Dublin, Ireland. Jan, 2020

Why ML Projects won’t take you to the promised land?

Intelligence Report

Machine Learning is powerful
Without Data, there’s no ML
It is easy to make mistakes
Small mistakes can result in large consequences
Mistakes are avoidable…

Executive Summary

This document analyses recurring mistakes and problems organizations face when conducting internal Machine Learning projects.

Machine Learning (ML) is a combination of technologies with the capability of providing insights from large amounts of data faster than a human. Its business value is currently in a state of inflated expectations, which can cause problems when organizations undertake internal projects without prior knowledge.

Thingbook.IO personnel have analysed the lessons learned during dozens of ML projects and undertaken a review of available literature, followed by a set of workshops with a sample group of customers. The following conclusions were drawn:

> For companies who decide to implement ML projects, there is a direct correlation between a miscalculation of Machine Learning value and the level of experience the organization has in it. Inflated expectations come from companies who lack ML knowledge without realizing that lack, at both the technical and managerial levels.
>
> Data related mistakes, such as not enough data, low quality of data, or biased data most severely impact the accuracy of results. This is a key differentiator with standard software development projects, where data related problems are usually identified during the development phase and are difficult, if not impossible, to foresee during project planning.
>
> It is difficult to accept the value of long-term Machine Learning projects, because there is no obvious Return on Investment (ROI) or the business case is unclear.

This document is structured as follows:

> An exhaustive review of the machine learning literature is presented including a list of both technical and managerial recurring mistakes. A managerial view of Machine Learning is also provided, as well as a short comparison between software development projects and Machine Learning projects.
>
> A “de facto” standard procedure for implementing machine learning projects is presented.
>
> The companies examined are split into groups based on their level of Machine Learning adoption, its maturity and the role of Machine Learning in their portfolio and processes. **Note**: *This document does not include companies commercializing Machine Learning Services or Data Analysis Products as their core business but focusses on how non-analytical companies can benefit from Machine Learning.*
>
> A conclusion is made about the correlation between identified mistakes and their impact on companies based on their categorization.

Intended Audience

For academic and theoretical purposes, this document is intended for the project management research stages of ML project implementation.

The conclusions outlined are useful for C-level executives in the early stages of implementing Machine Learning to optimize or automate their businesses.

Introduction

Context and Background

In a noted article published by Science Magazine in 2015, Jordan and Mitchell described Artificial intelligence (AI) as the discipline to address the question of how to build computers that can learn and improve automatically through experience. AI has also been defined as a science with the goal of making machines do things that would require intelligence if done by humans (Negnevitsky, 2005). Machine Learning, a technology to achieve AI, is also explained by Marsland (2015) as “adaptive mechanisms that enable computers to learn from experience, learn by adaption, and learn by analogy". With machine learning progressing rapidly over the past two decades, it is now used for computer vision, speech recognition, natural language processing, robot control and more (Jordan & Mitchell, 2015).

Over the last decade the term big data has emerged. It references the ability to manage large amounts of data used for Machine Learning. This emergence is due to technology being able to iterate through, and gain useful insights about the data, faster than humans.

The concept of artificial intelligence was first suggested by Alan Turing (1950), when he described the idea of machines acting as humans. Increased CPU and GPU speed with increased availability of data, has made the concept useful for companies in building business value. The AI trend has quickly grown from being a vast concept to an emerging technology that many companies are looking to integrate into their businesses (Gartner, 2019).

Every year Gartner provides a visual representation of “the maturity and adoption of technologies and applications", shown in figure 1.1 (Gartner, 2020). Machine Learning and Data Fabric are pervasively influencing or dominating the “Peak of inflated expectations" category.

With the recent rise of AI, the subject is heavily discussed but might not be living up to the expectations of some organizations. With Machine Learning on the peak of the hype cycle, organizations might not fully understand the technology and its drawbacks before starting implementation projects.

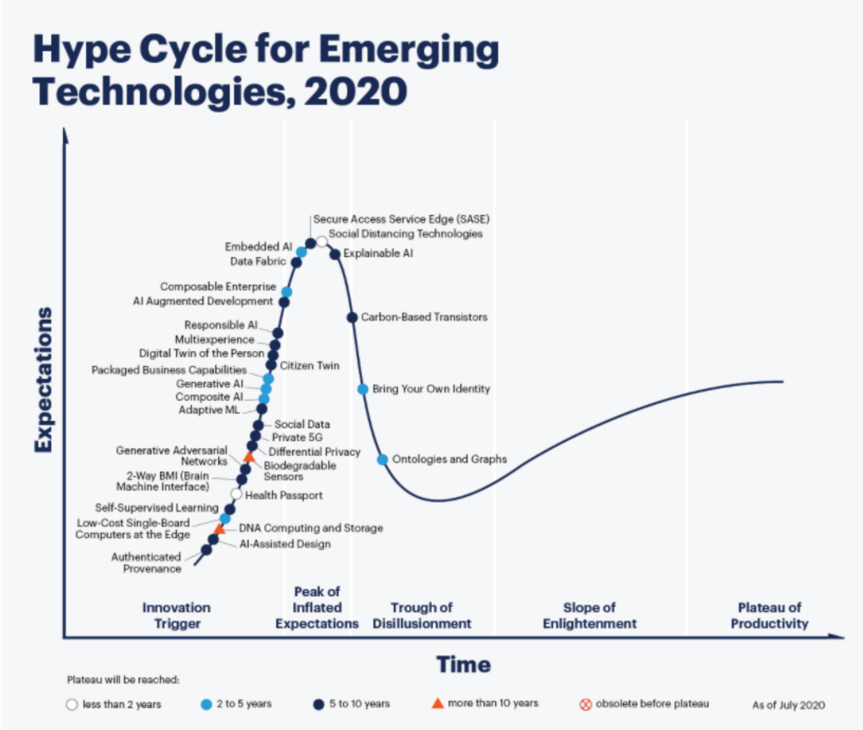

Figure 1.1: Gartner Hype Cycle for Emerging Technologies, 2020

Before entering the area of Artificial Intelligence and Machine Learning, organizations need to know which resources are required and which resources are beneficial for successful implementation and usage. When Gartner discusses inflated expectations within organizations, their reports highlighted the risk of companies neither possessing the knowledge nor realizing the lack of it. Equally important, is understanding the types of problem Machine Learning is useful in solving. Inflated expectations can also create unrealistic ideas about the problems Machine Learning can solve.

Because of the rise of ML technologies and the aforementioned factors, companies without previous experience are jumping into internal ML implementation projects. As such, mistakes are being made, which might otherwise have been avoided. Identifying the mistakes will make future implementations more likely to succeed.

Although Machine Learning is considered software implementation, the risks, mistakes, and problems are not necessarily the same, or even a subset of them. Therefore, it is necessary to understand how ML project mistakes might differ from standard software implementation mistakes.

Purpose

The purpose of this document is to recognize common misconceptions and mistakes that have been identified in 52 machine learning implementation projects and 8 technical and commercial workshops with technology-intensive companies. The companies have different levels of Machine Learning experience and implementation maturity. By identifying mistakes, future implementation projects can run more efficiently and avoid the generalist problems.

This document is an introduction to decision makers in the area of Machine Learning. It is designed to help them make better decisions on how much to invest, where to invest, the required internal knowledge, and more realistic ROI expectations.

Identifying mistakes in different stages of the implementation projects will help future project managers know which stages require more focus, which are not necessarily the same as software implementation projects, and therefore, do not correspond to previous experiences, as well as which have been proven to be underestimations of the Machine Learning factors of success.

Literature Review

This section describes some of the literature available and provides additional information about AI and ML as concepts. Later sections provide more technology intensive information about Machine Learning and a managerial perspective.

Machine Learning

Machine Learning is making computers modify or adapt actions, so they become more accurate based on experience. In other words, they are learning. The accuracy is measured by how well computer choices reflect correct decisions (Marsland, 2015).

Jordan and Mitchell (2015) say machine learning addresses the question of how to build computers that improve automatically through experience. Many developers of Machine Learning solutions recognize some problems are more easily solved by showing a machine examples of desired input-output behaviour, than programming it manually to respond to every input correctly (Jordan & Mitchell, 2015).

An example of a basic machine learning problem used for education by the data science platform Kaggle (2018), is predicting who would survive the Titanic catastrophe in 1912. Taking a set of data about a subset of the passengers including several features, such as their age, sex, whether they survived, ticket class, port of embarkation, number of family members on the boat, and so on, a machine can be shown this data and learn which of the dimensions have stronger influence in increasing the chances of survival. When the machine is then presented with another subset of data about other passengers, where the “survivor or not" indication is removed, it can, to a certain confidence level, predict who survived and who did not (Kaggle, 2018).

The previous two definitions imply there is a finite number of correct decisions, or answers, and, the same applies to the incorrect ones. They also imply there is a procedure to measure the correctness or incorrectness of the actions. The algorithms cannot process extra-data considerations, in consequence, the correctness or incorrectness of the proposed decision cannot be absolute, and will always be relative to the data, including a factor of bias in this process. Although this explanation might be thought obvious, the lack of understanding and explanation of the insights provided by ML, is a source of unrealistic expectations and distrust in Machine Learning. The following is a good example:

https://eu.usatoday.com/story/tech/news/2017/12/07/california-fires-navigation-apps-like-waze-sent-commuters-into-flames-drivers/930904001/

A popular navigation app was directing drivers to neighbourhoods where wildfires forced closures and evacuations in California, December 7, 2017. The algorithms are largely reliant on information supplied by drivers, when the Google-owned app notices gridlock on a user route, it reroutes the driver to quieter streets, even if they are empty because drivers have fled smoke and the threat of flames. So, the question arises, was the algorithm recommending the correct action? Unfortunately, the authors of the report do not have a correct answer to that question.

Stages of Machine Learning Implementations

Several proposals have been made in literature to describe the process followed by a Machine Learning project. Of course, all of them cover the process from a generic viewpoint and individual projects might require adjustments. For instance, the training and validation phases are only applicable to "Supervised Machine Learning algorithms", in contrast, "Unsupervised algorithms" do not have labelled examples and therefore, training and validation are not required.

Stephen Marsland describes in *Machine Learning: An Algorithmic Perspective* (2015) four stages of Machine Learning implementation projects:

Data Collection and Preparation is the process of collecting the data needed, as well as making sure the data sets are clean, meaning they do not have significant errors or missing data. The data could also need to be merged from different sources hosting the data on different forms. Part of this process is thereby ensuring the data is represented in the same place in the same format which could be a challenging task.

Also, the quantity of the data needs to be considered. Machine Learning usually is a computationally expensive process and with more data comes increased computational costs. Simultaneously, not having enough data makes the model less correct in its insights. There is a point at which there is enough data to build a representative model until satisfaction while using an amount of data that is not adding too much computational overhead to the process (Marsland, 2015). The balance between the amount of data to be considered for analysis and the associated computational cost is closely related with the problem to solve (the use case).

The second stage is **Feature Selection**. Features are the variables used for the algorithm, such as age, sex, or location. Feature selection is the process of selecting which features are (the most) useful for solving the problem and which are not useful (Marsland, 2015).

Algorithm Choice is choosing an algorithm that is appropriate to the problem. There are different algorithms that might suit different types of problems. For many algorithms, there are parameters that must be set manually, which happens in the stage Parameter and Model Selection (Marsland, 2015).

When the solution is implemented, the next stage is **Training**. This is "the use of computational resources in order to build a model of the data in order to predict the output of new data" (Marsland, 2015). Marsland also discusses a problem known as overfitting. Overfitting happens when the machine is over-trained on the training data and is not identifying a general generating function, but rather adapts the model to perfectly fit the training data. This reduces the generalization capacities of the model.

The last step is **Evaluation**, which involves testing the system on data it has not been trained on. This can include comparison with experts in the field results (Marsland, 2015).

Artificial Intelligence from a Management Perspective - an Analysis of the Future

Three studies have been made by Accenture, MIT in collaboration with Boston Consulting Group, and SAS (Ransbotham et al., 2017; Kolbjornsrud et al., 2016; SAS, 2017) regarding AI from a management perspective. The studies all focus on the status of artificial intelligence, future development and executive thoughts on this area of technology developing within their industries.

According to the studies, the expectation of AI solutions is higher than the actions taken to build or implement them. 85% of a global survey of over 3000 executives believe AI solutions will benefit their company in getting a competitive advantage while only 23% have incorporated AI in any part of their business (Ransbotham et al., 2017). SAS (2017) show that around two-thirds of the participants in the study claim to believe that “AI would have a wide-ranging effect in the next five to ten years". The study also shows that while optimism regarding AI was high among organizations, fewer thought their organization was ready to reap the benefits they were positive about, to which SAS suggest possible problems with execution. The MIT study shows there are large gaps between today's leaders who have adopted AI, who thereby have competence and understanding of the technology, and laggards who have not. One of the biggest gaps is the companies' approach to data. The research showed misconceptions about resources needed to train AI. It is shown that companies sometimes erroneously believe they have access to the data needed.

Challenges raised by the participants in the surveys include handling of changes in jobs because of AI, both creation and loss, developing trust in AI both internally and externally, and few own-industry use cases to learn from, and lack of competence (Ransbotham et al., 2017; Kolbjornsrud et al., 2016; SAS, 2017). A lack of trust in the output given by AI as a cultural challenge was the biggest challenge for the respondents, with 49% of the respondents' answers (SAS, 2017).

The findings also include the conclusion that executives need to start experimenting with AI and that the “wait and see" method is not affordable. Both SAS and MIT also raise the issue of a lack of ROI (Return on Investment) or unclear business case.

Ransbotham et al. (2017) show that the cultural resistance to AI approaches is among the top barriers to AI adoption, but on the lower half after survey respondents ranked them; around 30% believes the cultural challenge is among the top three challenges. Although, 49% of the survey respondents in the SAS study think the cultural challenge is the biggest.

Common Mistakes Implementing Machine Learning Projects

Even though Machine Learning still represents a challenge for many companies in the market, the literature review indicates there are not enough published studies about common mistakes in Machine Learning implementation projects.

There are studies on other software implementation projects from a generalist point of view. However, considering the substantial differences between Machine Learning and standard software development, it is clear companies lack the required information to avoid common mistakes.

Technical Mistakes

From Marsland's work exposed in Chapter 2, nine technical mistakes are identified:

- Not testing properly
- Wrong choice of algorithm
- Not enough data
- Data on wrong form
- Wrong features selected
- Wrong parameter and model selection Training data biased
- Incomplete testing
- Overfitting
- Statistical Results not aligned with business expectations

Non-technical Mistakes

From the studies conducted by SAS, MIT, and Accenture, four additional groups of mistakes were identified:

Lack of Competence

Author	Pitfall
Boehm	Personnel shortfalls
Boehm	Straining computer-science capabilities
Keil et al.	Lack of required knowledge/skills in the project personnel
Keil et al.	Insufficient/inappropriate staffing
Oz & Sosik	Inadequate skills and means

The first item brought up by three of the authors is lack of competence, meaning organizations having difficulty acquiring resources with the competence needed for Machine Learning implementation projects (not only Data Scientists but also Data Management Engineers and Data Science Executives).

The recurring mention of a lack of competence by the authors suggests finding a competent workforce might be difficult. Another well-accepted hypothesis covered by the MIT and Boston Group study, indicates many companies do not realize their lack of appropriate staffing and just convert software developers to data scientists. This is also raised in Boehm's list as Personnel shortfalls. Also, Oz and Sosik have a staffing-related item on their list, Inadequate skills and means, aiming at shortcomings in competence.

Poorly Estimated Schedules and/or Budgets

Author	Pitfall
Boehm	Unrealistic schedules and budgets
Oz & Sosik	Deviation from timetable/budget
Fairley & Willshire	Excessive schedule pressure

Boehm, Oz and Sosik, and Fairley and Willshire all raise the issue of planning and budgeting. Schedules and budgets not being estimated correctly is causing problems in the projects according to the authors. It is found on Boehm's list as Unrealistic schedules and budgets, on Oz and Sosik's list as Deviation from timetable/budget and on Fairley and Willshire's list as Excessive schedule pressure. Fairley and Willshire do not, as opposed to the others, bring up the budgeting problem, because budgeting and scheduling are dependent on each other.

Requirements Changes

Author	Pitfall
Boehm	Continuing stream of requirement changes
Keil et al.	Changing scope/objections
Keil et al.	Lack of frozen requirements
Fairley & Willshire	Lack of technical specifications
Fairley & Willshire	Requirements creep

Requirement issues are raised twice by two of the authors as similar items on their lists. Keil et al. state that both Changing scope or objections and Lack of frozen requirements are potential mistakes, and Fairley and Willshire claim that Lack of technical specification and Requirements creep might cause projects to fail. The issue with requirements is also touched on by Boehm who claims Continuing stream of requirement changes is a common problem.

Lack of Management Support or Stakeholder Buy-In

Author	Pitfall
Keil et al.	Lack of top management commitment to the project
Oz & Sosik	Lack of corporate leadership

Lack of management support is brought up as a common problem by two of the four authors. Oz and Sosik call it Lack of corporate leadership and Keil et al. mention it as Lack of top management commitment to the project. Although, through reports by the consultancy firm BCG and MIT as well as SAS, a lack of management support for projects is brought up as a challenge.

Companies Classification

This section presents the companies and provides context for their current status on Machine Learning implementations. The companies are active in the following verticals: Telcom, Insurance, Financing, Smart Agriculture and Industry 4.0. This study does not include companies actively selling Machine Learning solutions (Services and /or Products), but is focused on non-analytical companies, without considering the specific verticals in which they operate. The companies have been divided based on the following criteria:

- The maturity of their process in adopting Machine Learning
- The mistakes they encountered and how common those problems were for the different companies

The companies were divided into four groups. The **Non-Adopters (group 1)** includes companies who have yet to implement any Machine Learning solutions. The **Unstructured Adopters (Group 2)** includes companies who have realized some initiatives but lack a unified strategy to give Machine Learning an active role in their processes and portfolio. The last two groups are called **Machine Learning Enablers (group 3)** and, **Products and Machine learning Core Products (group 4)**.

Non-Adopters (Group 1)

According to the workshop and Thingbook.IO's experience with this company type, Machine Learning competence within the company is very small. Usually, there are no employees hired to work solely with Machine Learning solutions and no Machine Learning initiatives (or casually explorative) have been started. One reason is the customer-driven project strategy some companies operate, in that the customer defines the solutions they want. Currently, Machine Learning is unlikely to be specifically requested by customers, because they generally ask for new features and do not care about the technology used.

Machine Learning projects are, with the limited competence and experience, considered riskier and unlikely to provide ROI. Delivering solutions to customers generates their revenue. Therefore, the general perception is that unless a customer asks for Machine Learning solutions to a problem, no substantial time or financial investment will be made in that area. This group of companies usually fail to see proactively how ML can improve their products or processes.

Despite customer-driven project strategies and limited technical knowledge of Machine Learning, some companies in group 1 started initiatives around data analysis and machine Learning, and there is some level of understanding of the benefits. However, the unclear business case and uncertainty in the investment has discouraged decision makers from investing.

Unstructured Adopters (Group 2)

There is a small common set of well-defined characteristics among this group of companies. They are big, usually with more than 10.000 employees, well established companies in critical sectors, such as Telco. Generally, they have been operating for longer than 25 year's and their internal processes are generally heavy and slow, particularly for initiatives involving departments across the company.

It is not difficult to find many different Machine Learning initiatives within the company, with little or no interaction between them, because the departments are siloed. Accessibility to the data generated is a serious problem and the ML initiatives are imitated in their contributions, because of the issue of access to the full data.

The main ML applications in this group are customer relations management, churn detection, personalized offers, that is, predicting which customers will be interested in which products. In the Telecom space, Machine Learning is also intensively used within networks to measure and detect quality issues, cybersecurity, and to predict geographically where the next network expansion will be required. In retail logistics, Machine Learning technology has proven useful, such as an application to predict how and when products should be moved from central storage to stores and between stores to optimize storage handling.

Machine Learning Enablers (Group 3)

The companies in this category, use Machine Learning extensively in such areas as risk management and operational intelligence, which involves automatically detecting undesired behaviour in the form of fraud, risky behaviour, predicting failures and violation of terms and conditions, among others. Using ML models for analysis of behaviour patterns helps identify deviation from normal behaviour or suspicious activity.

Usually, these companies have a team of data scientists working with Machine Learning projects in all parts of the company. Some of the main internal use cases are risk management, predictive maintenance, self-operations, credit scoring, and predicting lifetime value for customers, that is predicting all future profits from individual customers and predicting the risk of the customer leaving (churn prediction).

A common factor between these companies is a continuous effort to identify additional areas where Machine Learning can be used to optimize processes. Even though these companies do not sell Machine Learning, they are "data-driven companies" and see the technology as a powerful enabler to make their products in the financial and insurance industry much more competitive.

Machine Learning Core Products (Group 4)

Group 4 is a subgroup of group 3 with some notable differences in the way they organize Machine Learning internally. Despite operating in the same spaces as group 3 and sometimes competing with them, the Machine Learning practice is not separated from the rest of employees, there is no such thing as a Data Science team. Instead, data scientists, with different levels of experience and knowledge, are spread throughout the organization. Usually, those companies see the way they extract insights from the data as their major competitive advantage, and Machine Learning is integrated in their core business. The use cases for these companies vary from online credit scoring, to intelligent marketing campaign management and real-time recommendations and personalization.

How Problems Affect the Different Company Types

For non-adopters, it is considered a risk to put a small group of people aside for a Machine Learning project that does not directly create revenue. This prevents these types of projects from being realized. It might be because of an old-fashioned way of thinking or the company might struggle financially. Company age might not be a direct factor, but the lack of a recent education in the area might, because projects are usually initiated by people with a recent education in data science or Machine Learning.

A general problem for Unstructured Adopters and some part of the Enablers adopters, is the belief that many problems are solvable with models and Machine Learning, which causes people to not consider the problem correctly, nor what kind of solution is required. Jumping to the conclusion that using AI over other tools or methods to solve a problem is common and considered a pitfall. Furthermore, there are often difficulties in transitioning an algorithm or a trained Machine Learning model from a pre-production to a production system.

Evaluating how much revenue a solution will generate is a challenge for Unstructured Adopters. The evaluation introduces complexity when deciding which projects should be considered next, because data quality is assessed, and has a big impact on whether to continue. These decisions consider both the chances of successful implementation and, if successful, the increases in business value.

A big pitfall is to not have the relevant data accessible before starting implementation. Besides accessibility, the quality of the data is also of high importance and is one of the biggest problems. The Unstructured Adopters are usually old and large companies, having many sources of data of varying quality. The sources are not connected, and data management solutions look different from case to case, which adds another layer of complexity to the solution.

Another challenge is people's tendency to believe Machine Learning and artificial intelligence are magic. Some people view the process as sending data to the data science team and receiving money in return.

How Technical Mistakes Impact the Companies

For Group 1 companies, who have not yet implemented any Machine Learning solutions, the question is limited to guesswork using the company's way of working and preconditions.

For group 4 companies, mistakes when collecting, normalizing and cleaning data, represent a huge problem, which usually consumes more than 60% of the project schedule. Deploying models to production is also a source of delays and poor performance, because it can create unforeseen problems.

After deployment, there is also a need for monitoring the performance of the model. The more models deployed to production, the bigger the task of monitoring and maintaining them. Machine Learning models do not break, but they do begin to perform badly, which requires more sophisticated monitoring and maintenance. For companies whose core business is far from Machine Learning, the maintenance of models in production represents a hidden cost, which makes many decision makers opt for alternatives to the internal data science groups.

For instance, if there is a change in customer behaviour, which is safe to consider usual in many industries, it can cause a model to perform badly. A change in pricing in some products can cause customer behaviour to change, resulting in a decreased performance of a model, because it was trained on old data. There is a challenge and a cost in discovering these scenarios as fast as possible.

Wrong Choice of Algorithm

For all companies included in the analysis, choosing the wrong algorithm is considered a big problem. Testing different algorithms before choosing one is not very time-consuming compared to, for example, managing the data. Dedicating more time to this stage will not considerably change the time plan for the project. A common problem is using advanced algorithms for simple problems, which results in poor performance in production and longer training times.

Not Enough Data or Data in Wrong Format or of Poor Quality

Not having enough data, data not accessible or of poor-quality is a big risk for all the companies, although it manifests in different ways. For group 1, the lack of initiatives around Machine Learning and where Machine Learning can be valuable, make the lack of data irrelevant and reduce the issue to pure speculation. For group 2, the lack of accessible data (or the lack of ownership of the data) is augmented by the difficulties in accessing data from other departments, as well as the form in which the data is stored. Clearly, when it comes to the amount of data required and access to that data, those in group 2 face the biggest challenges, particularly the larger companies.

For companies in groups 3 and 4, it is common that the data is pre-processed (clean, normalize and correlate) before getting into the storage or landing area. This is because the entire products were built with Machine Learning in mind, and the process from raw data to algorithm results was always fresh for the product architects. However, when companies from groups 3 and 4 decide to start a new Machine Learning initiative requiring new data, they might face the same problems as companies from group 2.

Companies from groups 2, 3 and 4 usually implement a common data lake, a way of storing all data in one place. When the data exists in a data lake, there is no longer a need to access a certain system to get a certain kind of data. Everything is in one place. But, working in a previously unexplored area, the data usually has not been processed and integrated into the data lake, which adds complexity to the project. This will result in vast differences between projects depending on the data availability. Data lake management and integrating data from sources to a centralized data lake is not a light or simple process. It is also important to highlight that many companies are reconsidering their policies on data lakes, because often there is no financial justification to storing all the data regardless of future analysis and business applications, which puts more pressure on justifying the data integration.

Biased training data and Incomplete testing are important factors for the success of any Machine Learning project, regardless of which group the company belongs to. As previously discussed, it is more likely problems will arise when models are deployed to the production systems. This is a key reason why Machine Learning success relies on constant monitoring of model performance.

How Non-Technical Issues Impact Companies

Not Developing Trust for AI Solutions in the Company, Cultural Challenge

Groups 3 and 4 do not have problems of trust for Machine Learning solutions. Employees understand the basics of Machine Learning and value results without questioning them, at developer level, managerial level and executive level.

It is common for group 3 and 4 companies to request an explanation of the decisions or predictions made, not because of a lack of trust, but because it eases discussion and decision understanding.

For Group 2, the trust of machine learning predictions can vary greatly between different parts of the company, possibly correlating to age and the number of years in the company. Despite that, there is an ongoing effort to make companies in group 2 more data driven.

High Deployment Cost and Lack of ROI

The understanding that Machine Learning projects are not always successful is high for groups 3 and 4, meaning a lack of clarity on ROI is not a showstopper. There is the perception that successful projects are profitable. Usually those companies have high deployment costs, but it is not something hindering the projects from being started.

For some companies in group 2, a lack of clarity in ROI was initially a problem. When entering a new area, it is hard to estimate ROI, but when doing similar projects, the estimations become more accurate with time because of experience.

Nevertheless, companies in group 2, perceive the uncertainty in the investment and poor results in the past as the main barrier to embracing Machine Learning.

Lack of Competence

All companies believe it is difficult to find Machine Learning and data science competence at both developer and managerial level in the current market. This means the cost of skilled resources is high, increasing the cost of internal Machine Learning initiatives. A well-known problem for some companies in group 2, and especially in group 1, is the absence of a defined set of competences required to successfully implement Machine Learning initiatives. They tend to assume the lack of knowledge is only technical and there is no need for specific knowledge at senior management level (VP Machine Learning / Chief Data Officer). This mistaken approach inevitably causes ML initiatives to get poor results, increasing the frustration at C-Level, exacerbating the feeling of burning money without any decent ROI.

Badly Estimated Schedules and Budgets

Group 3 and 4 companies understand the risk of Machine Learning projects not being executed as planned. There is a continuous iteration over the estimations to keep them up to date.

Early estimations of group 2 companies are often naïve. After projects have started it is often realized that it will take more time than initially planned. The main reason for the delay is the data required is not available from a single location, or it is spread around many siloes using very specific data adaptors, which take time to integrate and, in some cases, to build. It is also common for those companies to have organizational issues when it comes to accessing data from another department. All these factors make the Machine Learning phase of the project seriously reduced and the data access and preparation tasks take most of the schedule.

Estimation is especially difficult in new areas where data is not available, or the systems where the data might reside must be researched. There is also an issue with communication to stakeholders, with less understanding of Machine Learning, that more resources are needed.

Lack of Management Support

For groups 3 and 4, the lack of management support, when it happens, is mostly due to the impossibility of generating a credible business plan, because of the perception of funding projects that are too explorative and have a high level of uncertainty. In general, both executives and managers show a genuine belief that ML and Data Science are a key differentiator for their business.

For groups 1 and 2, there is a combination of unrealistic expectations and scepticism, which moves managers to prioritize short-term revenue over long term transformation.

For some companies in group 2, there is a genuine effort to become more data-driven through the whole organization, meaning senior management are supportive of Machine Learning even though they have little understanding of its effects and value. Though, it is necessary to be able to motivate in numbers what can be done with data to have full management support, which is understandable.

Differences in how Machine Learning Projects and other Software Development Projects Should be Treated

Unlike standard software development projects, internal Machine Learning should be more exploratory in the early stages. This is because of the difficulty, or the impossibility in some cases, of knowing exactly what can be done before starting the project, which is easier, or possible at least, in other software development projects.

Senior management of group 3 companies do not understand Machine Learning and, therefore, expect magic from data, without realizing the amount of work involved. This can be exacerbated by the data science team solving a problem for a department without imparting any information about how it was done.

For group 1 companies, projects are considered high-risk and more difficult. So far, machine learning projects are proof of concept and usually require being funded (totally or partially) by the customer.

Additional Considerations

With new GDPR in force, regulations are a big part of Machine Learning projects, which takes up time. This can stop ideas from being realized when sensitive or private data needs to be analysed.

Employees also require continuous education. This area of technology is fast-moving (and at times noisy), so employee education planning is an important factor for success.

Results

Through insights gained from the companies contacted and collaboration with Thingbook, the following main conclusions were reached:

- For companies who decided to implement Machine Learning, there is a correlation between an overestimation of the value of it and how much experience an organization has in it. Inflated expectations come from organizations who neither possess ML knowledge, nor realize the lack of it.

- Data related mistakes, such as not having enough data, low quality of the data, or biased data, have the most severe impact on results. This has been identified as a key differentiator with standard Software development projects, as data related problems are usually identified during the project, or even during the production phase, and are difficult to consider during project planning.

- Realizing the value of long-term solutions through Machine Learning projects is difficult, because of a lack of obvious ROI (Return on Investment) or an unclear business case.

Overestimation of Machine Learning Capabilities

Machine learning is categorized as being “on the peak of inflated expectations" according to Gartner. This is proven by the companies involved, from both the developer perspective and the managerial perspective.

The developer perspective is visible for Groups 2 and 3, because of a tendency to use too advanced models on small amounts of data, or using sophisticated techniques for simple problems, often resulting in overfitting and poor results in a production environment.

The managerial perspective is visible for Group 2 where there are difficulties experienced with accurate scheduling or acceptance when projects are not successful and an overall overestimation of what the technology can do.

Both these perspectives can be translated into a lack of understanding of the resources which a Machine Learning project needs to be successful. A more realistic view would be beneficial, rather than believing in a black box solution that works like magic. Based on the results of this study, indications point to overestimation varying with experience in Machine Learning.

Overestimation is not present for group 1, because of their unwillingness to start machine learning projects. With group 2 companies possessing some Machine Learning knowledge at developer level and insufficient at managerial level, the overestimation occurs in the managerial perspective discussed previously, which could be caused by their way of working with Machine Learning. When there is a team of internal data scientists implementing ML solutions in different departments, they might work in isolation from the department and not transfer knowledge, increasing the belief that Machine Learning works like magic.

For Group 4 companies, the opposite happens, where there is a continuous effort to educate employees, spreading the knowledge through the organization instead of isolating it within a group.

The belief that Machine Learning can solve complex problems with few resources and negligible risk, causes unrealistic expectations and needs to be managed. This could negatively affect requirements, scope, time planning, and so on, because of the unrealistic view.

Data Related Mistakes are the Most Severe

Ben Hamner, co-founder and CTO at the data science platform Kaggle, in his speech at the Strata conference in 2014, raised four problems in Machine Learning projects: data leakage, overfitting, data quality, and data sampling and splitting.

The results from Thingbook's research confirm Hammer's thesis and show that data management is the largest and most demanding part of Machine Learning projects. This includes collection, structuring, and cleaning of data. In other words, transitioning the data from its source to an analysis ready state. The problems encountered vary between storing the data in the wrong form or spreading it out over several sources, data traveling too fast and is too big to be stored (streaming analytics), not having enough data, low quality of the data, or biased data. Structuring the data was shown to be the largest problem of all the companies studied.

This could be because the data already collected has not been collected with Machine Learning in mind. This is understandable, because Machine Learning is only now at the top of Gartner's hype cycle. Because groups 3 and 4 are younger companies than 1 and 2 and have a large relative growth during the years when Machine Learning has been increasingly discussed, their growth is likely to be influenced by Machine Learning to a larger degree.

With more Machine Learning experience, data structuring seems to be less of a problem, but there is a broad understanding of it being a time-consuming task. The more experienced companies give more attention to the complex problems and issues that are difficult to detect, such as biased training data.

Correlation Between the Ability to Value Long-Term Solutions and Machine Learning Experience

A problem often expressed by developers and data scientists is the lack of long-term investment in Machine Learning. This is often the case for those companies where the adoption of Machine Learning is in its infancy and Machine Learning has not generated revenue for the company, making the investment necessary for long term plans, hard to justify. For group 1 companies, the strategy of not building Machine Learning solutions for problems because of a lack of concrete revenue, indicates a lack if prioritizing the long-term effects of internal Machine Learning.

This is not the case for companies belonging to group 4, where the previous financial success of Machine Learning implementations makes senior management more willing to embrace the uncertainty of investing. They also actively encourage education in Machine Learning and are continuously hiring new competence. The acceptance of projects not always being successful also indicates a long-term view on the learning processes.

In any case, it seems clear that the more experience and skill companies have with ML, the more willing they are to explore further applications and take the risk of a long-term investment.

Bibliography

- M. Turing (1950). "Computing Machinery and Intelligence". In: Mind - A Quarterly Review of Psychology and Philisophy 59.236.
- Blomkvist, Par and Anette Hallin (2015). Method for engineering students: degree projects using the 4-phase model.
- Boehm, Barry W. (1991). "Software Risk Management: Principles and Practices".
- In: IEEE Software 8.1, pp. 426{435.
- Clearinghouse | Definition of Clearinghouse by Merriam-Webster (2018). url:https://www.merriam- webster.com/dictionary/clearinghouse.
- Collis, Jill and Roger Hussey (2014). Business Research: A Practical Guide for Undergraduate & Postgraduate Students. 4th ed.
- Duncan, William R (1993). "The process of project management". In: Project Management Journal 360.3, pp. 5-10.
- Fairley, Richard E. and Mary Jane Willshire (2003). "Why the Vasa Sank: 10 Problems and Some Antidotes for Software Projects". In: IEEE Software March/April,pp. 18-25.
- Gartner (2017). Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017. url: https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/
- Gartner (2018). Hype Cycle Research Methodology | Gartner Inc. url: https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp
- Hamner, Ben (2014). Machine Learning Gremlins. Youtube. url: https://www.youtube.com/watch?v=tleeC-KlsKA.
- Jordan, M. I. and T. M. Mitchell (2015). "Machine learning: Trends, perspectives,and prospects". In: Science 360.6245, pp. 255{260. doi: 10 . 1126 / science. aac4520.
- Kaggle (2018). Titanic: Machine Learning from Disaster | Kaggle. url: https://www.kaggle.com/c/titanic (visited on 04/16/2018).
- Keil, Mark et al. (1998). \A Framework for Identifying Software Project Risks". In:COMMUNICATIONS OF THE ACM 41.11, pp. 76-83. doi: 10.1145/287831.287843.
- Kolbjornsrud, Vegard, Richard Amico, and Robert J Thomas (2016). "The promise of artificial intelligence Redefining management in the workforce of the future".Tech. rep. Accenture.

- Marsland, Stephen (2015). Machine Learning : an algorithmic perspective. 2nd ed.isbn: 1498759785.
- Negnevitsky, Michael (2005). Artificial Intelligence - A Guide to Intelligent Systems.2nd ed. isbn: 0321204662.
- Oz, Elfy and John J. Sosik (2000). "Why Information Systems Projects are Abandoned: A Leadership and Communication Theory and Exploratory Study". In: Journal of Computer Information Systems 41.1, pp. 66-78.
- Ransbotham, S. et al. (2017). Reshaping Business With Artificial Intelligence. Tech.rep. MIT Sloan Management Review and The Boston Consulting Group.
- SAS (2017). The Enterprise AI Promise: Path to Value. Tech. rep. SAS.(2018). 5 machine learning mistakes and how to avoid them | SAS. url: https://www.sas.com/en%7B%5C_%7Dus/insights/articles/big-data/5-machine-learning-mistakes.html
- Schmidt, Roy et al. (2001). "Identifying Software Project Risks: An International Delphi Study". In: Source Journal of Management Information Systems 17.4,pp. 5-36.
- Stewart, David B. (2001). "Twenty-Five Most Common Mistakes with Real-Time Software Development". In: Embedded Systems Conference. Boston.
- Thomas, David R (2006). "A General Inductive Approach for Analyzing Qualitative Evaluation Data". In: American Journal of Evaluation 27.2, pp. 237-246.
- Walliman, Nicholas (2011). Research Methods: The Basics.

About the Author

Roman Ferrando has numerous years of experience within some of the world's leading technical companies like Ericsson Research and Oracle Labs. He has focused its research activity on clustering and pattern recognition techniques and the scalability issues related to hight-peed generated data. He holds 15 patents related to the analysis of real-time data in telecommunications networks and have also produced several academic papers on the same topic published in well-respected journals. He has been also member of the 3GPP standard committee for several years. Roman holds a PhD from UC Berkeley in Machine Learning and also authored a book on the role of data analytics in the cloud and next generation telecom networks. Wiley & Sons published this during 2013. In 2016, Roman started Thingbook, a company created to facilitate the adoption of predictive analytics in high-speed data generation environments.