa, M. (2020). “Perceptions of Artificial Intelligence Among Healthcare Staff: A Qualitative Survey Study”. *Frontiers in Artificial Intelligence*.

Chen, J., Hong, H., Huang, M., Kubik, J. D. (2004). “Does Fund Size Erode Mutual Fund Performance? the Role of Liquidity and Organization.” T*he American Economic Review* 94 (5). pp. 1276–1302.

Clark, R.(1981). ‘The Four Stages of Capitalism’. *Harvard Law Review* 561.

Deloitte (2020). “Asset Management Survey: The Gold Rush for Route for Asset Managers”. *Deloitte*. pp. 5-9.

DeLone, W. H., McLean, E. R., (2003). “Model of information systems success: A ten-year updated”. *Journal of Management Information Systems*. 19. pp. 9–30.

Dhall, D., Kaur, R., Juneja, M. (2020). “Machine Learning: A Review of the Algorithms and Its Applications”. *Proceedings of ICRIC 2019. Lecture Notes in Electrical Engineering.* 597.

Dietvorst, B. J., Simmons, J. P., Massey, C. (2015). “Algorithm Aversion: People Errone-Ously Avoid Algorithms After Seeing Them Err.” *Journal of Experimental Psychology. General* 144 (1). pp. 114–126.

Du, J., Zhou, L. (2012). “Improving financial data quality using ontologies”. *Decision Support Systems*. 54. pp. 76–86.

European Parliament and Council, (2004). *European Union Directive 2004/39/EU.*

European Parliament and Council, (2014). *European Union Directive 2014/65/EU*.

Ferreiraa, H., Ruivob, P., Reis, C. (2021). “How do data scientists and managers influence machine learning value creation?”. *Procedia Computer Science.* 181. pp. 757–764.

Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2017). “An Introduction to Statistical Learning (8th ed.)”. *Springer Science+Business Media New York.*

Garg, A.K., Joubert, R.J.O., Pellissier, R. (2004). “Measuring business performance: A case study”. *Southern African Business Review*. 8 (1). pp. 7-21.

Greene, W. (2000). “Econometric Analysis: 4th ed”. *Upper Saddle River, NJ: Practice Hall.*

Grobys, K., Kolari, J. W., Niang, J. (2022). “Man versus machine: on artificial intelligence and hedge funds performance”. *Applied Economics*. pp. 3-5.

Harrell, F. E. (2001). “Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis”. *Springer New York, NY.*

Hayes, A., Cai, L. (2007). “Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation”. *Behaviour Research Methods*. 39 (4). pp. 709-722.

Helm, J.M., Swiergosz, A.M., Haeberle, H.S. (2020). “Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions”. *Curr Rev Musculoskelet Med.* 13. pp. 69–76.

Hinton, G., Sejnowski, T. (1999). “Unsupervised Learning: Foundations of Neural Computation”. *The MIT.*

IAM, (2022), “What is Asset Management?”, *The Institute of Asset Management,* [Online], Available at: https://theiam.org/what-is-asset-management.

Ipsos Mori (2017). “Public views of ML”. *The Royal Society*. pp. 77-82.

Ives, A. R. (2015). “For testing the significance of regression coefficients, go ahead and log-transform count data”. *Methods in Ecology and Evolution.* 6 (7). pp. 828-835.

Jarrahi, M. H. (2018). “Artificial intelligence and the future of work: Human-AI symbiosis in organisational decision making”. *Business Horizons*. 61. pp. 577–586.

Kietzmann, J., Pitt, L. F., (2020) “Artificial intelligence and machine learning: What managers need to know”. *Business Horizons*. 63. pp. 131–133.

Kuzey, C., Uyar, A., Delen, D. (2014). “The impact of multinationality on firm value: A comparative analysis of machine learning techniques”. *Decision Support Systems*. 59. pp. 127–142.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., Kruschwitz,N. (2011). “LaValle, Steve, *et al*. 2011 Big Data, analytics and the path from insights to value”. *MIT Sloan Management Review.* 52. pp. 21–31.

Lee, I., Shin, Y. J., (2020). “Machine learning for enterprises: Applications, algorithm selection, and challenges”. *Business Horizons*. 63. pp. 157–170.

Lee, Y. W., (2003). “Crafting Rules: Context-Reflective Data Quality Problem Solving”. *Journal of Management Information Systems*. 20. pp. 93–119.

Legendre, A. M. (1805), “Nouvelles méthodes pour la détermination des orbites des comètes”.

Lemke, L. *et al*. (2014). “Hedge Funds and Other Private Funds: Regulation and Compliance”. *Thomson West.*

Lismont, J., Vanthienen, J., Baesens, B., Lemahieu, W. (2017). “Defining analytics maturity indicators: A survey approach”. *International Journal of Information Management*. 37. pp. 114–124.

Liu, C., Zowghi, D., Talaei-Khoei, A. (2020). “An empirical study of the antecedents of data completeness in electronic medical records”. *International Journal of Information Management*. 50. pp. 155–170.

Long, J. S., Ervin, L. H. (2000). “Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model”. *The American Statistician*. 54 (3). pp. 217–224.

Lopez de Prado, M. (2020). “Machine Learning per Asset Manager (Elements in Quantitative Finance)”. *Cambridge: Cambridge University Press*.

McAfee, A., Brynjolfsson, E. (2017). “Big Data: The Management Revolution”. *Harvard Business Review*.

Merriman, M. (1877). “A List of Writings Relating to the Method of Least Squares”. *MacMillan and Co*.

Mohri, M., Rostamizadeh, A., Talwalkar, A. (2012) “Foundations of Machine Learning”, *The MIT.*

Moloney, N. (2012). “Liability of asset managers: a comment”. *Capital Markets Law Journal*, 7 (4). pp. 414–422.

Pearson, K. (1895). “Notes on regression and inheritance in the case of two parents”. *Proceedings of the Royal Society of London*. 58. pp. 240–242.

Perkins, S. E. (2014). “When does prior experience pay? Institutional experience and the multinational corporation”. *Administrative Science Quarterly*, 59 (1). pp. 145-181

Ragin, C., Amoroso, L. (2011). “Constructing Social Research: The Unity and Diversity of Method”. *London: Sage.*

Rai, A., Brown, P., Tang, X. (2009) “Organisational assimilation of electronic procurement innovations”, *Journal of Management Information Systems*. 26. pp. 257–296.

Ramamurthy, K., Sen, A., Sinha, A. P. (2008). “An empirical investigation of the key determinants of data warehouse adoption”. *Decision Support Systems*. 44. pp. 817–841.

Ruane, J., (2019) “What Machine Learning Will Mean for Asset Managers”, *Harvard Business Review*. pp. 2–7.

Ruivo, P., Oliveira, T., Johansson, B., Neto, M. (2013). “Differential effects on ERP post-adoption stages across Scandinavian and Iberian SMEs”. *Journal of Global Information Management*. 21. pp. 1–20.

Sandelowski, M. (1995). “Sample size in qualitative research”. *Res. Nurs. Health*. 18. pp. 179-183.

Sarker, S., Sarker, S., Sidorova, A. (2006) “Understanding business process change failure: An actor-network perspective”. *Journal of Management Information Systems*. 23. pp. 51–86.

Shrestha, N. (2020). “Detecting Multicollinearity in Regression Analysis”. *American Journal of Applied Mathematics and Statistics*. 8 (2). pp. 39-42.

Siemsen, E., Roth A., Oliveira P. (2010). “Common Method Bias in Regression Models With Linear, Quadratic, and Interaction Effects”. *Organizational Research Methods*. 13 (3). pp. 456-476.

Snow, D. (2020) *The Journal of Financial Data Science Winter.* 2 (1), pp. 10-23.

Stigler, S. M. (1981). “Gauss and the Invention of Least Squares”. *Ann. Stat*. 9 (3).

Stuart J. Russell, Peter Norvig (2010) “Artificial Intelligence: A Modern Approach, Third Edition”, *Prentice Hall.*

White, H. (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”. *Econometrica*. 48 (4). pp. 817–838.

Appendices

Appendix n.1: Information Sheet For Participants

INFORMATION SHEET FOR PARTICIPANTS

Ethical Clearance Reference Number: MRSU-21/22-32408

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of study

Machine Learning in Financial Asset Management: a challenge between perception and reality.

Invitation Paragraph

I would like to invite you to participate in this research project which forms part of my Postgraduate research. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

What is the purpose of the study?

The purpose of the study is to understand if there is a correlation between the use of Machine Learning and the size of companies and what are the drivers that lead, or not, a manager of an Asset Management company to use Machine Learning in his own activities.

Why have I been invited to take part?

You have been invited to participate in this study because you are a manager or a member of an Asset Management company who operates directly in investment choices and in the choice of tools for investment choices.

What will happen if I take part?

If you agree to take part you will complete a survey anonymously. The survey will ask you questions about use, knowledge, perception of Machine Learning and the size of the company in which you work. The survey will take approximately 5 minutes to complete and each participant is required to complete only one survey.

Do I have to take part?

Participation is completely voluntary. You should only take part if you want to and choosing not to take part will not disadvantage you in anyway. If you choose to take part you will be asked to provide your consent. To do this you will be asked to indicate that you have read and understand the information provided and that you consent to your anonymous data being used for the purposes explained.

You are free to withdraw at any point during completion of the survey, without having to give a reason. Withdrawing from the study will not affect you in any way. Once you submit the survey, it will no longer be possible to withdraw from the study because the data will be fully anonymous. Please do not include any personal identifiable information in your responses.

Data handling and confidentiality

This research is anonymous. This means that nobody, including the researchers, will be aware of your identity, and that nobody will be able to connect you to the answers you provide, even indirectly. Your answers will nevertheless be treated confidentially and the information you provide will not allow you to be identified in any research outputs/publications. Your data will be held securely, completely anonymised, stored in the KCL SharePoint until 30/11/2022 (date on which the research project will be concluded and evaluated).

What will happen to the results of the study?

The results of the study will be summarised in a Postgraduate thesis, collectively processed in order to make statistical inferences and the anonymous dataset will not be shared with third parties or made publicly available.

Who should I contact for further information?

If you have any questions or require more information about this study, please contact me using the following contact details:

Simone Maria Caranci
Email address: simone.caranci@kcl.ac.uk

What if I have further questions, or if something goes wrong?

If this study has harmed you in any way or if you wish to make a complaint about the conduct of the study you can contact King's College London using the details below for further advice and information:

Prof. Keith Brouthers
Email address: keith.brouthers@kcl.ac.uk

Thank you for reading this information sheet and for considering taking part in this research.

Appendix n.2: Survey Questions

AREAS	QUESTIONS	ANSWERS	SOURCES
USAGE OF ML	How many applications of Machine Learning have you come across in your work?	5-point Likert scales with description at both ends (None - More than three)	Adapted from Ipsos Moris (2017) and Grobys, Kolari and Niang (2022)
	How often does your firm use Machine Learning?	5-point Likert scales with description at both ends (Never - On a daily basis)	
	(Intro) According to a recent study (Grobys, 2022), investment funds can be divided into 4 categories depending on the use of the ML. Which of these 4 categories best represents your company? (please click just one)	Traditional: they do not use Machine Learning Discretionary: they are based on both Machine Learning and managerial choices with a greater focus on managerial choices Combined: they are based on both Machine Learning and managerial choices in equal measure Systematic: they are based on both Machine Learning and managerial choices with a greater focus on ML.	
KNOWLEDGE OF ML	Have you ever heard of "Machine Learning"?	Yes No	Adapted from Ipsos Moris (2017) and Castagno and Khalifa (2020)
	How much, if anything, would you say you know about "Machine Learning"? A great deal = You have read about it extensively and feel comfortable talking about it with experts and you have used it before. A fair amount = You read and got informed about it and you feel comfortable in talking about it with non-experts Heard of, know nothing about = You have heard of it but you have no idea what it is Not at all = You don't know what it is and you've never even heard of it	5-point Likert scales with description at both ends (Not at all - A great deal)	
COMPANY DIMENSION	Please indicate the size of your company according to the total value of the Assets Under Management (AUM)	Bank for actual number	Adapted from Deloitte (2020)
	Please indicate the size of your company according to the number of employees	Blank for actual number	
	Please indicate the size of your company according to the number of countries in which it operates	Blank for actual number	
	In what year was your firm started?	Blank for actual number	
	In which country is your firm headquartered?	Blank for the Country	
PERCEPTION OF ML PERFORMANCE	How useful do you think Machine Learning could be in your area of work?	5-point Likert scales with description at both ends (Of no use at all - Extremely useful)	Adapted from Ipsos Moris (2017) and Castagno and Khalifa (2020)
	(Intro) Some suggest that Machine Learning can improve investment performance by allowing computers to add to what people can already do. Others say there are risks, because a computer's learning process isn't always perfect, which can present possible dangers if a computer makes a decision rather than a human being. Which of the following is closest to your view on the balance between risks and benefits?	5-point Likert scales with description at both ends (The risks are much bigger than the benefits - The benefits are much bigger than the risks)	
	(Intro) Imagine working in a company that uses Machine Learning and uses a computer that, using mathematical models, can adapt to the financial market to invest the money you are managing where it thinks it will provide the best returns: **After choosing where to invest your money, it automatically invests it without a human being authorizing it to do so.** To what extent, if at all, would you feel comfortable or uncomfortable with this process?	5-point Likert scales with description at both ends (Not at all comfortable - Very comfortable)	
	(Intro) After choosing where to invest your money, the computer asks you to decide whether or not to authorize the investment. To what extent, if at all, would you feel comfortable or uncomfortable with this process?	5-point Likert scales with description at both ends (Not at all comfortable - Very comfortable)	

INVESTORS PERCEPTION OF ML	In your choice of whether or not to use Machine Learning while investing, how important is the investor's opinion of the reliability of Machine Learning to you? **(Intro) Imagine working in a company that uses Machine Learning and uses a computer that, using mathematical models, can adapt to the financial market to invest the money you are managing where it thinks it will provide the best returns:** **After choosing where to invest your money, it automatically invests it without a human being authorizing it to do so.** To what extent, if at all, would investors investing in your company feel comfortable or uncomfortable with this process? **(Intro) After choosing where to invest your money, the computer asks you to decide whether or not to authorize the investment.** To what extent, if at all, would investors investing in your company feel comfortable or uncomfortable with this process?	5-point Likert scales with description at both ends (Not important at all - Extremely important) 5-point Likert scales with description at both ends (Not at all comfortable - Very comfortable) 5-point Likert scales with description at both ends (Not at all comfortable - Very comfortable)	Adapted from Ipsos Moris (2017) and Castagno and Khalifa (2020)
WHICH ML METHOD IS USED	You are going to read a list of several technologies that use Machine Learning. For each, could you tell me if you use one or more of these technologies by checking them?	(see checklist below)	Adapted from Ipsos Moris (2017)

	X
Artificial Neural Networks	
Decision Trees	
Support-Vector Machines	
Regression Analysis	
Bayesian Networks	
Genetic Algorithms	
Swarm Intelligence	
Federated learning	
Others: _________	
None of them	

Summary

Introduction

The use of Machine Learning (ML) within the world of Asset Management has become increasingly popular (Dhall *et al*., 2020): the main reason for this is that the best investment strategies are based on economic and financial theories. ML is a powerful tool to implement them and discover new ones (Lopez de Prado, 2020). The main activities of which Asset Management is composed and within which in recent years there has been increasing use of ML are portfolio construction, risk management, capital management, infrastructure and distribution, and sales and marketing (Snow, 2020). Both the use of these techniques and their studies have significantly increased in recent years. Recent studies in the literature tried to analyse the phenomenon of ML use within the Asset Management sector, and two interesting points of reflection have emerged from them: First, there has been a growing concern in the literature about the use of ML in many aspects of life, including investors' perception of its use in finance (Ipsos Moris, 2017). Other recent studies conducted on hedge funds have also shown a positive correlation between the use of ML and Performance, but it is little used, mainly in small businesses (Grobys, 2022). However, so far, the literature does not present an answer about the reasons behind these phenomena and an overview of the state of the art of ML use and its perception at a more enlarged level. Therefore, this study aims at providing some new insights into this field. First, I will investigate further if there is indeed a negative correlation between company size and the use of ML (Grobys *et al*., 2022); Second, I will extend prior research to find out what are the reasons behind the use or not of ML within companies operating in this sector, analysing its use in relation to three other variables, which are: Knowledge of the ML, Perception of the impact of the ML on the Performance and perception of managers about how much the ML appeals to investors (Ipsos Moris, 2017); Third, I will contribute to a more holistic understanding of the phenomenon analysing not only Hedge Funds but to broaden the research to the larger category of Asset Management companies (Grobys, 2022). The rest of the work is organised as follows: After presenting the analytical framework, I will verify the four hypotheses related to ML use in the Asset Management sector. The model tests the interdependencies between the use of this technology and the four variables mentioned above. This work will conclude with the analyses of the obtained results, a conclusion on the overall topic, and a suggestion for further research.

Literature Review

This literature review aims to give an overview of the topics analysed within the research work, examining the existing literature on ML use within the Asset Management sector to lay the foundations for the completion of the purpose of this study.

Machine Learning and Company Size

The first part of the Literature Review aims to analyse the information on the relationship between the use of ML by companies operating in the Asset Management sector and their size. Studies show a predominant use of ML in small businesses operating in the Asset Management sector compared to larger ones (Grobys *et al*., 2022). Two possible explanations for this "size factor" are investor aversion to new technologies that replace human judgment (Dietvorst, Simmons & Massey, 2015) and that the use of ML leads to the creation of unique strategies that are difficult to scale (Chen *et al*., 2004). Grobys' study was conducted solely on hedge funds and concluded that small ones use ML more than large ones and perform better. But the reasons may be other: the greater risk appetite of small funds and their being more liquid (Ammann and Moerth, 2005). Therefore, taking into account the above, the first contribution that this work wants to provide is to understand the actual relationships between the use of ML and the size of companies, shifting the focus from hedge funds only, to the broader category of Asset Management.

Managers and Machine Learning Knowledge

Previous studies have shown that the relationship between data scientists and managers and knowledge of ML in managerial decisions can positively influence company results (Ferreiraa *et al*., 2021). However, what is not present is a response to the presence, or not, of a correlation between the knowledge of ML and its greater use to understand if it is one of the reasons for its adoption. Starting from this question, this analysis will be the second contribution this work aims to give to research in this field.

Managers and Machine Learning Performance

In their article on hedge funds, Grobys *et al*. (2022) divide funds into four categories depending on how much ML is integrated into the business processes: Traditional, Discretionary, Combined and Systematic. They conclude that funds that use ML in investment decision-making (systematic funds) outperform those that don't. However, they also point out that funds that use an average level of ML and human decision together (Combined Funds) perform worse than those driven only by the instinct of managers (Traditional Funds) and those driven by the calculation of ML algorithms (which, as mentioned, show the best results). Ferreiraa *et al*. (2021) show how using platforms that exploit ML can improve the financial and organisational performance of the company. They also point out that many factors influence the outcome: the maturity of the platform, the feasibility of the project, the quality of the data, the intensity of the information and the compatibility between existing systems within the organisation. When studying the impact of performance, what is not examined is the perception that managers have on the practical implications for performance in companies operating in the asset management sector and how this affects the choices of these platforms and technologies usage. The research aims to study how managers perceive the benefits of using ML on performance concerning the risks that these methodologies may entail. I also try to understand if the degree of participation of the manager in the decision-making process in investment choices (Supervised/Unsupervised Learning) plays a role in the frequency of use of ML in business activities.

Managers' and Investors' Perception of Machine Learning

The last part of the literature review will analyse the existing literature on investor perceptions of using ML in investing. It will also be analysed how and if this affects the choices of use of this technology by the managers operating in the Asset Management sector. In 2017, the British market research company Ipsos Mori conducted a study on public opinion on using ML in various fields of application, including finance, on behalf of the Royal Society (Ipsos Mori, 2017). It emerged that most of the participants were against the use of a machine that checks their portfolio daily: out of a sample of 978 adults, when asked if they were in favour of computers adaptation to the financial market and investment of money in their behalf, only 18% answered affirmatively, 41% replied negatively, and 30% replied to remain indifferent. The remaining 11% did not know how to answer (Ipsos Mori, 2017). From a regulatory point of view, the relationship between investors and asset managers is regulated by MiFID I and II to avoid liability problems for the Asset Managers: the latter are obliged by law to pursue the interests of investors without exposing them to large significant risks and listening to their needs. To the best knowledge of the author, there is no study in the literature on the relationship between investor concerns about the use of ML in investment choices and the use of this technology.

Analytical Framework

Taking into account all the above, the following hypotheses can be formulated:

- ***Hypothesis 1****: Small companies are more likely to use ML than larger companies.*
- ***Hypothesis 2****: Managers who have more knowledge of ML are more likely to use it than managers who have less knowledge of it.*
- ***Hypothesis 3****: Managers who believe that using ML leads to better performance are likelier to use it than Managers who think it leads to lower performance.*
- ***Hypothesis 4****: Managers who believe investors like ML are more likely to use it than Managers who believe investors don't like it*.

The relationships between the use of ML and the four variables just illustrated will be studied through a regression model on the data collected through a survey sent to companies operating in the asset management sector. This model will be better explained in the next chapter of the document.

Research Methodology

Sample

The study and analysis were conducted through the administration of an anonymous survey. Respondents were allowed to respond only once. This survey's target sample is managers from companies operating in the Asset Management sector. The survey was answered by 53 managers out of a sample of 150: 35 managers came from companies based in Europe, 15 in America, 2 in Africa and one in Asia; One company was founded in the eighteenth century, 8 in the nineteenth century, 24 in the twentieth century and 20 after 2000; 32 companies have an AUM of less than 100 billion, 11 companies between 100 and 500 billion and 10 companies above 500 billion; 9 companies are traditional, 13 discretionary, 15 combined and 16 systematic.

Measures

The model used in this analysis is a multiple linear regression that uses the least squares method (OLS) to study the relationship between the dependent variable "Use of the ML" and the independent variables. The dependent variable is the use that managers make of ML within the decision-making process in companies operating in the Asset Management sector. The study of interactions with independent variables allows us to understand the reasons for its use. To gather information on this variable, in the first part of the survey, a question was asked about the frequency of use by the respondent manager. To measure the frequency of use in the survey, a 5-point Likert scale was adopted with a value of 1 "Never" and 5 "On a daily basis". In the previous chapters, four areas have been identified that can be considered as the reason and, therefore, the object of the study for the use of ML: (1) the Company Size, (2) the Knowledge, (3) the Manager's Perception of Performance and (4) the Manager's Perception of investor concerns. To study these phenomena, within the survey, more questions were asked by the research area to find the most significant ones, refine the model and go deeper into the question (Siemsen, 2010). This study will use the answers to these questions as independent variables. Various control variables have been implemented to control for other influencing factors. To this end, a series of questions were asked in the questionnaire: Use of the ML by the company where the manager works (5-point Likert scale); Company category (dummy variable that identifies the four categories framed by Grobys); Foundation Year; Headquarters by continent.

Results and Analysis

Descriptive Statistics and Correlation

After exposing the Descriptive Statistics of the variables (the number of cases, the minimum, the maximum, the mean and the standard deviation), I performed a correlation matrix that shows the interaction between the single variables. Looking at the first column, which shows the correlation between the dependent, independent and control variables, we have that:

Positive and significant variables: Knowledge, Usefulness of ML, Risks/Benefits, Unsupervised ML - Manager, Supervised ML - Manager, Unsupervised ML - Investor, Supervised ML – Investor, N. Application Firm, Company Category, America.

Positive and non-significant variables: N. Employees, N. Countries, Foundation Year, Asia.

Negative and significant variables: Investor Importance, Europe.

Negative and non-significant variables: AUM, Africa.

<u>Test of Hypotheses</u>

I run sixteen different multiple regression models using a different number of independent variables simultaneously. Specifically, the models are constructed by analysing all the independent variables individually, in the groups relating to the study areas connected to the hypotheses and, finally, all together. All models are controlled for the factors previously presented. The dependent variable is how often managers use ML. None of the regression coefficients of the independent variables is significant, except for the number of employees in Models 5 and 16. The reason for this can be given by the multicollinearity of the variables present in the model (Greene, 2000) and by the large number of variables inserted. Therefore, I excluded from the model the variables that were not significant in the correlation study and then conducted a study on the multicollinearity of the remaining variables by studying the Variance Inflation Factor (VIF). I thus eliminated the resulting multicollinear variables and created five other regression models. The remaining variables are: Knowledge, Usefulness, Investor Importance, Asia, Africa and Europe.

Results for Control Variables

Analysing the remaining control variables in the five final models, Africa always has a negative beta coefficient and is only significant in models 1, 2 and 4. From this, it can be deduced that the condition in which the interviewed manager lives and works on the African continent is a deterrent to ML's use in decision-making. Its non-significance in models 3 and 5 can probably be linked to the small number of observations presenting this variable positive; Asia, on the other hand, has positive and non-significant beta coefficients in all five models. Therefore, it is impossible to determine a relationship between it and the dependent variable. Finally, Europe always has a negative beta and is significant only in models 2 and 4. In significant surveys, the interpretation that can be given is, as in the case of Africa, that the presence of the company in Europe is a factor which determines a lower use of ML compared to other continents.

Results for Machine Learning and Company Size

The three variables that measure the size of the company (AUM, number of employees and number of countries in which the company is present) are all non-significant and have therefore been excluded from the analysis.

Results for Managers and Machine Learning Knowledge

The Knowledge of the ML variable is significant in the models in which it is present (models 2 and 5) and has a positive beta coefficient. It is possible to deduce that as a manager's knowledge of ML increases, its use also tends to increase.

Results for Managers and Machine Learning Performance

Among the variables that measure the manager's perception of the impact on financial performance, only the level of perceived utility was included in the second set of models since the other three presented multicollinearity that displaced the model. The Usefulness variable is present in models 3 and 5 and is positive and significant at the 0.01 level. It means that as the utility perceived by the manager of the use of the ML increases (and therefore the better performance obtained from its use), its use also increases.

Results for Managers' and Investors' Perception of Machine Learning

Finally, the only independent variable representing the relationship between the manager and the investor left in the five final models is the level of importance that the manager attaches to the investor's opinion on using the ML. The variable in models 4 and 5 is significant, and its betas assume negative values. It means that as the manager attaches importance to investor opinion on ML, its use in decision-making processes increases.

Robustness

I have conducted several robustness tests to verify the results. First, I calculated the values of the Unstandardised Residuals squared and substituted them for the managers' ML frequency of use as a dependent variable (Hayes and Cai, 2007). Second, I used general linear models to ensure that the model does not show heteroskedasticity, which are: the White Test (White, 1980), the Breusch-Pagan test (Breusch and Pagan, 1979), and the modified Breusch-Pagan test (Bolakale and Oyeyemi, 2021). Finally, I compared the results of the analysis using the regression models with the Parameters Estimates values resulting from the use of Robust Standard Errors (Hayes and Cai, 2007). To calculate the estimates of the parameters with robust standard errors, I used the HC3 method (Long and Ervin, 2000). From the overall analysis of all five tests, it can be concluded that the final five models are robust.

Conclusions, Limitations, and Further Research

This paper provides an empirical analysis of the use of ML in the managerial choices of companies operating in the asset management sector and the reasons that influence this phenomenon. The findings, controlled for the territorial variable in which the company has its headquarters, reveal important implications for using ML. The previous literature analyses the relationship between the size of companies and the use of ML by hedge funds (Grobys, 2022) without expanding the study to the broader asset management category and studying the possible reasons behind this phenomenon. This study does not reveal a significant correlation between these two factors, thus finding a massive use of this technology within small boutiques and medium-sized firms and giants on the market. The studies in the literature have also analysed the impact of AI on performance (Ferreiraa, Ruivob and Reis, 2021) without considering the managers' perception of its use and how it affects it. It has been studied how more excellent ML knowledge by managers and working with data analysts improves the company's financial and organisational performance (Lee and Shin, 2020). Instead, this work investigates how managers' understanding of ML technologies and platforms increases their use. A significant correlation between this effect was found. The managers' perception of their effect has not been studied, as well as how the trust placed in these methodologies affects the more frequent application of ML. From the analysis carried out in this thesis, the greater the manager's perception that ML positively impacts financial results, the greater they will use it. However, regarding the difference in the degree of human participation in the decision-making process, the data collected are not statistically significant. Another area covered by the literature is how these technologies are perceived by the public and by investors (Ipsos Mori, 2017). It emerges that they are, on average, opposed to its use in investment decision-making, considering it dangerous and thinking it reduces their freedom of choice (Dietvorst, Simmons and Massey, 2015). Other literature studies the relationship between the manager and the investor by analysing the relevant international regulations (European Parliament and Council, 2004 and 2014). What is missing again is a study of how the investors' concerns about the use of ML affect the decisions on its use by managers. The analysis revealed an inverse relationship between managers' perception of investor concerns and the ML use. It means that the more important the manager thinks the investor's thinking on the ML topic is, the less he will use it. This finding is consistent with the literature (Ipsos Mori, 2017) as, as mentioned, investors are averse to its use (Dietvorst, Simmons and Massey, 2015). So, if the manager listens to their ideas, he will be led not to use ML. Once again, the difference between the presence or absence of human intervention in the decision-making process is not statistically significant. Using multiple regression (Mansfield, 1877), the study extends the existing literature on using new technologies in the financial field. Therefore, combining the literature with my results, it is possible to state that ML should be more studied and used as knowledge increases its application: the greater the knowledge, the greater the application; the greater the application, the better the financial results obtained.

This thesis has several limitations. First of all, it presents a relatively small sample. A more extensive data collection could lead to expanding the number of significant correlations within the model. It is linked to the second limitation: a large part of the initial set of variables was excluded as not statistically significant or because they presented multicollinearity (Greene, 2000). Increasing the number of cases would have been possible to correct these problems and raise the number of factors studied. Furthermore, the model explains only some of the phenomena that can influence the use of ML by managers in the decision-making process.

Further research could be performed from this starting point, expanding the sample size and collecting data from a higher number of companies in a higher number of countries. The analysis could therefore be extended to more areas searching for a broader and more complete set of reasons behind the phenomenon. Some possible interesting variables to analyse could be the relationship between the use of ML and the countries of origin of the managers and, above all, the age of the managers, which could profoundly influence the choice of using ML in the financial field.