**Haute Ecole**
**Groupe ICHEC - ISC St-Louis - ISFSC**

Enseignement supérieur de type long de niveau universitaire

# The impact of artificial intelligence on companies' financial forecasting process

Mémoire présenté par
**Michael Stukkens**

Pour l'obtention du diplôme de
**Master en Gestion de l'Entreprise- MIBM-120**

Academic year 2018-2019

Promoter:
Mister Benoit Stevens

Boulevard Brand Whitlock 2 - 1150 Brussels

**Haute Ecole**

**Groupe ICHEC - ISC St-Louis - ISFSC**

Enseignement supérieur de type long de niveau universitaire

# The impact of artificial intelligence on companies' financial forecasting process


Mémoire présenté par
**Michael Stukkens**


Pour l'obtention du diplôme de
**Master en Gestion de l'Entreprise- MIBM-120**

Academic year 2018-2019

Promoter:
Mister Benoit Stevens

Boulevard Brand Whitlock 2 - 1150 Brussels

## Acknowledgements

I would like to express my gratitude to my promoter Mister Benoit Stevens for his guidance and help all along the writing of the thesis. Mister Benoit Stevens was always available and willing to read my thesis progression in order to give me feed-back, further advice and guidance. Sir Benoit Stevens helped me to ask myself the right questions and write a coherent thesis. I also thank him for putting me in contact with the pharmaceutical company UCB to see how predictive analytics works in practice.

Besides, I want to express my appreciation to both my internship supervisors, Gregory Schmidt and Frederic Manseau. They gave me the opportunity to experience two amazing internships into the field of financial reporting and ERP implementation. Frederic Manseau also shared his knowledge about his financial forecasting experience and let me participate to the financial forecasting process. It gave me a more practical view on how financial forecasting is executed into a services company. Thanks to this internship experience, I could formulate my research question and consequently spend time on this exciting topic.

Moreover, I would like to thank my internship-thesis coordinator Madam Marie-Lise Heinen for her support during the whole internship-thesis process. Since the beginning of this challenging task, she helped me to achieve my goals by giving me relevant feed-backs and continuous advices after many presentations during the year.

In addition, I want to express my gratitude to all the experts I was able to talk to informally or/and formally about my topic. I'm thankful to Guillaume Cote, Nicolas Coallier, Morad Masnaoui, Jas Brar, Antoine Cosse, Arnaud Lieutenant and Malek Djili for all the information they were able to provide me with.

Last but not least, I would like to thank my family and friends for their continuous support during the whole year. They gave me strength and determination which helped me to take up this challenge.

TABLE OF CONTENTS

TABLE OF FIGURES

# Introduction

The technological revolution has already disrupted many times our markets, way of working, way of thinking and proved to be a key asset for companies' success. When speaking about information technology, we directly think about the ever-growing emergence of data. Companies are pushed to incorporate a data driven strategy because inside the available data relies very useful information even if it isn't always easy to find out. Last year, multiple technologies have seen the day to help companies to take advantage of the available data. Nevertheless, there are still a lot of people that do not realize the importance and the real added value that relies into internal and external data.

When the author did his internship within the financial department at Keyrus, he had the opportunity to take a closer look at their financial reporting process. Making these reports and updating them are time consuming. When making the financial reporting, a section is dedicated to the company financial forecast which is certainly the most impactful and important section from a strategic point of view. Forecasting is about finding patterns within historical information in order to apply it to new data and forecast future results. This task isn't easy to do at all. At that moment, the author began to look at IT tools that could find patterns within the data and which could be used to forecast. When he realized the potential of artificial intelligence (AI), he began to question himself about the use of artificial intelligence for financial forecasting purposes. The author strongly believes that all future and current businessmen, businesswomen should gain knowledge about the use and impact of new technologies as it often enables companies to solve very important issues they are facing. It doesn't mean that business people should be able to code themselves but they should be able to understand the potential of technological tools, identify the problems it could solve, and understand what is happening without necessarily knowing all the technical aspects of the technology. Why wouldn't companies implement technology that enables them to face problems? What are the capabilities of artificial intelligence? Are there other technologies linked to AI? Is it beneficial for all companies? Why would companies use it? What is the impact of it on the financial forecasting approach? What are the solutions already available on the market? What is the journey to go through in order to be able to implement a machine learning model to forecast financial figures? These are the questions that the author asked himself and that he wants to answer to through this thesis.

After making some research and asking himself hundreds of questions, the author decided to challenge himself to answer the following question:

*"Why would companies use artificial intelligence into their financial forecasting process, what solutions are currently available on the market and how do companies implement a predictive model in order to forecast financial figures?"*

The objective of the author is to demonstrate the importance of leveraging data to predict future outcomes by finding patterns within the data. Companies operate into an uncertain environment and could use data to back-up their decision takings. Current forecasting techniques aren't always able to plainly take advantage of the available data which isn't the case when using machine learning and data mining methods to find patterns within the data. Multiple artificial intelligence, machine learning solutions will be presented in order to give to the readers an overview of the current market solutions. Last but not least, the different steps of a data science project using artificial intelligence will be demonstrated in order to give a practical view of how companies need to implement artificial intelligence into their financial forecasting process.

This thesis question will be answered within 4 main chapters. The first chapter is related to the importance of financial forecasting. The current most common forecasting techniques will be pointed out, the added value of a reliable forecast will be identified and the shortcomings of current financial forecasting processes will be explained. Finding patterns in huge amounts of historical data isn't easy and this is where artificial intelligence can become very interesting.

The second chapter of this thesis will explain to the readers how artificial intelligence and data mining can help to break out the limits of current forecasting methods. The foundations of artificial intelligence will be explained, the advantages and drawbacks of predictive analytics will be pointed out in order to provide a clear understanding of its added value to businesses. Afterwards, the author will point out the on-going challenges of using artificial intelligence and will compare the artificial intelligence concept with the business intelligence concept. These two terms are often mixed up but are very different. The multiple AI use cases within companies will be identified to show the different area where AI can have a positive impact. The type of machine learning algorithms required to forecast will be identified and its impact on the financial forecasting approach of companies will also be explained. Furthermore, the author is going to make an analysis of the type of business models required with the conditions to be fulfilled in order to assess either a company is in a favorable position or not to take advantage of predictive analytics to forecast. Last but not least, at the end of this chapter, the different machine learning solutions on the market for forecasting purposes will be exposed to make readers aware that companies could already take advantage of such solution nowadays.

The third chapter will focus on the journey companies need to go through to take advantage of the solutions available on the market. The different steps to implement a machine learning model will be demonstrated in order to give a more pragmatic view on how to implement such a technology to forecast financial figures. The advantages and stakes of implementing machine learning will be identified in order to balance the pros and cons.  A coin has two sides, which means that everything always possesses positive and negative aspects. Even when it's about using artificial intelligence to find patterns within the data and to forecast future outcomes, there are potential drawbacks. The fourth chapter will provide a feedback of the use of machine learning within UCB financial forecasting process. Theory and practice will be confronted.

Concerning the methodology used to write this thesis, the author tried to diversify as much as possible the sources and methods used to collect qualitative and quantitative data. The information collected for this thesis comes from ICHEC courses, scientific articles, a book, business articles, informal discussions and some interviews with experts from the field. First of all, he wanted to define the financial forecasting concept, compare it with budgeting, explain current forecasting methods because he is willing to measure the impact of artificial intelligence on it. Therefore, he used some of his Ichec courses such as management accounting control and supply chain. In order to bring more insight, the author also used scientific articles. Afterwards, the author wanted to point out the importance of a reliable financial forecasting and he based his reasoning on a survey of KPMG because that survey gathers the opinion of hundreds of companies. At the end of the first chapter, the author pointed out the shortcomings of today's financial forecasting processes because he strongly believes that some of these shortcomings could be overcome by using artificial intelligence. He used scientific articles and collected the opinion of some experts from the field in order to get both a theoretical and a practical list of today's financial forecasting process shortcomings.

In the second chapter, the author explained the concept of artificial intelligence and its foundations in order to provide the reader with a good understanding of what this technology implies and to better analyze its potential for businesses. Afterwards, he pointed out the advantages and drawbacks of predictive analytics because it's the sub-technology of artificial intelligence that enables companies to predict their financial figures. The complementarity between predictive analytics and business intelligence has been explained because the author believes that these two concepts are often mixed up because both share the same purpose but are completely different. The on-going challenges have also been pointed out to show that artificial intelligence isn't perfect and that there are still challenges when using artificial intelligence. The author decided to mention the AI business cases to demonstrate where it could bring value to companies and then pointed out which one is used when speaking about adopting artificial intelligence to forecast financial figures. Afterwards, he described the type of machine learning algorithms required for prediction because these are the ones at the basis of predictive machine learning models that will be used for financial forecasting. To provide all this information, the author used multiple scientific articles, informal discussion and content from interviews. Using machine learning models to forecast also changes the financial forecasting approach. This is the reason why the author decided to explain it into his thesis. Moreover, he analyzed the required business models because he believes this will enable the readers of this thesis to identify either or not a certain company could take advantage of predictive analytics before even knowing how to implement it. In order to assess the required business models, the author inspired himself from his readings, informal discussions, interviews with a CFO's, BI consultants and data scientists. Last but not least, the author ended this chapter with the explanation of the different types of solutions available on the market in order to inform the readers about the solutions that could be potentially used to implement predictive analytics into a company financial forecasting process. However, implementing predictive analytics into a company financial forecasting process isn't just about buying a solution, it requires the company to go through a whole process where it builds its predictive model with the use of these machine learning solutions. To write this last part, the author used scientific articles, the websites of the services providers and interviews with data scientists in order to better understand the different types of solutions.

The third chapter is dedicated to the demonstration of the whole implementation process. To do so, the author mainly based it on the several interviews he had with experts. He also used a book about data mining and some scientific articles written by experts in artificial intelligence. This part is very exciting, and the author wanted to demonstrate the implementation process in order to provide the readers with a more pragmatic view on how to take advantage of artificial intelligence and more specifically predictive analytics into their financial forecasting process.

The fourth chapter will provide a practical feedback of the use of machine learning for financial forecasting purposes at UCB. It will provide readers with an overview of UCB predictive analytics experience and a comparison between the theory and the reality.

In a nutshell, the author always tried to follow a logical path in his thesis using the best methods to collect the information and write his thesis. To do so, he diversifies as much as possible the data collection sources in order to answer to his research question.

Regarding the hypothesis made at the beginning of this thesis, the author thought that using artificial intelligence within a financial forecasting process would:

- Automate the whole financial forecasting process;
- Improve companies financial forecasting accuracy;
- Artificial intelligence could be implemented by all companies;
- Machine learning automates the whole implementation process, no human intervention anymore;
- People willing to implement a predictive model need coding skills.

This thesis will be helpful and interesting for whoever wants to understand the advantages of using artificial intelligence for financial forecasting purposes (more specifically machine learning and data mining), discover the potential solutions on the market and how to take advantage of these AI solutions into their financial forecasting process. This thesis gives an overview on the importance of leveraging data, using artificial intelligence to perform human tasks, current AI forecasting solutions and the journey a company needs to go through in order to implement a machine learning model to find patterns within the data and forecast financial figures.

# 1. Financial forecasting

## 1.1 Introduction

Financial forecasting has always been very important within businesses because of its strategic importance and the need to anticipate future outcomes. Organizations are operating in a more than ever changing environment which makes it more and more difficult to anticipate and forecast accurately. Forecasting financial figures is obviously a difficult task but when it's done well done, it provides companies with a significant competitive advantage over those with poor financial forecasting accuracy. Nowadays, around 84% of companies strongly believe that investing into AI will lead to competitive advantage. (Columbus, 2018) Companies financial forecasting process is essential because it enables to give direction to them. Without the implementation of a good financial forecasting process, it means that companies are moving ahead with blind eyes and become vulnerable. There are many quotes that shows the importance but also the difficulty to find patterns within historical information and consequently to forecast future outcomes. Even if it isn't with the expected accuracy, forecasting financial figures is essential. Companies should always try to look ahead in order to anticipate what is coming next and act consequently. (Hyndman & Athanasopoulos, 2018)

Hereunder are some famous quotes about the difficulty and importance of forecasting:

- "Prediction is very difficult, especially if it's about the future" (Bohr, 2007)
- "Some things are so unexpected that no one is prepared for them" (Rosten, 1980)
- "It's far better to foresee even without certainty than not to foresee at all" (Pointcaré, 2015)

These quotes highlight rightfully the fact that predicting future results isn't an easy task. However, it's important to do it in order to know where the company is going and what could be done to improve the business performance.

In this chapter, the concept of financial forecasting will be defined and compared with the budgeting concept. The existing forecasting methods will be pointed out in order to know where most companies stand today in terms of forecasting methods. Moreover, the importance of a reliable financial forecasting process will be explained followed by an analysis that highlights the limits of today's financial forecasting processes.

## 1.2 Financial forecasting concept

Financial forecasting is about finding patterns within historical financial data and predict the future outcomes as accurate as possible taking into account the historical data and future potential events that could impact the forecasted results. It will provide a realistic and projected outcome taking into account the business environment changes and the last view of expected underlying performance. It will help management teams to make the right decisions and interventions on time. Decisions will be taken based on the understanding of their performance and the gap to their targets. (Stevens, 2018)

There are some specific characteristics defining what financial forecasting is. These characteristics are the following one's: (Stevens, 2018)

- Financial forecasting is about estimating what will be achieved
- Financial forecasting is most of the time done for revenue and expense lines
- Financial forecasting is regularly updated. Depending the company, it's done on a monthly or quarterly basis
- Financial forecasting could be used to support short-term and long-term decisions
- There shouldn't be a comparison between the forecasted results and the actuals because this is done in the budget. So, there is no variance analysis in the financial forecasting.

Forecasting financial figures is mainly about forecasting the income statement. The first step of predicting the income statement is to forecast future revenues for the company products or services based on historical figures and trends. (Corporate Finance Institute, 2019) Afterwards, the second step consists at forecasting the gross margin which is often expressed in percentage of revenues. Therefore, companies will predict the cost of goods and subtract it to the forecasted revenues. Once the company has predicted revenues, COGS and the gross margin, it will predict the SG&A expenses in order to be able to predict the operating margin. G&A expenses is about calculating general and administration costs such as IT, management, administration, marketing, sales and general overheads costs. (F. Manseau, personal communication, February, 2019) So, the forecasted operating margin can be calculated by subtracting SG&A expenses from the gross margin. Besides, companies can also forecast the balance sheet without completing the cash. This will be completed once the cash flow statement will be predicted. These 3 statements, the income statement, the balance sheet and the cash flow statement create a financial model. (Corporate Finance Institute, 2019)

Moreover, it's important to point out that when companies are forecasting their future revenues, they are trying to monetize the future clients demand. (F. Manseau, personal communication, February, 2019) It's important for companies to be able to answer to the clients demand because if they can't answer to the clients' demand, it means that there will be potential loss of sales, and thus a loss of revenue. If the forecasted demand was accurate, the company could have answered to the customer demand and increase its revenue. (Vandeput, 2018)

## 1.3 Complementarity between forecasting and budgeting

Forecasting and budgeting are often confusing and used together but there are significant differences between them. Budgeting is a task done every year by the management team where they are quantifying expectations for what the company wants to achieve. It's all about encouraging behaviors and accountability to make sure that the strategy is well executed. (Stevens, 2018)

In order to distinguish both budgeting from financial forecasting concepts, the main characteristics of budgeting are listed below: (Stevens, 2018)

- Budgeting is a representation of the future results, cash flows and financial position that the management wants the company to achieve for a certain period of time.
- Budgeting isn't regularly updated such as financial forecasting. It is a yearly task.
- In the budgeting process, there will be a comparison between the budget and the actuals to point out the potential difference between the expected performance and the actual performance of the company.
- The management team will find the required solution to align the actual results with the budget.

To conclude this comparison between these two important processes, we can state that budgeting is more about setting up a plan to determine the direction where the business wants to go. Besides, financial forecasting is more an estimation of where the company is actually going. In addition, financial forecasting should be seen as a real competitive tool and should be integrated in the decision-making activities of the management team as its impacting and playing an important role in barely all areas of a company. (Hyndman & Athanasopoulos, 2018)

## 1.4 Financial forecasting needed for critical business decisions

Now that the terms financial forecasting and budgeting have been defined, it is important to define why the company need to forecast. When companies are setting their yearly budget, there are taking accountability to achieve the fixed budget. (F. Manseau, personal communication, February, 2019) During the year, the company will forecast revenues, costs, margins etc. in order to see if they are going to achieve or not their budget. As a consequence, they will be able to take decision and adapt themselves to still try achieving their budget.

There are 3 types of time-horizons in forecasting: forecasting for short-term decisions, forecasting for medium-term decisions and forecasting for long-term decisions. (Vandeput, 2018)

- Short-term forecasting is used for operational decisions. In this case, there is a small-time horizon with high precision. These types of forecasting are used to determine for example how much quantities need to be shipped, how many quantities need to be produced, how much work hours need to be performed. With short-term forecasting, companies can define the scheduling of production, transportation and personnel. (Vandeput, 2018)

- Medium-term forecasting is used for tactical decisions. In this case, we have a medium-time horizon with a lower precision. This type of forecasting is needed to determine future resource requirements such as the purchase of raw material, machinery and equipment or the quantity of personnel to hire. (Vandeput, 2018)

- Long-term forecasting is used for strategic decisions. In this case, we have a high-time horizon with low precision. This should take into account external factors, internal resources and market opportunities. A few examples would be: should this new product be launched on the market? How much should be invested in production/inventory? (Vandeput, 2018)

From the author point of view, it's important that companies develop an adequate financial forecasting process to predict future financial figures and take into account several elements such as their business needs, the available data, the market prosperity... Nevertheless, this requires a very good understanding of the environment, identify potential financial forecasting problems, applying a range of forecasting methods to compare them, being able to select the appropriate forecasting method and monitor it through time in order to redefine the forecasting methods if needed.

## 1.5 The existing traditional forecasting methods

When companies are asking themselves the question about which forecasting method to use, it mostly depends on the type and quantity of data available. There are two types of forecasting methods spread among companies to forecast their figures. On one hand, companies could use judgmental forecasting and on the other hand they could use statistical methods. There are 4 different types of quantitative methods.

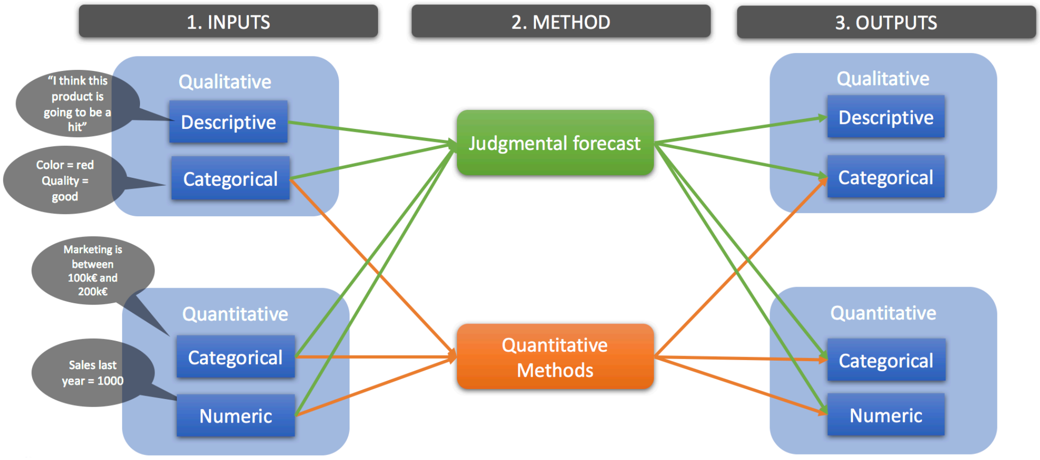


*Figure 1: Two type of forecasting methods (Vandeput, 2018)*

However, the question that companies need to ask themselves is which of these methods is the most adequate for their forecasting needs. Therefore, companies need to take a closer look at the data available and the different relationships between the input and their output data of their financial forecasting process.

Let’s take the example of a company that is forecasting the revenue of a product but is missing historical data because that product has been recently launched on the market. Consequently, it would be more relevant to use the judgmental forecasting method. When using the judgmental forecasting method, it could implicate the use of qualitative and/or quantitative input data. It isn’t just about guessing future revenues, there are well defined and structured approaches that enable companies to obtain good forecasted results without the use of quantitative methods.

Nevertheless, when a company decides to use the judgmental method for financial forecasting, it should pay attention to three types of pressure that could be harmful to the accuracy of the forecasted results. (Vandeput, 2018)

Below, the 3 pressures are listed: (Vandeput, 2018)

- Social, hierarchical pressures. When a manager is taking care of the development of a certain product on the market for example, he will be influenced by the fact that he is the one that is in charge of that product on the market. Moreover, if this manager is asking his “n-1” point of view, the employee might say that the product will be successful because he isn’t willing to disapprove what is superior says. In a nutshell, this social and hierarchical pressures need to be taken into account and could be mitigated by making anonymous surveys about the potential successfulness and revenues that the product could bring.
- Cognitive bias. What is meant with cognitive bias is that when human beings are confronted with a situation where they are trying to prove that they are right,

they will ask the opinion of people that will confirm what they are believing and this can bring bias in the forecasting of some product or services revenues.

- Budget, bonus pressure. What is meant by the author with budget pressure is that some forecasters will increase the forecasted amount to make sure that it follows the budget and that the company will be able to align with the budget and some forecasters will lower the forecast amount to make sure that it's achieved and that they will receive their bonus.

If companies want to make sure that these elements aren't influencing their judgmental forecasting they should consider two elements. The first one is to make sure that financial forecasting are not done by the user. If a manager is leading a project, he shouldn't do the financial forecast of that project alone because he would be influenced by the fact that he is the manager of that project. The second element that could mitigate the risk of financial forecasting error is to anonymize people's opinion with the Delphi method. (Vandeput, 2018) Questionnaires are sent to experts in order to gather anonymous answers, the results are shared among all experts. Since there are multiple rounds of asked questions, the Delphi method tries to reach a correct answer through consensus. (Twin, 2019)

The second type of method that companies could use to forecast revenues are quantitative models based on statistics and mathematics. These quantitative methods could be useful and provide more accuracy than judgmental forecasting if two conditions are fulfilled. The first one is that the company has enough historical data to train the model. The second condition is that the model has proven to be reliable in the past and that it has proven to reach good results. In order to create a quantitative forecasting model, there are 4 steps to follow. (Vandeput, 2018)

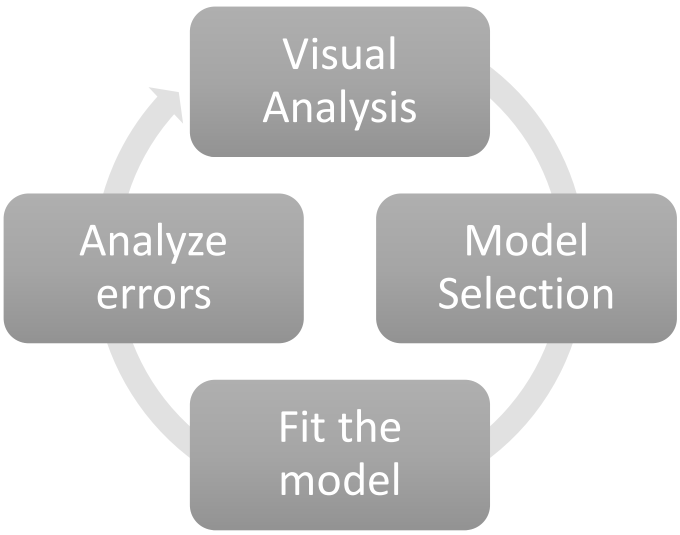


*Figure 2: How to build a statistical model (Vandeput, 2018)*

The first step consists at analyzing the data by plotting the revenue of a certain product over time and if possible plotting also the revenue against explanatory variables such as pricing, holidays or marketing. Then, the forecaster will have to see if there are trends, seasonality or similarities between different products or locations.

The second step consist at choosing the adequate model. Before choosing the model, it's important to first select the kind of data. Either using endogenous variables such as the revenue over time (time-series) or the exogenous variables that integrate explanatory variables such as the price, marketing budget… If the relationship between the explanatory variables and the revenue isn't very clear and understandable, then it's better to just use endogenous variables.

As mentioned before, there exist 4 statistical models that enable companies to forecast future revenues based on historical data. Hereunder, the four most known quantitative forecasting models: (Bista, 2016) (Vandeput, 2018)

- Simple Models which is about calculating the average of previous revenue points
- Linear regression using exogenous inputs such as prices and marketing expenses to forecast revenue
- Exponential smoothing which is a model using exponential smoothing of the previous revenue
- ARIMA models which are the most complicated model combining exponential smoothing and multiple linear regressions to forecast. Here we have a combination of endogenous and exogenous variables to improve the forecasting accuracy.

The third and fourth steps consist at training the model with historical data to make sure that the model is accurate enough and will be efficient at predicting future financial figures such as revenues. This third step need to be completed by analyzing the errors and making sure that these errors won't happen again. (Vandeput, 2018)

It's rather clear that an accurate financial forecasting process within a company is essential and primordial to give direction and help companies to increase their corporate performance. Below, a scheme showing the link between the strategic planning of a company, its budget, its planning and its financial forecasting process.

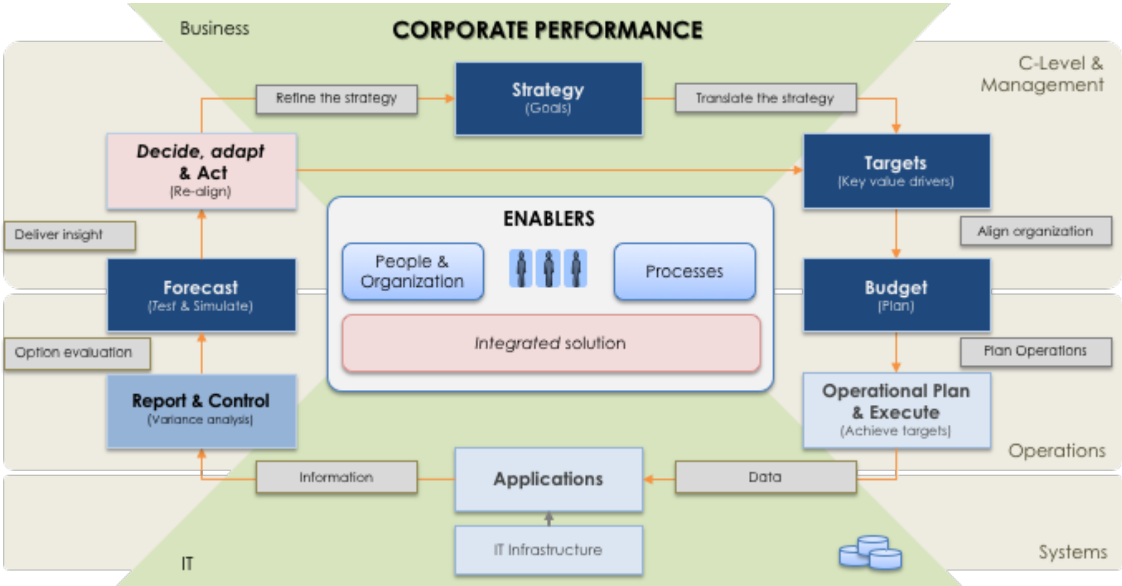


*Figure 3: Corporate performance (Keyrus, 2019)*

Companies are making their strategic planning where they are setting their future goals such as organic growth or acquisitions, portfolio expansion… Based on their strategic planning they will set some targets which are key drivers of value. Afterwards, they will elaborate their budget of their revenues, costs, margins, cash flows in order to communicate what the company wants to achieve. People within the company are accountable in regard to the budget. The budget is set up through a collaboration between C-level and operations. Companies execute their operations with a goal to achieve the budget elaborated at the beginning of the year. Every month, the company will forecast the financial figures of the next 12 months in order to know if the outlook is favorable or not to the achievement of the fixed budget. When forecasting every month, the coming 12 months, it's called a rolling forecasting. This will help the management teams to adapt, act in consequence and take the best decisions possible. (F. Manseau, personal communication, March, 2019)

In addition, both statistical and judgmental forecasting methods can be used by companies to increase the accuracy of their forecasting process. First, they'll find a good statistical model to forecast the financial figures and then they bring more insights to the forecast by starting what is called a collaborative forecast where different stakeholders give their feedback. Clients are for example consulted to know about their intentions to buy the company products or services. Sales representatives are asked what the opportunities are and if they agree or not with the forecasted baseline. And last but not least, the management team has to make sure what has been forecasted is in favor of the company's strategy. This process enables to gather key players of the company, to inform them about the budget situation and put the company in the right direction again by taking the right decision.

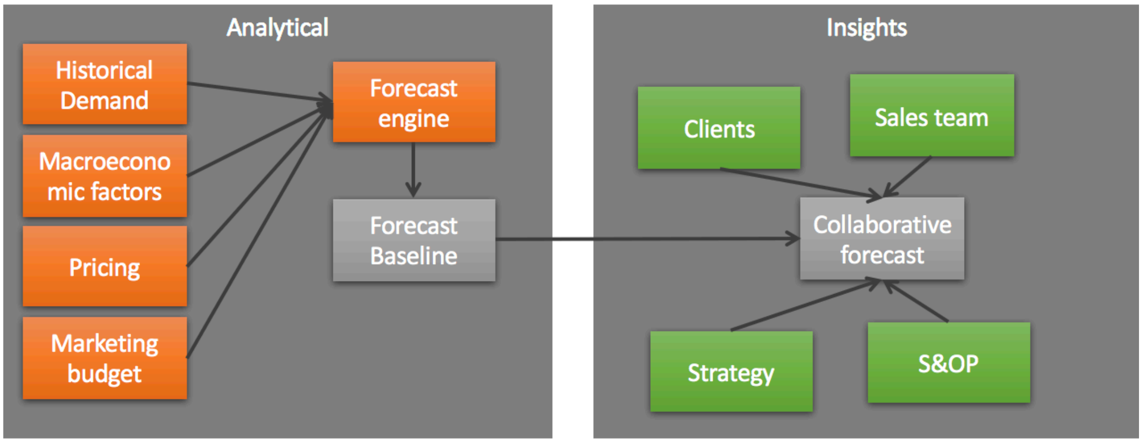


*Figure 4: A collaborative forecasting approach (Vandeput, 2018)*

From the author point of view, companies are moving alongside a learning curve. At the bottom of the learning curve, there is the judgmental forecasting method which could be useful in case of a lack of historical data and input, output data relationship understanding. At the top of the curve there are some complex statistical methods that enable companies to improve their forecasting accuracy but requiring a lot of historical data and an advanced understanding of current patterns which isn't always easy when products/services are new on the market. Moreover, if companies want to increase their forecasting accuracy, they need to shorter the forecasting horizon, aggregate more as it is making forecasting easier and try to collect the most reliable data possible.

## 1.6 A reliable financial forecast increases business value

Financial forecasting is considered as very important for most companies because of its ability to create and sustain business value. It's a key pillar to manage a company's performance and provide executives with important information to adapt quickly to a changing environment and communicate with their external stakeholders. CFO's are always willing to improve their view on future financial performance and this is the reason why leading companies are investing a lot of time and efforts for its improvement. (PwC, 2011) As a matter of facts, forecasting plays an important role within a company and most of the time impacts all the areas of the company. Forecasting plays an important role in the annual budgeting process, the improvement of the strategic planning, the improvement of the ongoing performance management, managing cash requirements or even as a way to improve communication with investors. Below, a graph shows the important roles of financial forecasting within a company. (Lies, Parker, & Reader, 2017)

*In which of the following does your organization's forecast play an important role?*

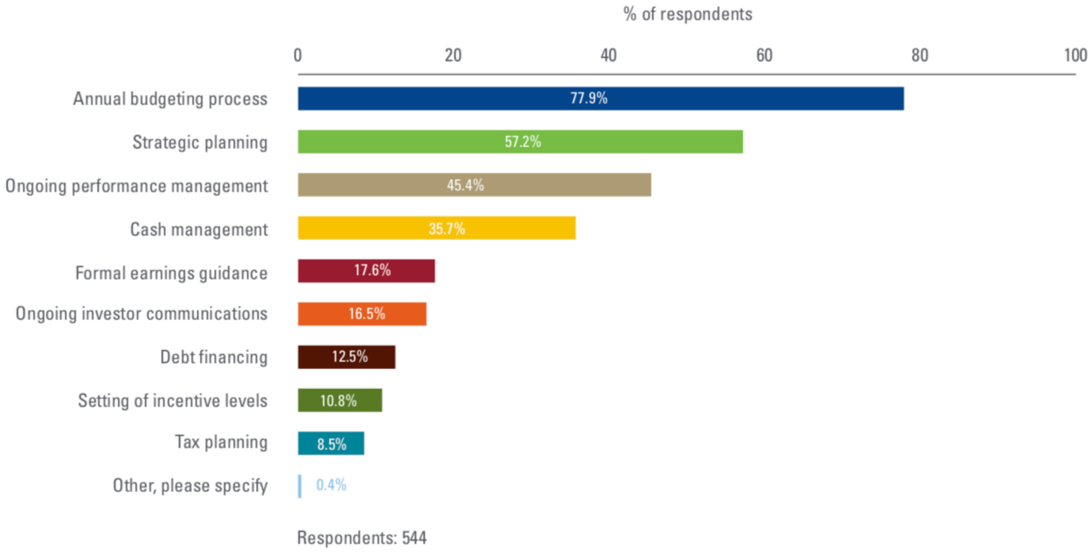


*Figure 5: The important roles of financial forecasting (Lies, Parker, & Reader, 2017)*

From the author point of view, this clearly shows that financial forecasting is essential to create value for businesses. In fact, it is a key task that explains where the company is going and it enables companies to improve the corporate performance by evaluating the current situation and delivering insight from the bottom line to the C-level and management of the company.

Moreover, it's important to point out that a reliable financial forecasting process enables companies to increase their capabilities over time. A more accurate financial forecasting gives companies the means to faster identify opportunities for growth, to increase their ability to identify risks to their strategic plan, to increase performance for the business units and to improve relationships with investors. The implementation of a reliable financial forecasting process within a company needs to be seen as crucial because it really adds value to businesses and empower companies' capabilities. Below, a graph identifies the main benefits coming along with the improvement of a financial forecasting process accuracy. (Lies, Parker, & Reader, 2017)

*What do you anticipate would be the main benefits of better forecasting within your organization?*

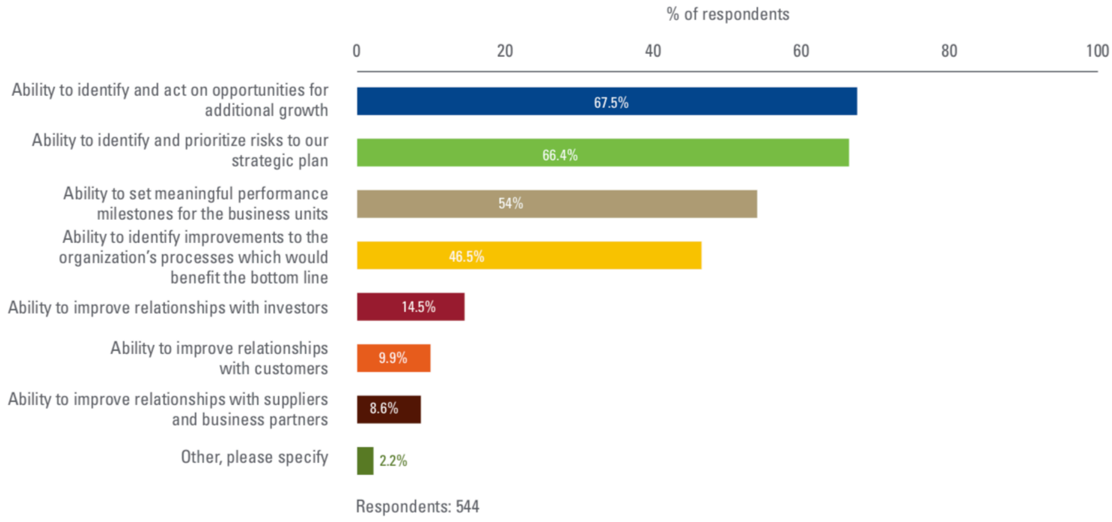


*Figure 6: The benefits of a better financial forecasting process (Lies, Parker, & Reader, 2017)*

If we look at the impact of forecasting accuracy on the share prices of listed companies, it also proves that unreliable forecasting has a negative impact on companies. Market capitalization is a good indicator to measure the impact of the forecasting accuracy. In 2012 for example, Walgreens affirmed that with its Boots merger, it predicts a profit of $9 billion in 2016. Two years later, the CEO of Walgreen communicated that the company wouldn't meet its profit forecast of $9 billion and would achieve instead a $7.2 billion profit. This shocked investors and the stock dropped by roughly 14%. (Valinsky, 2018)

If we analyze the graph below, we can notice that poor financial forecasting cost money to companies. The average cost of forecasting errors over the last three years reaches 6% of the share price for companies. This cost represents an important amount of market capital. Moreover, it has been proven that accurate financial forecasting leads to a more important increase of the share prices than those having a less accurate financial forecasting. The difference of share prices increase could reach 5%. (Lies, Parker, & Reader, 2017)

*When considering the impact of forecasting errors, approximatively how much do you estimate that these have cost your organization over the last three years in terms of share price?*

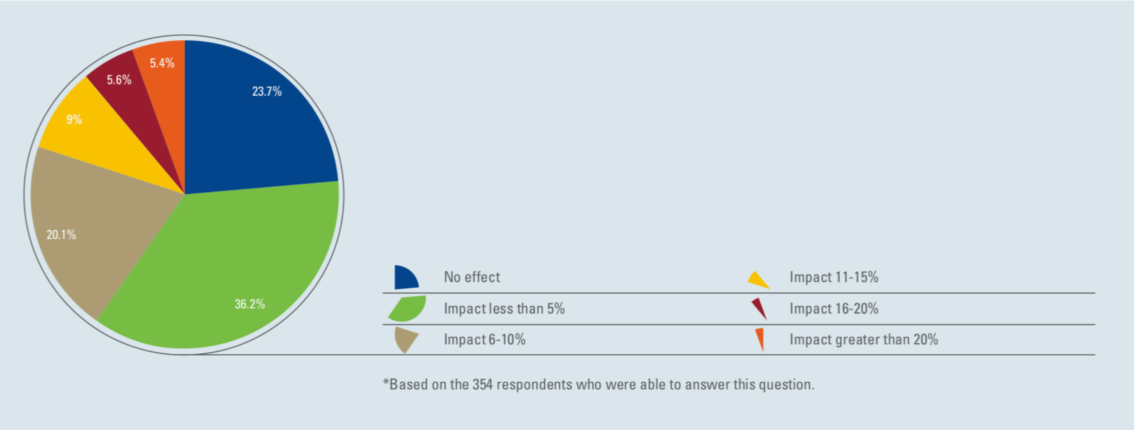


*Figure 7: The cost of forecasting errors in terms of share price (Lies, Parker, & Reader, 2017)*

Companies acknowledge that a reliable forecasting process could increase their business value by boosting their overall capabilities to perform and increasing their share prices. A reliable forecasting process at the heart of companies' performance management process is essential. If managers neglect the importance of their forecasting process, it will cost money to the company and will also blur the direction in which the company is going.

On one hand, the forecasted business information will not be reliable because of its lack of insight. Consequently, managers won't use the information to take decisions and drive performance. On the other hand, top management teams of companies won't get a clear view of the business direction, its opportunities and risks. This could also lead to shareholders problems as they are requiring nowadays more and more information about the look out of future performance. Financial forecasting should be definitely seen as a core business capacity rather than a responsibility of the finance function with small contribution to the business cycle. (Lies, Parker, & Reader, 2017) So, the importance of a reliable financial forecasting process has been previously explained by pointing out its potential impact on the business performance and on the business value increase.

As a matter of fact, it seems rather important to point out the different areas in which companies could invest time, efforts and money in order to increase the reliability and accuracy of their financial forecasting process.

*Which of the following would/has give(n) your organization most benefit in improving the confidence of forecasts?*

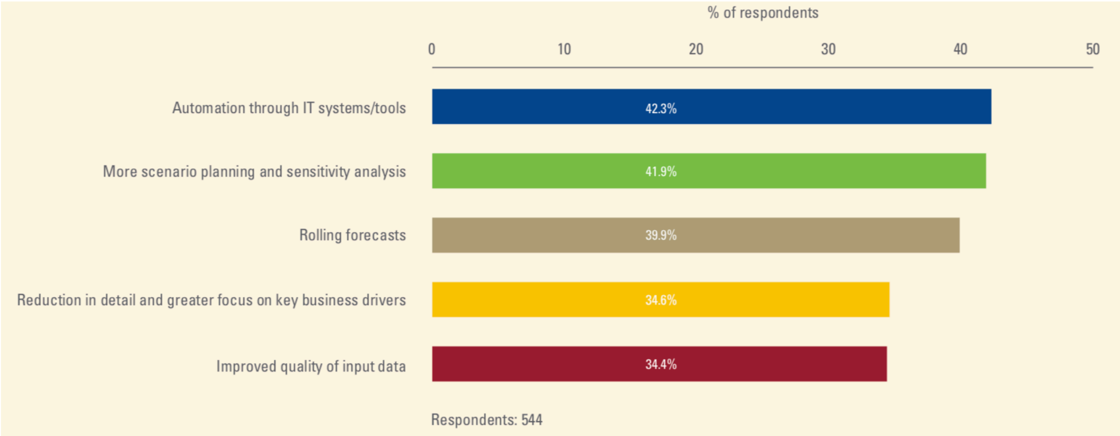


*Figure 8: Points for forecasting improvement (Lies, Parker, & Reader, 2017)*

Above, the graph shows the top 5 areas in which companies could invest to increase their forecast reliability and accuracy. The difference in percentage between the 5 areas is minimal which shows that whatever area a company wants to improve, they are all more or less equally important and will all bring additional value to companies financial forecasting process. The number one area concerns the implementation of technology that would automate partially or totally the forecasting process. The top 3 is completed by an increase of scenario planning and the use of rolling forecasting that enables a company to keep pace with rapidly changing business environment. The fourth area concerns the importance to focus on key business drivers. The top 5 improvement measures are completed with the improvement of the input data quality. The author strongly believes that input data quality and automation through IT are going hand in hand. Both should be linked together because automation through IT without quality input data isn't going to improve the forecasted results. This will be more detailed later in this thesis.

From the author point of view, companies are operating into a world where technologies became part of their daily business life. Companies integrating technology into their processes enable them to outperform competitors by gaining efficiency and thus gaining a considerable competitive advantage. There are many examples of IT tools such as enterprise resource planning, customer relationship management, big data analytics, business intelligence, chatbot, digital communication tools and many more. All these technologies have disrupted the way companies operate but it enables them to increase their efficiency and performance by bringing more insight into their ecosystem. In today's business world, a lot of people speak about artificial intelligence and its huge capacity to potentially increase companies' capabilities in many areas such as financial forecasting.

## 1.7 Shortcomings of today's financial forecasting processes.

Developing accurate and fast financial forecasts isn't an easy task for financial teams. Moreover, low forecasting accuracy has far-reaching implications for companies' performance. It's logic that many CFO's and its finance team look at finding solutions to improve their financial forecasting process because of the many shortcomings of traditional forecasting processes.

When the author did his internship at Keyrus, he was able to take a closer look at the monthly forecasting process of the company and update the financial reporting file that had to be presented to the management team. This process was heavy and took a lot of time and efforts. Getting people together such as sales representatives, business unit directors, finance team members was essential in order to set up a good financial forecast. Forecasting and updating the financial reports through spreadsheets wasn't always efficient and asked a lot of time and manual work. During his internship, the author saw that there were some potential shortcomings with today's financial forecasting processes. These shortcomings could be improved by using artificial intelligence.

Below, a list with some shortcomings concerning today's traditional forecasting processes: (Murray , 2018) (Masnaoui, 2019)

- Lots of manual work driven by spreadsheets. It requires considerable people, time and efforts to produce but also to update forecasts files.
- Human and organizational bias. (Vandeput, 2018)
- Traditional forecasting methods cannot handle huge amounts of data that would be available. (Cote, 2019)
- Low communication and integration with sales and operations
- Financial forecasting models that aren't accurate enough because they use current revenues multiplied by arbitrary growth rates to forecast rather than using the available data and business drivers within the company.
- Lack of insight within the company's data and therefore a lack of pattern discovery. (Cote, 2019)
- Some companies rely on limited data sources (F. Manseau, personal communication, February, 2019)
- Some statistical methods used to forecast have difficulties to deal with non-linearity. (Coallier, 2019)

If we take the example of Microsoft Benelux, they were truly willing to invest into artificial intelligence for revenue prediction. This would automate their financial forecasting process and enable them to get a real-time financial forecasting process. They identified the revenue drivers, used internal and external data in order to forecast their revenues with the use of a predictive machine learning model. They were able to improve their revenue forecasting accuracy by 2%. After implementing AI into their financial forecasting process, they could achieve around 98% forecasting accuracy. An improvement of 2% seems quite small but this represents millions of euros for Microsoft Benelux. (J. Brar, personal communication, November, 2018)

The author is of the opinion that traditional forecasting methods result in a very high time-consuming process in order to produce and update forecasts that will be more of a qualitative than a quantitative nature. Financial forecasting is often impacted by human and organizational bias and thus it often ends with less accuracy in the forecasted results. Moreover, traditional forecasting method cannot handle and take advantage of huge amounts of available data. As a consequence, companies don't have the means to forecast financial figures based on historical data which reflects the company's reality. Last but not least, it impacts decision takers from the C-level of companies who need to take critical business decisions while relying on information that isn't reflecting the company's reality. The information isn't accurate enough and do not necessarily provide business insight.

## 1.8 Conclusion

Financial forecasting is about predicting future outcomes which indicates where the company is going. Most of the time, it's done on a monthly or quarterly basis. Financial forecasting is part of the corporate performance process and helps the management team to better manage the company and make the best decisions possible.

The four main traditional forecasting methods that companies use have been pointed out and shortly explained. These methods are the simple method, linear regressions method, exponential smoothing method and ARIMA model.

Financial forecasting plays an important role within companies' annual budget process, strategic planning, ongoing performance management and cash management. A reliable financial forecast is essential for companies if they want to see where they are going and correctly anticipate what is coming next. To do so, companies could invest into IT tools and using quality data to forecast.

However, current forecasting processes possess some shortcomings that make financial forecasting a heavy and time-consuming task. Some of these traditional forecasting methods shortcomings would be the quantity of manual work, human & organizational bias, difficulties to deal with non-linearity, difficulties to find patterns within huge amounts of data.

The author strongly believes that artificial intelligence and more specifically predictive analytics could help companies to improve their financial forecasting process by taking advantage of the historical data available by finding patterns and insight within it.

# 2. Using artificial intelligence for financial forecasting

## 2.1 Introduction

The term "Artificial Intelligence" already exists since 1956 when scientists where looking how computers could solve problems on their own. These last years, the concept has become more and more popular because of the increase of data volume, improvement of computer power and advanced algorithms. (SAS, 2019)

Since the 1950's, AI research already focused on topics such as machine problem solving. In the 1960's, the US department of Defense already began to train machines by integrating human reasoning in order to give computers the means to think like them. In the 1970's, the Defense Advanced Research Projects Agency achieved street mapping projects thanks to AI. Artificial Intelligence is often interpreted as human-like robots that are taking over the world but the real AI technology isn't about robots but more about algorithms that enables companies to improve their decisions-takings and increase their efficiency by thinking such as humans and producing intelligent output. (SAS, 2019)

In this chapter of the thesis, the concept of artificial intelligence and its foundations is going to be explained. The advantages and drawbacks of predictive analytics will be detailed in order to provide readers with a good understanding of it. Artificial intelligence and more specifically predictive analytics still got some challenges that are going to be pointed out. Besides, predictive analytics and business intelligence concepts will be compared because of their common objective. Afterwards, the use cases of AI will be shortly explained and the impact of predictive analytics on the financial forecasting approach will be analyzed more into detail. At the end of this chapter, readers will be aware of which type of business models are in the best position to adopt predictive analytics within their financial forecasting process. To conclude this chapter, readers will be showed the different type of machine learning solutions currently available on the market.

## 2.2 Understanding artificial intelligence

In order to understand where AI stands today and what business value the technology could provide in terms of financial forecasting, it's important to define the concept and its foundation.

Artificial Intelligence makes it possible for machines to learn from given inputs, past experiences and to perform tasks usually done by humans. Artificial intelligence Is considered as the science of training machines to perform human tasks. (SAS, 2018) When humans take decisions, there is a whole process going on in their brains to decide what is the best outcome to a given problem. The idea is that artificial intelligence enables computers to replicate this human thinking. When we look at today's AI applications such as computers playing chest, autonomous driving cars, robotic process automation, predictive models… there are considerable technologies on which artificial intelligence relies. These sub-technologies are machine learning, natural language processing, computer vision and predictive analytics. Using these underlying technologies enable for example computers to be trained to process huge amounts of data and recognize patterns in the data in order to achieve human-like tasks. This is what happens when companies use artificial intelligence for financial forecasting purposes. (SAS, 2018)

When algorithms process the data in order to find insight and patterns within the data, it is also called "data mining". Data mining is all about mining the data in order to extract knowledge from it and turn it into useful information for the company. (Masnaoui, 2019) In order to discover insight within the data, there are multiples data mining techniques that could be used such as linear regressions, decision trees, random forest, neural networks, KNN… These techniques merge many fields such as mathematics, statistics, machine learning, deep learning. The decision about which data mining technique to use will depend on the task that need to be performed and the available data. The main data mining techniques will be explained in the implementation chapter of this thesis in order to show the readers which one could be used for regression tasks such as financial forecasting. (cf. infra p.70)(T. Larose & D. Larose, 2015)

When using these algorithms and data mining techniques, it's part of a concept called "data science". (New Generation Applications, 2017) Data science is a well-known word used to point out the exploration and extraction of knowledge from the data by using different data mining techniques. When machine learning algorithms are finding patterns within the data, they are using data mining techniques to do so. In fact, they are literally mining the data, meaning that they try to find useful information within it. (Scherbak, 2019) Moreover, in the case of financial forecasting, data science is also about finding the best data mining techniques to use and build the best predictive models possible.

In other words, artificial intelligence is a science where algorithms are trained to emulate human tasks. Artificial intelligence is part of the data science concept and is using data mining techniques to learn and extract knowledge from historical data. (Masnaoui, 2019) So, at the heart of AI, there is the ability for algorithms to gain a deep understanding of complex datasets in order to apply human logic and reasoning. The algorithm receives huge amounts of historical data and uses data mining techniques to identify relationships and patterns within the dataset. Once humans have determined the key features, that an analysis method has been determined, that the algorithm has processed the input data, that the code to execute the analysis has been written, the required output will be produced automatically. Once operational, this is all done through an automated process which will need minimal human intervention. (SAS, 2018)

From the author point of view, artificial intelligence must be seen has a technology that will give the means to companies to take better decision and improve their overall performance. Artificial intelligence will come along with business value for companies. In the case of financial forecasting, algorithms will be able to analyze historical data, find patterns within it and predict future financial figures such as revenues and costs. It will disrupt the working market by eliminating jobs but it will compensate the losses by creating new jobs. Companies should bear in mind that artificial intelligence has a huge potential to increase the businesses efficiency and performance. Hereunder, the figure shows the future outlook of artificial intelligence on the market.

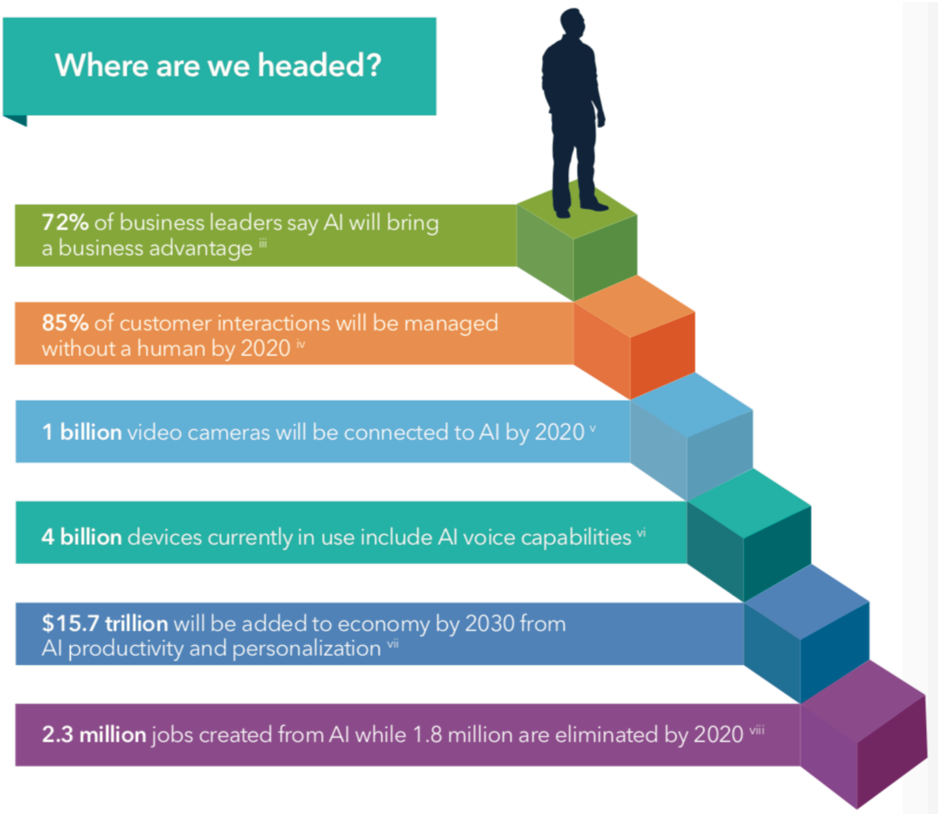


*Figure 9: The use of artificial intelligence in the future (SAS, 2019)*

From the author point of view, the goal of AI is to provide companies with human-thinking algorithms in order to support their decision-makings for specific tasks such as financial forecasting. The use of artificial intelligence shouldn't be seen as replacing human jobs but should be seen as helping them to be more efficient and take better decisions. AI algorithms have the capacity to learn more in depth than humans which enable them to discover patterns and make relationships within the dataset that goes sometimes behind human understanding. It will improve business capabilities and make them better at what they are doing.

## 2.3 How AI works and its foundations

As mentioned before, AI is mainly based on 4 core technologies which are machine learning, computer vision, natural language processing and predictive analytics.

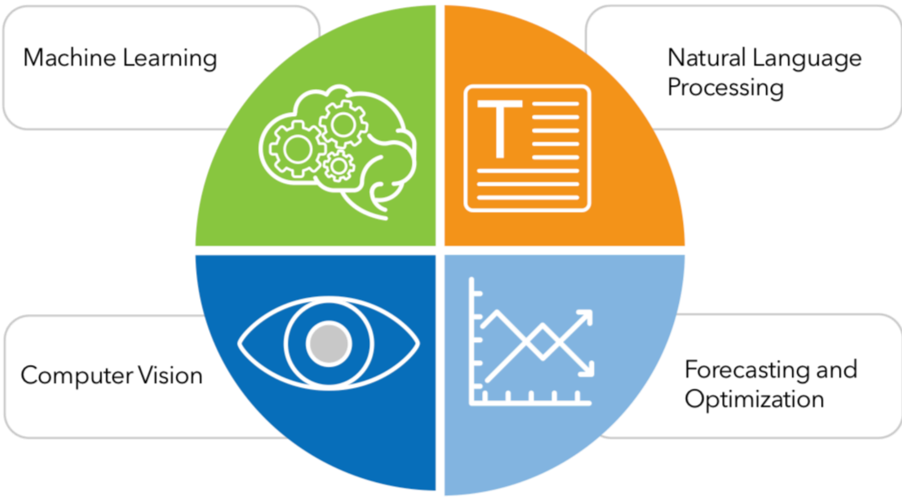


*Figure 10: The foundations of artificial intelligence (SAS, 2018)*

AI works by analyzing large amounts of data with advanced, fast and intelligent algorithms allowing the software to learn from the features and discover patterns in the data. Multiple technologies are used such as "**Machine learning**" which train the model how to learn. While analyzing the data, machine learning models are looking for patterns within the data and try to draw up conclusions. Machine learning is often used to predict future results by analyzing key features, predictor variables and assessing its impact on the output. (Cote, 2019) The models aren't explicitly programmed by humans but are given examples which will help the model to learn on its own by using data mining techniques. It's much easier for humans to give examples. So, machine learning models are using methods from mathematics, statistics, physics, machine learning, data mining to get insight from the data without being explicitly programmed in advance by humans. They learn on their own based on huge amount of data given to them. (SAS, 2018)

Sometimes, data scientists use the word "**Deep learning**". Deep learning is a sub-category of "machine learning" but used when there is more complexity and when the quantity of data to analyze is more important. It uses neural networks techniques to learn from the data and requires more computer power. Deep learning will enable the algorithms to understand and learn more in depth some very complex patterns within the dataset. There is an evolution based on the computer power, the amount of data and the variety of data to learn from. Hereunder, there is a figure showing the evolution of the "neural network" concept across time with the different variables impacting it. (SAS, 2019)

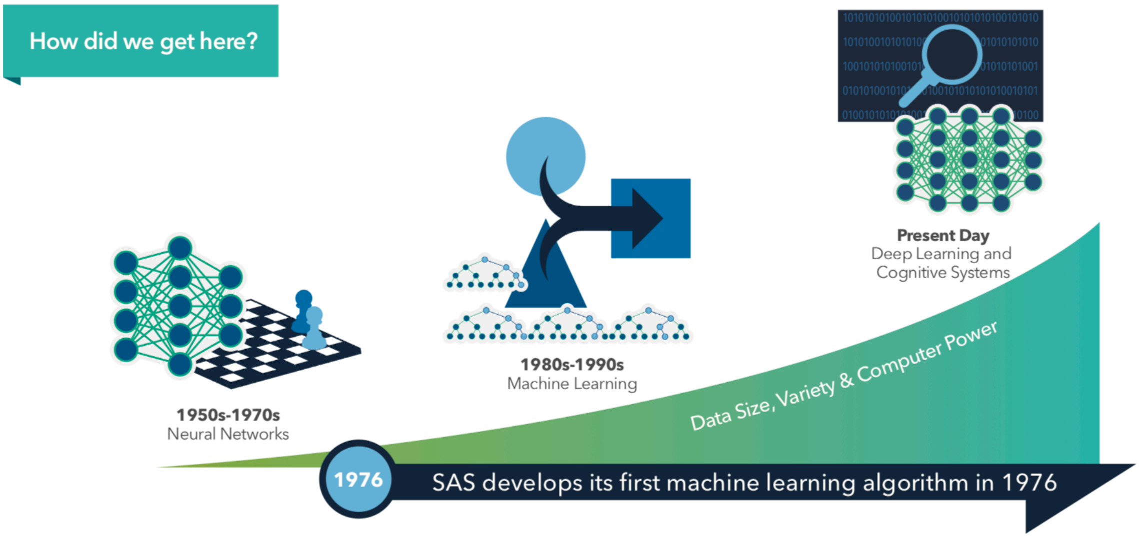


*Figure 11: The algorithms evolution through time (SAS, 2019)*

The second technology underlying the concept of AI is "**computer vision**". It means that machines learn to recognize and gain a very good understanding of what is in a picture or in a video thanks to pattern recognition and deep learning. If machines can process, analyze and understand images or videos, they will be able to capture pictures or videos in real time and understand its environment. Machines can understand and react to what they see. (Boscacci, 2018) There are daily examples such as face recognition where a phone can identify if it's the right person. In the football sector, computers can perform an analysis either or not the goal should be allowed based on real-time images of the ball crossing the goal line.

The third technology at the heart of AI is "**natural language processing**". This gives machines the ability to understand and generate human language, including speech. Humans will be able to communicate with computers which use a normal language in order to perform a task. The real added value of NLP is being able to elaborate a communication between computers and humans with a natural language. Computers can analyze a text, understand a speech, measure sentiment and gain insight by retaining what is important. (Dr. Garbade, 2018)

Last but not least, another underlying technology of AI is forecasting and optimization. It enables companies to predict future outcomes and potentially point out the best action given the multiple resource constraints. These predictive machine learning algorithms used to discover patterns and predict future results is also called "predictive analytics". When machine learning models within companies are well trained with the use of diversified and quality historical data, it could enable companies to forecast financial figures with high accuracy. Machine learning models understand what they need to forecast and drill down into historical data and try to define patterns by understand the relationship between the different input and output variables. Besides, optimization is when a predictive machine learning model is capable to point out the important actions to take in order to achieve the forecasted results. (SAS, 2019) Predictive analytics and machine learning are the sub-technologies used for financial forecasting purposes.

### 2.3.1 The added value of predictive analytics

Now that the concept of artificial intelligence has been explained and that the underlying technologies have been described, let's take a closer look at the added values of using predictive machine learning models for financial forecasting purposes. (SAS, 2019)

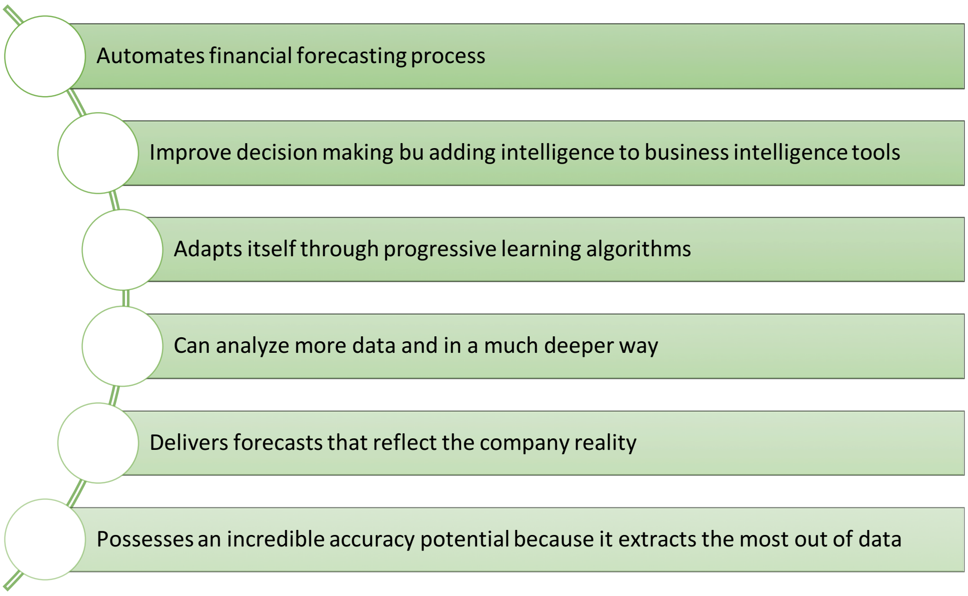


*Figure 12: The benefits of predictive analytics*

First of all, **predictive analytics automates forecasting processes through discovery within the data**. AI objective has never been to automate manual tasks, it has been created to perform high-volume and frequent computerized tasks with a high degree of reliability and without interruption. (SAS, 2019) Wouldn't it be nice for a company to get a real-time financial forecasting process totally automated which is based on the company's data? Besides, building and monitoring a predictive machine learning model will always require humans' participation to define the right machine learning question, to pre-process the data, to build the predictive machine learning model and to deploy it. Humans will always be needed even if companies use artificial intelligence into their financial forecasting process. (Van der Meulen, 2019) Business people are needed to understand the business problem and to define the required input data. In addition, data scientists are also essential because they are the one's helping business people to identify the required data, cleaning the dataset in order to provide quality data to the model, make explanatory analysis and build/determine the right machine learning algorithm to use. All these tasks from the definition of the machine learning problems to the building of the machine learning model have as objective to improve companies financial forecasting processes which will be based on quantifiable data. (Coallier, 2019)

The second advantage is that **predictive analytics will improve decision making by adding intelligence to business intelligence tools that might already be in place**. In fact, using AI to forecast will help companies to take better decisions and complete BI tools. Not long ago, the leading technology was business intelligence which is a reporting tool that provide companies with a clear view of how it performed before and how it performs today. This advantage is explained more into detailed in the next sub-chapter because it is essential to distinguish both technologies. (cf. infra p.37) (TopTal Research, 2018)

The third advantage coming along with the use of **predictive analytics is that predictive algorithms adapt themselves through progressive learning.** This means that predictive machine learning models aren't programmed explicitly since the beginning and that the data provided to it partially determine the performance of the predictive machine learning model. The algorithms acquire knowledge by finding patterns and relationships within the dataset. Moreover, predictive machine learning models have the ability to adapt itself to new given data. So, just as the algorithm can learn how to play chess against you, it can learn to forecast future financial figures such as revenues or costs based on new input data provided to it. (SAS, 2019)

The fourth advantage of **predictive analytics is that it can analyze more financial data and in a much deeper way** by using supervised machine learning models. Using predictive analytics has been made possible thanks to the increase of computer power and big data. In fact, to train predictive machine learning models, there need to be a lot of data available because the model directly learns from it. Moreover, it's essential that the data provided to the model is reliable and of quality in order to not biased the learning process of the predictive model. The more reliable the data given to the model, the more accurate it will become at predicting future results. (Vandeput, 2018)

The fifth advantage is that **predictive analytics delivers financial forecasts that reflect the company reality.** There are two reasons why a financial forecast made with a predictive machine learning model is reflecting the company reality. The first reason is that the predictive models will predict future results based on internal and external data available to the company. Companies will get a more quantitative forecast. The second reason is that the financial forecasting process will be faster and provide forecasted results in real-time. Companies won't need days or even weeks to forecast their financial figures where they forecast in fact outdated financial figures. The financial forecast will be done in real time thanks to a real-time predictive machine learning model. (Lieutenant, 2019)

Last but not least, **predictive analytics has an incredible accuracy potential because it extracts the most out of the data**. Machine learning algorithms can deal with non-linearity and huge amounts of data. (Coallier, 2019) When looking at the most important element that enables algorithms to learn, it's clearly data. So, data become a real asset for companies. Developing a data strategy becomes more and more important for businesses because data brings insight and answers to business matters. For most companies, implementing artificial intelligence provides a competitive advantage. (Columbus, 2018) Companies just need to apply artificial intelligence to the right data in order to get the right answers to a specific matter. The company which provides the most accurate data to the best predictive model will acquire a competitive advantage in terms of financial forecasting efficiency. (SAS, 2019)

In a nutshell, the author strongly believes that predictive analytics has the ability to perform forecasting tasks done by humans. Companies always look at increasing their efficiency, productivity, enhance business performance and results. Predictive analytics could help companies to achieve it. Predictive machine learning models seem to be the technology that could aggregate firms to the next level in terms of financial forecasting. Nevertheless, using predictive machine learning models comes not only with advantages for companies. There are also limits and drawbacks to consider when using predictive analytics for financial forecasting purposes within a business.

### 2.3.2 Drawbacks of predictive analytics

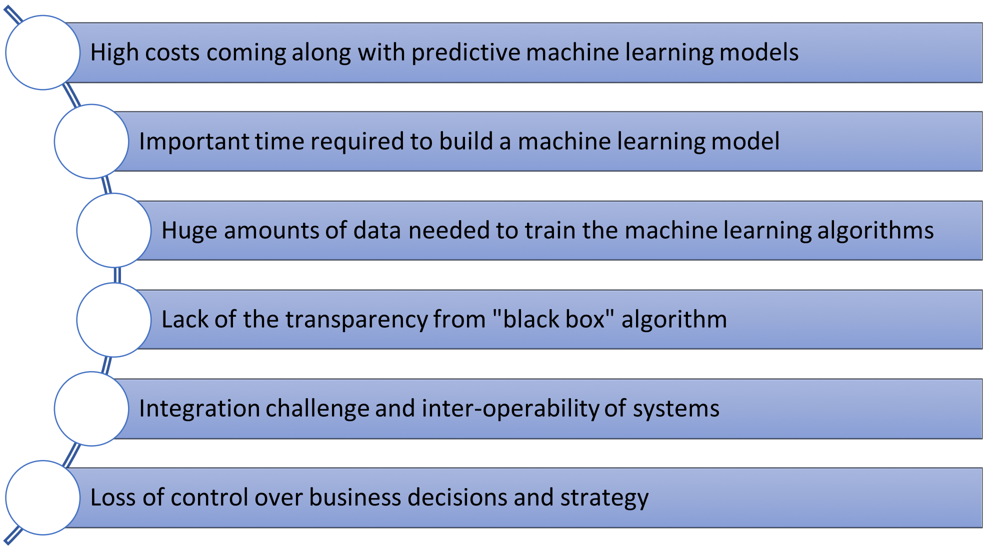


*Figure 13: The drawbacks of predictive analytics*

The first drawback of using predictive analytics is the **cost** that comes with it. Implementing a predictive machine learning model comes with an important cost due to many pre-implementation conditions and the need for on-going supervision during and after the implementation phase. Predictive models need to be trained with huge amount of data and the implementation process requires people, a lot of efforts and time. Moreover, companies integrate more and more technological solutions into their businesses which increase their dependency on the functioning of their IT tools. By implementing predictive machine learning models, companies rely once again a little bit more on IT and increase the risk of IT issues within their business. (AI Business

Information, 2018) Nevertheless, companies are obliged to follow the trend and implement emerging technologies otherwise they would lower their ability to compete and consequently increase the risk to be outperformed by competitors.

The second drawback of using predictive analytics is **the time required to build predictive machine learning models**. It can often be lengthy because the algorithms need to be trained before being used for current business situations such as the prediction of the profit and loss top lines. If we analyze for example a predictive model that forecast the revenues of an ice cream shop based on predictor variables such as the weather, the price, the location, the number of waiters… it means that the algorithm used by the predictive model needs to be trained thanks to **huge amount of historical data** before being able to predict future revenues for real. These steps of collecting the data, cleaning and transforming the data, finding patterns within the data, deploying the predictive model can take time. At UCB, they needed around 10 to 12 months in order to implement their predictive models for financial forecasting purposes. (Lieutenant, 2019) Once the whole implementation process is done, the predictive machine learning model will be able to be used for real financial figures predictions. (Cote, 2019)

A third drawback of predictive analytics is the problem of **the "black box" understanding.** It's a term used to highlight the fact that the algorithms are forecasting results without explaining the reasoning behind. The connections made between the data are sometimes going behind human understanding which makes it very difficult for financial forecasters to find a logic behind the forecasted results. In other words, the predictive algorithms show a lack of transparency. If there is no solution found to bring more insight from what algorithms forecast, it means that companies that implemented predictive analytics could one day rely somehow on an intelligent technology that they do not fully understand because of that lack of transparency. These companies would play a very dangerous game because it means that they take for granted the prediction of a model without knowing how it came to this output. Companies would follow with blind eyes the direction and forecasted results provided by the predictive model. (Vandeput, 2018)

The fourth limit of predictive analytics is that companies should **consider the integration challenge and the inter-operability between predictive analytics and systems already implemented or that should be implemented into the company**. Earlier in the thesis, it has been explained that companies need a lot of reliable data to train predictive algorithms. This pushes companies to implement very good information systems that enable the company to collect, store, analyze huge quantities of data otherwise it would be difficult to train the predictive algorithms to forecast financial figures. In fact, using predictive analytics is the top of the learning curve and before even thinking about predicting financial figures with machine learning, companies need to go through a digital transformation that enables them to collect the required data and fully take advantage of predictive analytics. (cf. infra p.45) (A. Cosse, personal communication, Mars, 2019)

Last but not least, when a company intend to carry out predictive analytics in its financial forecasting process, it should also consider **the possible loss of control over its business decisions and strategy**. It is still a sensitive topic but using predictive machine learning models means a decrease of human intervention within the business making process. In the future, predictive analytics should even be able to recommend future actions in order to achieve the forecasted results. Therefore, if algorithms do provide more transparency, it's possible that companies will follow forecasted situations and recommendations expressed by these predictive machine learning models. In addition, there is also a rise of numerous ethical concerns such as the potential loss of jobs with the increase of predictive analytics use cases, what will happen the day that a financial forecasting process relies on predictive analytics and that the algorithm makes an important error in the forecasted results…who's responsible? (AI Business Information, 2018)

In a nutshell, a coin has always two sides. When looking at predictive analytics, there are benefits but also some drawbacks that come along with the technology. These drawbacks should clearly be beard in mind in order to avoid future ambiguous situations with the use of predictive analytics. Nevertheless, the author strongly believes that companies should implement predictive machine learning models into their financial forecasting process because it has a huge potential to create business value with accurate and real-time financial forecasts based on the collected data within and outside the company.

### 2.3.3 Predictive analytics completes business intelligence capabilities

Since a decade, technological tools are disrupting the way companies are operating. Companies using digital tools increase their performance and profitability. Some years ago, everyone was speaking about "business intelligence", known as "BI". Nowadays, artificial intelligence already takes the lead on business intelligence. But what is the difference between business intelligence and artificial intelligence? Lots of people confuse these two terms thinking that it is roughly the same meanwhile these two terms are everything but similar. (TopTal Research, 2018)

When speaking of "business intelligence", it refers to collecting, processing, analyzing and reporting business data. It gives the opportunity to companies to increase the quality of the collected data and present it in a user-friendly way. As a consequence, companies will be able to improve their decision-takings because of a real-time image of the company past and current situation. In fact, BI tools should be seen as an answer to the question "what was the situation of my company and what is it now". (K. Pratt, 2017) The magic of BI is that it turns unstructured data into relevant and structured information in order to give a clear picture of where the company was and where it stands today. Nevertheless, BI tools aren't explaining how to use the displayed information in order to predict what is coming in the future and how take the best decision possible. In other words, BI should be seen a virtual dashboard where employees of a company can create graphs, charts, performance metrics by organizing the data in the way they want in order to fasten and improve the decision-making process. (J. Brar, personal communication, November, 2018) These type of tools sold on the market by Microsoft, Oracle and Tableau have known a real success in the business world because of the value it adds in terms of decision-making and management performance. In the last three years, the adoption of BI tools in businesses has increased by 50%. (TopTal Research, 2018)

Hereunder, a figure shows the benefits from integrating a BI tool into a business. It's a whole process beginning with an easy access and sharing of the data. This enables managers to know what is happening in the company in real time. They can reduce the risks of output constraints by identifying faster the bottleneck. In a nutshell, managers better know the company and the environment in which it operates. Consequently, managers can improve their decision takings. (Roberson, 2014)


*Figure 14: What does business intelligence provide to companies (Roberson, 2014)*

When making a comparison between BI and AI, we can notice that they share the same goal which is to improve decision making process. However, the way they add value is different. On one hand, we have a BI tool that enables to collect, process, analyze and report data in order to improve decision-taking and operational efficiency. On the other hand, we have algorithms that replicate human-decision making by learning autonomously from data provided to it. One of the most important development in the AI area has been using predictive analytics on very large datasets to get insight from it and predict future results. Companies operate more and more into a competitive environment where huge quantities of data could be collected and ready to be used. Machine learning models are going through an iterative process where it processes the data in order to detect patterns from it. It will be able to predict future results such as products revenues or costs. Predictive models should even be able to provide recommendations on how to achieve the predicted results. Predictive machine learning algorithms aren't explicitly programmed in advanced by humans, but will instead be trained and improved with quality data provided to it. The added value of predictive machine learning models is also that it can adapt itself to new input data. This will improve the efficiency of the model over time. (McKinsey & Company, 2019)

There are 3 types of analytics which are descriptive, predictive and prescriptive. They are classified in their complexity order with description being the less complex analytic and prescriptive the most complex one. BI is typically descriptive analytics. The focus of machine learning is predictive and prescriptive analytics. Hereunder, a figure points out the differences between these 3 types of analytics and which companies are using it.

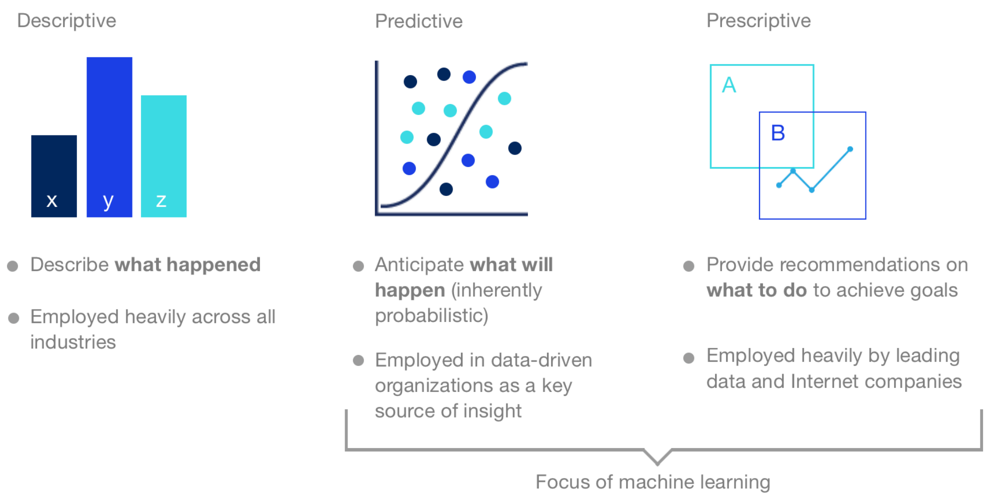


*Figure 15: The different type of analytics (McKinsey & Company, 2019)*

The author strongly believes that these two technological tools could be very complementary. It could create a strong team together because predictive analytics requires a lot of reliable data to learn and predict accurately future results. Using BI could be helpful to collect unstructured data and transform it into structured information to provide to the predictive algorithms. Artificial intelligence, but more precisely predictive analytics could be the key to use intelligently the data from the BI tool and unlock the power of AI for prediction matters. Companies shouldn't look at business intelligence and predictive analytics as completely separate technologies but rather invest in a way to fully take advantage of the potential of both technological tools. The use of these two technological tools will improve for sure the overall decision-making process of companies.

### 2.3.4 The on-going challenges of using predictive analytics

Predictive analytics has enormous capabilities and will without doubt disrupt the market in a few years if it is not already doing it. We can already notice some predictive analytics interventions in our daily life but this will even increase in the future. Nevertheless, there are nowadays still challenges facing the use of predictive analytics.

The major limitation of predictive analytics is the fact that it learns from the data it receives. One's should bear in mind that predictive algorithms aren't explicitly programmed by humans at the beginning. The knowledge of these algorithms increases through time by learning from the data provided to it. As a consequence, it obviously means that if predictive algorithms receive inaccurate data, it will learn from biased data and it will impact the forecasted results of the algorithms. This is the reason why data scientists and data engineers spend most of their time to pre-process the data. It is a very important step to make sure that the algorithms aren't biased as from the beginning of their learning process. (Cote, 2019) There is a well-known quote that says, "garbage in, garbage out". (T. Larose & D. Larose, 2015) It highlights the fact that the predicted results of the algorithm strongly depend on the quality of the data given to it.

The second challenge that needs to be highlighted is that predictive models are trained to perform a specific task, which means that predictive algorithms are specialized for some task solving and haven't the capacity to adapt to other given-tasks. If for example a predictive machine learning model is built to give legal advice, it won't have the capacity to forecast future revenue and costs of a company or predict the time when a machine needs a maintenance. So, once the system is accurate for a determined task, it won't be able to perform another prediction task. The iterative process of training a model will need to be done again with new data. (SAS, 2019)

Last but not least, predictive models aren't totally autonomous. They are only to some extent. In the learning process for example, they don't need human intervention but they still need the intervention of humans to define the right machine learning question, to feed them with reliable data in order to be able to learn quickly and efficiently, to build the predictive machine learning model, to deploy it and to monitor its accuracy once it has been deployed. (A. Cosse, personal communication, Mars, 2019) In fact, situations in movie where AI technologies are totally replacing the human is still a science fiction. This isn't reality! However, algorithms learning to perform human-tasks by processing very complex data becomes quite common.

## 2.4 Using artificial intelligence for business cases

Companies can use AI for several business cases such as improving medical diagnoses, automate low value tasks, better understand retail customers, communicate with clients and even forecast financial figures. AI applications within companies fall into one of the three following buckets: process automation, cognitive insight or cognitive engagement. (Davenport & Ronanki, 2018)

Process automation is the most common AI application for the moment and it refers to the replacement of human back-office and administrative functions. (Davenport & Ronanki, 2018) At UCB, they did implement process automation to handle back-office tasks such as incoming invoices treatment. (Lieutenant, 2019) Besides, we also have cognitive insight AI applications which is seen as analytics brought to a higher level. It's more complex than process automation because it means that AI algorithms learn from the data, extract knowledge and become better over time. This type of AI algorithms could predict future revenue, costs, margins, customer demand... Using this type of AI application could be very useful to improve companies' financial forecasting process. Last but not least, there is also "cognitive engagement" applications where AI algorithms are directly in contact with customers or employees. It's the case when companies use chatbots to communicate and answers to clients' questions on their website. (Davenport & Ronanki, 2018)

Hereunder, a pie chart reflects the current adoption rate of these 3 AI applications into companies.

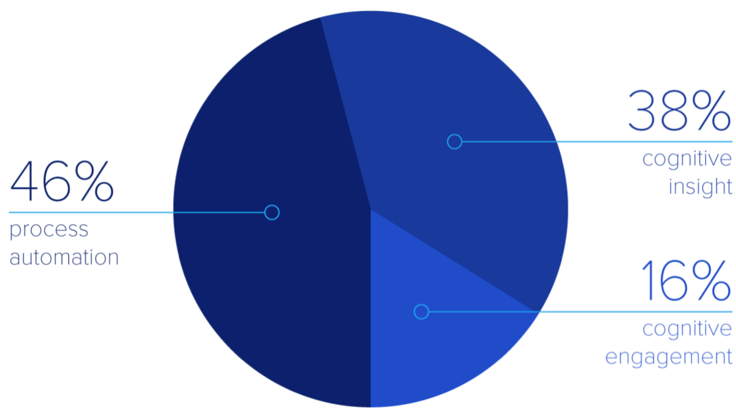


*Figure 16: The three type of AI applications (Davenport & Ronanki, 2018)*

Cognitive insight is backed by machine learning, sub-technology within AI which enables to discover patterns within the data and predict future results. Machine learning algorithms can answer to different prediction tasks such as classification, regression, clustering and anomaly detection. (Microsoft, 2019) This lead us to the following question: "which type of machine learning algorithms is required to forecast financial figures?"

## 2.5 Defining the type of machine learning algorithms required to forecast

There are three type of machine learning algorithms and each of them have different purposes. One type of algorithm cannot perform all tasks and solve all machine learning problems. So, it's important to first define the purpose of the algorithm in order to choose the right type of machine learning algorithms. There are three type of machine learning algorithms which are supervised, unsupervised and reinforced machine learning algorithms. (Malik, Machine learning algorithms comparison, 2018) Hereunder, a graph shows the purpose of each type of machine learning algorithms. This will bring more clearance about which type of machine learning algorithms to use in order to forecast financial figures.

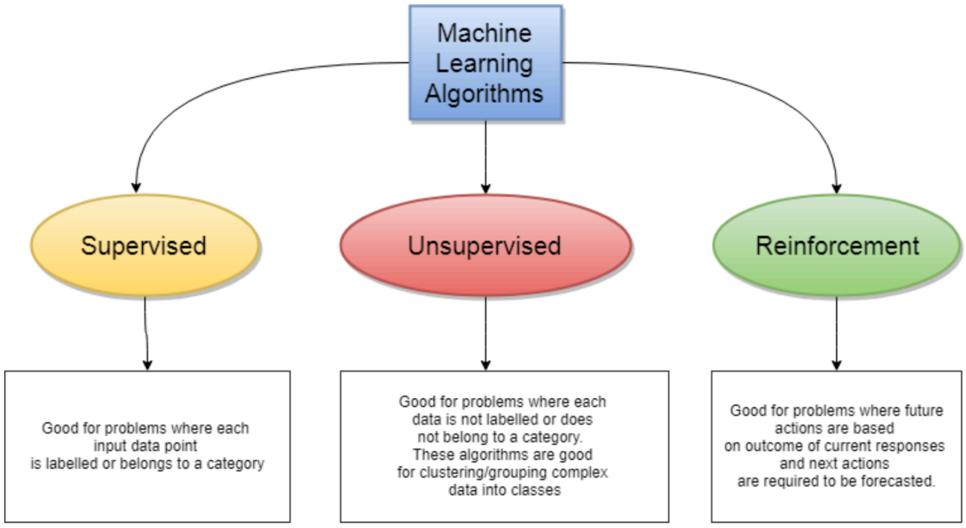


*Figure 17: The different type of machine learning algorithms (Malik, Machine learning algorithms comparison, 2018)*

A supervised machine learning algorithm is trained with given input and output data. Based on the input features and the targeted output value, the algorithm will look for patterns within the data. Moreover, the data from the dataset that is provided to the algorithm is labeled. It means that the representation of each row or column is known. This type of algorithm is used for classification or regression tasks. So, in the case of financial forecasting, companies do predict continuous values which is a regression task. The required type of machine learning algorithms would be the supervised machine learning algorithms. This type of algorithms would be trained with training data where the input and output data are given. It will give the means to the algorithm to learn and discover patterns from the dataset. Moreover, supervised machine learning models leverage data mining techniques in order to discover those patterns within the dataset and predict future results. (T. Larose & D. Larose, 2015)

## 2.6 A different financial forecasting approach with a machine learning model

With a machine learning model, the adopted methodology to predict future results isn't the same. Both traditional and machine learning methods are used to predict future results but both methods aren't proceeding the same way. The unknown isn't at the same place.

With the classical approach, we have input data that is provided to a statistical model, which is explicitly programmed by humans in order to forecast the required output. In this case, humans understand the model and its interpretation of reality (Vandeput, 2018). Nevertheless, the model cannot adapt itself to new data. Moreover, when a company tries to forecast revenues of a single product, the model will only take into account that specific product and not the numerous products from the dataset. The drawback of it is that the revenue of a certain product could impact another product revenue from the same company. (Vandeput, 2018) Below, a representation of the classical forecasting approach. (Nikkhah, 2018)


*Figure 18: Traditional forecasting approach (Nikkhah, 2018)*

With the emergence of machine learning for forecasting purposes, there is a model that learns on its own by finding out patterns within the data (connections made between the input data and the output data). The input and output data is provided by humans. The issue is that it's often difficult to understand the patterns found by the algorithm within the data. This is the reason why predictive machine learning models are currently considered as a black-box. (Lieutenant, 2019) The connections made are sometimes going behind human understanding. As a consequence, the relationship between the input and output data is unclear and non-transparent. The advantage is that the model has proven to adapt itself to the data provided to it. Machine learning models take into account the whole dataset. As for example, the model will take into consideration the impact between the revenue/sales of different products from the same company. (Vandeput, 2018) Hereunder, a representation of the machine learning approach to forecast future financial results. (Nikkhah, 2018)

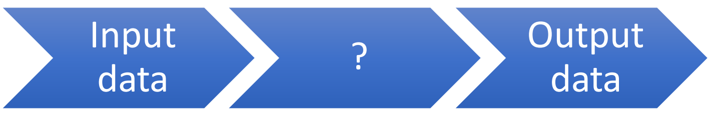


*Figure 19: Machine learning forecasting approach (Nikkhah, 2018)*

For financial forecasting purposes, companies are going to provide as input data some revenue drivers also called predictor variables or features. For a shop which sells ice creams, these features could be the location, the number of waiters, number of parking places, the number of ice cream varieties, the weather… All these features will potentially impact the revenue of the ice cream shop. Companies could also use external factors that could impact the products revenues. These external factors could be the "households' disposable income per capita, social allowance distribution, consumer confidence, household indebtedness, households savings rate, households' financial net worth, unemployment rate". (OECD, 2019) Moreover, the corresponding output will also be given to the predictive model. For the example of an ice cream shop, the output could be the revenue of the ice cream sales. During the training phase, both the input and output data given to the model is coming from historical data gathered from internal or external sources.

Let's take a closer look at the different type of machine learning solutions on the market that companies could adopt to get the means to create their own predictive machine learning model to forecast their financial figures.

## 2.7 Analysis of the required business model to build a predictive model

In the first place, people could think about making an analysis either a company could implement predictive analytics or not based on their size. However, from the author point of view, it makes no sense to look at the size of a company in order to determine if it could implement or not predictive analytics for financial forecasting purposes. It could happen that small companies are in a better position to implement predictive machine learning models to forecast their financial figures than large companies' due to the specifics of their business model. In order to determine if a company would be a good fit to implement such a technology into its financial forecasting process, people should take a closer look at 4 important questions.

These four questions are the following one's:

- Does the company possess historical data and is it a transaction intensive business?
- Is the company a B2B or a B2C company?
- Does the company already undergo its digital transformation?
- Does the company has enough money and time available to build a predictive machine learning model?

All these questions will enable companies to understand and assess their potential fit to implement a predictive machine learning model into their financial forecasting process.

Possessing huge amounts of historical data and being a transaction intensive business is key if companies want to implement a predictive machine learning model. The reason is very simple, in order to train and test a machine learning model, companies need to gather a lot of historical data which could come from internal and external data sources. The more transactions, the more data available to the company to train and test their machine learning model. (Cote, 2019)(F. Manseau, personal communication, Mars, 2019)

Let's take the example of Uber, it's a service company that operates for 10 years in the passenger transport sector and already counts 5 billion trips worldwide. Each day, this amount increases by 15 million. (Iqbal, 2019) Uber is typically a company that could be defined as a transaction intensive business. All these transactions are recorded via the Uber application. Uber can use the historical transactions to train their predictive machine learning model. On the contrary, a company that exists for 1 year which sells around 10 products per month won't probably be a good fit for building a machine learning model to predict future financial figures. The reason being of course that the company won't have enough historical data available to build and train its predictive machine learning model. Moreover, companies with moderate historical data might be able to build a machine learning model. However, it might take more time to get a predictive model that will forecast the financial values with the expected accuracy. It isn't 100% sure that these companies with moderate historical data availability will be able to build a predictive model that lives up to their accuracy expectations. If this question is positively answered, meaning that the company has a lot of historical data and that it's a transaction intensive company, a second question needs to be answered. Is the business a B2B or a B2C business?

From the author point of view, both business models could build predictive machine learning models to predict future outcomes but not to answer the same question. When it's the case of a company willing to predict its future revenues or costs, companies that possess a B2C business model will be a better fit for the implementation of a machine learning model due to two main reasons.

The first reason is the fact that most B2C companies have a higher transaction level than B2B companies. Most of B2B companies will provide long-term services that will last several weeks or months, meaning that the transaction level will be lower than B2C companies. On the contrary, B2C companies will sell lots of products or services every day to their final customers. The level of transactions is supposed to be higher.

The second reason is that predicting services revenues for B2B companies is based on a lead-opportunity analysis. So, service companies will look at their leads (100% closed deal) and opportunities (between 50 and 100% chances of closing the deal) to predict their revenues of the following months. It means that a services company that currently has low leads and opportunities, already knows that the revenues of the next 3 months, also called backlog, will be low. (F. Manseau, personal communication, March, 2019) It's logic because the company revenue is a consequence of the closed deals and opportunities from previous months. There is less uncertainty in the prediction of services revenues but it still remains difficult to predict accurately future services revenues. However, B2C companies selling products cannot anticipate revenues such as these B2B companies selling services on the long-term. The reason is that it's more difficult to anticipate a customer coming into one of their shops to buy a product. In fact, it's more difficult for B2C companies selling products or services to predict revenues because of the unpredicted one-shot sales. This isn't often the case for services companies. Most of the time, services companies recognize their revenue when the service is performed. It can be split on several months following the deal.

Nevertheless, the author believes that B2B companies that provide services could use machine learning models to predict either an opportunity is going to be transformed into a lead or not. It will still be a prediction task done with supervised machine learning algorithms but the purpose and the prediction task isn't the same anymore. It will be a classification task and no more a regression task. It isn't about predicting a continuous value anymore but about classifying an opportunity in two brackets, "yes it's going to be turned into a lead" or "no it's not going to be turned into a lead". (Coallier, 2019)

To cut it short, both business models could build a predictive machine learning model but for different purposes. In the case of financial forecasting, it would be a fit for B2C companies selling products or services. B2B companies which sell long-term services should adopt machine learning models to predict either an opportunity is going to be turned into a lead or not.

The third question that companies need to ask themselves is if they already undergo a data and digital transformation in order to capture the required data. It's essential for a company to first go through a digital transformation phase before even thinking about implementing predictive analytics to forecast its financial figures. Without a digital transformation that enables the company to capture the required historical data, the company won't be able to fully take advantage of predictive analytics within its company. Those typical required IT tools to collect data about the historical transactions are for example an adapted enterprise resource planning system, a website, social networks, a mobile application... All these tools give companies the means to collect data about the customer or the transaction which could be useful to better understand what happened in the past and where the company stands today. Digital transformation is essential in a company to improve its performance and productivity. It's a must before even thinking about implementing a machine learning model to forecast financial figures (Bughin, Chui, & Mccarthy, 2017)

Last but not least, companies should also bear in mind that building a machine learning model is expensive and requires a lot of time. However, once the predictive machine learning model is build and deployed, companies will be able to benefit from a faster and more accurate financial forecasting process. Companies will get real-time financial forecasted results based on an automated workflow from the collection of the data to the forecasting of the financial results. (Cote, 2019)

Below, the table describes the fit of companies to implement a predictive machine learning model to forecast their revenues, costs, margins... This compatibility analysis is based on four criteria's which are the historical data & transaction level, the type of business, the presence of a digital transformation in the company and the time & money availability for the project of building a predictive machine learning model.

| | Best fit | Possible fit | No fit |
|---|---|---|---|
| Historical data & transaction level | High | Medium | Low |
| B2B vs B2C | B2C | B2C | B2B |
| Digital transformation | Yes | Yes | No |
| Time and money | Yes | Yes | No |

*Figure 20: Conditions required for machine learning implementation*

The author strongly believes that companies need to analyse the fulfilment of these four criteria's in order to determine if they are a good fit to use a machine learning model to forecast their financial figures. He also thinks that companies which are in the best position to take advantage of machine learning models for financial forecasting are companies that have huge amounts of historical data & huge amounts of transactions every day. Those companies are also B2C businesses and already have achieved a digital transformation (adequate cloud ERP, big data, BI tools, website, mobile application…). The fourth element that those companies fulfil is the time and money required because implementing predictive analytics and monitor it could be expensive and take some time. If these four criteria are met, then companies should think about adopting machine learning models within their financial forecasting process.

## 2.8 Machine learning solutions available on the market

Using predictive machine learning models for financial forecasting purposes has proven to possess a bright future because of its capabilities to find patterns within the data and as a result provide a data-driven prediction of future financial results. Nowadays, technological firms are already selling some machine learning solutions that companies could leverage in order to forecast financial results.

If companies want to build their own machine learning model to forecast their financial figures, there are two types of machine learning solutions. There is one type of solution which is a platform to build customized models and there are also semi-automated solutions which provide some pre-build algorithms. Both have as aim to build a candidate model to predict future results but these two types of solutions are quite different. (Coallier, 2019)

The first type of solution is a platform where data scientists can code with python, java script, R in order to build their own predictive algorithms. It's a kind of book with white pages where data scientists will build their algorithms by writing some codes coming from a coding library. These algorithms are needed to build and train the candidate model. By using this type of solution, it brings more flexibility and enable data scientists to build more customized algorithms for their prediction task. (Altexsoft, 2018)

There are different solutions of this type on the market such as:

- Amazon Sage Maker (Amazon, 2019)
- Azure ML Services (Microsoft, 2018)
- Google Cloud ML Engine (Google, 2019)
- IBM Watson Studio, which trying to make the transition to semi-automated solutions (IBM, 2019) (Masnaoui, 2019)

From the author point of view, this type of solution should be adopted by companies which are willing to build some tailor-made algorithms that will be at the basis of their predictive model. In fact, this type of solutions is perfect for companies seeking flexibility in the building and deployment of their algorithms. Moreover, companies are often willing to create their own algorithms because if they create a better one than its competitors, it could be a potential competitive advantage. (Masnaoui, 2019) However, companies should bear in mind that adopting this type of solution will lengthen the process of building a predictive machine learning model. The reason is that there are no pre-build algorithms such as it would be the case with semi-automated solutions.

The other type of machine learning solution is a semi-automated solution which provide companies with a package of pre-build algorithms. The whole process, workflow can easily be visualized thanks to a graphical interface. The whole process is aimed to be user friendly with a graphical drag and drop interface. This type of solution should be accessible to people willing to build their predictive model without having pure coding skills. Nevertheless, companies will get less or no flexibility to build their own algorithms because the solutions already includes hundreds of pre-build algorithms that only can be configured to some extent. (Altexsoft, 2018)

Today, there are mainly six semi-automated solutions on the market that could be used for financial forecasting purposes. Hereunder, a list of these semi-automated solutions:

- Azure ML Studio (Microsoft, 2019)
- Google Cloud AutoML (Google, 2019)
- IBM Watson ML Model Builder (IBM, 2019)
- Amazon ML Frameworks (Amazon, 2019)
- SAP Leonardo ML (SAP, 2019)
- SAS Visual Data Mining and Machine learning (SAS, 2019)

Most of these solutions can be use on premise which means that companies buying the solution will run the solution on their own servers. It will require very good IT infrastructures that can support the computer power needed. If companies aren't willing to invest into powerful infrastructure, they also can adopt the solutions on the cloud. This means that the solution will run on the infrastructure of the service provider which could be Microsoft, Google, IBM, Amazon, SAS or SAP.

Moreover, the second type of machine learning solutions provides companies with pre-processing modules to clean and transform the data. It also incorporates a set of machine learning algorithms that will be at the basis of their predictive machine learning model. The predictive model will be trained and tested with a prepared dataset with an aim to get a candidate model accurate enough to predict financial figures. Most of these solutions also incorporate API's to give applications access to the machine learning model that has been developed. (Chappell)

Once the model is considered as accurate enough, it will be deployed in order to use it for real financial forecasting purposes. Nevertheless, everyone should bear in mind that both solutions only provide companies with the means to create an accurate predictive machine learning model. Buying these machine learning solutions won't give companies the opportunity to directly forecast their financial figures. The implication of data scientists and the use of quality data remains primordial to build an accurate candidate model. (Ait Amir & El Mahrsi)

Below, a figure shows the intervention of Azure machine learning within the building process of a predictive machine learning model.

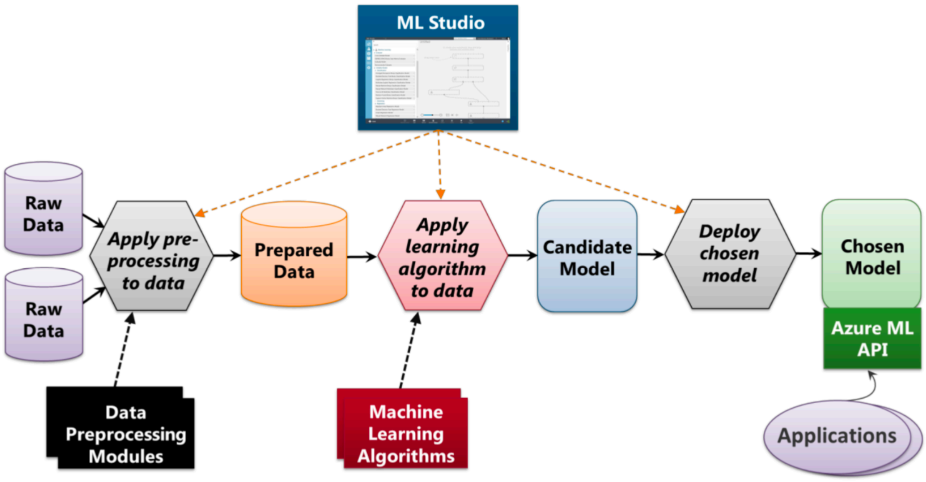


*Figure 21: ML Studio support in the implementation process (Chappell)*

Some people tend to forget that implementing machine learning models to forecast financial figures isn't just about buying a solution with pre-build machine learning algorithms that will forecast whatever financial figure you want. Going through a data science project is heavy and time consuming. Companies need to follow numerous steps before getting a predictive model accurate enough to forecast their financial figures such as revenues or costs.

When considering the data collection and data preparation phases of the implementation of a machine learning model, it is estimated at 80% of the whole implementation process.

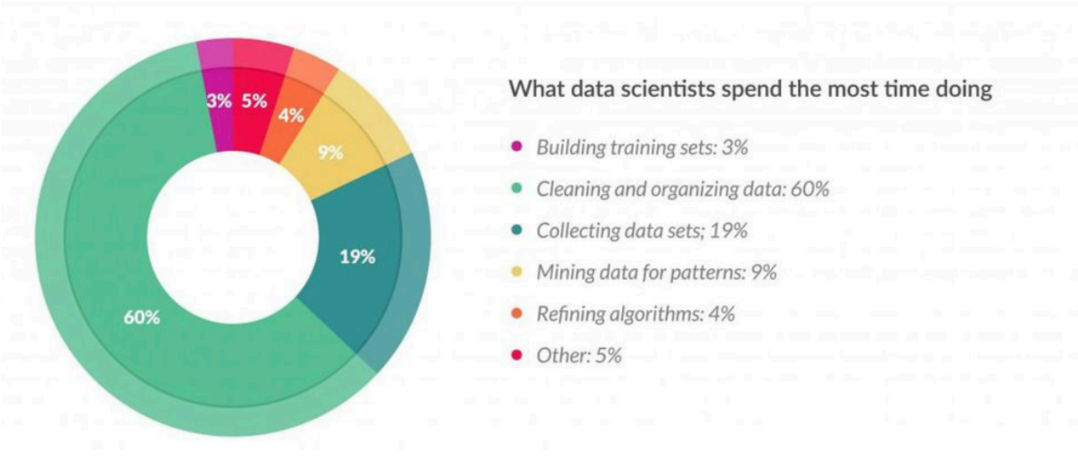


*Figure 22: What data scientists spend the most time doing (Press, 2016)*

Nowadays, people often hear the word "data scientist". This new job appeared in the 21 centuries and becomes more and more important with the emergence of huge amounts of data. Data is seen as an asset for companies because it possesses very useful information within it. That data need to be analyzed and structured in order to provide insight and drive profit. (SAS, 2019) A data scientist is a person that supports companies during the whole process of collecting data, cleaning and transforming the data, train, evaluate and deploy the predictive machine learning model. As highlighted in the figure above, one of his most time-consuming tasks is to collect and clean the data. He needs to make sure that the data provided to the machine learning model is of high quality. So, data scientists who are pre-processing the data have a direct impact on the learning process of the predictive model. The efficiency of the predictive model to discover patterns and forecast future financial figures is strongly impacted by the quality of the data. (Malik, Processing data to improve learning models accuracy, 2018)

The data scientist has multiple roles all along the machine learning process. His several roles can be listed such as: (SAS, 2019)

1. Collecting large amount of unstructured data and transform it into structured and qualitative data (Masnaoui, 2019)
2. Using data mining techniques to solve concrete business matters
3. Working with multiple coding programmes such as Python, R
4. Excellent knowledge of statistics and mathematics
5. Working with analytical techniques such as machine learning, deep learning
6. Being able to communicate with both IT and business
7. Looking for patterns and trends in the data

## 2.9 Conclusion

Artificial intelligence is about machines being able to replicate human thinking and perform human tasks. We saw that the underlying technologies of artificial intelligence that are used for prediction tasks are “predictive analytics” and “machine learning”. These machine learning algorithms are trained with historical data in order to discover patterns. These algorithms will use and apply what they have learned to new input data in order to predict future financial figures.

There are some advantages for companies to use predictive machine learning algorithms into their financial forecasting process. Firstly, companies will benefit from an automated financial forecasting process which will reflect the company’s reality. No more human or organization bias. Secondly, predictive machine learning algorithms will complete and bring more intelligence to business intelligence tools. With BI tools, companies can answer to the question “what happened and what is happening”. With predictive analytics, companies will be able in most cases to forecast accurately revenues, costs, margins… Secondly, machine learning algorithms can handle huge quantities of data and will extract the most out of it. Machine learning algorithms are going to look after patterns within the historical data and learn by themselves progressively in order to become better by time. Within the data relies tremendous and useful information that predictive machine learning algorithms can leverage. This is why companies using machine learning algorithms to forecast their financial figures will benefit from a more accurate financial forecasting than the one's not using it.

However, there are also drawbacks when using predictive analytics. One of them is the cost and time required to build and run a predictive machine learning model. The second drawback, which is also a limit of predictive analytics, is the fact that companies need huge amounts of historical data in order to train the predictive model. Without historical data, the model won't be able to be trained and to predict accurately. The third drawback would be the understanding of the "black-box". The algorithms will forecast revenues and costs but won't explain how it came to those results. It will sometimes discover patterns and relationships that humans cannot always understand. The fourth disadvantage is that companies will lose some power over their business decisions because they will, for example, consider a revenue or a cost predicted by the algorithm. Human intervention decreases and IT intervention increases.

There are some limits when it comes to using machine learning algorithms for financial forecasting. The first one is the huge amounts of data required to train and test the predictive model. Moreover, humans have an important impact on the learning process of the model because the model learns from the data provided to it. If humans provide wrong data to the model, it will learn from erroneous data and will predict wrong financial figures. The second limit is that the predictive model is built to perform a specific prediction task, once build its impossible for the model to predict something else than what is has been configured for. The third limit is that predictive models aren't totally autonomous. People will always be needed to collect the data, pre-process the data, build the predictive model and deploy it. Even after the implementation of the predictive model, people will always need to assess the accuracy of the model over time and retrain it with new data.

When training a predictive model, companies should use supervised machine learning. This means that the data provided to the model is labeled and known. Moreover, this also means that the predictive model will receive the input data and the output data in order to learn. The model sees itself given the input data and the output data and needs to find a pattern that will enable it to understand the relationship between the input and output data. As a consequence, the financial forecasting approach changes when using a predictive machine learning model. The question mark isn't the predicted result anymore but the predictive model itself because there is a real lack of transparency from the algorithm predictions. In the case of financial forecasting, the model will receive as input data variables that are impacting the revenue (i.e. the weather, the type of product, the price, the number of waiters) and will be asked to give as output the forecasted revenue. The same could happen for cost prediction.

From the author point of view, companies that are in the best position to implement machine learning algorithms within their financial forecasting process are B2C companies selling products or services on short-term periods. These companies also have a transaction intensive business with huge amounts of historical data in order to train and test their predictive machine learning model. These companies already undergo a data & digital transformation and as consequence do possess the required IT tools (such as cloud ERP, big data infrastructure, BI tools, websites, mobile applications) to capture the data. Last but not least, these companies possess money and time to build their machine learning model.

Two types of solutions available on the market have been pointed out. The first type of solution enables companies to build their predictive machine learning model with their own custom-build algorithms. The second type of solution is a semi-automated solution where pre-build algorithms are already available within the package in order to pre-process the data, build and deploy their predictive machine learning model.

However, companies buying these solutions won't be able to directly use the solution within their financial forecasting process. They'll first need to collect and clean the data, train and test their potential predictive machine learning model with their historical data. The next chapter will demonstrate the process a company need to go through in order to build and deploy their predictive model for financial forecasting purposes.

# 3. How to implement a predictive machine learning model

## *3.1 Introduction*

Nowadays, there are machine learning solutions on the market that enable companies to leverage their data in order to predict future results by discovering patterns within the data. Nevertheless, it isn't just about buying a solution that will find insight into the data and predict autonomously without the intervention of humans. In this chapter, the steps in order to implement a predictive model will be pointed out and explained with some examples. Humans still need to be involved all along the data science project to supervise and define the workflow during the implementation. There is a well-known process that companies should follow to correctly use machine learning algorithms to predict future outcomes: this process is called the CRISP-DM, which stands for "cross-industry standard process for data mining". This process is neutral and could be used for any data mining project. It gives a clear direction to companies and points out the important steps to follow in order to build a machine learning model for financial forecasting purposes. (T. Larose & D. Larose, 2015)

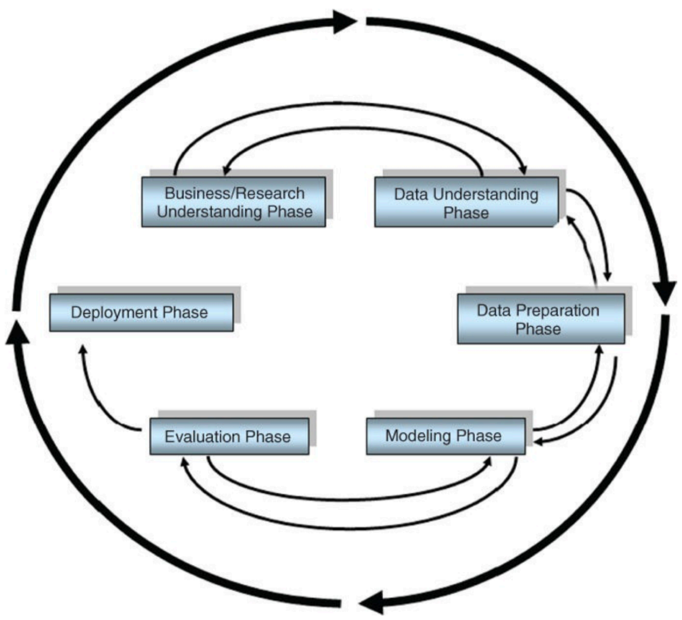


*Figure 23: The CRISP-DM process (T. Larose & D. Larose, 2015)*

Above, the figure shows the 6 steps process that companies need to carry on in order to build and use a machine learning model to forecast. This process is seen as an adaptive process because all the steps impact each other. A modification into previous steps requires to adapt the following steps. An example would be a company that trains its model in the 4$^{th}$ phase and then suddenly realizes that the data provided to the model isn't good enough because of a lack of predictor variables or data errors. This implies that the company will need to prepare again the data and change or rebuild a model afterwards. Moreover, the process is also seen as an iterative process because whatever data mining problem a company faces, it will always follow this 6-step process. (T. Larose & D. Larose, 2015)

Hereunder, the figure shows more specifically the machine learning process that companies need to go through in order to build a machine learning model to forecast future financial figures. To do so, companies need to use one of the machine learning solutions specified earlier.

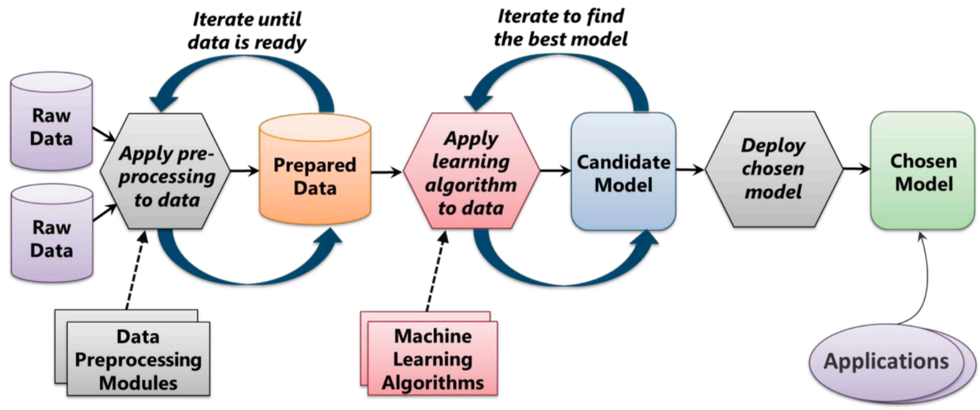


*Figure 24: The implementation process (Chappell)*

## *3.2 Business/research understanding phase*

### 3.2.1 Defining the objective, task to perform

When companies are willing to lead a project within the company, it's always essential to figure out and define what are the objectives of the project. Understanding the current situation and knowing what the company wants to achieve is a very important step. A lot of people tend to forget to spend time on this step.

When the author did his internship, he took part to the implementation of a new financial ERP system for the company. It requires in the first place to determine what the problem was with the current financial ERP system. What for requirements should the new financial ERP answer to? This analysis was done in order to point out the objectives of the project and the purpose of adopting a new ERP system. This business understanding phase could also be compared with the moment where students are asking themselves the objectives they want to achieve through their thesis. When building a predictive machine learning model, there are different algorithms that can be used. This implies that companies first need to determine what they want to find out. (Microsoft) In function of the output required, the company will need to decide which type of algorithm will be used to train the machine learning model and to achieve what is required from the model.

Hereunder, a list with different type of data mining tasks that could be performed: (Microsoft)

- *Anomaly detection*: identify and predict unusual data points
- *Classification*: the categorical targeted variable is estimated and afterwards classified in a bucket
- *Clustering*: grouping observations, data points into classes of similar characteristics. The difference with classification is that there is no targeted value.
- ***Regression***: forecast future numerical values based on patterns recognition

One's should bear in mind that at the basis of machine learning, we have algorithms that could be used for several prediction tasks. However, when looking at the research question of the thesis, it's typically about using an algorithm for regression tasks as the aim is to predict financial figures such as revenues, costs, margins... which are numerical values.

### 3.2.2 Formulating a data mining problem

Once the data mining tasks is defined, companies also need to translate these objectives into a clear and structured data mining problem definition. This phase is very important because it defines and impacts all the steps of the process that follow. (T. Larose & D. Larose, 2015)

If we consider a regression data mining task, different question could be asked such as for example:

- Predicting companies' products or services revenues
- Predict the marketing costs
- Predict the clinical costs
- Predicting the COGS
- Predicting the products or services demand (helped indirectly to forecast revenues)

Before even starting about collecting data and building a machine learning model, it's important to define the objectives of the project and to translate them into a data mining problem statement. (T. Larose & D. Larose, 2015) A problem statement that machine learning algorithms will be asked to deal with by finding patterns within the prepared dataset given to it. In order to feed predictive models with data, it implies that business people and data scientists need to identify and collect the required data. This task isn't an easy task and takes a lot of time. The next step of the process points out the defining of the required data and the collection of it.

## *3.3 Data understanding phase*

Once the problem has been defined and that the type of data mining task has been identified, it's important to define the required data and collect it. So, the second step consists at understanding what type of data is needed and collect it in order to train the model. The first question to begin with is to define what input data is required, which are the predictor variables, also called features impacting the output. Data scientists and business people should look if it's possible to collect them or to create them based on other kind of data. Therefore, the output data should also clearly be defined in order to find the best predictor variables. A close collaboration between business people and data scientists in this step is essential as business people certainly know what are the main revenues or costs drivers. (Coallier, 2019)

Two problems could be encountered within this phase making the collection of the data difficult. The first issue could be that companies haven't enough data to collect in order to pre-process and train the model. In that case, there isn't a lot to do but to wait until the company has enough example, data to provide to the model. Earlier, it has been pointed out that historical data is a requirement in order to build a predictive model. (cf. supra p.46) The second problem that companies could encounter is that they don't have the required digital tools to capture the data and drill down within the details. (Cote, 2019)

When the author worked at Keyrus, he was working on a ERP project that would give the company the means to capture the historical data of the company. They could capture historical revenues and costs by business unit, by service, by clients… Their current ERP system is a system that enables them to keep track of their general ledger, payables, receivables but it does not give them the means to do analytics. Analytics would give the opportunity to drill down within the data and provide a clear picture of the past and present situation of the company. It's called business intelligence and it plays a role within the digital transformation of a company, which is crucial before even thinking about implementing machine learning models into a financial forecasting process. Without those digital tools that give companies the means to capture the historical and on-going data, the collection phase will be much more expensive and time consuming. It might even be impossible to execute. If the data isn't available or too aggregated companies aren't able to capture the data as they need it.

From the author point of view, it's also important that companies determine the forecasting improvement they seek and how they are going to measure it. It's often said that without being able to measure the improvement, there is no improvement. This point will be demonstrated in the fifth phase of the implementation process by explaining the two metrics mostly used to measure the forecasting accuracy improvement. (cf. infra p.76)

Besides, the time required to collect the right data in order to train the model is around 20% of the whole machine learning process.

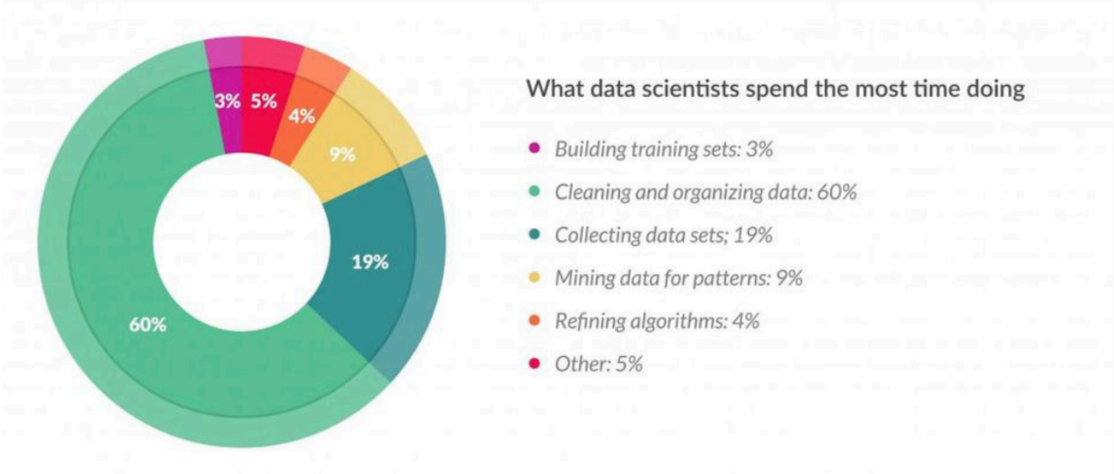


*Figure 25: What data scientists spend the most time doing (Press, 2016)*

## 3.4 Data preparation phase

### 3.4.1 Setting up a prepared dataset

Some people tend to believe that machine learning models do completely act independently from humans. The data preparation phase, commonly called "pre-processing data phase" is the most important and time consuming one. This step proves that the intervention of humans is still very important. The feeding with data of the machine learning algorithm will always be performed by humans and thus makes the intervention of humans' essential. Data scientists are the ones making sure about the relevance and quality of the data provided to the model. They do transform the raw data into a prepared data that will be provided to the model.

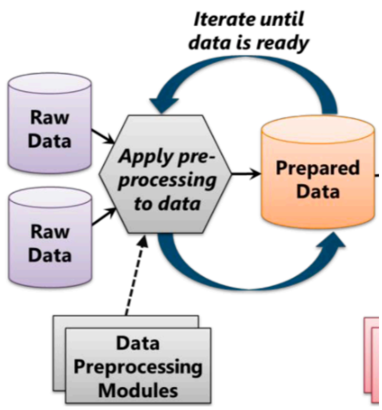


*Figure 26: Pre-processing the data (Chappell)*

From the author point of view, pre-processing data and improving the quality of the data is a key element in order to improve the accuracy of the predictive machine learning model. When data scientists are collecting raw data, it's often noisy and inaccurate because the data hasn't been look at for many years. Previous forecasts might have been done without data that nowadays are expired fields values or missing values… If the data provided to the machine learning model isn't cleaned and accurate, as will be the forecasted result from the machine learning model.

There is a well-known expression used by data scientists to put the finger on the importance of pre-processing data which is "garbage in, garbage out". (T. Larose & D. Larose, 2015) This means that there is a positive correlation between the quality of the input data provided to the model and the predicted result of the model. If the input data is garbage, the forecasted results provided by the model will also be garbage. The objective of pre-processing data is thus to minimize the garbage provided to the machine learning model that could potentially decrease the forecasted results. If the garbage going into the model is minimized, then the output provided by the model will get a better quality. There will be an overall accuracy increase of the forecasted result. (Cote, 2019)

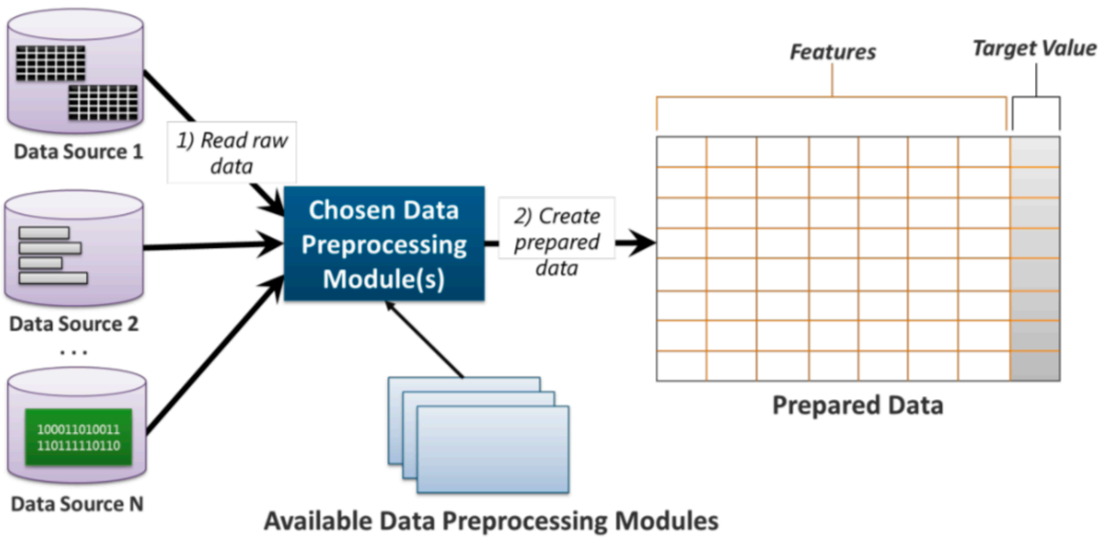


*Figure 27: Creating a prepared dataset (Chappell)*

The figure above shows that data scientists are collecting raw data and apply pre-processing modules to create a clean and prepared dataset to provide to the machine learning model. Within a prepared dataset, the features expressed in columns correspond to the input data and the target value column correspond to the output data. But what does happen when using these pre-processing modules? (Chappell)

Nowadays, data scientists are often pre-processing data that haven't been looked at for many years because their previous forecasts weren't necessarily based on that data. So, the data is often inaccurate and data scientists need to clean the dataset in order to increase the prepared dataset quality. By addressing this issue, they inevitably help the predictive model to learn from a clean and accurate dataset which will increase its forecasting accuracy.

### 3.4.2 Data pre-processing tasks

The main data cleaning and transformation tasks are performed in order to solve the following dataset issues: (Malik, Processing data to improve learning models accuracy, 2018) (Coallier, 2019) (Cote, 2019) (Masnaoui, 2019)

1. A dataset that possesses missing values.
2. A dataset that contains erroneous values.
3. A dataset that possesses some outliers.
4. A dataset that includes categorical values that need to be transformed into numerical values or reclassified in larger field values
5. A dataset where all features do not have the same value importance. This means that some features are over or under valued. The features need to be normalized in order to find an equal importance between the different features.
6. A dataset that possesses to much dimensions. Data scientists will have to reduce the number of dimensions because some of them could be strongly correlated.
7. Last but not least, split the dataset into training and testing data (this isn't an issue but a necessary step to train and test later on the predictive machine learning model).

Let's analyze the following dataset with some features that impact the revenue stream of an ice cream chain. The data within this dataset is known because its historical data (both input and output data).

| Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|
| Period of the year | Weather | Location | # of waiters | Parking places | Dogs admission | Revenue |
| Q1 | 5 | Uccle | 3 | -5 | no | **12000** |
| Q1 | 10 | Laeken | 1 | 0 | yes | **10000** |
| Q2 | 20 | Woluwe | 4 | 10 | yes | **20000** |
| Q2 | | Schaerbeek | 1 | 0 | no | **15000** |
| Q3 | 25 | Woluwe | 9 | 10 | yes | **35000** |
| Q3 | 30 | Wemmel | 3 | 4 | yes | **33000** |
| Q5 | 16 | Laeken | 2 | 0 | yes | **16000** |
| Q4 | 12 | Uccle | 3 | 7 | no | **18000** |

*Figure 28: A noisy dataset*

The idea is to identify the noise into the dataset and clean it. First, we are going to look at possible missing values. Most of the features data aren't missing except for the weather. We can notice that we are missing a value at the fourth row. It might be due to a data collection error, a blank within a survey, blank spaces… In order to clean the dataset and take away the missing value, there are 3 main solutions. The first one is to delete the sample (row) that contains the missing value. Eliminating the sample is risky because there could be relevant information relying in it. (Coallier, 2019) The second solution is to replace the missing value by the mean of the respective column. And the third solution consists at replacing the missing value by an imputed value based on the characteristics of the other features. Here, we can see that the period of the year is "quarter 2", which means that we can deduct that the weather is approximatively between "15" and "20". (T. Larose & D. Larose, 2015)

Now, we are going to look at the possible errors within the dataset. There are potentially 2 errors in the dataset. The first one concerns the period of time column and more precisely the 7th row. The mentioned period is "Q5" which isn't possible because there are only 4 quarters in a year. As a consequence, data scientists are going to try to find the right quarter elsewhere or deduct its value by comparing with the other samples. Another error that could be identified is the number of parking places in the first row. We can notice that it indicates "-5" parking place which is also impossible. The minimum parking places being "0". The error can be managed by looking at the characteristics of the row. The characteristics of the same row could give a hint for the erroneous value that need to be corrected. In this case, we can think that the number of parking places was "5" but that there was a coding error where a minus has been added unintentionally. Finding and correct erroneous values isn't easy because the distinction between an error and an outlier could be confusing. To clarify, outliers are data points that aren't following the trend but that aren't necessary an error. (Masnaoui, 2019) It literally signifies a point that is out of line. But in this financial dataset, the time period "Q5" and the number of parking places "-5" should clearly be considered as an error and not outliers because it isn't possible at all. (T. Larose & D. Larose, 2015)

Concerning the identification of numerical outliers, the technique mostly used is to create a histogram for each predictor variable/feature in order to get a clear picture of the samples and identify the potential outliers. (Vandeput, 2018)

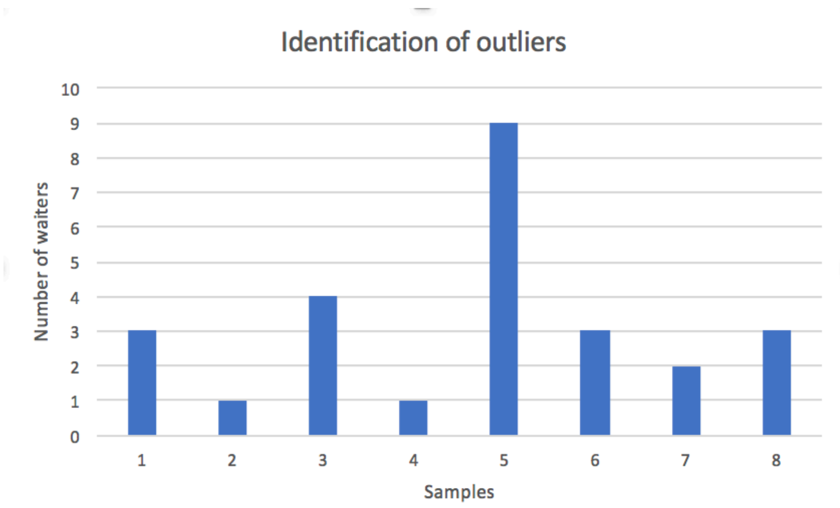


*Figure 29: Identification of outliers*

Within the graph above, we can clearly notice that the fifth sample isn't following the trend of the other five samples. Sometimes the outliers can be a data entry error and sometimes they exist because of a certain reason. It's important to identify the nature of the outlier because it will impact the accuracy of the predictive machine learning model. The decision to take away an outlier isn't easy. However, when there is no clear reason behind, the outliers are taken away in order to not affect the accuracy of the forecasted result. (T. Larose & D. Larose, 2015)

Within the dataset, we can notice that one of the features represents the location. This is a categorical value (no numbers). Machine learning models are very bad at finding patterns and predicting future numerical value when there are features with lots of categorical values. This is why the categorical value need to be reclassified into larger field values. In this case for example, we could classify the multiple locations into Flanders and Brussels. By doing this, it will improve the learning process and the accuracy of the predictive model. (T. Larose & D. Larose, 2015) Another method often used by data scientists is to transform each categorical value into a feature and then indicate within each row a "0" or a "1" either the new feature is respected or not. (Coallier, 2019) (Roman , 2018)

Another very important element in the pre-processing phase of the data is to normalize the features which means that the same scale is going to be applied to all features. This will increase the accuracy of the machine learning algorithm as it performs better with features with the same scale. When doing normalization, all the features are going to be rescaled to a range between (0,1). The min-max method is going to be applied to all the values of each feature column. (Roman , 2018) So, all the values of the dataset are going to be ranged between "0" and "1". The formula used for the normalization is the following:

$$Xchanged = \frac{X - Xmin}{Xmax - Xmin}$$

If we take the example of the normalization of the weather, the 6th sample will become 1 as

$$\frac{30 - 5}{30 - 5}$$

The 1st sample will become “0” as

$$\frac{5 - 5}{30 - 5}$$

All the other sample will be range between “0” and “1”. In the weather feature, the 8$^{th}$ sample will become “0,28” as

$$\frac{12 - 5}{30 - 5}$$

Last but not least, data scientists will need to extract features which means that they will have to remove existing features to avoid over-fitting. Over-fitting happens when the model is to precise within the training dataset but won’t be able to draw up patterns within new data. To avoid over-fitting, data scientists increase the training data and decrease the features. A correlation matrix could be done to discover the correlation between the features and take some of them away because of their similar impact on the output. Additionally, the least important features can be excluded as their impact on the output variable is minimal. (Malik, Processing data to improve learning models accuracy, 2018) In addition, it’s also possible to create a new feature that would be more significant by merging two or more features that are strongly correlated. (Coallier, 2019) (Roman , 2018)

From the author point of view, it’s very important to train as much as possible machine learning models with huge amounts of data in order to increase its accuracy. However, it’s even more important to make sure that the dataset given to the predictive machine learning model is of high quality. The reason is that machine learning models trained with poor quality and erroneous data will forecast poor results. People should always remember “Garbage in, garbage out”. (T. Larose & D. Larose, 2015) The word “data science” takes on its full meaning within this phase. Collecting, analyzing, pre-processing data is crucial and it’s an essential part of the data science project. These pre-processing modules enable companies to take the most out of the data in order to increase the efficiency of machine learning models and to drive future profits for companies. This pre-processing phase within the machine learning process counts in average for 60%. (Press, 2016)

Once all these tasks to improve the dataset quality have been performed, it's primordial to split the dataset in two categories. There will be a training dataset and a testing dataset. (Coallier, 2019) (Cote, 2019)

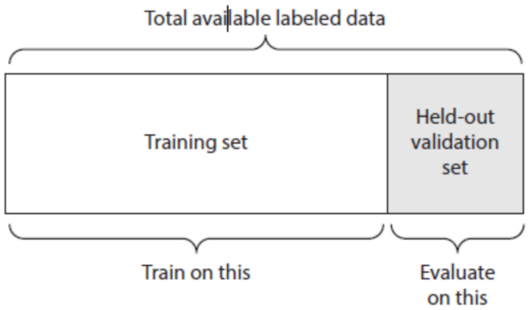


*Figure 30: Splitting the prepared dataset (Roman , 2018)*

One's should bear in mind that the labeled data is known, which implies that the input variables and the corresponding output are known. The difference between both datasets is their purpose. Their respective names already strongly point out their purposes. The training dataset is going to be used to train the machine learning model in order to make it learn and find patterns within the data (both input and output are given to the model). Meanwhile the testing dataset is going to be used to test the prediction accuracy of the model. Input variables will be given but the targeted value won't be provided to the machine learning model.

## *3.5 Modelling phase*

### 3.5.1 The pipeline of a machine learning model

The data has been collected and pre-processed in order to provide the machine learning model with an accurate dataset. In this phase, data scientists are going to train different machine learning algorithms to build a predictive model. Below, the picture gives a clear overview of the standard pipeline of building a predictive machine learning model. Data preparation and feature extraction has already been done. This phase typically concerns the "model building" and "model training" wheels. The selection of the machine learning algorithm and the feeding of the model with the prepared data (model training) will enable companies to build a machine learning model to forecast financial figures such as revenues and costs. Afterwards, the predictive machine learning model will be tested with testing data in order to measure its forecasting accuracy.

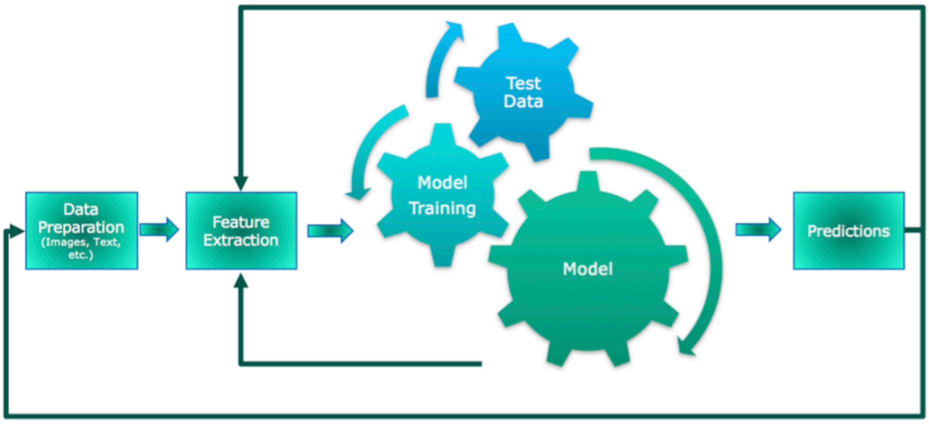


*Figure 31: Modelling pipeline (Zhou, 2018)*

In our case, the model will be trained with the supervised machine learning method. The modelling phase will give the means and will condition the predictive model to find patterns within the data and predict future results. The model will be given some input data and asked to predict a determined output such as a revenue, a cost…

### 3.5.2 Supervised machine learning for regression tasks

Training a model is one of the most exciting steps of the whole process. To do so, humans will provide the model with the labelled input data and corresponding output data from the training dataset. The whole training dataset provided to the model is known which means that we know what the different input features and what the predicted numerical value represent. Moreover, every data value from the dataset is labelled to a feature. Let's take the example of predicting the house prices, the labelled input data will be features such as the number of rooms, the location, the square meters, the number of facades, the number of parking places… The different input variables will be linked with the corresponding output data. Once well trained, the model should be able to predict the targeted value. This is how supervised machine learning works. The model see itself given different independent variables (the input data) and given the corresponding dependent variable (numerical value that we want to predict). (McKinsey & Company, 2019)

During the training phase, as the model will see itself provided with input and the output data, it will try to discover patterns within the data and find out possible relationships between both input and output data. It could be represented with the following mathematical formula: (Brownlee, How machine learning algorithms work (they learn a mapping of input to output), 2016)

$$Y = f(x)$$

- "Y" is the dependent variable that the model is asked to predict (output data).
- "X" are the different independent predictor variable, features that impact the predicted result (input data).
- "F ()" is the targeted function that the model needs to learn in order to best map input features with the output variable.

In fact, the model will look for patterns, trends that enables it in the future to predict the output variable "Y" (numerical value), once it's given new input variables (X).

The training phase of the model is an iterative process, meaning that it happens several times until the machine learning model is considered as accurate enough. (T. Larose & D. Larose, 2015) Once the training phase is done, the machine learning algorithm is asked to apply what it has learned to new data in order to predict the targeted value. In fact, the model will apply the patterns it has learned to the new data which will give him a certain output value. The test phase is as important as the training phase because it enables to see the result given by the model and assess if the model was close at predicting the right targeted value. The evaluation phase will be explained in the next sub-chapter. (cf. infra p. 76)

Hereunder, a table to illustrate how it works with an example of the house prices prediction.

| Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|
| Period of time | # of rooms | Location | Tot of square meters | Parking places | # of facades | House prices |
| Q1 | 5 | Uccle | 400 | 2 | 4 | 2000/m2 |
| Q1 | 2 | Schaerbeek | 150 | 1 | 2 | 800/m2 |
| Q1 | 4 | Waterloo | 300 | 3 | 4 | 1900/m2 |
| Q1 | 1 | Etterbeek | 110 | 0 | 2 | 600/m2 |
| Q1 | 2 | Woluwe | 165 | 2 | 2 | 1500/m2 |
| Q1 | 5 | Wemmel | 800 | 4 | 4 | 3200/m2 |

*Figure 32: Prepared dataset for house prices prediction*

We can notice that there are five features within the input data. So, the input data is the number of rooms, the location of the house, the total number of square meters, the parking places and the number of facades. The output data, which is the targeted value that need to be predicted by the machine learning model is the house prices in euros per square meter. So, when we have as input data a 4-façades house of 400 square meters with 5 rooms and 2 parking places in Uccle, the predicted price of the house should be 2000 euros the square meter. This will be done with each row until the model is able to find out patterns.

If we want to predict future house prices, we have to take as much as possible historical data with both the labelled input data and the corresponding output data in order train the model. As the we are using historical data, the company should exactly know the input and the output data and this is what enables them to train the model. Showing it the features and asking to predict the result corresponding to it. The following table shows how the machine learning algorithm is trained by time.

| | Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|---|
| | Period of time | # of rooms | Location | Tot of square meters | Parking places | # of facades | House prices |
| Loop 1 | Q1 | 5 | Uccle | 400 | 2 | 4 | 2000/m2 |
| Loop 1 | Q1 | 2 | Schaerbeek | 150 | 1 | 2 | 800/m2 |
| Loop 1 | Q1 | 4 | Waterloo | 300 | 3 | 4 | 1900/m2 |
| Loop 1 | Q1 | 1 | Etterbeek | 110 | 0 | 2 | 600/m2 |
| Loop 1 | Q1 | 2 | Woluwe | 165 | 2 | 2 | 1500/m2 |
| Loop 1 | Q1 | 5 | Wemmel | 800 | 4 | 4 | 3200/m2 |
| Loop 2 | Q2 | 5 | Meise | 350 | 3 | 4 | 2000/m2 |
| Loop 2 | Q2 | 4 | Schaerbeek | 300 | 1 | 2 | 1000/m2 |
| Loop 2 | Q2 | 3 | Waterloo | 180 | 2 | 4 | 1600/m2 |
| Loop 2 | Q2 | 2 | Ixelles | 150 | 1 | 2 | 900/m2 |
| Loop 2 | Q2 | 2 | Kraainem | 400 | 3 | 4 | 3500/m2 |
| Loop 2 | Q2 | 4 | Wemmel | 300 | 2 | 3 | 2300/m2 |
| ... | Q3 | ... | ... | ... | ... | ... | ... |

*Figure 33: Prepared dataset through time*

In this example, the machine learning model will be trained every quarter with new data once the input variables and the corresponding targeted value are known. Remember that the candidate model will be trained through time, even after if the predictive model is accurate. By time, the model will always be given historical data in order to improve it by finding new patterns and increase its prediction accuracy.

In the case of financial forecasting, the same methodology should be applied to train the model at predicting future revenues for example. Let's take the example of the company that wants to predict its future ice cream revenues.

| | Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|---|
| | Period of the year | Weather | Location | # of waiters | Parking places | Dogs admission | Revenue |
| Loop 1 | Q1 | 5 | Uccle | 3 | 5 | no | **12000** |
| Loop 1 | Q1 | 10 | Laeken | 1 | 0 | yes | **10000** |
| Loop 1 | Q2 | 20 | Woluwe | 4 | 10 | yes | **20000** |
| Loop 1 | Q2 | 15 | Schaerbeek | 1 | 0 | no | **15000** |
| Loop 1 | Q3 | 25 | Woluwe | 9 | 10 | yes | **35000** |
| Loop 1 | Q3 | 30 | Wemmel | 3 | 4 | yes | **33000** |
| Loop 1 | Q4 | 16 | Laeken | 2 | 0 | yes | **16000** |
| Loop 1 | Q4 | 12 | Uccle | 3 | 7 | no | **18000** |
| Loop 2 | Q1 | ... | ... | ... | ... | ... | ... |
| Loop 2 | Q1 | ... | ... | ... | ... | ... | ... |
| ... | Q2 | ... | ... | ... | ... | ... | ... |

*Figure 34: A prepared dataset for revenue prediction*

In the training dataset above, there are 6 input variables (input data) and we have our targeted value which is the expected revenue (output data). The candidate model will be told for the first line that if the period of the year is Q1, that the weather is 5 degrees, that there are 3 waiters, 5 parking places available and that dogs aren't admitted, then the predicted revenue should be 12 000 euros. This will happen with all the other samples from the different periods. Time is moving forward, once the next quarter is passed, the company should train the model again with the new input and output data that is known. The more data available, the more the candidate model can be trained. Moreover, a predictive model trained with a lot of data will increase its prediction accuracy.

As we are in the training phase, one's should remember that not all the prepared dataset is used to train the model. A certain percentage of the dataset will be preserved to test the candidate model. In general, data scientists use around 70-75% of the dataset to train the model and held back around 25-30% of the prepared dataset for the testing phase. (Cote, 2019)

Below, the figure shows the process to get a candidate model. The left of the figure shows how a prepared dataset looks like. That prepared training dataset will be used to train the candidate model. We can notice that the prepared training dataset looks exactly the same as the examples provided earlier. However, the predictive model need an algorithm to learn from the data provided to it. Without an underlying algorithm, the candidate model won't be able to extract patterns within the dataset and predict numerical values.

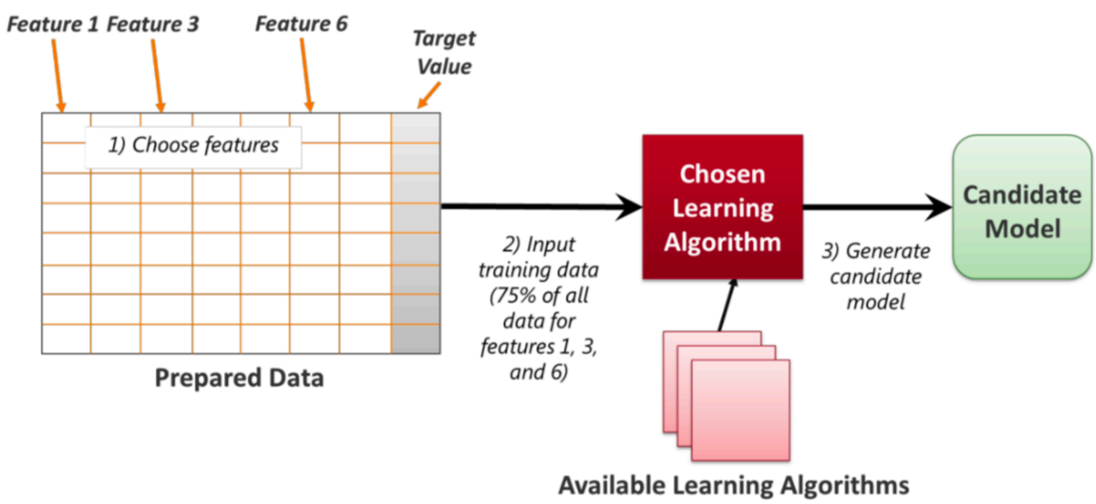


*Figure 35: Train a model with training data (Chappell)*

For companies, having a prepared dataset of quality is crucial to get an accurate candidate model. Nevertheless, data scientists still need to perform the training and evaluation phase with different data mining techniques, algorithms in order to find the best algorithm for prediction task. The next sub-chapter will highlight the most known algorithms which are at the heart of a candidate model. These algorithms enable models to learn, find patterns and predict future numerical values.

### 3.5.3 Choosing the underlying algorithm of the candidate model

In order to train the candidate model, data scientists need to configure the algorithm. It's the algorithm at the basis of the predictive model that will discover patterns within the data and predict the targeted value. When a company adopted a semi-automated machine learning solution, it will already get pre-configured machine learning algorithms and will only need to select the desired algorithm. However, if a company chose to adopt a solution offering a coding platform, data scientists need to create their own machine learning algorithms. In general, they need to choose which algorithm; data mining technique will be the best fit for the expected prediction task. The selection of the algorithm will strongly depend on prediction task and the prepared dataset. (cf. supra p.54-57) Is it a clustering task? anomaly detection task? classification task? regression task?

In the case of financial forecasting, the targeted value is a continuous value which means that it's a regression task. Numerous machine learning algorithms could be used to learn from a dataset in order to predict future results such as revenues, costs, margins… In this thesis, the four-main type of data mining algorithms will be pointed out and briefly explained to provide the readers with an overview of what for algorithms could be used. (Coallier, 2019)

These four types of algorithms are: (T. Larose & D. Larose, 2015)

- Linear regression methods (Linear regression, Lasso, Ridge)
- Boosted decision trees
- K-Nearest Neighbor
- Recurrent Neural Network (Deep Learning)

#### *Linear regression methods*

These methods are used when there is an assumption of linearity between the independent variables and the dependent variable. Models build with these linear methods demonstrate high bias because they are based on the assumption of linearity. If this assumption of linearity between the input variables and the output variable isn't right, the model won't be accurate. Moreover, these models have a low variance. It means that they are inflexible due to low parameter possibilities. On the contrary, with non-linear predictive models, it is possible to fine tune the models by improving the parameters. (Data Science, s.d.) Companies should use these linear predictive models when they have few data available. If companies do use non-linear algorithms with few data, it will lead to overfitting which means that the model won't be able to generalize what it has learned and won't be able to find patterns within new input data. Besides, companies should only use linear models when there is an indication of linearity between features and outcome. If not, they shouldn't use these models at all. (Nasyrov, 2017)

There are mainly 3 methods which are the simple linear method, lasso method and the ridge method. (Coallier, 2019) These three methods use the same parameters. However, they differ in the setting of constraints. With a simple linear model, there are no constraints. Lasso will ignore some features by setting some coefficients to zero and Ridge will lower the coefficients close to zero so that all independent variables have a minimal impact on the predicted result. (Nasyrov, 2017)

If a company has huge amounts of historical data and that they believe that there is non-linearity between the input features and the targeted output, they could use methods such as boosted decision trees, k nearest neighbor or recurrent neural networks. These methods are going to be explained in the following lines.

### *Boosted decision trees*

One of the machine learning algorithms that enables to build a candidate model is the "decision tree" algorithm. People need to visualize a tree in which each branch splitting represents a question about a feature in order to learn about the data and try to find patterns. In real life, a decision tree as a lot of analogies which can visually show how decision are taken. Decision tree algorithms for prediction tasks is often used to train candidate models because the time needed to train the model is fast and provide most of the time a good accuracy. (Brownlee, A tour of machine learning algorithms, 2013)

In practice, decision tree algorithms will basically ask closed questions about different input features that have been provided to the candidate model. In the previous case of predicting the house prices, the algorithm will ask questions about the location, the total square meters and the number of rooms. As a matter of facts, the questions will depend on the features that are provided to the predictive algorithm. A decision tree is drawn up from the top the bottom with the root at the top.

Within the decision tree from the next page, each bold text represents a question about a specific feature, also called the internal nodes. Based on internal nodes, the decision tree will be divided into branches, also defined as "edges". The last branch which won't split anymore is called the leaf and represents the end of the algorithm learning process. (Sayad) At that stage, the final decision is taken and the targeted value will be predicted. In this case, it will predict the future house prices taking into account all the features that have been provided to the algorithm. (Gupta, 2017)

Let’s take a closer look at the prepared dataset of the house prices prediction. Each node will ask a closed question for a specific feature. Let’s suppose that the 3 features provided to the algorithm are location, total square meters and number of rooms. Only 3 features are provided in our example because otherwise the drawing of the decision tree would become too complex.

| | Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|---|
| | Period of time | # of rooms | Location | Tot of square meters | Parking places | # of facades | House prices |
| Loop 1 | Q1 | 5 | Uccle | 400 | 2 | 4 | 2000/m2 |
| Loop 1 | Q1 | 2 | Schaerbeek | 150 | 1 | 2 | 800/m2 |
| Loop 1 | Q1 | 4 | Waterloo | 300 | 3 | 4 | 1900/m2 |
| Loop 1 | Q1 | 1 | Etterbeek | 110 | 0 | 2 | 600/m2 |
| Loop 1 | Q1 | 2 | Woluwe | 165 | 2 | 2 | 1500/m2 |
| Loop 1 | Q1 | 5 | Wemmel | 800 | 4 | 4 | 3200/m2 |
| Loop 2 | Q2 | 5 | Meise | 350 | 3 | 4 | 2000/m2 |
| Loop 2 | Q2 | 4 | Schaerbeek | 300 | 1 | 2 | 1000/m2 |
| Loop 2 | Q2 | 3 | Waterloo | 180 | 2 | 4 | 1600/m2 |
| Loop 2 | Q2 | 2 | Ixelles | 150 | 1 | 2 | 900/m2 |
| Loop 2 | Q2 | 2 | Kraainem | 400 | 3 | 4 | 3500/m2 |
| Loop 2 | Q2 | 4 | Wemmel | 300 | 2 | 3 | 2300/m2 |
| ... | Q3 | ... | ... | ... | ... | ... | ... |

*Figure 36: Prepared dataset for house prices prediction*

Hereunder, the decision tree represents the prediction of the house prices based on the three features. The bold text represents a question about a specific input feature. Moreover, when there is a branch splitting, the left of the branch represents a “yes” answer and the right represents a “no” answer. (Vandeput, 2018)

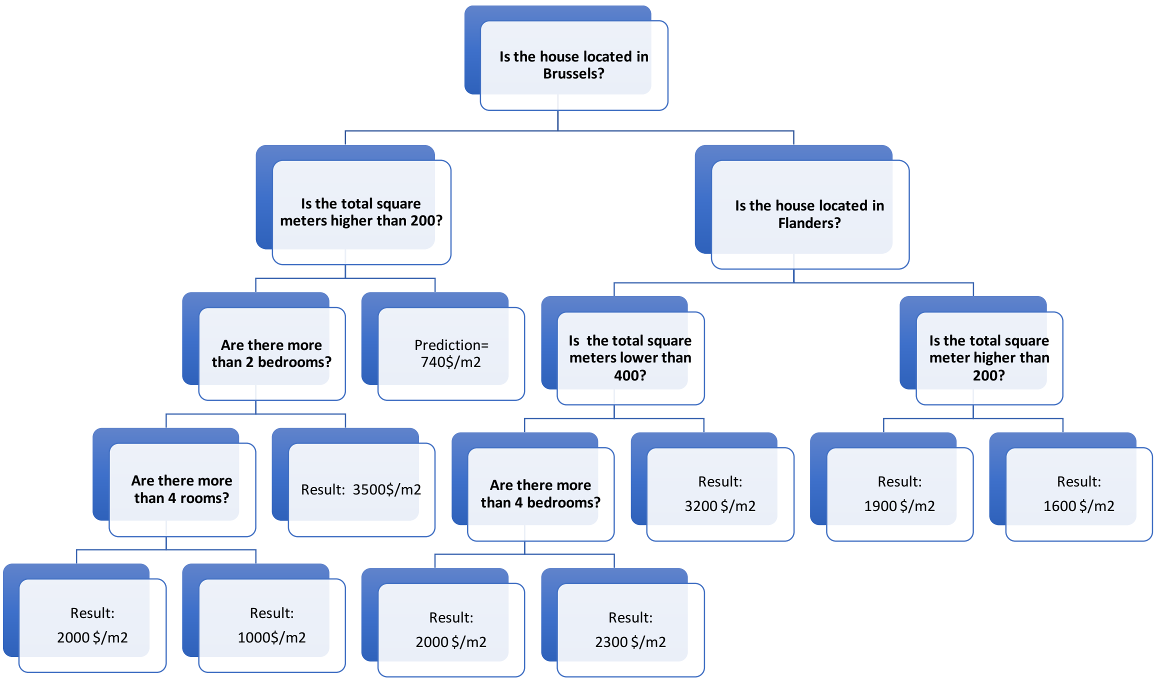


*Figure 37: Decision tree for house prices prediction*

Now, let's apply this methodology to the financial forecasting example where an ice cream chain tries to predict its future revenues. The same technique will be applied in order to get insight and learn patterns from the given dataset. Hereunder, the dataset provided to the algorithm (only location, weather and number of waiters will be provided to the model as features in this case).

| Historical input data | | | | | | Output data |
|---|---|---|---|---|---|---|
| Period of the year | Weather | Location | # of waiters | Parking places | Dogs admission | Revenue |
| Q1 | 5 | Uccle | 3 | 5 | no | **12000** |
| Q1 | 10 | Laeken | 1 | 0 | yes | **10000** |
| Q2 | 20 | Woluwe | 4 | 10 | yes | **20000** |
| Q2 | 15 | Schaerbeek | 1 | 0 | no | **15000** |
| Q3 | 25 | Woluwe | 9 | 10 | yes | **35000** |
| Q3 | 30 | Wemmel | 3 | 4 | yes | **33000** |

*Figure 38: Prepared dataset for revenue prediction*

Hereunder, a decision tree representing the prediction of revenue based on the three chosen features. In bold, the questions based on the three features. Moreover, again, when branches split, the left of the branch represents a "yes" answer and the right represents a "no" answer. (Vandeput, 2018)

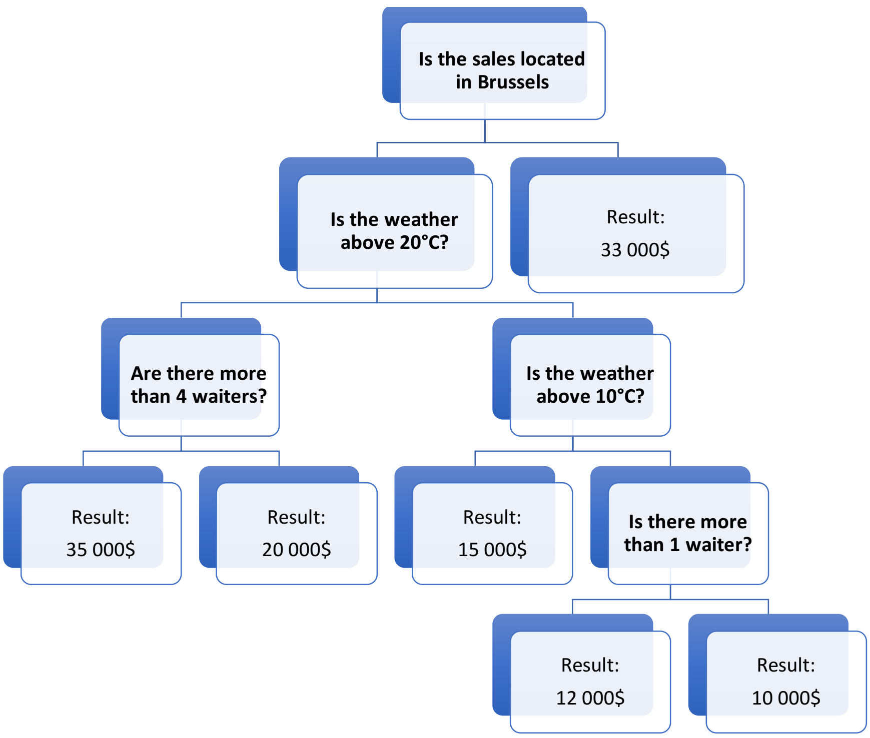


*Figure 39: Decision tree for revenue prediction*

One's should bear in mind that in real life, the number of features and the size of the decision trees will be much more important. However, the objective of these two examples is to show how decision tree algorithms function in order to learn and predict future results. The complexity of these decision trees is relatively simple but it's only based on three features. What if there are lots of features taken into account to predict a targeted value?

This clearly highlight the importance of the feature extraction phase where the most impactful features are retained. This stage has been mentioned in the process of building a machine learning model. The two decision trees above are called “regression trees” because the targeted value is a continuous value such as revenues, costs, margins, prices… For information, the decision tree algorithm can also be used for classification tasks which focus is to classify the input data into different categories. (Gupta, 2017)

Let’s take a closer look at the background of our decision tree procedure. It has been explained that decision tree questions which enable the split of our dataset are based on the features. People need to bear in mind that the choice of the features order and the decision to stop the splitting isn’t done randomly. There is a well-known technique used for the splitting and it’s called the “recursive binary splitting”. (Gupta, 2017) When using the recursive splitting technique in a decision tree procedure, it means that all the provided features are considered when a branch split occurs within the decision tree. At each split point, all features are tried and tested based on a cost function. The cost function looks at the accuracy loss of each split. In other words, it looks at the standard deviation of each split. The algorithm will choose the feature with the lowest cost, which means the one with the highest standard deviation reduction. In fact, the cost function tries to find out which are the most homogeneous branches. Branches which have groups with similar responses. The reason behind is that the probability that the test data (new input data) will follow the same path is higher. The splitting is seen as recursive because this cost function is done repetitively after each split decision in order to determine the right targeted value. (Gupta, 2017)

In addition, it sometimes happens that there are too many features which lead to a very large decision tree with a high number of branch splits. Very complex trees are bad because it could produce a model that is overfitting. (cf. supra p.68) So, it’s important to know when to stop the splitting of the tree branches. There exist two techniques that could help companies to manage the complexity of a decision tree. On one hand, it’s possible to determine a minimum amount of training data to use on each leaf. On the other hand, it’s also possible to determine the maximum depth of the tree which refers to the maximum length between the top and the bottom of the tree (between the root and the leaf of the tree). It’s very important to avoid overfitting because it means that the prediction will fail on the testing data. (Coallier, 2019)

Last but not least, the performance of the decision tree algorithm could be improved by pruning the model. It implies that branches based on features with low importance are removed from the decision tree. At the end of the day, it will lower the complexity of the decision tree and strengthen the avoidance of overfitting and increase its overall prediction accuracy. (Gupta, 2017)

There is a variance of this algorithm which is called "random forest". The idea is that there are multiple decision trees that form a forest. All the decision trees are trained with the same prepared dataset but where the training data/testing data split isn't exactly the same in all decision trees. (Donges, The random forest algorithm, 2018) This is done intentionally to add diversity and predict a continuous value from different decision tree and confirm what has been predicted. (Masnaoui, 2019) At UCB, this one of the algorithms that they are using to predict their product revenues. (Lieutenant, 2019)

In a nutshell, the regression decision tree is one way a machine learning algorithm can learn and predict. Closed questions are asked based on the input and output features which will enable the algorithm to discover some patterns within the dataset and predict the targeted value. When the number of features is too high, it can lead to overfitting. This need to be absolutely avoid because it means that the algorithm isn't able to generalize what he has learned and won't be able to forecast accurately future results when its confronted with new input data. In addition, the model can be improved by fine tuning the parameters or by removing some branches based on features with low importance.

### *K Nearest Neighbor*

K Nearest Neighbor is an algorithm that will try to predict a continuous value by using the features and predicted value of the nearest neighbors. In order to do so, the algorithm is going to be told how much neighbor he is going to consider in his calculation of the predicted value. This is where "k", a parameter of the model is determined. Most of the time "k" is determined by calculating the root square of the total number of samples. If a company has 1 million lines in its prepared dataset, the "k" will be 100 because $k = \sqrt{1\ 000\ 000}$ (Simplilearn, 2018)

The distance is going to be calculated between the new data point and each historical data point in order to identify the hundred closest neighbors. The formula used to calculate the distances is the Euclidian formula. The formula is the following: (T. Larose & D. Larose, 2015)

$$d_{Euclidian}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Once the distances are calculated, the algorithm will be able to point out the 100 closest neighbor that will be used to predict the targeted value. The second formula used to predict the targeted value is the following: (T. Larose & D. Larose, 2015)

$$y_{new} = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

where:

$Y_{new}$ = the predicted targeted value of the new data point

$W_i = \frac{1}{d(new,x_i)^2}$

$Y_i$ = the historical targeted values that are known

However, this type of algorithm is rarely used to predict continuous value such as revenues or costs because this type of algorithm cannot always deal with complexity and huge amounts of data. It loses accuracy when there are a lot of predictor variables which is more than probable when predicting financial figures such as revenues and costs. (Masnaoui, 2019)

From the author point of view, this algorithm should only be used for prediction tasks where they aren't a lot of features and thus where complexity isn't too high. Otherwise, he would recommend using boosted decision trees, random forest or the RNN algorithm which is going to be explained in the next lines.

### *Deep learning*

Deep learning is a sub-category of machine learning that can handle more important quantities of data. Deep learning is done through a network of neural called "neural Network". The network replicates the way our human brain is thinking and is solving problems. (Patience, 2016) In fact, neural network is seen as an alternative and upgrade of machine learning. The advantages of deep learning compared to other machine learning methods is that it can handle unstructured data and that it possesses a long-term memory.

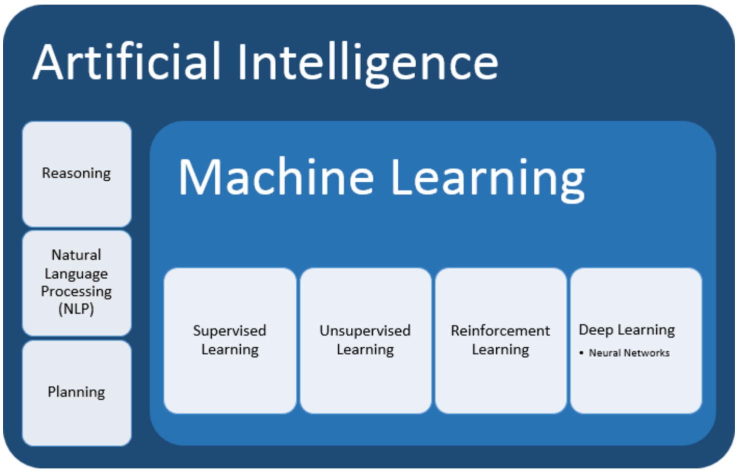


*Figure 40: AI, ML, DL concepts (Networking technologies, 2018)*

When working with sequential data and dealing with huge amounts of data, using "recurrent neural network" seems to be the best option to predict future outcomes. It's a strong and robust type of neural network with long-term memory that has incredible potential for predicting future numerical values. As it possesses long-term memory, it's one of the reasons why it's used for time series with financial data… Recurrent neural network gives the possibility to have a very good and much deeper understanding of the context of a sequence than other type of machine learning algorithms. (Donges, Recurrent Neural Networks and LSTM, 2018)

Let's take a closer look at the structure of a recurrent neural network and how it's working in order to predict future results. First let's illustrate a neural network with the input layer, the hidden layers and the output layer.

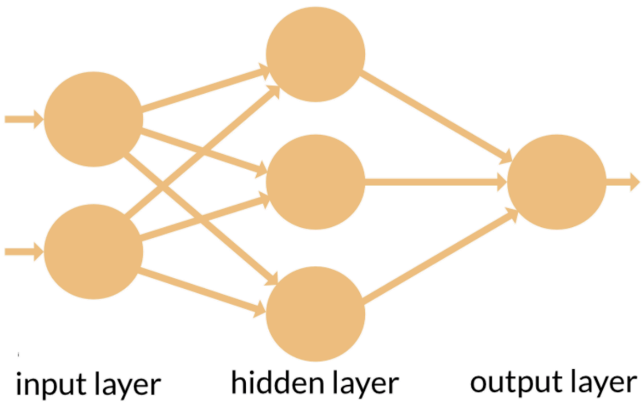


*Figure 41: Neural Network representation (Donges, Recurrent Neural Networks and LSTM, 2018)*

The model will be given a certain number of inputs that will be treated in the hidden layers with very complex computations using randomly initialized variables called weights. This will give a determined output that will be compared with the expected result. The prediction error will be measured and back-propagated into the neural network. The error will be communicated to the different variables of the recurrent neural network so that they could adapt themselves by changing the respective weights.

The prediction accuracy will increase by taking into account the back-propagated error through time. This procedure is repeated several times until the variables are well weighted and that we are sure that the RNN will forecast accurately future outcomes. Once the RNN is ready, it will be applied to unseen data. The numerous steps enumerated above is the way how a recurrent neural network is working in order to predict results. (Skymind, n.d)

Moreover, the type of neural network used for time series is called "recurrent neural network" because the information cycles through a loop. It means that when the neural network processes an output, it takes into consideration the current input data but also what it has learned from previous input data. The scheme hereunder shows the difference between a recurrent and standard neural network. On the left graph, it shows that not only current input data is used but also past experiences and input data is used for current prediction task.

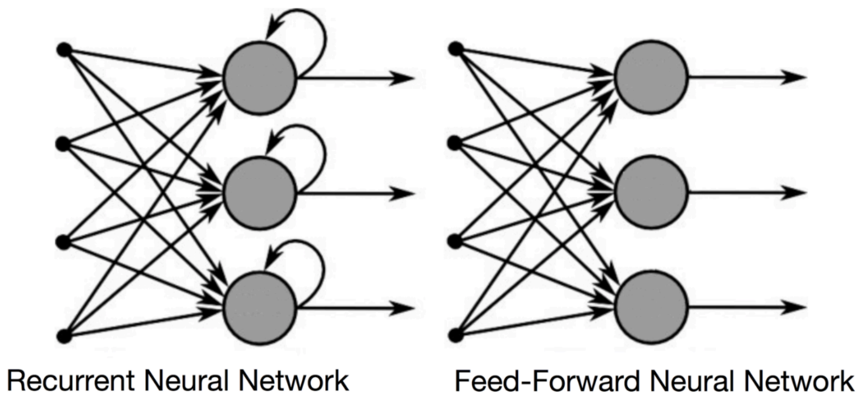


*Figure 42: Comparison between RNN and FFNN (Donges, Recurrent Neural Networks and LSTM, 2018)*

## 3.6 Evaluation of the model

### 3.6.1 Assess each candidate model accuracy

After the modelling phase, it's essential to assess the accuracy of each machine learning model that has been built. This step is also called "scoring the predictive models". At this stage, it implies that the machine learning algorithm has been selected and that the candidate model has already been trained with the training data. But can the model accurately predict the targeted value if we give it new input data? The idea of the evaluation phase is to prove the accuracy of the model on unseen data before using it for real business predictions.

The testing data is known and comes from the prepared dataset. Let's remind within the pre-processing phase, the dataset has been divided in two categories, the training and the testing dataset. (cf. supra p.62) (Roman , 2018)

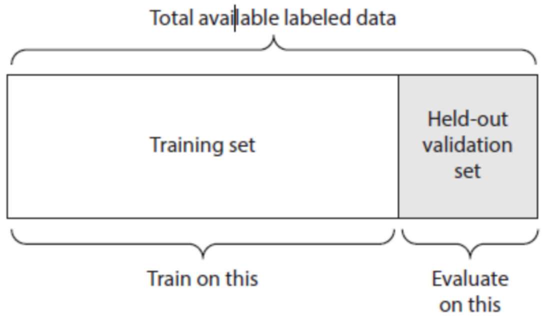


*Figure 43: Splitting the prepared dataset (Roman , 2018)*

Above, the figure points out that there is still around 25-30% of the prepared dataset that has been left over to test the accuracy of the machine learning model. So, the remaining 25-30% of testing data will be provided as input data to the model and the candidate model will be asked to predict the targeted value. The new input data will be structured in the same way than it has been done to train the model. (Cote, 2019) In contradiction with the training phase, the machine learning model won't be given the targeted value. It will only receive the input data and will be asked to predict the targeted value. This step is called the scoring step. Afterwards, the predicted targeted value will be used to assess the model prediction accuracy. (Coallier, 2019) The scoring step is thus essential to assess the accuracy of each candidate model.

Below, the figure provides a good representation of how the testing phase is working.

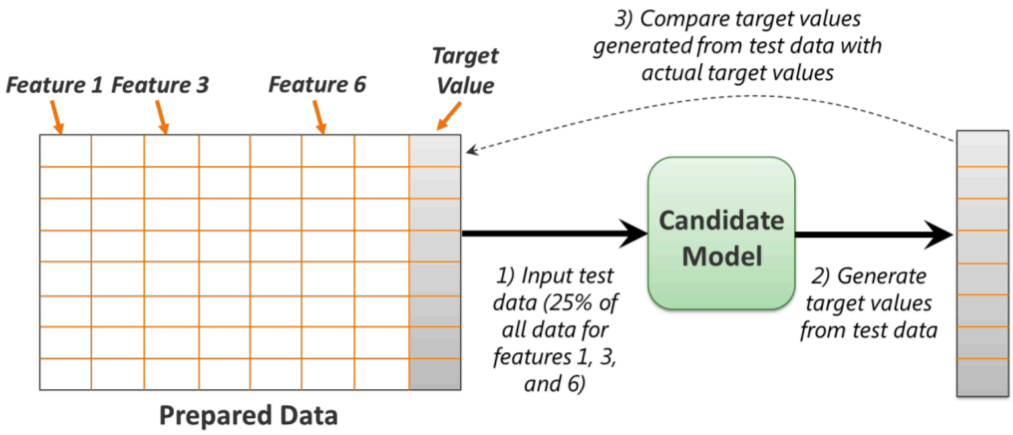


*Figure 44: Test the model with testing data (Chappell)*

First of all, the features that will be used to test the candidate model need to be pointed out. These features are already identified because they are supposed to be the same as the one used to train the candidate model. Afterwards, the 25% remaining testing data will be used and provided to the candidate model. If we have a prepared dataset with 1 million samples (which are the lines), 750 thousand lines will be used to train the candidate model and the 250 thousand remaining lines will be used as testing data. The candidate model will then, based on the testing data predict for each sample the targeted value. Last but not least, the predicted value will be compared to the actual targeted value in order to assess the accuracy of the candidate model.

To sum up, the 3 steps within the test phase are: (Chappell)

1. Provide the testing data as input data to the candidate model (25% of the dataset and no mention of the targeted value that needs to be predicted).
2. The candidate model predicts the targeted value based on the test data provided to it.
3. The predicted values are compared to the actual targets values so that the accuracy of the candidate model can be assessed.

When a candidate model accuracy is measured, there are mainly two methods that could be used. Each metrics has its advantages and drawbacks. At this stage, companies will be able to assess the potential improvement of their financial forecasting process by testing the candidate model with unseen data. A well-known quote highlights the fact that “if you can’t measure it, you can’t improve it”. (Safdari, 2018) In the second phase of the model implementation process, companies needed to find out the required data but also to define the financial forecasting improvement expectations. (cf. supra p.56) (Safdari, 2018)

From the author point of view, the accuracy assessment of the candidate model is one of the most fundamental steps within the process before using it for real financial forecasting situations. If this step is neglected, companies could end up with a unworthy candidate model that doesn’t provide the expected accuracy. This evaluation step is important in order to be sure that the model won’t predict inaccurate results within real financial forecasting situations. (T. Larose & D. Larose, 2015)

The first metrics that could be used to measure the candidate model prediction accuracy is “MAE”, which is the abbreviation of “mean absolute error”. This metrics gives the average error between the actual targeted value and the predicted targeted value. (Vandeput, 2018) Moreover, it provides an equal weight on each error. Nevertheless, this metric won’t clarify the direction of the error because it’s taking the absolute error. So, it will be impossible to understand if the candidate model is over predicting or under predicting. (Mishra, 2018)

The mathematical formula of MAE is the following:

$$MeanAbsoluteError = \frac{1}{N}\sum_{j=1}^{N}\left|y_j - z_j\right|$$

Taking into account that:
N = the number of samples (i.e lines)
$y_j$ = the predicted targeted value
$z_j$ = the actual targeted value

The second metric that is often used to measure the prediction error of the candidate model is "RMSE", which stands for "root mean square error". This metric also measures the average magnitude of the error but weight it differently. The advantage of using the RMSE metric is that errors are squared before being averaged. It implies that the weight is higher for large errors. So, if large errors are willing to be avoided, RMSE could be more useful. (Towards Data Science, 2016)

The mathematical formula of RMSE is the following:

$$RootMeanSquareError = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y_j - z_j)^2}$$

Taking into account that:
N = the number of samples (i.e lines)
$y_j$ = the predicted targeted value
$z_j$ = the actual targeted value

To compare both metrics, we are going to take the example of the revenue prediction of the ice cream chain. In order to measure the predicted accuracy both the predicted and the actual targeted value are going to be compared by using the MAE and RMSE metrics. (Towards Data Science, 2016)

Below in the table, you can notice that when the variance of error is more or less equally distributed within the sample. As a consequence, the calculated average error of both metrics is quite similar because there are no outliers with huge errors.

| Actual targeted value | Predicted targeted value | MAE | RMSE |
|---|---|---|---|
| **12000** | 11500 | 500 | 250000 |
| **10000** | 9500 | 500 | 250000 |
| **20000** | 20500 | 500 | 250000 |
| **15000** | 14800 | 200 | 40000 |
| **35000** | 36000 | 1000 | 1000000 |
| **33000** | 32300 | 700 | 490000 |
| **16000** | 15600 | 400 | 160000 |
| **18000** | 17700 | 300 | 90000 |
| ERROR | | 512,5 | 562,36 |

*Figure 45:Comparing MAE and RMSE with equal error distribution*

However, in the second table below, we can point out that the error isn't equally distributed and that there are some outliers where there is an important prediction error. This means that the poor accuracy of the model will be more important and highlighted while using the RMSE metric. If both tables are compared, it's possible to highlight the fact that the overall error is the same but the error distribution is different. RSME is helpful to discover some important outliers. (Towards Data Science, 2016)

| Actual targeted value | Predicted targeted value | MAE | RMSE |
|---|---|---|---|
| **12000** | 12000 | 0 | 0 |
| **10000** | 10000 | 0 | 0 |
| **20000** | 20000 | 0 | 0 |
| **15000** | 14900 | 100 | 10000 |
| **35000** | 35000 | 0 | 0 |
| **33000** | 32000 | 2000 | 4000000 |
| **16000** | 18000 | 2000 | 4000000 |
| **18000** | 18000 | 0 | 0 |
| ERROR | | 512,5 | 1000,62 |

*Figure 46: Comparing MAE and RMSE without equal error distribution*

## 3.6.2 Compare the different candidate models' accuracy

It has been demonstrated that companies can use numerous algorithms to create and train a candidate model. Each of these algorithms have their own learning methodology, advantages and drawbacks. However, in most situations, the modelling and testing phases is done with several algorithms in order to get multiple candidate models. This means that each candidate model uses a different underlying algorithm to find out patterns and predict future results. Once the candidate models are trained, tested and evaluated, the accuracy of each of them is compared to make sure that the right candidate model is used to predict the targeted value. (Coallier, 2019) In the business case, it will be explained that UCB also uses different algorithms to forecast their product revenues, clinical and marketing costs. This analysis has been done by UCB to make sure that they use the more accurate predictive model for each item line. (Lieutenant, 2019)

Hereunder, the table compares four candidate models where each of them uses a different underlying algorithm. In this case, the best model to use would be the neural network model because it has the lowest average prediction error. (Microsoft, 2014)

| Evaluation of the models | MAE | RMSE |
|---|---|---|
| Decision tree model | 80,87 | 97,54 |
| Neural Network model | 51,34 | 88,56 |
| KNN model | 124,54 | 140,27 |
| Linear model | 130,9 | 145,78 |

*Figure 47: Comparing the different models' accuracy*

## 3.7 Deployment phase

Now that the best candidate model has been chosen, companies need to deploy the predictive model in order to be able to use it for their daily, monthly or quarterly business forecasting situations. This phase is the last phase of the implementation process of a machine learning model. Getting an accurate candidate model is essential but without deploying it, companies won't be able to use it to forecast in a near future.

The first step is to upload the predictive machine learning models on the cloud to make it accessible. There are several providers that offers their cloud services such as:

- Microsoft with Microsoft Azure Platform
- Amazon with AWS Lambda
- Google with Google Cloud
- SAP with SAP Cloud

These companies created cloud platforms to give companies the opportunity to deploy their machine learning model on the web. In fact, once the model is build, uploaded on the cloud and accessible for applications, it's seen as a "web service". In order for the application to access to the model and to the data warehouse, an API needs to be implemented. (Amazon, 2019) API stands for "application programming interface" and it's an interface that acts as the front door for applications to access to the machine learning model and the data warehouse. (Gazarov, 2016)

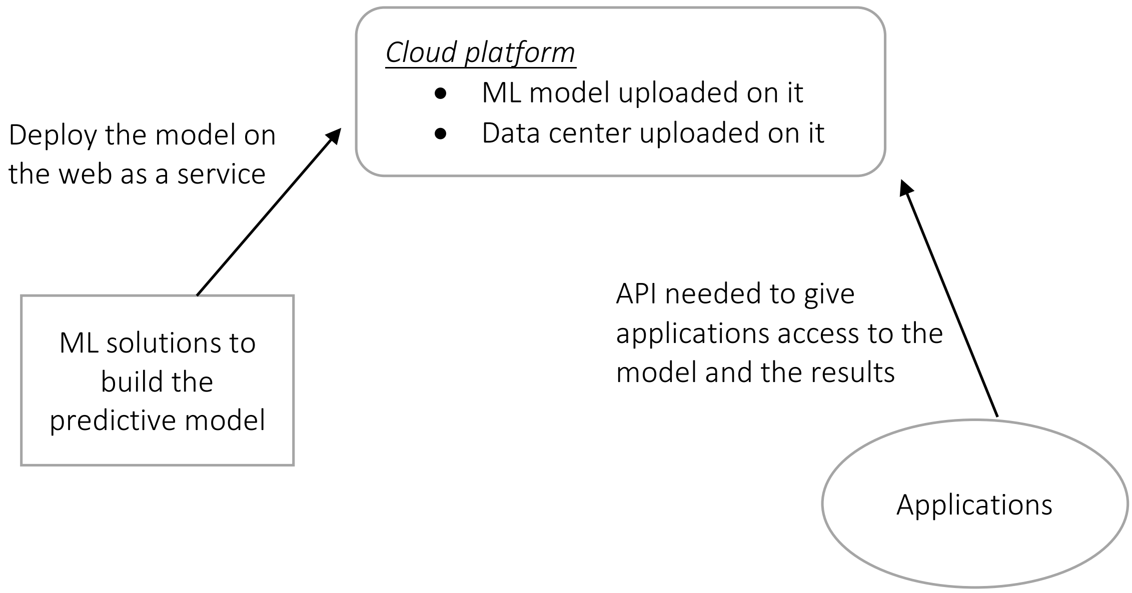


*Figure 48: Deploying the predictive model in the cloud (Chappell) (Masnaoui, 2019)*

Above, the scheme clearly highlights the process happening in the deployment phase. The candidate model created with the machine learning solutions (Microsoft Azure, Google AutoML, AWS ML…) will be deploy on the cloud to make the model available as a web service. Afterwards, applications have access to the model thanks to API's. API's being the connections between the application and the machine learning model on the cloud. (Amazon, 2019)

But how does the application manage the financial prediction request to the model?  In fact, in the earlier phases, it has been explained that the model need to receive the input features in order to predict the targeted value. The whole process of receiving the input key (could be a specific product), looking for the right input features, sending the input features to the model and providing the targeted value will be managed by the application. (Chappell) (Coallier, 2019)

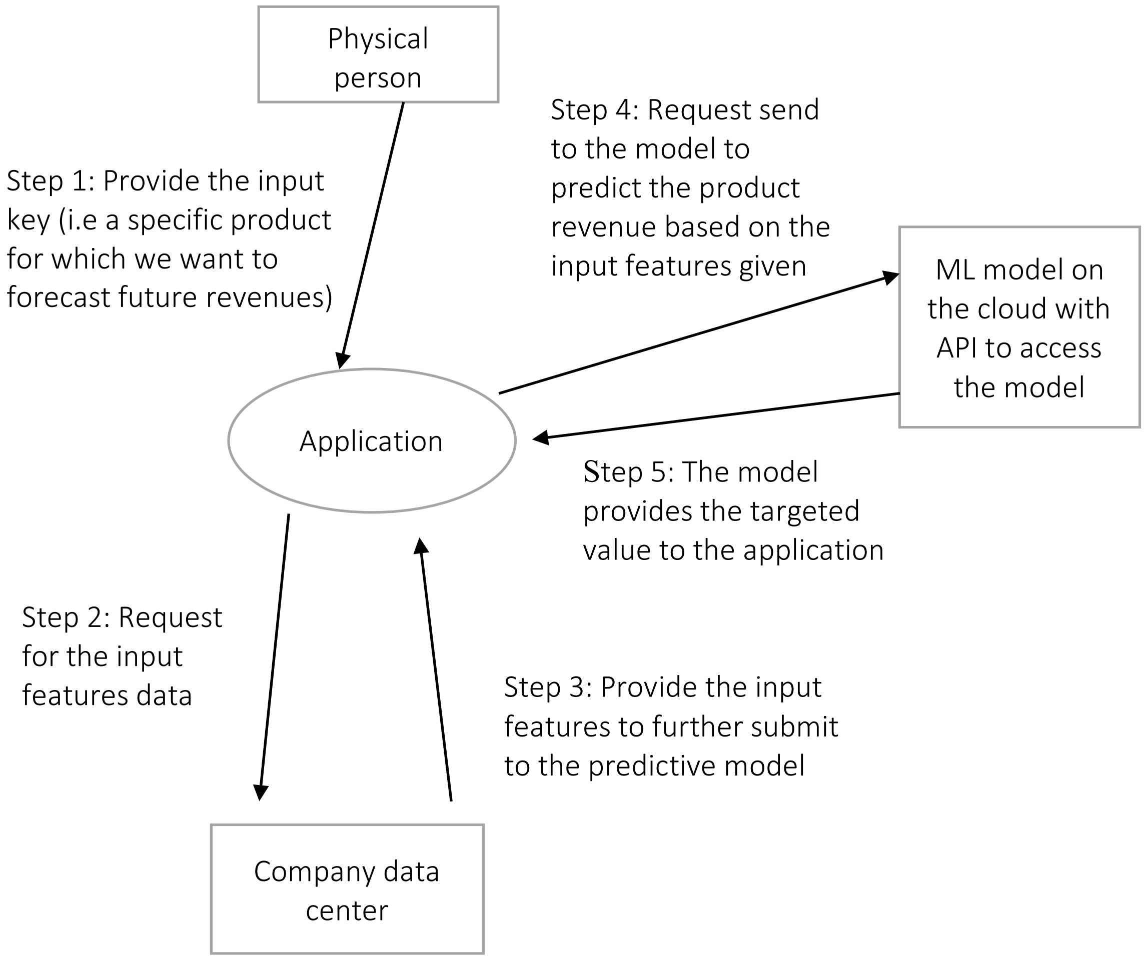


*Figure 49: The steps for real-time prediction (Chappell)*

The whole process of collecting the required input features, provide it to the machine learning model and expose the results via the application is fully automated. It gives companies the opportunity to forecast their revenues in real time. (Cote, 2019)

## *3.8 The advantages and stakes of implementing a predictive machine learning model*

Before taking whatever decision concerning a certain matter, it's always important to strike a balance between the pros and cons. In this section, the advantages of building a predictive machine learning model for financial forecasting purposes will be pointed out. However, predictive analytics also comes along with drawbacks which will also be explained to bring a critical point of view about the use of machine learning models for financial forecasting. When looking at the advantages of machine learning for forecasting, it's possible to analyze it from different point of views. There are multiple reasons why businesses should integrate and experience the use of machine learning models into their financial forecasting process.

From a pure financial forecasting performance point of view, machine learning has incredible accuracy capabilities. With the technological revolution, companies can collect and use huge amounts of data. An opportunity that wasn't necessary possible before. Today's traditional forecasting methods can't always handle the forecasting complexity and the overheated, massive amounts of data. Overcoming these limits could be the key to an increase of a financial forecasting performance. (Muddassir, 2018) It isn't easy to discover patterns within huge amounts of data where the relationship between input and output variables is complex; However, machine learning algorithms give the opportunity to companies to do so. (Cote, 2019) It would be non-sense for a company to not use the available data because of the lack of their current forecasting method capabilities. This is one of the reasons why companies should be aware that machine learning could be beneficial within their financial forecasting process. It represents a real added value for them. (Owen, 2019)

Another advantage of machine learning models is the fact that it can provide deep insight from the business data. Therefore, it will forecast accurately companies' revenues, costs and reflect the companies' reality. With the use of predictive machine learning models, the most important input features are taken into account for the forecast of financial figures. This isn't always the case when companies do use traditional forecasting methods such as exponential smoothing or growth rate. From the author point of view, businesses can get real insight thanks to the use of their historical business data, better leverage it, discover patterns and anticipate future revenues or costs. The real drivers of the companies' revenue and costs are identified and used to forecast future financial figures. (Owen, 2019)

A third advantage that can be highlighted is the fact that the financial forecasting process will be improved by time because the machine learning models become more accurate by time thanks to its learning and memory capacity. Each example and experience provided to the model is kept in its memory and used when its confronted with new data. Whatever type of algorithm is used; boosted decision trees, random forest, KNN, lasso or recurrent neural network, all of them are very good at finding relationships between the data in order to forecast future financial figures. Drilling down within the data and extracting insight isn't easy with traditional methods. At least not as deep as with machine learning algorithms. (William McGhee, s.d.)

A fourth advantage of using predictive machine learning models is that the model can make connection between the different products, services from the dataset. When companies are for example forecasting future products revenues, traditional forecasting methods will forecast separately each product revenue. On the contrary, machine learning models will take into account the possible impact of a product revenue on the revenue of other products. In other words, machine learning algorithms make links between the numerous products of the whole dataset which is isn't the case with traditional forecasting methods forecasting individually each product revenue. (Vandeput, 2018)

A fifth advantage, from an operational point of view, is that machine learning models can indirectly improve supply chain coordination for manufacturing or distribution companies. (Vandeput, 2018) Being able to forecast accurately product revenues and costs is key in order to determine the coordination of the company's supply chain. Machine learning models enable companies to get a better coordinate and point out the supply chain needs in order to fulfil customer demand. Some companies even use machine learning forecasting to predict the required inventory stock, the time for machine maintenance…

Last but not least, implementing machine learning models seems to be promising because it automates the whole financial forecasting process and provide some very good forecasting accuracy. A forecasting that could be higher than traditional forecasting methods. (Cote, 2019) Nevertheless, this requires a model that is trained with quality data. If we take the example of Microsoft Benelux, they could increase their forecasting accuracy to 98% thanks to the implementation a predictive machine learning model into their financial forecasting process. This represented an increase of roughly 2%. This seems to be a small increase but that small percentage on their total turnover represents a huge amount of money. Therefore, businesses could be more efficient and thus increase their overall profit. (J. Brar, personal communication, November, 2018) From the author point of view, all companies should try to improve their financial forecasting capabilities because it provides a real competitive advantage. Using machine learning for financial forecasting purposes makes it easier for companies to increase their overall corporate performance.

From a strategic point of view, being able to forecast in real-time and accurately future financial figures such as revenues, costs, margin is crucial because it enables to give a clear direction where the company is going and what it needs to do in order to achieve its objectives. The financial forecasting process will reflect the company reality thanks to the use of internal data but also external factors that could impact the predicted targeted value.

Nevertheless, implementing a predictive machine learning model doesn't just comes with benefits. It's essential to also identify some drawbacks coming along with the implementation of predictive analytics.

The first drawback would be the increase of risk related to IT. By implementing a predictive machine learning model, companies once again rely a bit more on IT and are more vulnerable to IT errors. (Bottefeux, 2018) Furthermore, it's the machine learning algorithm that discover patterns and make links between the dataset, which sometimes goes behind human understanding. It increases the risk from a strategic point of view, because it means that companies are following with a kind of blind eye the targeted value given by the predictive machine learning model. From the author point of view, this black box issue is the real issue that makes it for the moment impossible for companies to totally rely on an automated machine learning forecasting process. As far as everything is running smoothly, the algorithm isn't questioned. However, it's important to assess the risk of implementing predictive algorithms and finding possible mitigation if an algorithm would make an error. In this case, companies could for example gather all the historical data in a back-up system that could be used in parallel with a traditional forecasting method in case the machine learning model would go of the road. (Masnaoui, 2019) From the author point of view, all companies should try to get a "B" plan in case the algorithm would go side track and predict uncoherent results in the future. Moreover, in the case of financial forecasting, it means that companies would delegate strategic tasks to computers. It remains a sensitive topic because it isn't just about asking to an algorithm to process an incoming invoice.

A second drawback of machine learning is the time and money requested to deploy the candidate model for financial forecasting purposes. The whole process to build a predictive machine learning model has been explained and there are lots of steps to follow in order for companies to take advantage of it. From the problem definition to the deployment phase, a lot of issue can be encountered such a lack of available data to provide to the machine learning algorithms, large cleaning and transformation modules to apply, choosing the right machine learning algorithm, possible high forecasting accuracy requirements, and also the need of high computation power. Using a predictive machine learning model to predict future revenues for example will only come true once the model has proven to be accurate enough with the prepared dataset (training + testing data). In a nutshell, the author wants to point out the fact that companies should be aware that running and experiencing the use of machine learning models could be very expensive and time consuming. (William McGhee, s.d.)

Finally, a third drawback would be the dependency of the predictive machine learning model on the quality of the historical data. This is one of the most important drawbacks because of its important impact on the accuracy of the predictive machine learning model. It has been explained that the higher the quality of the data provided to the model, the higher the accuracy of the forecasted outcome. As the data is provided by humans, the model partly relies and is impacted by humans' performance. If humans are providing biased data to the model, the model will learn from an erroneous dataset and will be susceptible to forecast wrong future outcomes. This is why data scientist and business people are still, and will always be very important when using a predictive machine learning model. Moreover, in order to achieve a very high machine learning model accuracy, the company needs to get a lot of data to train the model. The more data available, the better the model will be able to learn and find out patterns. If not, the model will take much more time to learn and will probably get a smaller forecasting accuracy. (Cote, 2019)

## 3.9 Conclusion

The numerous steps to implement a predictive machine learning model have been demonstrated in order to give the readers a practical view on how a predictive model is implemented.

Machine learning algorithms can be used for different prediction purposes such as classification, clustering, anomaly detection and regression. So, the first step consists at defining the prediction task. In the case of financial forecasting, it's will be a regression tasks and companies will need supervised machine learning algorithms. A good understanding of the business is essential to formulate a data mining task that will be performed by the predictive machine learning model.

The second step consists at finding out the required data in order to provide the candidate model with the best data possible. At that stage, companies will identify the revenue drivers for example that will be labelled as input data, predictor variables. This second step happens with a close collaboration between business people and data scientists. Moreover, people will also highlight the forecasting accuracy improvement wanted.

The third step consists at pre-processing the data with an aim at getting a prepared dataset which is totally cleaned. Several elements will be treated such as errors, outliers, categorical value, features normalization, dimensions reduction and split the prepared dataset in training and testing data.

The fourth step is about training the predictive model. To do so, the predictive model will be given as input data the different features such as the weather, the time period, the number of waiters, the average revenue of the location… and also be given the targeted value which in our case was the revenue. During the modelling phase, the predictive model will receive the historical input and output data in order to find patterns within the data. Different models are created with different algorithms at the basis of each model. Algorithms such as linear regression, decision trees, random forest, k nearest neighbour and recurrent neural network could be used.

The fifth step represents the evaluation phase where the accuracy of each candidate model will be assessed and compared against each other. After the training phase, the algorithm will apply what it has learned to unseen input data. The new input data will be structured in the same way. To do so, the predictive model will be given the testing data where only the input data will be given to it and will predict the targeted value based on the input data. Afterwards, the predicted targeted value will be compared with the actual targeted value. There are mainly two metrics used to assess the accuracy of the predictive models which are the MAE and RMSE metrics

The last step consists at deploying the predictive model as a web service. The predictive model will be deployed on a cloud platform such as Microsoft Azure platform, AWS Lambda or Google Cloud and application will get access to the predictive model thanks to an API. Companies will get a real-time forecasting of their financial figures on their applications.

Last but not least, an analysis has been made in order to strike the balance between the pros and cons of the implementation of a predictive machine learning model for financial forecasting purposes. The author believes that using machine learning models to forecast future revenues, costs or margins could be very useful and could drastically increase forecasting accuracy. Many advantages could be linked with the use of predictive machine learning models into their financial forecasting process such as being able to capture all the available data, bring very deep insight from the company's data, use the revenues drivers, memorize and learn autonomously which improves accuracy by time, learn from the dataset as a whole (no forecast of different products but a forecast with many products that are interconnected), get an automated financial forecasting process, get a real-time financial forecasting process.

Nevertheless, companies should bear in mind that some drawback are also present when using predictive machine learning models. Drawbacks would be the Increase of IT risk, relying on a black-box technology, huge amounts of quality data required, lots of money and time consuming before getting an accurate machine learning model. These elements could potentially affect the implementing and use of a machine learning model. The objective was to demonstrate that a well-trained machine learning model can bring a lot of value into the financial forecasting process of companies but it also comes with drawbacks that companies need to be aware of and consider before and after implementing it into their financial forecasting process.

# 4. Business case: UCB

## 4.1 Introduction

It's interesting to analyze the case of a company that already has experienced the use of predictive analytics into their financial forecasting process. Within this business case, the company that will be analyzed is UCB. UCB is a Belgian pharmaceutical company which total revenue grew to 4.6 billion euros in 2018. (UCB, 2018) The company counts around 7500 employees within 40 countries. (UCB, 2018) UCB mission is to help patients with neurology and immunology disorders. UCB is experiencing the implementation of machine learning models into their financial forecasting process since last year. For UCB, implementing such technologies is essential in order to develop their competitive advantage and attract talents. (Lieutenant, 2019)

Every quarter, they need to release their performance figures and communicate their prediction for the following months and quarters. In addition, they also need to provide financial figures internally to the management team. With the implementation of predictive analytics, their aim was to get a more accurate, faster financial forecast and improve the decision-making process. At the time being, UCB use predictive analytics to give direction but isn't totally replacing financial analysts and forecasters. They still use financial analysts to provide insight and try to justify what has been predicted by the predictive algorithms. According to Arnaud Lieutenant, head of AI research, pharmaceutical companies do possess a good business design for predictive analytics because once a product is on the market for 6 months, they can more or less predict easily what will be the revenue curve of that product. This makes the prediction of revenues easier for pharmaceutical companies. (Predictive analytics at UCB, 2019) Moreover, as they are trying to curse patients with chronical diseases, they have the advantage that they approximatively know how much of these patients need to be cursed. In addition, they have important information about their competitors' products which provide a good hint concerning future revenues of their products. Predicting their future products revenue still remain difficult but these elements make it easier for them to predict the revenue curve of their product and use predictive analytics (Lieutenant, 2019)

## 4.2 The transition to predictive analytics

Before adopting predictive analytics within their financial forecasting process, they were contracting around 80 financial analysts to generate these financial forecasts. It is a heavy task that requires a lot of time. These people were spending all their time at producing the forecast instead of trying to analyze the results of the forecasted figures produced and understand where the company is going. Since they implemented predictive analytics, they can produce forecasts very quickly. However, these predictive models will only provide the forecasted revenues or/and costs but won't explain the logic behind the forecasted result. The path taken by the algorithm won't be communicated by the algorithm. This is considered at UCB as a real drawback of predictive analytics because they cannot just take into account a random forecasted result without understanding what's behind. Today, financial analysts at UCB use predictive analytics to get a fast-forecasted baseline and spend most of their time trying to understand and provide insight from what has been forecasted by the predictive algorithms. Currently, UCB believes that artificial intelligence isn't mature enough to completely replace financial analysts and it won't be the case shortly. (Lieutenant, 2019)

## 4.3 What item lines are forecasted

They predict the top lines of their profit and loss. It's also called a short profit and loss statement. The most important elements that UCB forecasts are the revenues by product and by subsidiary. Besides, they notified that it was possible to also do it by nature of products. Another very important element that they do predict with their predictive machine learning models are the marketing and clinical costs. These two costs are the most significant one's for UCB. For the prediction of their costs, it can be more complex than predicting revenues because they are doing business in 40 countries and there are variables such as the cost of raw material, the variation of product prices and the cost of people that could vary from country to country and by time. Consequently, it impacts the overall prediction of costs. (Lieutenant, 2019)

## 4.4 The journey of UCB

In order to build their predictive models, they use a lot of historical data. They could collect the required historical data thanks to their global wide SAP enterprise resource planning application. In theory, one of the conditions to take advantage of predictive analytics was the completion of a digital transformation. UCB went through a data & digital journey before implementing predictive analytics and it took them around 5 to 6 years to implement a global wide ERP system for UCB Group. (Lieutenant, 2019) According to Arnaud Lieutenant, UCB can currently take advantage of predictive analytics because they began their date and digital journey many years ago. It enables them to capture huge amounts of historical data (from previous profit and losses and exogenous drivers) and use it for predictive machine learning purposes. (Predictive analytics at UCB, 2019) They could capture the profit and losses of all products and subsidiaries from the past 6 years (36 months). All these profit and loss data were very clean and structured because they had the same structure. UCB could get very complete and accurate prepared datasets. As a consequence, at UCB, there were very little pre-processing tasks to perform on their prepared datasets which was a real advantage for them. This pre-processing task could take a lot of time when the captured information isn't well structured and complete.

## 4.5 What solution is used and how did they build their predictive models

There are many ways to build a predictive model for financial forecasting purposes. UCB decided to use "R" and "python" frameworks on a Microsoft machine learning services platform. The first reason is that a lot of data scientists do work on these frameworks. The second reason is that it brought them more flexibility and enabled them to build tailor-made algorithms. Nowadays, these algorithms are at the basis of their predictive models. In order to build their predictive models, UCB hired a data scientists' team. It took them more or less 10-12 months to get some pretty accurate predictive models for their products revenue, clinical and marketing costs. It's important to notice that they do not use the same predictive model for all their top lines. They applied all their predictive algorithms to all their top lines in order to adopt the best predictive algorithm for each item line. An example of algorithm they are using is "random forest" which is part of the decision tree algorithms family. (Lieutenant, 2019)

Concerning the input variables taken into account to predict their future revenues, they use the historical revenues of each product and some impactful features such as information about their competitors' products, the evolution of their market shares, their marketing spending's… From a theoretical and data science point of view, revenues should only be predicted based on exogenous features that impact the revenue prediction. UCB tried to build predictive models only based on features, business drives without previous revenues. However, these predictive models became too complex and UCB experienced a decrease of forecasting accuracy. As a consequence, UCB isn't only using predictor variables as input data because they notified that time-series are more accurate. It means that the predictive machine learning models use historical product revenues as input data and some important predictor variables previously mentioned. (Lieutenant, 2019) Here, it's possible to point out that there is a divergence between the theory and the practice.

Below, the table shows the logic UCB uses to forecast their product revenues. The numbers (in millions of euros) aren't reflecting the real numbers of UCB. They have been set up to provide an understanding of UCB financial forecasting approach.

| | | Historical products revenues | | | | Future products revenue | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Y1 | | | | Y2 | | | |
| | Products/Time of the year | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| Loop 1 | Product 1 | 60 | 50 | 45 | 53 | **47** | ? | ? | ? |
| Loop 1 | Product 2 | 12 | 17 | 30 | 19 | **25** | ? | ? | ? |
| Loop 1 | Product 3 | 4 | 2 | 7 | 4 | **10** | ? | ? | ? |
| Loop 1 | Product 4 | 75 | 80 | 82 | 74 | **65** | ? | ? | ? |
| Loop 2 | Product 1 | | 50 | 45 | 53 | 47 | **45** | ? | ? |
| Loop 2 | Product 2 | | 17 | 30 | 19 | 25 | **22** | ? | ? |
| Loop 2 | Product 3 | | 2 | 7 | 4 | 10 | **3** | ? | ? |
| Loop 2 | Product 4 | | 80 | 82 | 74 | 65 | **72** | ? | ? |
| Loop 3 | Product 1 | | | 45 | 53 | 47 | 45 | **?** | ? |
| Loop 3 | Product 2 | | | 30 | 19 | 25 | 22 | **?** | ? |
| Loop 3 | Product 3 | | | 7 | 4 | 10 | 3 | **?** | ? |
| Loop 3 | Product 4 | | | 82 | 74 | 65 | 72 | **?** | ? |

*Figure 50: Financial forecasting methodology of UCB*

In the first loop, the model will be told that when it is given as input data the product revenues of Q1, Q2, Q3, Q4 of Y1 and the value of the different predictor variables (which are binary features), it should provide as output data the product revenue of Q1 of Y2 (numbers in bold). The same will happen the next quarter. In the second loop, the predictive model will be told to predict the product revenue of quarter 2 of year 2 (number in bold) based on the input data which is again the historical product revenue of Q2, Q3, Q4 of Y1, Q1 of Y2 and the value of predictors variables such as product maturity, market shares, competitor products' information... Once the predictive model is well trained, it will be tested with testing data. This is what happens in loop 3. The model is given new data but with the same input data variables and the model needs to predict quarter 3 of year 2. (Lieutenant, 2019) It's important to point out that they always do provide as input data some predictor variables that are strongly impacting their revenues prediction such as information about the maturity of their products, information about their competitors, market shares… Below, a figure shows the design of UCB advanced forecasting process.

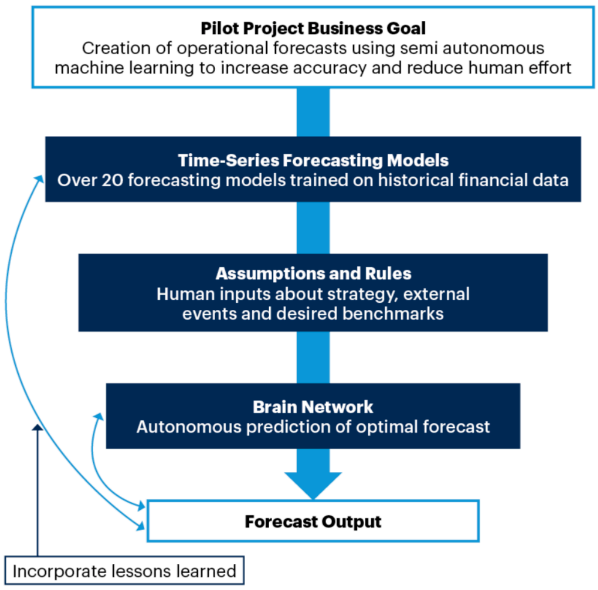


*Figure 51: UCB advanced forecasting process (Gartner, 2019)*

In a nutshell, they are using time-series forecasting models which receive as input data historical revenue data and predictor variables. They do not only use exogenous predictor variables as input data because then their predictive models become too complex and predict with a lower accuracy. So, UCB does not necessarily use the most complex predictive models but it is done on purpose because they notified that simple models forecasting accuracy often is better than complex models. (Lieutenant, 2019)

## 4.6 The impact of predictive analytics on their financial forecasting process

Predictive analytics has proven to be a real added-value for UCB. They can forecast the top lines of their profit and loss such as their product revenues and their clinical, marketing costs. According to Arnaud Lieutenant, predictive analytics do in average forecast more accurately than traditional forecasting methods. However, the increase of forecasting accuracy couldn't be communicated for confidentiality reasons. (Predictive analytics at UCB, 2019) Predictive analytics is used as a forecasting baseline and is discussed with business people to understand the reasoning behind and determine the final forecasted results. At UCB, predictive analytics is seen as a way of giving direction and help to faster forecast financial figures. Financial forecasters get hints if the trend is increasing, stagnating or decreasing. According to UCB, it is important to point out that it isn't there to replace all financial analysts but it is there to improve the forecasting process. For the pharmaceutical company, financial analysts are still needed and used at UCB to understand and explain what has been forecasted by machine learning models. However, there is still a lack of transparency and insight from what is forecasted when using machine learning models. At UCB, keeping financial analysts and using predictive analytics are going hand in hand. They believe that both are needed but do not believe that predictive analytics will totally replace the job of a financial analysts shortly. (Lieutenant, 2019)

However, there are some limits and drawbacks at using predictive analytics at UCB. The first drawback is that they sometimes experienced important gaps between what has been forecasted by the predictive models and what has been predicted by financial analysts. There was lots of internal discussion about what forecasted result is right? Which forecasted result is the one taken into account? In most of the cases, the algorithms were closer to the right result. However, it happens that the algorithms do predict wrong and that the error is far more important than an error made by financial analysts. This is also the reason why their financial forecasting process does not entirely rely on predictive analytics. They cannot take for granted what has been predicted by the predictive algorithms. These errors could cost them millions of euros if this happens on their most profitable markets such as the United States. In addition, they could experience that their predictive models only could forecast with a time-horizon of 2 to 3 years. Behind that time-horizon, their predictive models do perform with very low accuracy because time-series isn't complex enough. They'll need to add more exogenous predictor variables (Lieutenant, 2019)

Secondly, there is still the problem of the black-box. It means that the algorithms provide forecasted results without providing insight on how it came to these forecasted results. Financial forecasters at UCB spend their time at understanding and finding insight within the reasoning of the predictive algorithms. This is an essential reason why UCB cannot totally rely on predictive analytics. According to Arnaud Lieutenant, the technology isn't mature enough for UCB to totally rely on it.

And last but not least, using predictive analytics do challenge the way managers within a company are incentivised. Today, all companies do incentivise its employees to reach their objectives and budget. As a consequence, management teams within companies take this into account when they forecast their revenues and costs. When using predictive analytics, this type of social pressure that exists within companies isn't taken into account. There could be a difference between what people want to forecast as financial figures and what predictive analytics do forecast. This creates tension and potential conflicts of interest between the way companies are motivating, incentivizing managers to perform and achieve their objectives. (Lieutenant, 2019)

## 4.7 Conclusion

UCB didn't take a final decision concerning the final role of predictive analytics within their financial forecasting process because they are still in a learning path. However, they do believe that predictive analytics improves their financial forecasting process even if there are still some matters discussed internally. They are of the opinion that they will always need both predictive analytics and financial analysts to get the best forecasted results possible. However, they didn't strike a balance between the quantity of financial analysts needed and the use of predictive analytics. According to Arnaud Lieutenant, UCB will need some time to better understand the place of predictive analytics within their financial forecasting process but they strongly believe that predictive analytics has a place within their financial forecasting process and they will for sure continue to invest in it. (Predictive analytics at UCB, 2019)

## 5. Conclusion

The objective of the author through this thesis was to demonstrate how important it is to leverage data and machine learning algorithms in order to predict financial figures. These algorithms will learn on its own by time and will find patterns within the data. Companies are operating into an uncertain environment and could use data to back-up their decision takings. Current forecasting techniques aren't always able to plainly take advantage of the available data which isn't the case when using machine learning algorithms which enable companies to deal with huge amounts of data and find patterns within the data. Moreover, the author wanted to point out and explain the two types of machine learning solutions in order to provide the readers with an overview of the current market solutions. Last but not least, the author was willing to demonstrate the different steps of a data science project in order to implement a predictive machine learning model. This would give a practical view on how companies could implement machine learning into their financial forecasting process. Last but not least, the author decided to analyze the case of UCB which has implemented predictive analytics last year. It provides readers with a practical feed-back of the use of predictive analytics for financial forecasting purposes and highlights the advantages but also the on-going challenges of it.

Companies are always looking to increase their overall performance. With the use of machine learning into their financial forecasting process, companies are acquiring a real competitive advantage. The finance function is becoming a real business partner that helps the top management team to improve the decision making because it won't spend time anymore at generating the information. Financial figures will be generated by the predictive models and human will be able to spend time to perform more added-value tasks and think about business decisions that need to be taken. Using machine learning provides companies with a real-time and automated financial forecasting process which reflects the company reality. To do so, machine learning algorithms will forecast financial figures based on the historical data which could be internal and external data. There will be less human and organizational bias into the forecasting process because machine learning algorithms will only rely on the provided data. However, companies still need to be very careful about the quality of data provided to the models when it's trained. Moreover, machine learning algorithms are progressively learning which means that it will perform better by time because it will remember what it has learned before and because it will be trained with more data. The more data, the better the predictive models become. In addition, machine learning models will add some intelligence to existing business intelligent tools in order to take better decisions.

Companies that believe machine learning could improve their financial forecasting process should know that there are two types of solutions on the market that enable them to build a predictive machine learning model. These two solutions have the same purpose which is to build a predictive model but what these two solutions offer in order to achieve this objective is completely different.
On one hand, there is a type of solution that offers a platform where data scientists can build their own algorithms and need to define the whole workflow from the business understanding phase to the deployment phase. This type of solution provides companies with a lot of flexibility but will also require more time to build the model. Solution such as Azure machine learning services, Amazon sage maker, Google cloud ML engine could be adopted. On the other hand, there is a type of solution that provides a semi-automated solution where pre-built algorithms are already available. It makes the workflow easier and accessible to business people that do not necessary possess coding skills. However, from the author point of view, it's always advised to hire some data scientists when building a predictive model.

Moreover, the author pointed out that using machine learning for financial forecasting purposes isn't a good fit for all companies. There are some conditions that need to be fulfilled to take the most out of predictive analytics. The first condition is to get huge quantities of historical data and being a transaction intensive company. The second condition is about being a B2C company because these companies have huge uncertainty into their financial figures forecasting. This isn't necessary the case with B2B companies which provide services on the long-term because their revenues for example will be based on a lead/opportunity approach and thus their future revenues are a result of the "leads" and "opportunities" from previous months. From the author point of view, these B2B companies providing services should better implement a machine learning model to predict either or not an opportunity is going to be closed. The third condition is that these companies already undergo a digital transformation that gives them the means to capture the required data to feed the model. In the UCB case, it has been mentioned that they needed 5 to 6 years to implement an ERP system that enables them to capture the historical data about their products' revenue, marketing and clinical costs from the past 6 years. And last but not least, these companies should get money and time because going through the whole process of building a predictive model can take some time. Moreover, once predictive analytics is implemented, companies will still need time and efforts to face challenges coming along with it. With the UCB case, it has been pointed out that there are some drawbacks that create some difficulties. Even after the implementation phase, companies need to experience the use of it and take decisions about its place within the company. In a nutshell, using machine learning to forecast financial figures isn't something that all companies could take advantage of as long as they do not fulfil the previous mentioned conditions.

Once companies are aware of the benefits that machine learning could bring into their financial forecasting process and that they have analyzed the fit between the requirements and their business, they can identify the solution that could be adopted within their company financial forecasting process. However, these companies need to go through a whole process in order to build their predictive machine learning model with the use of their solution.

The first step is about understanding the business and determining what type of tasks will be performed. This needs to be translated into a specific data mining task such as for example; predicting the revenue of the ice cream for a certain period.

Afterwards, companies need to understand the data required and define the features that will be used as input data. Afterwards, these companies will collect those data. The third step consists at pre-processing the data. Companies will clean and transform the data to get a prepared dataset to provide to the predictive model. Multiple tasks will be done to get a clean dataset such as removing errors, analyzing outliers, fulfilling missing values, transforming categorical values into numerical values, normalizing the data, adapting the number of features and splitting the dataset into training and testing data. This second and third step is essential to make sure that the predictive model will learn from quality data. This will increase drastically its prediction accuracy.

The next step is called the modeling phase where companies select a specific algorithm for each candidate model and train it with the training data. The features are provided as input data to the model and are given the corresponding output data. At UCB, they predict future revenues for each product and each subsidiary. In order to do so, they provide as input data to their predictive models; previous revenues and some predictor variables such as market share, product maturity, competitors product information... This will give the means to the predictive model to discover patterns within the historical data. Afterwards, the model will be tested. To do so, the model will be given new data from the testing dataset and will be asked to predict the product or subsidiary revenues (which is the output value). The predicted targeted value is compared with the actual value in order to determine the model accuracy. The training and testing phase has been done with different models using different underlying algorithms such as linear models, KNN, decision trees, recurrent neural networks… These models will be compared to each other's so that companies can adopt the most accurate one. The most accurate predictive model is the one that predicted a targeted value that is the closest to the actual targeted value. At UCB, they tested all their algorithms on each item line in order to use the best algorithm possible for each item line. This proves that UCB really compared their different models for each top line of their profit and loss.

The last implementation step consists at deploying the model to use for real business predictions. The predictive model is uploaded on the cloud and application gets an access to it through an API. However, this step is still a challenge because sometimes there are issues of code compatibility between the machine learning solution used to build the model and the cloud platform where the predictive model is hosted.

Concerning the hypothesis made by the author at the beginning of his thesis, they haven't all been confirmed:

- "Once implemented, it would automate the whole financial forecasting process". This is certainly the final purpose. In theory, it is totally possible. However, there are still practical issues today that make it impossible for companies to rely on a fully automated machine learning forecasting process. At UCB there are issues faced concerning change management, lack of transparency from the results forecasted by the algorithms, situations where the algorithm and financial analysts do forecast a completely different result. This proves that companies can't today fully rely on predictive analytics. The technology isn't mature enough.
- "It will improve companies financial forecasting accuracy". The author believes it is true but it is still difficult to quantify it because few companies already implemented it and they aren't willing to share their forecasting accuracy improvement for confidentiality reasons.
- "Artificial intelligence could be implemented by all companies". It is false, there are specific requirements that need to be fulfilled such as the transaction level, historical data availability, the type of business model, having engaged a digital transformation and having enough money and time at disposal.
- "Machine learning automates the whole implementation process, no human intervention anymore during the implementation". This is completely false because humans need to define the workflow, define the required data, collect the data, pre-process the data, select the algorithms, deploy it and monitor it once implemented. Machine learning need supervision during the whole implementation process. Even after the implementation process, humans need to monitor the predictive models.
- "People willing to implement a predictive model need coding skills". In theory, it isn't true because there exists drag and drop solutions to make predictive analytics available to business people. These solutions have a low learning curve and are accessible for people without coding skills. However, in practice, it's recommended to hire a data scientist team to pre-process the data, build algorithms and train the predictive models.

From the author point of view, there is a very important aspect concerning the use of a predictive machine learning model that could lead to further investigation. It would be the challenge of change management and the buy-in of business people with the use of machine learning for financial forecasting purposes. As it has been mentioned within the business case of UCB, financial forecasting remains a strategic task for companies and machine learning algorithms don't provide enough transparency to let business people trust what the predictive model has forecasted. Business people shouldn't blindly trust a predictive model that forecasts a result without explaining how that result has been achieved. Moreover, diminishing human intervention into the decision-making process is a very sensitive subject that could lead to conflicts of interest. It has been explained that in practice, it could go in opposition with the incentive system on which companies rely today. Moreover, is it ethical to use algorithms to predict strategic elements such as revenues, costs… Knowing that these wrong predictions could impact companies' and employees' wealth. What would happen the day a company completely relies on a predictive model and that it makes a completely wrong prediction? What if it impacts the company and its employees? Who's responsible in the case of a wrong prediction? Humans can do errors, why wouldn't machine be authorized to do errors? The other strongly believes that all companies should try to experience the use of predictive analytics within their financial forecasting process and determine its position within their financial forecasting process. Implementing predictive analytics and knowing its place within a company takes time and all companies should invest in it as soon as possible.

## 6. Critical analysis and limits of the thesis

The author really enjoyed acquiring knowledge about the use of artificial intelligence, more specifically on "predictive analytics" and its impact on the financial forecasting process of companies. As for every thesis, there are always points that can be improved and the author is aware of it. After writing this thesis, the author puts his work into perspective in order to highlight some improvement points. From the author point of view, some improvement points could be:

- The shortcomings of today's financial forecasting processes that have been enumerated into this thesis are based on my internship experience and the experience of the interviewed experts. A survey could be done to get a larger sample of opinions. However, the author thought that the collected information was already enough to point out the most relevant shortcomings of today's financial forecasting processes;

- There are two types of machine learning solutions that have been explained and for each of them, several machine learning solutions have been enumerated. However, there is no analysis comparing the capabilities of the specific machine learning solutions. The author didn't make one because he was more willing to show that there are different types of solutions available. His intention wasn't to make a presentation about the characteristics of each machine learning solution because all of them can be used for regression tasks;

- Collect more quantitative data about the impact of predictive analytics on the financial forecasting accuracy. However, nowadays, it's very difficult because there are still very few companies using predictive analytics into their financial forecasting process. The author strongly believes that machine learning algorithms can improve financial forecasting processes but it's still very difficult to assess its impact in terms of accuracy due to the low adoption rate by companies at the time being;

- The data understanding phase could be further improved with a concrete business case. The author didn't have the chance to implement a predictive model so he couldn't go more into details for that step. That step will widely vary in function of the company and its business. Revenues and costs drivers strongly depend on the sector, the type of products/services and the strategy of the company. In the future, if a student would do a project management thesis related to the implementation of a machine learning model for financial forecasting purposes, he may detail this phase by identifying the input variables (internal or external factors) used by the company as revenues or costs predictor variables;

- ⇨ The phase where the different algorithms are pointed out and explained could be improved by doing further research about the divergence between the main algorithms and their advantages. This hasn't been done because it would require some data science skills with a very good understanding on how to build an algorithm and how to configure its parameters;

- ⇨ The author tries to demonstrate through this thesis how to implement a predictive model for financial forecasting purposes. The limit would be that there is no explanation on how to code and build an algorithm. However, since the beginning, this wasn't the intention of the thesis. If companies do adopt semi-automated solutions, they won't need to code because pre-built algorithms are already available;

- ⇨ One of the biggest challenges about the use of predictive analytics for financial forecasting purposes is certainly change management and the buy-in of people within a company. This could have been a subject for further investigation. How should companies communicate their data strategy and intention to implement predictive analytics into their financial forecasting process? How to find a common ground between the use of machine learning models and the way companies do incentivize their managers? What would happen if one day AI predicts a wrong targeted value for a real-life case? Who is responsible? These types of questions require further investigation and could be for sure the topic of another thesis.

# 7. Bibliography


AI Business Information. (2018). *Risks and limitations of AI in business*. Retrieved from AI Business Information: https://www.nibusinessinfo.co.uk/content/risks-and-limitations-artificial-intelligence-business

Ait Amir, B., & El Mahrsi, K. (n.d.). *Machine learning: attention aux solutions clé en main.* Retrieved from Keyrus: https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-model-builder.html

Altexsoft. (2018, July 19). *Comparing machine learning as a service: Amazon, Microsoft Azure, Google Cloud AI, IBM Watson*. Retrieved from Altexsoft: https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai-ibm-watson/

Amazon. (2019). *Amazon API Gateway*. Retrieved from Amazon: https://aws.amazon.com/api-gateway/?nc1=h_ls

Amazon. (2019). *Amazon SageMaker*. Retrieved from Amazon: https://aws.amazon.com/sagemaker/

Amazon. (2019). *ML frameworks*. Retrieved from Amazon: https://aws.amazon.com/machine-learning/

Bista. (2016, Mai 31). *5 statistical methods for forecasting quantitative time series*. Retrieved from Bista : https://www.bistasolutions.com/resources/blogs/5-statistical-methods-for-forecasting-quantitative-time-series/

Bohr, N. (2007, July 15). The perils of prediction. Washington: The Economist.

Boscacci, R. (2018, Decembre 16). *What even is computer vision?* Retrieved from Towards Data Science: https://towardsdatascience.com/what-even-is-computer-vision-531e4f07d7d0

Bottefeux, M. (2018). *Importance of IT.* Brussels.

Brownlee, J. (2013, November 25). *A tour of machine learning algorithms.* Retrieved from Machine learning mastery: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Brownlee, J. (2016, March 11). *How machine learning algorithms work (they learn a mapping of input to output).* Retrieved from Machine learning mastery: https://machinelearningmastery.com/how-machine-learning-algorithms-work/

Bughin, J., Chui, M., & Mccarthy, B. (2017, August). *How to make AI work for your business*. Retrieved from Harvard Business Review: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-to-make-ai-work-for-your-business

Chappell, D. (n.d.). *Introduction to a machine learning* . San Francisco: Microsoft.

Coallier, N. (2019, April 4). Data Scientist.

Columbus, L. (2018, January 12). *10 charts that will change your perspective on artificial intelligence's growth*. Retrieved from Forbes:

https://www.forbes.com/sites/louiscolumbus/2018/01/12/10-charts-that-will-change-your-perspective-on-artificial-intelligences-growth/#a928b0e47583

Corporate Finance Institute. (2019). *3 Statement Model*. Retrieved from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/knowledge/modeling/3-statement-model/

Corporate Finance Institute. (2019). *Financial forecasting*. Retrieved from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/knowledge/modeling/financial-forecasting-guide/

Cote, G. (2019, March 27). Business Intelligence Consultant. (M. Stukkens, Interviewer)

Data Science. (n.d.). *Minear prediction models*. Retrieved from Data Science: https://www.datascienceblog.net/tags/linear-model/

Davenport, T., & Ronanki, R. (2018, February). *Artificial Intelligence for the real world*. Retrieved from harvard Business review: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

Donges, N. (2018, February 26). *Recurrent Neural Networks and LSTM*. Retrieved from Towards Data Science: https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5

Donges, N. (2018, February 22). *The random forest algorithm*. Retrieved from Towards Data Science: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

Dr. Garbade, M. (2018, Octobre 15). *A simple introduction to natural language processing*. Retrieved from Medium: https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32

Gartner. (2019). *3 machine learning myths for forecasters.* Gartner.

Gazarov, P. (2016, August 13). *What is an API? In English, please.* Retrieved from Medium: https://medium.freecodecamp.org/what-is-an-api-in-english-please-b880a3214a82

Google. (2019). *Cloud Auto ML*. Retrieved from Google: https://cloud.google.com/automl/

Google. (2019). *Cloud machine learning engine*. Retrieved from Google: https://cloud.google.com/ml-engine/

Gupta, P. (2017, May 17). *Decision Trees in machine learning.* Retrieved from Towards Data Science: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice.*

IBM. (2019). *IBM Watson Studio*. Retrieved from IBM: https://www.ibm.com/be-en/marketplace/watson-studio/details

IBM. (2019, February 28). *Model builder overview*. Retrieved from IBM: https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-model-builder.html

Iqbal, M. (2019, February 27). *Uber revenue and usage statistics*. Retrieved from Uber: http://www.businessofapps.com/data/uber-statistics/

K. Pratt, M. (2017, Septembre 1). *What is BI? Business intelligence strategies and solutions*. Retrieved from CIO: https://www.cio.com/article/2439504/business-intelligence/business-intelligence-definition-and-solutions.html

Keyrus. (2019). Corporate performance. Brussels.

Lies, S., Parker, S., & Reader, G. (2017). *Forecasting with confidence.* KPMG.

Lieutenant, A. (2019, May 7). Predictive analytics at UCB. (M. Stukkens, Interviewer)

Malik, F. (2018, August 27). *Machine learning algorithms comparison*. Retrieved from Medium: https://medium.com/fintechexplained/machine-learning-algorithm-comparison-f14ce372b855

Malik, F. (2018, Novembre 29). *Processing data to improve learning models accuracy*. Retrieved from Medium: https://medium.com/fintechexplained/processing-data-to-improve-machine-learning-models-accuracy-de17c655dc8e

Masnaoui, M. (2019, April 30). Limits of forecasts and building a machine learning model. (M. Stukkens, Interviewer)

McKinsey & Company. (2019). *An Executive's guide to AI*. Retrieved from McKinsey & Company: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai

McKinsey & Company. (2019). *How it works.* Retrieved from McKinsey & Company: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai

McKinsey & Company. (2019). *Machine Learning*. Retrieved from McKinsey & Company: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai

Microsoft. (2014, September 2). *Regression: Demand forecasting*. Retrieved from Microsoft: https://gallery.azure.ai/Experiment/Regression-Demand-estimation-4

Microsoft. (2018, April 4). *What is Azure machine learnign service?* Retrieved from Microsoft: https://docs.microsoft.com/en-us/azure/machine-learning/service/overview-what-is-azure-ml

Microsoft. (2019, April 3). *Machine learning algorithm*. Retrieved from Microsoft: https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-cheat-sheet

Microsoft. (2019, April 20). *What is azure machine learning studio?* Retrieved from Microsoft: https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio

Microsoft. (n.d.). *Dig deep with azure machine learning.* Retrieved from Microsoft: https://docs.microsoft.com/en-us/azure/machine-learning/studio/basics-infographic-with-algorithm-examples

Mishra, A. (2018, February 24). *Metrics to evaluate your machine learning algorithm*. Retrieved from Towards Data Science: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

Muddassir, A. (2018). *5 reasons why machine learning forecasting is better than traditional forecasting techniques*. Retrieved from SCM: https://www.scmdojo.com/machine-learning-forecasting-better/

Murray , M. (2018). *Protiviti*. Retrieved from Seeing the future more clearly: how machine learning can transform the financial froecasting process: https://www.protiviti.com/US-en/insights/how-machine-learning-can-transform-financial-forecasting-process

Nasyrov, D. (2017, September 19). *Machine learning linear models*. Retrieved from Medium: https://medium.com/pharos-production/machine-learning-linear-models-part-1-312757aab7bc

Networking technologies. (2018, October 2). *What is machine learning.* Retrieved from Networking technologies: https://www.net-cloud.com/blog/what-is-machine-learning/

New Generation Applications. (2017, Novembre 7). *Artificial Intelligence vs Machine Learning vs Data Science*. Retrieved from New Generation Applications: https://www.newgenapps.com/blog/artificial-intelligence-vs-machine-learning-vs-data-science

Nikkhah, M. (2018). *How machine learning finds network trouble faster than anyone*. Retrieved from Cisco: https://www.cisco.com/c/m/en_us/network-intelligence/service-provider/digital-transformation/get-to-know-machine-learning.html

OECD. (2019). *Households' economic well-being: teh OECD dashboard*. Retrieved from OECD: https://www.oecd.org/sdd/na/household-dashboard.htm

Owen, K. (2019, January 30). *Why machine learning forecasting*. Retrieved from Projects Small: https://smallaprojects.com/why-machine-learning-forecasting/

Patience, N. (2016, December 20). *Artificial intelligence, machine learning and deep learning.* Retrieved from Nick Patience: https://twitter.com/nickpatience/status/811247390770376704

Pointcaré, H. (2015). *The foundations of Sciences.* Scholar's Choice.

Press, G. (2016, March 23). *Cleaning big data: most time-consuming, least enjoyable data science task*. Retrieved from Forbes: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#143da8916f63

PwC. (2011). *Financial planning: Realizing the value of budgeting and forecasting.* Retrieved from PwC: https://www.pwc.com/my/en/assets/services/realizing-the-value-of-budgeting-n-forecasting.pdf

Roberson, N. (2014, February 8). *How business intelligence helps you understand your consumer*. Retrieved from Business to Community: https://www.business2community.com/business-intelligence/business-intelligence-helps-understand-consumer-0770764

Roman , V. (2018, Decembre 23). *How to develop a machine learning model from scratch.* Retrieved from Towards data science: https://towardsdatascience.com/machine-learning-general-process-8f1b510bd8af

Rosten, L. (1980). Forecasting quotes. Retrieved from https://allauthor.com/quotes/106691/

Safdari, N. (2018, November 26). *If you can't measure it, you can't improve it!* Retrieved from Towards Data Science: https://towardsdatascience.com/if-you-cant-measure-it-you-can-t-improve-it-5c059014faad

SAP. (2019). *SAP Leonardo Machine Learning Foundation*. Retrieved from SAP: https://www.sap.com/products/machine-learning-foundation.html#product-overview

SAS. (2018). *Artificial intelligence for Executives: Integrating AI into your organization.* Retrieved from SAS: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/artificial-intelligence-for-executives-109066.pdf

SAS. (2018). *Visual Forecasting Benefits*. Retrieved from SAS: https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-visual-forecasting-109226.pdf

SAS. (2019). *Artificial Intelligence History*. Retrieved from SAS: https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html#used

SAS. (2019). *Deep Learning*. Retrieved from SAS: https://www.sas.com/en_us/insights/analytics/deep-learning.html

SAS. (2019). *Forecasting & Optimization*. Retrieved from SAS: https://www.sas.com/en_ca/solutions/ai.html

SAS. (2019). *SAS Visual Data Mining and Machine Learning*. Retrieved from SAS: https://www.sas.com/en_ca/software/visual-data-mining-machine-learning.html

SAS. (2019). *What are the challenges of using artificial intelligence*. Retrieved from SAS: https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html#used

SAS. (2019). *What happens when you add learning and automation to analytics?* Retrieved from SAS:

https://www.sas.com/content/dam/SAS/en_us/doc/infographic/artificial-intelligence-109282.pdf
SAS. (2019). *What is a data scientist?* Retrieved from SAS: https://www.sas.com/en_ca/insights/analytics/what-is-a-data-scientist.html
SAS. (2019). *Why is AI important?* Retrieved from SAS: https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html#howitworks
Sayad, S. (n.d.). *Decision tree regression.* Retrieved from Saed Sayad: http://www.saedsayad.com/decision_tree_reg.htm
Scherbak, M. (2019, March 25). *Data science vs Business intelligence: same but completely different*. Retrieved from Towards Data Science: https://towardsdatascience.com/data-science-vs-business-intelligence-same-but-completely-different-1d5900c9cc95?fbclid=IwAR1cx-YcK4euLXVotIIss1xfSl99Umvnyi1N4mnUF6fzh5_QdHiRcRLN39c
Simplilearn (Director). (2018). *How KNN algorithm works* [Motion Picture]. Retrieved from https://www.youtube.com/watch?v=4HKqjENq9OU
Skymind. (n.d). *A beginner's guide to LSTMs and recurrent neural networks*. Retrieved from Skymind: https://skymind.ai/wiki/lstm
Stevens, B. (2018, April). Management Accounting Control. *Planning, Budgeting and Forecasting*. Brussels, Belgium.
T. Larose, D., & D. Larose, C. (2015). *Data Mining and predictive Analytics.* Wiley.
TopTal Research. (2018). *Ai vs BI: differences and synergies*. Retrieved from TopTal Research: https://www.toptal.com/insights/innovation/ai-vs-bi-differences-and-synergies
Towards Data Science. (2016, March 23). *MAE and RMSE - Which metric is better?* Retrieved from Towards Data Science: https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d
Twin, A. (2019, April 19). *Delphi method*. Retrieved from Investopedia: https://www.investopedia.com/terms/d/delphi-method.asp
UCB. (2018). *Our company*. Retrieved from UCB: https://www.ucb.com/our-company/
Valinsky, J. (2018, Septembre 28). *Walgreens knews its profit forecast was wrong but didn't tell investors*. Retrieved from CNN Business: https://money.cnn.com/2018/09/28/news/companies/walgreens-sec-settlement/index.html
Van der Meulen, R. (2019, April 2). *Financial forecasters should beware 3 machine learning myths.* Retrieved from Gartner: https://www.gartner.com/smarterwithgartner/financial-forecasters-should-beware-3-machine-learning-myths/
Vandeput, N. (2018). Forecasting and unlocking AI power. Brussels.

William McGhee, L. (n.d.). *What are the pros and cons of machine learning?* Retrieved from SevenTablets: https://seventablets.com/blog/what-are-the-pros-and-cons-of-machine-learning

Zhou, L. (2018, Septembre 5). *How to build a better machine learning pipeline.* Retrieved from Datanami: https://www.datanami.com/2018/09/05/how-to-build-a-better-machine-learning-pipeline/