Universiteit Leiden

ICT in Business and the Public Sector

Decision Support Systems and Business Intelligence
for agile portfolio management

Name: Hamdy Michael Ayas
Student-no: s2055201

Date: 02/08/2019

1st supervisor: **Dr Christoph J. Stettina**
2nd supervisor: **Dr Arno Knobbe**

MASTER'S THESIS DRAFT

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Acknowledgements

First of all, I would like to thank my thesis supervisor Dr Christoph J. Stettina. The door of Dr Stettina was always open, and during the period of developing this thesis he provided a set of very valuable insights and guidelines. The professionalism and knowledge of Dr Stettina was very inspiring and motivating. I would also like to acknowledge professor Arno Knobbe as the second supervisor of this thesis. I am gratefully indebted to him for his very valuable comments on this thesis. As a young researcher, I am grateful that I had the chance to work with two great researchers.

Most importantly, I would like to express my gratitude to my friends and my beloved family. My parents Mohammad Ayas and Eleni Michael Ayas, my dearest siblings Sandy and Malek and of course my grandparents: Vasiliki, Kiriakos, Khadija, Ahmad for believing in me and for their continuous support and encouragement. Thank you all for the unconditional love and support. Nothing would have been possible without you.

Hamdy Michael Ayas.

Abstract

Technological trends are effectively changing the world and organisations are forced to operate in novel types of digital markets and ecosystems. Organisations need to coordinate all their initiatives in order to deliver experiences that give them the edge in the modern economy. This leads to the need of starting more and more initiatives, but also to the need of managing a portfolio of diverse and very often software-intensive initiatives effectively. In addition, as organisations increasingly adopt outside-in perspectives in their operations, the amount and importance of agile teams increases, and agile principles influence all operational levels. The growth of the number of projects that organisations undertake leads to the need of continuously responding to fast paced changes. Otherwise it is not possible to compete and develop new competitive products and services in a timely manner. Therefore, the right methodologies and tools are needed to be in place, that will enable organisational agility in all levels in a systematic, consistent and data-driven manner.

The ways in which organisations typically manage their portfolios, do not always take into account the fast-paced circumstances that technology-driven elements of projects bring. In addition, the available tools that organisations use to measure and evaluate portfolio performance are mostly capable of describing their business and not predicting potential outcomes. Also, in the ever-changing environments that organisations operate, the desired business outputs are very dynamic and costly to manually keep track of with the current portfolio management support tools. Decision-Support tools that are currently in place, accommodate the measurement of performance metrics that describe the present and the past but do not intend to give solutions on what actions are needed in the future.

Thus, more emphasis needs to be given in different ways of analysing projects, but also in considering the potential impact of projects' elements in the outcome, during decision-making. It is intended to fill this gap with a Decision-Support System that evaluates projects and predicts their success or failure. This takes place by analysing agile projects in networked organisations or marketplaces. Specifically, via a set of machine learning models that learn from the digital footprint of 378,661 projects and 7,200 mobile applications, we (1) evaluate projects based on certain patterns and (2) predict the outcome of projects with a certain degree of accuracy (between 60% to 70%). This can assist the decision-making process of Project Portfolio Managers in regard to 1) project selection and 2) portfolio configuration. The machine learning models act as a decision-support system and automate parts of project evaluation and selection in the early stages of projects in a portfolio. As a project evolves through the different phases of portfolio management, it has different types of decision-making milestones and different types of analytics can support them. We focus on the phase of Evaluating projects from a project Funnel and we predict their likelihood of successfully achieving their budgeting target or their customer satisfaction target. Therefore, our models try to imitate the decisions that investors of projects make or model the assignment of ratings from users.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

1 Introduction

1.1 Background

Technological trends like the Internet of Things (IoT), mobile applications and digital services are effectively changing the world and organisations are forced to operate in novel types of markets and ecosystems (Anttiroiko, Valkama, & Bailey, 2014). Business development is facilitated through new digital offerings (McAfee & Brynjolfsson, 2017) and participation in digital marketplaces and software ecosystems is becoming the norm amongst leaders in many industries (Mayer-Schönberger & Ramge, 2018). Moreover, customers in all industries are shifting from valuing stand-alone products and services to valuing integrated experiences that consist of multiple products or services and are customised to their needs (Meyer & Schwager, 2007). Therefore, organisations need to coordinate all their digital and physical offerings and initiatives, in order to deliver experiences that give them the edge in the digital economy (McAfee & Brynjolfsson, 2017). This leads not only to the need of creating more and more software, but also to the need of managing a portfolio of diverse and very often software-intensive initiatives effectively, under a single but frequently changing strategy (Teece, Pisano, & Shuen, 1997; Yeow, Soh, & Hansen, 2018). Consequently, the roles of decision-makers in organisations are changing and the actual activities that managers do are becoming more evident (Mintzberg, 1997).

Therefore, very often organisations and their leaders struggle with continuous and systematic prioritisation because they need to focus on their strategic objectives and not on constant transformation initiatives that require additional work, more resources and development of new expertise. As they are increasingly finding themselves in the middle of dynamically developing or implementing software for a variety of individual projects or programmes, changes in business processes and business models are increasingly taking place (McAfee & Brynjolfsson, 2017). Therefore, organisations adopt digital strategies that require the dynamic alignment of software initiatives with business in order to achieve their strategic objectives (e.g. growth and/or operational excellence) (Yeow et al., 2018). As a result, it is very common nowadays to have transformation programmes in place that allow the effective acceleration of innovation processes. Moreover, functional departments are shrinking, processes are getting reengineered, start-up spin-offs or acquisitions take place, and business development is relied on new types of software-enabled services (Mayer-Schönberger & Ramge, 2018). Keeping up with such dynamic changes and effectively aligning with digital strategies is a challenging task (Yeow et al., 2018). Also, making systematically the right judgement-calls about the projects can be the difference between achieving and not achieving strategic objectives or even surviving and not surviving (McAfee & Brynjolfsson, 2017).

Hence, agile management of projects and agile teams are increasingly becoming a vital element of organisations. Specifically, projects are usually developed and accommodated by agile teams that are characterised by nimbleness and effective change, based on users’ inputs (Dingsøyr, Nerur, Balijepally, & Moe, 2012). Similarly, the management of such projects and such implementation teams is increasingly taking agile forms as well (Kuusinen et al., 2017). However, organisational effectiveness depends not only on how the organisation is led but also on situational factors, that are not always measurable or expected (Vroom & Jaago, 2007). Leading the project portfolio of an organisation is a highly dynamic and multi-dimensional task that often has many independent stakeholders involved, with different interests, that often compete with each other (Archer & Ghasemzadeh, 1999; Meskendahl, 2010). As a result, it can be characterised as a highly incentivised setting in which is needed to take the right decisions. Therefore, the right processes for decision-making are needed, in order to enable decision-makers to achieve the organisational effectiveness required for achieving their strategic objectives (de Bruijn & ten Heuvelhof, 2002).

In this study, we propose a decision-support system which facilitates more effective and efficient judgement calls in agile portfolio management. Decision-making for projects takes place mostly in top-bottom processes but the adoption of agile methodologies is usually from bottom-up initiatives that are supported eventually by management (Stettina & Hörz, 2015). Theoretically, these dynamics are changing the ways that projects are coordinated, due to the new business models that large-scale agile predisposes (Strode, Huff, Hope, & Link, 2012). As agile teams become essential for delivering products and services, organisations need to manage portfolios of multiple projects and steer them according to their strategic objectives (Yeow et al., 2018). Very often, the coordination place of teams for these projects is in tools that generate insightful data about the projects. However, these coordination tools do not sufficiently address the needs of stakeholders with the knowledge and information they provide. Specifically, the analytical capabilities of the tools are mainly in describing projects and then human intervention is needed for analysis, interpretation and judgement. This is viable when there are not many projects, but it can get very costly in large numbers of projects. This study intends to navigate the ways on which large numbers of independently developed projects and agile teams can be efficiently managed. Hence, it is intended to formulate a set of models for supporting the evaluation and selection of projects in portfolios of networked and self-managed teams, using prediction.

1.2 Research Objective

We intend to fill this gap by analysing parts of the digital footprint of projects in agile organisations and then, use a set of Machine Learning techniques to formulate a classification, prediction machine that helps portfolio managers in project evaluation and selection. The objective of this study is to investigate how agile portfolio management can be enhanced through predictions and how key success factors can be integrated to predict the performance of projects. We believe that more emphasis needs to be given in different ways of analysing elements of projects that influence the outcome, but also in considering during decision-making the potential impact of such elements in predicting the outcome. It is intended to fill this gap with a Decision-Support System that evaluates projects and predicts their success or failure through classification. This takes place by analysing agile projects in two case studies. Specifically, via a set of machine learning models that learn from the digital footprint of 378,661 projects and 7,200 mobile applications, we (1) evaluate projects based on certain patterns and (2) predict the outcome of projects with a certain degree of accuracy in the range of 60% to 70%. This assists the decision-making process of Project Portfolio Managers in regard to 1) project evaluation and selection and 2) portfolio configuration. The machine learning models can act as a decision-support system and automate parts of project evaluation and selection in the early stages of projects in a portfolio.

1.3 Research Relevance

1.3.1 Theoretical Relevance

Designing and engineering new (software) products and services predisposes an increased pressure to product developers, service designers and portfolio managers on making many crucial decisions that influence the resulted business outcome. However, there is a significant gap in literature and practice about how software engineers, product/service designers and portfolio managers can go through such decision-making processes in an integrated manner (Razavian, Paech, & Tang, 2019; Vliet & Tang, 2016). In addition, there are not many methodologies and tools for supporting these decisions using prediction machines or analytics that are more advanced than describing the current and past state of projects. Decision-making in managing portfolios of autonomous teams can be improved and the alignment between their strategies and organisational strategy needs to be facilitated in the level of portfolio management. Especially when there is a large number of projects to be evaluated and filtered. Hence, we approach the project evaluation and project selection process in portfolio management as a prediction problem and we develop a system to enhance the process with prediction. For this to be realised, it is required to formulate a way for portfolio solution management to quantify the performance of projects, in terms of likelihood to succeed and thus, classify projects with their probability of success and failure. Additionally, it is required to enable portfolio management and agile teams to evaluate the performance of projects and dynamically make judgments about projects.

1.3.2 Practical Relevance

Project evaluation and selection is becoming increasingly costly with the increasing number of projects that need to be evaluated. As organisations have many undergoing projects, the need to analyse them and take decisions about them is important. The form of existing project evaluation in large scale agile is mostly in the form of principles and self-assessment checklists which rely on individuals to consider several factors and teams to assess their situation and adopt specific practices (Baxter & Sommerville, 2011). Such methodologies possibly work in many contexts but in a dynamic, highly paced and extremely practical environment, they have many shortcomings. Therefore, in this study it is proposed a more practical approach on assisting teams to take the best possible decisions. This includes the application of a set of Machine Learning models that can be used in addition to a project portfolio management tool. The models can assess the performance of a project in its early stages and predict its potential outcome. The predictions will improve the support of project portfolio management systems to the decision-maker. Consequently, it is intended in this study to evolve in a practical way the knowledge developed about decision-making processes of agile portfolio management.

1.4 Research Question

As a result, the central research question to be addressed is:

"How can Machine Learning support decision-making processes in Agile Project Portfolio Management?".

Some guiding research questions are:

1. "To what extent Decision Support Systems (and Business Intelligence) can influence the management of project portfolios?"
2. "How can parts of the decision-making process in Agile Portfolio Management be accelerated with prediction?"
3. "How can performance of projects in agile portfolio management be measured using metrics for success factors, obstacles or challenges?"
4. "How do decision support systems assist Agile Portfolio Management of software-intensive projects?"

1.5 Research Scope

This study answers the main research question by finding the ways in which decision-making in portfolio management can be facilitated and supported via prediction machines. Specifically, it is intended to model parts of the decision-making process in portfolio management. To do so, the decision-making process in portfolio management needs to be defined and decomposed into its detailed activities. Then, it will be determined which activities are best to be executed from portfolio managers and which activities are best to be automated (project evaluation and project selection). As a result, a Machine Learning model will be developed for those activities, that analyses elements of portfolio management through a specific set of criteria. Then, the model will provide its results to managers in order to add their judgement and reach a final decision.

1.6 Thesis Overview

Consequently, this study first starts with the decomposition of the relevant concepts, then a methodology is defined, and an analysis takes place. In the second chapter, we provide an overview of the application domain which is Agile Portfolio Management. In addition, we cover in depth the anatomy of Computer-Aided Decision-Making and the technology elements required to achieve data-driven decision-making. The third chapter includes a detailed description of the research methodology that is followed, and it takes the reader through the data analytics process for creating the decision-support system that introduces predictive analytics in agile portfolio management. Then, in chapter 4 an analysis is provided with the state-of-the-art tools that are used to support decision-making in the level of portfolio management. The fifth chapter describes the analysis that took place in order to develop the Machine Learning models. In chapter six the results from the chosen methodologies and performed research are described. In the seventh chapter, we initiate a discussion for interpreting the results and provide some imitations of the developed decision-support system. Finally, we conclude in the chapter 8 with the key takeaways of this study, a concluding summary and future recommendations for research and practice.

2 Literature Review & Related Work

2.1 Agile Portfolio Management

2.1.1 Portfolio Management

Organisations set their strategies with a very specific aim, which is to generate a business output that creates value to their customers, in exchange of resources (for private organisations) or a well-functioning society (for public institutions). As organisations grow, it is becoming challenging to actually implement their strategy and make the increasing amount of undertaken initiatives contribute to it (Hrebiniak, 2006). Portfolio management can facilitate a solution to this challenge and act as a means for strategy execution (Agyapong, Guitton, Fairdoon, & Lasar, 2016). In this section it is intended to provide a clear picture of what Portfolio Management is, why it is important and what are the elements that it contains. This means that we will describe the particularities of strategy execution and how portfolio management is essential for its realisation; provide definitions and objectives of portfolio management; and describe how projects and programmes across organisations need to be coordinated to achieve strategic objectives.

It is clear by now that organisations need to develop a strategy in order to gain competitive advantage and survive or thrive (M. E. Porter, 2008). But most importantly, organisations need to undertake a set of initiatives in order to implement strategy and achieve their objectives (Hrebiniak, 2006). Developing a strategy is not the same as actually implementing it, which is, in fact, much more difficult and challenging (Meskendahl, 2010). Specifically, there is a set of typical obstacles that are usually present against effectively implementing strategy (Hrebiniak, 2006). The most important ones, as Hrebiniak (2006) describes, are inability to manage change, poor or vague strategy, inadequate guidelines or model to guide implementation efforts, poor or inadequate information sharing, unclear responsibility and accountability, and working against the organizational power structure. Therefore, a linking element between the strategy and the several initiatives is required to make strategy implementation possible (Meskendahl, 2010).

Portfolio Management is the linking element, that navigates through the different projects and initiatives of an organisation, in order to maximise the potential contributions to the determined strategy (Anyosa Soca, 2009; Meskendahl, 2010). This is achieved with its underlying elements which are Projects and Programs (PMI, 2013). Organisations in all industries undertake projects in order to achieve the successful realisation of initiatives (Agyapong et al., 2016). These projects are usually grouped in sequenced programmes (or roadmaps) that have a longer-term implementation scope (PMI, 2013). More often than not, even though the undertaken projects and programmes are (more or less) equally important, they do not contribute equally to the strategic objectives of the organisation and thus, project portfolio selection needs to take place (Archer & Ghasemzadeh, 1999). Ultimately, the objective of a portfolio is to steer multiple (connected or disconnected) initiatives towards the maximum possible returns, with the minimum possible risks and with the best possible contribution to the organisation's objectives (PMI, 2013). Therefore, it is important that the criteria of the selection process are based on linkage to strategy, maximization of the portfolio's potential financial outcomes and balancing the load between projects (Martinsuo & Lehtonen, 2007). In addition, establishing a systematic way to evaluating, selecting and executing projects, is essential in order to achieving a decision-making process that allows making the right judgement calls by design within a project portfolio (Archer & Ghasemzadeh, 1999). Such a process enables better decision-making, since projects and programmes often exceed the available implementation capacity (e.g. due to limited resources) and hence, they compete each other for priority (Meskendahl, 2010). Furthermore, structured ways of executing projects and formalisations of Idea-to-Launch processes can be achieved, with distinct decision-making milestones (Cooper, 2008). Such deterministic approaches for implementation can be key for the success of predetermined projects (Archer & Ghasemzadeh, 1999).

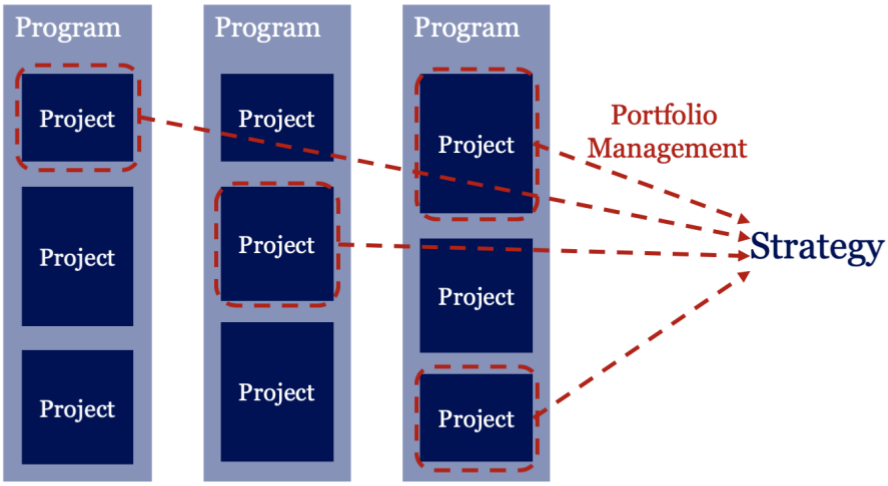


Figure 2.1: *Project, Program and Portfolio Management*

However, deterministic control of projects' implementation does not always lead to maximising economic returns and value delivery (Svejvig & Andersen, 2015). Specifically, elements of projects make their management a complex system with strong interdependencies that are sometimes difficult to describe in a deterministic manner (Baxter & Sommerville, 2011). Furthermore, in Portfolio Management is required to aim for both successful project implementation, but also for what delivers more value to the organisation (Martinsuo & Lehtonen, 2007). Therefore, a different view might need to be taken regarding projects and what their role should be, leading to a shift from Classical Project Management towards Rethinking Project Management (RPM) (Svejvig & Andersen, 2015; Walker & Lloyd-Walker, 2016). Specifically, in our interpretation of RPM and (Svejvig & Andersen, 2015), projects change from being a means to implement a strategy that is developed at a top-management level and become a means of dynamically responding to circumstances for achieving strategic objectives. Such an approach is essential for the ever-changing circumstances that organisations find themselves in because of technological developments (Yeow et al., 2018). Additionally, this also calls for establishing mechanisms about keeping track of the creation, delivery and capturing of business value (Souza et al., 2018). Most importantly, establishing such a value-driven mind-set can lead to decentralised organisational structures and thus, more effective decision-making against uncertainties (Martinsuo, Korhonen, & Laine, 2014). And this already addresses most of the strategy implementation obstacles that were identified in the first place by (Hrebiniak, 2006).

2.1.2 Agile Portfolio Management

As digitalization is becoming more influential, organisations are going through significant changes on the ways they operate, in order to innovate (Christensen & Overdorf, 2000). Such operational settings have many advantages that contribute to the achievement of start-up like agility and nimbleness in big organisations as well as more customer centred approaches and room to innovate (Brown, 2008; Brown & Wyatt, 2001). The amount of software projects implemented in organisations is continuously growing and such projects are increasingly gaining criticality for organisations (M. Bloch, Blumberg, & Laartz, 2012). Therefore, managing the success and the risk of such projects is essential and a good way to do so, comes with the application of agile methodologies in their development and implementation (Rasnacis & Berzisa, 2016).

Specifically, organisational responsiveness and agility can enable innovation and business development through more effective management of new software products, (digital) services and solutions development (Vlaanderen, Jansen, Brinkkemper, & Jaspers, 2011). In addition, as organisations increasingly adopt outside-in perspectives in their operations, the amount and importance of agile teams increases, and agile principles influence all operational levels (Conforto, Amaral, da Silva, Di Felippo, & Kamikawachi, 2016). The growth of software development projects leads to the need of organisations to continuously respond against fast paced changes. Otherwise, they are not able to compete and develop new products and services in a timely manner (Michael E Porter & Heppelmann, 2015). Therefore, organisations need to find ways to respond against this agility in a systematic and continuous manner and thus, adopt more dynamic ways of operating (Christensen & Overdorf, 2000; Yeow et al., 2018).

First of all, agile methodologies are proven to be effective for individual, small-scale projects because they deal well with the risks that are presented in software development (Strode et al., 2012). Teams, that use agile as their main operational model, develop capabilities of changing the development direction and influencing the scope of projects. This is a successful approach in enabling the development of competitive Digital Services (Yeow et al., 2018) because of the required swift responses to the forces that shape strategy (Michael E Porter & Heppelmann, 2015). Some of the reason that agile teams are able to respond more effectively are because of characteristics that are key to their success, such as better coordination, communication, balance of member contribution, mutual support, effort and cohesion within teams (Lindsjørn, Sjøberg, Dingsøyr, Bergersen, & Dybå, 2016).

However, on a level where multiple projects take place it is required to coordinate between the different teams of agile software development (Moe, Dingsøyr, & Moe, 2013; Strode et al., 2012). These sets of individual projects are not always coordinated under common strategic objectives and their nature of agile development results on significant uncertainty and risks for their overall management in a large scale (Dikert, Paasivaara, & Lassenius, 2016). Specifically, agile project management of agile teams, very often comes in pair with implications like decoupling tendency of autonomous teams from their parent organisation (Patanakul, Chen, & Lynn, 2012). Within a competitive business context, uncertainty is not well regarded and changing of directions are not always in line with strategic objectives of organisations (Yeow et al., 2018).

In addition, in the dominating result-oriented culture of organisations, performance of agile is mostly measured by a pre-determined business output and this is not always a good fit with the dynamic nature of agile teams (Lindsjørn et al., 2016). In the ever-changing environments that organisations operate, the desired business outputs are changing dynamically and thus, more flexibility is needed (Teece et al., 1997; Yeow et al., 2018). Consequently, more emphasis needs to be given on the ways in which strategic objectives are collectively achieved in order to dynamically respond to market forces (Strode et al., 2012).

Organisations increasingly become complex systems that need to facilitate the development of many products or services. Therefore, a shift of focus is needed towards methodologies that comprehensively consider the inherited complexity and try to embed in the portfolio management process mechanisms to cope with it (Svejvig & Andersen, 2015). Systems development is complementing itself on that manner, via ways to cope with uncertainty or complexity and socio-technical approaches on systems development gain momentum (Patnayakuni & Ruppel, 2010). However, these approaches are not always taking place in pragmatic or practical manners (Baxter & Sommerville, 2011) and thus, there is a gap in practically embedding them in the operations of managing project portfolios.

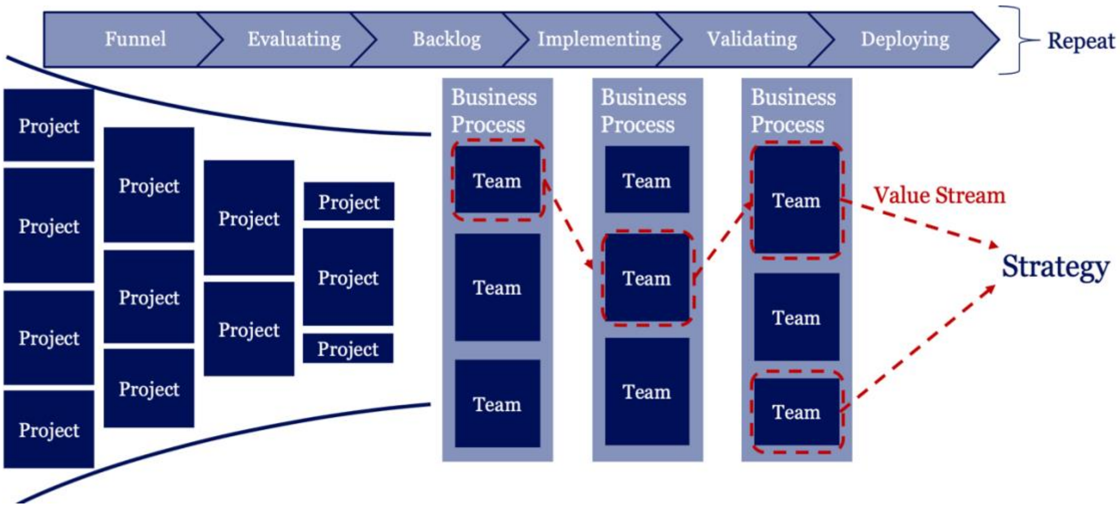


Figure 2.2: Agile Portfolio Management

2.2 Computer-Aided Decision Making

2.2.1 Decision-Making Process

Implementing strategy and relating all initiatives that take place in an organisation with it is a highly challenging task (Hrebiniak, 2006; Meskendahl, 2010) whose weight falls on the shoulders of the leaders that need to manage this. Therefore, managers that operate and make decisions need to lead the organisation towards organisational effectiveness (Vroom & Jaago, 2007). While doing so, according to Mintzberg (1997), managers' roles fall into 3 main categories that are dependent on the activities that managers do, as seen also in figure 2.2. These roles are (1) *Interpersonal Roles* that help them build the required relations to be effective, (2) *Informational Roles* that make them a central figure in gathering and distributing knowledge across the organisation and (3) *Decisional Roles* that is how relationships and knowledge are utilised with taking actual courses of action (Mintzberg, 1997).

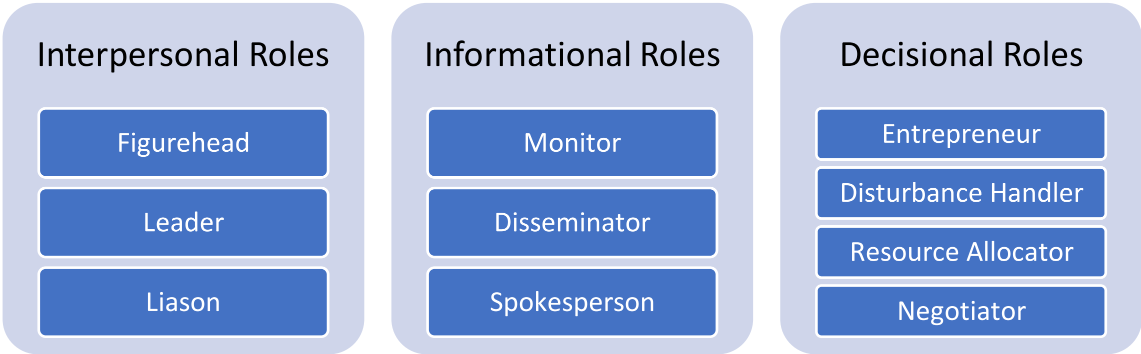


Figure 2.3: *The Manager's Roles (Mintzberg, 1997)*

The conventional notion of decision making in the management of many businesses and industries, is that of a talent in one's intuition and also that of intuition-related skills that come out of experiences (Turban, Sharda, & Delen, 2014). This could be evident in the existence of a variety of management styles and roles (Mintzberg, 1997), that would allow flexibility for effectively dealing with the different situations that can be at hand (Vroom & Jaago, 2007). Consequently, this leads to decision-making processes that are not standardised into objective methodologies but adjusted to personal preferences of individual managers. Therefore, it has always been challenging to systematically adopt quantitative methods grounded in factual approaches (Turban et al., 2014). With the emergence of Big Data and economically viable ways to derive actionable knowledge from data, this is changing (Agrawal, Joshua, & Goldfarb, 2018). Therefore, we are already in the era in which humans are collaborating with data powered machines in order to improve our processes (McAfee & Brynjolfsson, 2017).

Nearly three decades now, human mind and information systems are collaboratively operating together, with the intention to optimize business processes and organisational structures (Hammer & Champy, 1993). Even though reactions to such developments were mixed, the re-engineering of business processes, many times due to developments in Information Technologies, was influential nevertheless (Neill & Sohal, 1999). For example, with the adoption of Enterprise Resource Planning (ERP) systems, organisations started accommodating many processes, functions and operations into fewer software systems and the potentials of computer-aided work execution seemed to be very high (Esteves, 2001). These potentials even seemed to be delivering on their promise, as adopters of Information Systems and ERPs were gaining competitive advantages over non-adopters (Hunton, Lippincott, & Reck, 2003). Furthermore, it eventually became evident that the impact of adopting

Information Technology systems on the enterprise level, is significant for improving operational performance over time (McAfee, 2002).

This naturally led to a wave (that lasts until this study's time) of designing and implementing business processes and operational roles, in order to make the best possible work combination of human and systems (Zammuto et al., 2007). In fact, a society-wide paradigm shift is taking place, on the work that people do, with professions being automated or changed entirely (West, 2018). An early example of such a change took place as ERP and other software systems have been adopted by corporations and institutions. For example, the typical work of civil servants, as we know them today, is transforming from work that takes place on a street level, to work that takes place on a screen level and eventually towards work that takes place on a systems development level (Bovens & Zouridis, 2002). In another example, the role of accountants changed significantly, with their activities becoming more about interpreting data and consulting on their interpretations instead of gathering data. Consequently, the responsibilities of accountants changed as well, since they were considered fit for more strategic decision-making, business management and information technology (IT) initiatives (Caglio, 2003). In a more recent example, the role of lawyers in the delivery of legal services is transformed by Machine Intelligence. Important areas of lawyers' work like discovery, legal search, document generation, brief generation, and prediction of case outcomes is expected to be provided by machine intelligence in the present and near future (McGinnis & Pearce, 2014).

As a result of a large amount of such examples, the generally accepted and dominating perspective of human-machine collaboration is that during the execution of work, computer systems and humans have complementary merits (McAfee & Brynjolfsson, 2017). Computer systems take over the repetitive, non-value-adding tasks (e.g. data gathering in the accountant example) and humans take over the intellectual tasks (e.g. data interpretation and consultation in the accountant example) (Hammer & Champy, 1993). One of the underlying assumptions that lead to this shift of professional roles was that machines can develop abilities that complement those of humans and that humans actually are better in some qualities, like judgement, creativity and decision-making. According to (Hammer & Champy, 1993), *"People working in a reengineered process are, of necessity, empowered. As process team workers they are both permitted and required to think, interact, use judgement, and make decisions"*. However, even though this assumption contains a large dose of truth about the nature of decision-making, it does not explain ways in which optimal decision-making can be achieved. Specifically, humans are perceived indeed to be more creative and have better judgement, but this is not always the case. In fact, our judgement is bounded with certain natural limitations (Simon, 2000) and we have a large set of biases in our natural decision-making processes (Kahneman, 2011). Therefore, a more comprehensive perception is that even in decision making, there is room for machines to contribute resulting to the fact that all human, social, organizational and technical aspects are intertwined with technological developments (Baxter & Sommerville, 2011).

A decision making process can be described as the application of judgement on a prediction for a particular outcome and then acting accordingly to achieve the desired outcome (Agrawal et al., 2018). However, in order to have the right actions taking place, knowledge is required and thus, experience and information are needed to be entered in a training process (McAfee & Brynjolfsson, 2017). Also, as the action takes place, it needs to be evaluated, usually with a comparison of the action's outcome against the desired outcome and consequently generate more experience and knowledge. Therefore, for this sequence to be successfully executed in a typical decision-making process, Agrawal, Joshua & Goldfarb (2018) are pointing out that there are some distinct elements that constitute a decision-making task. These are (as shown in figure 2.4), the input of information through a judgement activity, a prediction activity and a training activity. Specifically, judgement is applied on an educated prediction, leading to an action that results to a specific outcome. The resulted outcome provides feedback to training, in order to further expand the available knowledge and improve the quality of decision making (Agrawal et al., 2018).

Consequently, the types of decision-making elements, as defined by (Agrawal et al., 2018) are basically activities and data. First of all, there are the activities of prediction, judgement, action and of course the resulted outcome. However, these activities are essentially valuable only when they can be described with data that provide feedback to train the decision-making process and build experience (Janssen, van der Voort, & Wahyudi, 2017). This built experience is what allows predictions to take place in similar situations that might appear in the future. Educated predictions in combination with judgement from people is what actually leads to actions and decisions (Turban et al., 2014). Until recently, activities of predictions had limited ways to influence the overall decision-making because of the simple economic idea that their returns did not surpass the required investments of effort and resources. Specifically, it was either too expensive to have because significant human effort in analysis was required, either not accurate enough that would not be of significant value because not sufficient data were available (Agrawal et al., 2018). However, this changes with the emergence of datafication (Mayer-Schönberger & Ramge, 2018). Hence, the economics of information goods, like predictions, are increasingly becoming more profitable for organisations and industries (McAfee & Brynjolfsson, 2017).

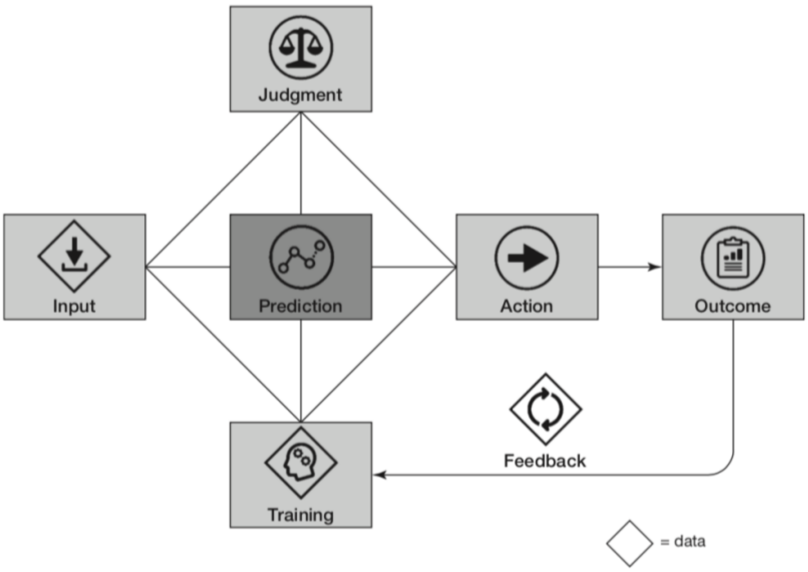


Figure 2.4: *Anatomy of a Decision-Making task, from (Agrawal et al., 2018)*

2.2.2 Data-Driven Decision Making

2.2.2.1 Decision Support Systems & Business Intelligence: two sides of the same coin

One of the main reasons that makes the economics of predictions work is the abundance of data. Specifically, now we have in our environments technologies like IoT (Atzori, Iera, & Morabito, 2010; Gubbi, Buyya, Marusic, & Palaniswami, 2013) and organisations adopt systems that digitally document, measure and capture their entire business (McAfee, 2002). This enables businesses with an unprecedented understanding on what is actually going on in their operations and thus, they are able to make important improvements that stem from large amounts of data (McAfee & Brynjolfsson, 2012). These large amounts of data are distinct for their volume, velocity, veracity, variety and value as it is repeatedly described in many studies (Sivarajah, Kamal, Irani, & Weerakkody, 2017) and they evolved during the years from the emergence of technology driven economic models (Mayer-Schönberger & Ramge, 2018; McAfee & Brynjolfsson, 2017) and business models (Schroeder, 2016).

Therefore, this resulted to a simultaneous evolution of Business Intelligence and Analytics which is the process of transforming the data generated into value (Gandomi & Haider, 2015). Specifically, data are valuable when they can have economic impact and *"The potential value of Big Data is solved simply when leveraged to the drive decision-making process"* (Sivarajah et al., 2017). This leads to the pursue of Business Intelligence, which is what enables and realises this impact, as (Chen, Chiang, & Storey, 2012) describes. Specifically, Business Intelligence and analytics at first was the process of creating actionable knowledge from structured data that were stored in systems' Data Bases and hence, in its essence was mainly taking place in Data Base Management Systems. However, with the emergence of Web-Based applications and interactive systems like social media and collaboration tools, valuable data increasingly became unstructured and therefore, new techniques and approaches needed to be in place in order to cope with such challenges. Finally, we are on the verge of the digitalisation of everything and the generation of data from interconnected objects and mobile devices, leads to a new wave of Business Intelligence and Analytics, that needs to cope also with complex networks of sensory data (Chen et al., 2012).

Decision-Support Systems evolved in a similar way and with the common objective to transform large amounts of data into actionable knowledge, in order to support decision-making (Turban et al., 2014). Additionally, as organisations are increasingly adopting Data-Driven operational structures and business models, the elements of data, information and analytics are gaining a more significant role in their operations (Delen & Demirkan, 2013). The increased amount of data made it challenging for Decision-Support Systems to generate valuable insights and hence, novel methodologies needed to come into place that would allow the transformation of data into knowledge suitable to support decision-making (Elgendy & Elragal, 2016). These new ways lead to the development of methodologies that usually follow the steps of 1) Data Collecting, 2) Data Preparing, 3) Data Analysing and 4) Decision Making (Janssen et al., 2017). These steps are what constitute the Data Mining process in decision-making and as Sivarajah et al. (2017) describe, the types of analytical methods that are commonly in place can fall into the categories of Descriptive Analytics, Inquisitive Analytics, Predictive Analytics, Prescriptive Analytics and Pre-Emptive Analytics (Sivarajah et al., 2017), as shown in figure 2.3.

Descriptive Analytics are about establishing ways to describe through historical data patterns of past behaviour and the current state of business processes and operations (Delen & Demirkan, 2013). Having established descriptive analytics methods, organisations are basically targeting to answer questions of *"What happened in the Business?"* (Sivarajah et al., 2017). This means that in descriptive analytics is needed to have digital representations of the business and thus, description of entities like processes, workflows, installed base and Customers' accounts in an integrated manner. Therefore, data management is a big part of descriptive analytics and it covers the development of definitions for data in order to have a common ground in describing business entities digitally (Assunção, Calheiros, Bianchi, Netto, & Buyya, 2015). According to Assunção et al. (2015), this requires having internal communication between different systems and hence, it predisposes systems integration for a systematic management of data. Data are often in a variety of structures and this raises a set of integration challenges that can be addressed with defining global data models and data cleansing strategies, or analytics in unstructured data (Chen et al., 2012). When the question of "What is happening in the Business?" is answered, it is possible to use the developed information to address the question of "Why is happening?" with inquisitive (or diagnostic) analytics, which is more insightful.

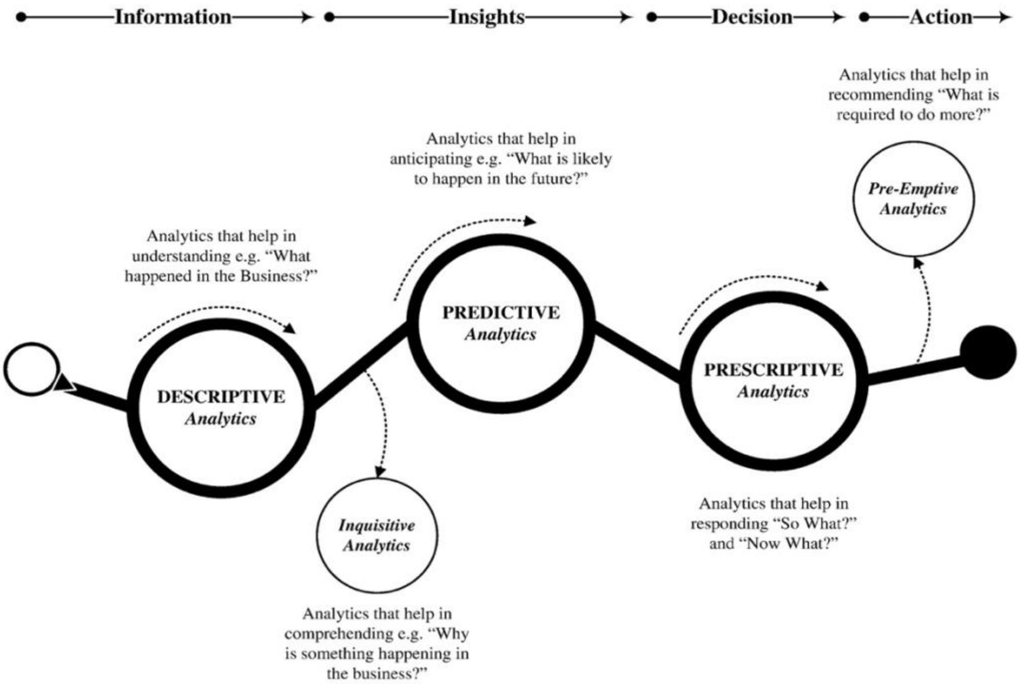


Figure 2.5: *Types of Big Data analytical methods (Sivarajah et al., 2017)*

Specifically, inquisitive analytics are about deriving information from data and being able to see patterns of information in these data (Delen & Demirkan, 2013). Additionally, through logical combination and aggregation of data, new relations can be observed as well as correlations (Assunção et al., 2015). Usually, the results of descriptive analytics is searchable databases, visualisation of metrics in the form of reports and dashboard applications that assist decision-making (Watson, 2014). Therefore, interpretation in inquisitive analytics rely mostly on human input (e.g. from Business Analysts) for monitoring, interpretation and judgement (Banerjee, Bandyopadhyay, & Acharya, 2013). The results from inquisitive analytics provide summarisation of knowledge in statistical methods as well as correlations and then human judgement is needed to translate the developed insight into actions (Rehman, Chang, Batool, & Wah, 2016). Further, with the right interpretations in place, the next question that analytics methods intend to answer is that of "What is likely to happen in the future?". This means that after it was possible to understand the past and the present, analytics methods attempt to transform this information into the knowledge of what might possibly happen in the future (Sivarajah et al., 2017).

Therefore, predictive analytics are about managing the knowledge derived from information and predicting possible outcomes based on the analysis of historical information (Rehman et al., 2016). The idea is that by making sense of what has happened in the past, through understanding patterns and relationships, organisations can develop knowledge of what might happen in the future (Gandomi & Haider, 2015). In order to achieve the development of such knowledge, Gandomi & Haider (2015) relate the analytics techniques and methods that can be used into a set of categories, depending on the nature of the data and the desired outcome. This comes in line with the notion in which prediction machines can be used depending on the nature of the prediction at hand and the desired economic and business outcome (Agrawal et al., 2018). For example, for business elements that are represented by time-series (e.g. cycle time of epics in portfolio management), the most suitable prediction techniques are of the likes of moving averages that identify historical patterns and extrapolate them to the future. For business elements that have a direct causal effect (e.g. critical success factors cause specific outcome in Portfolio Management), techniques like linear regression are more suitable, since they can capture and exploit interdependencies between explanatory variables and outcome variables, to conclude in a prediction (Gandomi & Haider, 2015). However, only developing knowledge with data-driven predictions does not necessarily mean that added value is created (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017).

The developed knowledge needs to lead into business actions in order to be valuable. Prescriptive analytics is what utilises the developed knowledge to enable the recommendation of actions and courses of actions that will achieve comprehensive business process optimisation (Sivarajah et al., 2017). Specifically, prescriptive analytics not only indicate what is likely to happen, but also estimate the business implications of potential actions that can come in response (Chaphalkar, Iyer, & Patil, 2015). Therefore, in prescriptive analytics a real-world system is comprehensively simulated digitally, and the knowledge created is beyond unidimensional predictions of specific variables. It is about the digital representation of the evolution of a business process during time and potential actions, in order to optimise it for achieving strategic objectives (Banerjee et al., 2013).

Furthermore, with pre-emptive analytics, the best possible set of actions can be evaluated and recommended (Banerjee et al., 2013; Delen & Demirkan, 2013; Gandomi & Haider, 2015). Hence, these types of analytics enable an organisation to transform knowledge and information into actions to take. For example, prescriptive and pre-emptive analytics enable the development of recommendation systems and provide answers to the questions “So what?” and “What is required to do more?”.

2.3 Computer-Aided Decision Making in Portfolio Management

2.3.1 Predictive Models for project selection

The data that can be gathered from projects can have a valuable contribution to making the decision-making process in portfolio management data-driven and evidence-based. In order to do so, it is required to develop more methods that process them, analyse them and transform them into comprehensions that are meaningful in their specific context (Gandomi & Haider, 2015). Such a method is the Artificial Neural Network that is developed based on a set of critical success factors and it can relate or map specific attributes of project management with the potential outcome of projects (Costantino, Di Gravio, & Nonino, 2015).

The Artificial Neural Network that Costantino et al. (2015) developed, provides a good example on how to define critical success factors and metrics but also how to quantify them and use them for prediction. In addition, different outcomes of projects that determine success can be predicted with statistical significance from the degree of effort in agile planning, moderated by vision or goal quality, experience of the team and complexity of the project (Serrador & Pinto, 2015). These studies and in combination with (Dikert et al., 2016; Martinsuo & Lehtonen, 2007) provide a set of predictors to project outcomes, as shown in table 2.1.

Study	***Predictors***
(Costantino et al., 2015; Dikert et al., 2016; Martinsuo & Lehtonen, 2007; Serrador & Pinto, 2015)	Quality of project vision, mission, goals
(Costantino et al., 2015; Dikert et al., 2016; Martinsuo & Lehtonen, 2007)	Systematic decision-making and support from top management
(Costantino et al., 2015; Dikert et al., 2016; Serrador & Pinto, 2015)	Effort in (agile) project scheduling and planning
(Costantino et al., 2015; Dikert et al., 2016)	Client Consultation
(Costantino et al., 2015; Dikert et al., 2016; Serrador & Pinto, 2015)	Personnel and team experience
(Costantino et al., 2015; Dikert et al., 2016)	Project Complexity / Technical Tasks
(Costantino et al., 2015; Dikert et al., 2016)	Client Acceptance
(Costantino et al., 2015; Dikert et al., 2016; Martinsuo & Lehtonen, 2007)	Availability of information for monitoring & feedback
(Costantino et al., 2015; Dikert et al., 2016)	Communication
(Costantino et al., 2015; Dikert et al., 2016)	Troubleshooting

Table 2.1: *Predictors of project outcomes from literature*

Furthermore, there is evidence that efficiency in single project management is correlated with the efficiency of the overall portfolio management. Specifically, the clarity of specifying goals, the availability to decision-makers of single projects information and decision making in a systematic manner are positively related to portfolio management efficiency (Martinsuo & Lehtonen, 2007). Consequently, in the evolution of a portfolio during time, there are a set of decision-making milestones related to projects, that need to systematically take place, as illustrated in figure 2.4.

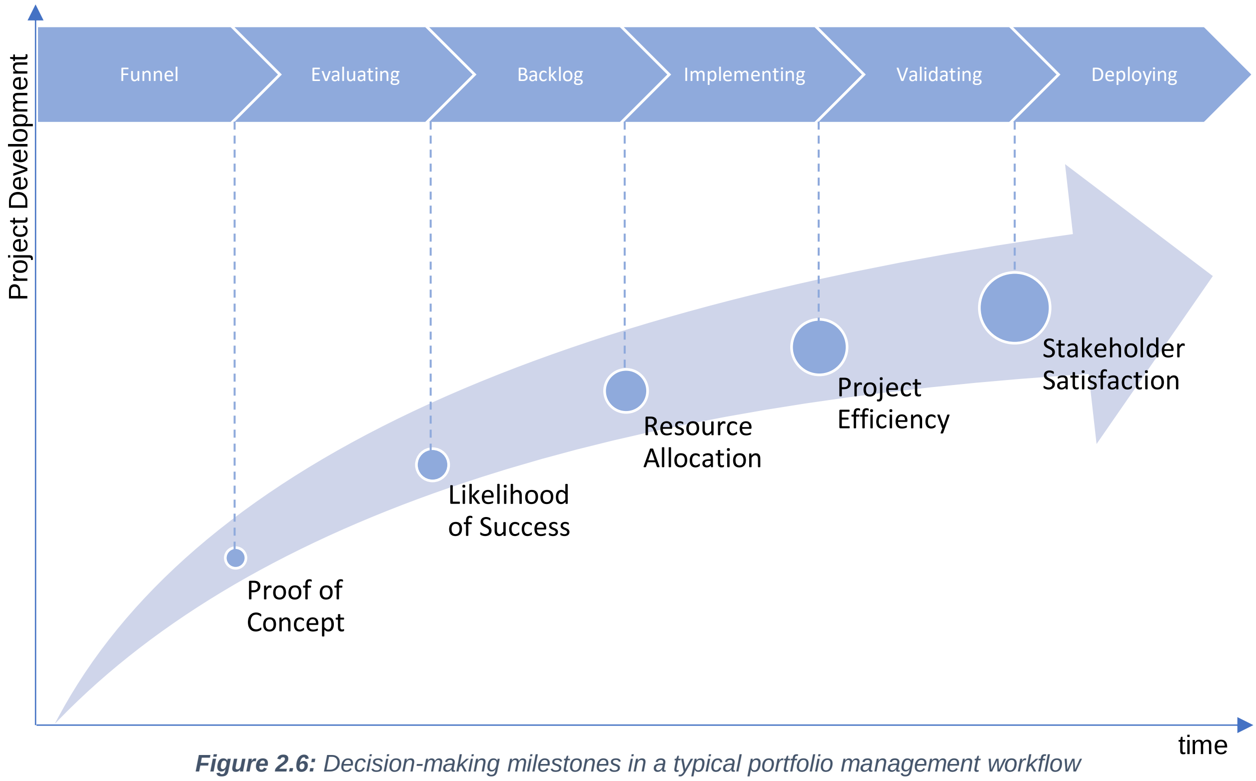


Figure 2.6: *Decision-making milestones in a typical portfolio management workflow*

2.3.2 Project Success Factors

Different outcomes of projects that determine their success, such as overall project success, project efficiency and stakeholder satisfaction can be predicted with different predictors (Serrador & Pinto, 2015). Therefore, it is important to also list the set of factors that predispose project success. Specifically, the outcome of software-intensive projects is determined from the Institutional Context that consists of People and Action, Development Processes and Project Content (McLeod & MacDonell, 2011). These determine a set of dimensions in projects according to McLeod & MacDonell (2011) and the most relevant ones for this study are technical, economic, behavioural, psychological and political. Additionally, according to Serrador & Pinto (2015), the success outcomes of projects can be measured with measures like the success rating of project sponsors and stakeholders, budgeting goals, time boundaries, scope and requirements and satisfaction of the team, client and end users. A summary of the factors that show project success are shown in table 2.2.

Study	***Factors that describe success***
(McLeod & MacDonell, 2011)	Technical factors
(McLeod & MacDonell, 2011)	Economic factors
(Serrador & Pinto, 2015)	Success rating of project stakeholders
(Serrador & Pinto, 2015)	Budget goals

(Serrador & Pinto, 2015)	Time goals
(Serrador & Pinto, 2015)	Scope and requirements goals
(Serrador & Pinto, 2015)	Team's satisfaction
(Serrador & Pinto, 2015)	Client's satisfaction
(Serrador & Pinto, 2015)	End users' satisfaction

Table 2.2: *Factors that measure project outcome*

2.4 Summary of Literature & Gap Analysis

Organisations often face the challenge that agile project management of agile teams, is very often decoupled from the strategy of their parent organisation (Patanakul et al., 2012). Therefore, portfolio management needs to be brought up to speed with new agile operational settings. In addition, establishing a systematic way to evaluating, selecting and executing projects, is essential in order to achieving a decision-making process that allows making the right judgement calls by design within a project portfolio (Archer & Ghasemzadeh, 1999). Specifically, there is a large amount of knowledge for structures that target agility within teams with small-scale agile methodologies. However, there is a general lack of proven methodologies for continuous and systematic organisational agility. Mechanisms for large-scale agility are not that advanced but increasingly necessary (Moe, Dingsøyr, & Dybå, 2010; Yeow et al., 2018).

Managing the complexity of large-scale agility is usually challenging for the scope and capabilities that typical agile teams operate on (Moe et al., 2013; Strode et al., 2012). There are ways to enhancing them in order to contribute in organizational strategies of systematic response (Patnayakuni & Ruppel, 2010) and existing research suggests that scaling the scope of agile teams in organisations is challenging but possible (Dikert et al., 2016). Additionally, even though there are plenty of experience reports, there is little academic research on cases that scaling of agile practices for organisational agility was successfully done (Dikert et al., 2016). Furthermore, the practical approaches that are in place are not optimised for achieving the full potential of computer-aided decision-making that prediction machines can assist with. There are not many predictive tools that help agile portfolio management. Possibly, this is due to the ambiguity of the criteria that are in place during decision-making processes on higher management levels and consequently, most of the tools are descriptive of projects.

Descriptive analytics rely solely on Business Analysts to monitor, interpret and make judgements (Banerjee et al., 2013). Hence, it can be characterized as decision-making based on intuition, since hindsight is what will make a Business Analyst search for something, monitor something or interpret something. Most of the tools available for project management are on the descriptive analytics side and the need for more predictive methodologies is addressed and tackled partly with predictive models for project selection (Costantino et al., 2015). However, in the developed solutions for predicting the outcomes of projects, there are some limitations and potential for further improvement. Even though there are known predictors of project success that increasingly gain research momentum (Serrador & Pinto, 2015), there are not many tools and methodologies that utilise them for predicting project success and incorporating these predictions in project portfolio management (Costantino et al., 2015). Furthermore, the predictive analytics models that are developed are limited to a small number of implementations and a variety of predictive techniques are yet to be tried and tested.

As a result, there is a gap, in literature and practice, of decision-support systems and tools that actively help project portfolio managers in decision-making. Furthermore, the quantification of the success factors takes place with data mostly from interviews but there is an abundance of project related data that can be analysed with smart methods. Such data are computationally demanding to analyse and the cost of analysing them often exceeds the benefits of the knowledge that can be developed (Agrawal et al., 2018). In addition, the data used are usually from a very specific type of projects in a niche industry and therefore, it is not accurately representative for portfolio management in general. Finally, even though several tools have been developed for predicting outcome of projects based on specific metrics, most of the models do not sufficiently generate strong correlations between specific variables and the outcome. Therefore, they can be characterized as spurious correlations and therefore, not accurate predictors (Gandomi & Haider, 2015).

3 Research Method

3.1 Research Approach

One of the main objectives of this research is to evolve the analytical capabilities that decision-support systems provide in project portfolio management. Specifically, we aim to introduce predictive analytics in the decision-making process of project portfolio managers. We do so by studying two cases in which Machine Learning models can develop as prediction machines. The two case studies are analysed, and the steps of the research approach used are shown in figure 3.1. Our developed models are able to predict the likelihood of projects in each case study for achieving their specific targets. We illustrate this with the prediction of achieving budgeting/funding targets and user satisfaction targets. The two cases show the possibilities of assisting the decision-making of portfolio managers using Machine Learning.

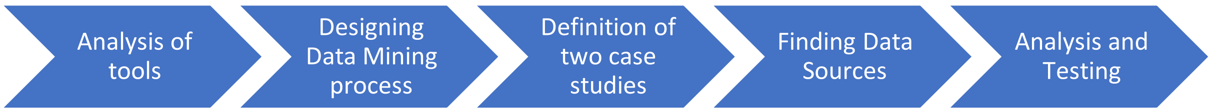


Figure 3.1: *Research approach for two case studies*

First, we analyse a set of project portfolio management tools, and we identify the analytical gap they have, based on the theoretical framework developed from the related literature. Then, we design and develop a set of models that can enhance the available tools and methodologies with prediction. Specifically, we do so with data from two different cases, of independent projects that correspond to initiatives within agile portfolio management. We attempt to model measures of project success as well as different predictor and moderator measurements of project success with operational data of projects. Then, we process and analyse these data with the goal to utilise them and develop prediction machines that use Machine Learning models. These models can be used to assist decision-making of project portfolio managers by analysing and predicting the outcome of a large number of projects that otherwise would be costly and time consuming to analyse. Finally, we interpret the resulted outcome and identify future work for evolving on the topic. This will focus on what are the next steps to increase the analytical capacity and data-driven decision-making.

The intention is to embed in the process of Agile Project Portfolio Management analytical capabilities of prediction. First, this takes place by enhancing the analysis of projects that are in the Project Funnel and Evaluating phase. Specifically, our Decision-Support-System contributes with the accelerated evaluation and selection of projects that will evolve to the Backlog, as indicated in figure 3.2. We intend to add value in the process of Agile Portfolio Management by enabling the repetition of project evaluation and selection with analyses that does not require extensive human intervention. For this to be realised, we develop prediction machines that learn from operational data of projects. With the prediction machines, the process can be repeated multiple times from the phase of Evaluating in a more systematic, seamless and efficient manner. Moreover, the same method can be applied in other decisions of the agile project portfolio management process.

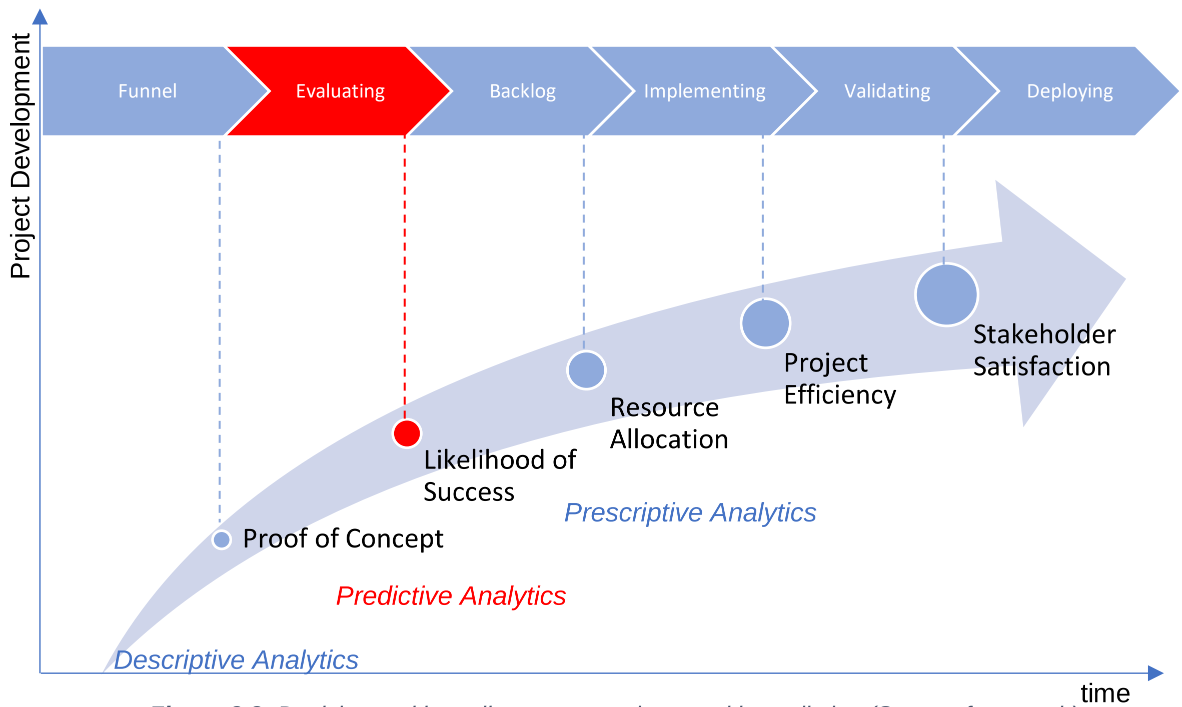


Figure 3.2: *Decision-making milestones to enhance with prediction (Scope of research)*

3.2 Analysis of Decision Support Systems for portfolio management

In the attempt of organisations to manage in the best ways possible their project portfolios, they use Project Portfolio Management tools in their operations that help them adopt the right operational models. Such tools complement the widely known collaboration tools that act on the project level, like JIRA and Trello. Specifically, portfolio management tools are targeted to support the decision-making process and the evaluation of the progress of projects and teams in regard to performance, contribution to strategic objectives but also added business value. In addition, such tools generate operational data that represent the organisations' business models and business processes. The generated data from these tools, if analysed can provide many insights during the different phases that a project goes through in portfolio management.

Therefore, data analytics capabilities are needed to increasingly enhance existing Project Portfolio Management tools that are used in practice. In this section, we intend to provide an analysis of a set of tools, that are used for (Agile) Project Portfolio Management. The methodology of our analysis is based on the analytical abilities that the tools have on one hand and the effectiveness in supporting decision-making with usable interactions on the other hand. As a result, since the tools to be analysed are decision support tools for project portfolio management, the qualities illustrated in our analysis are on the basis of the following criteria:

1) *Intelligence* derived from data (analytical capabilities), and
2) *Usability* for effective decision-making (capabilities for maximising the utilisation of the available data, information and analytics).

The categories that the tools and their dashboard can fall in terms of the criterion "intelligence" stem from the Big Data analytical methods of (Sivarajah et al., 2017) and are:

1) *Descriptive Analytics* which is considered to be fit for what we call *Intuition-driven decision making*,
2) *Predictive Analytics* which is considered to be fit for what we call *Data-driven decision making,*
3) *Prescriptive Analytics* which is considered to be fit for what we call *Knowledge-driven decision making*.

The categories that a dashboard can fall in terms of the criterion "usability" are based on the ability to visualise actionable knowledge. Specifically, they are:

1) *Static Data Presentation* which is considered to be fit for what we call *Individual Decision Making,*
2) *Dynamic Information provision and collaboration* which is considered to be fit for what we call *Collaborative Decision Making* in a team,
3) *Interactive and Actionable Business advice and coordination* which is considered to be fit for what we call *Coordinated Decision Making* through multiple teams.

The analytical framework that is described above is visualised in figure 3.3. In our analysis we include the tools FlightMap and TargetProcess but some other tools that are used in similar use cases are Workboard, ASANA, Aha.io and CELOXIS. In general, all tools have similar analytical capabilities (mainly Descriptive Analytics) and they mainly differ in their User Experience.

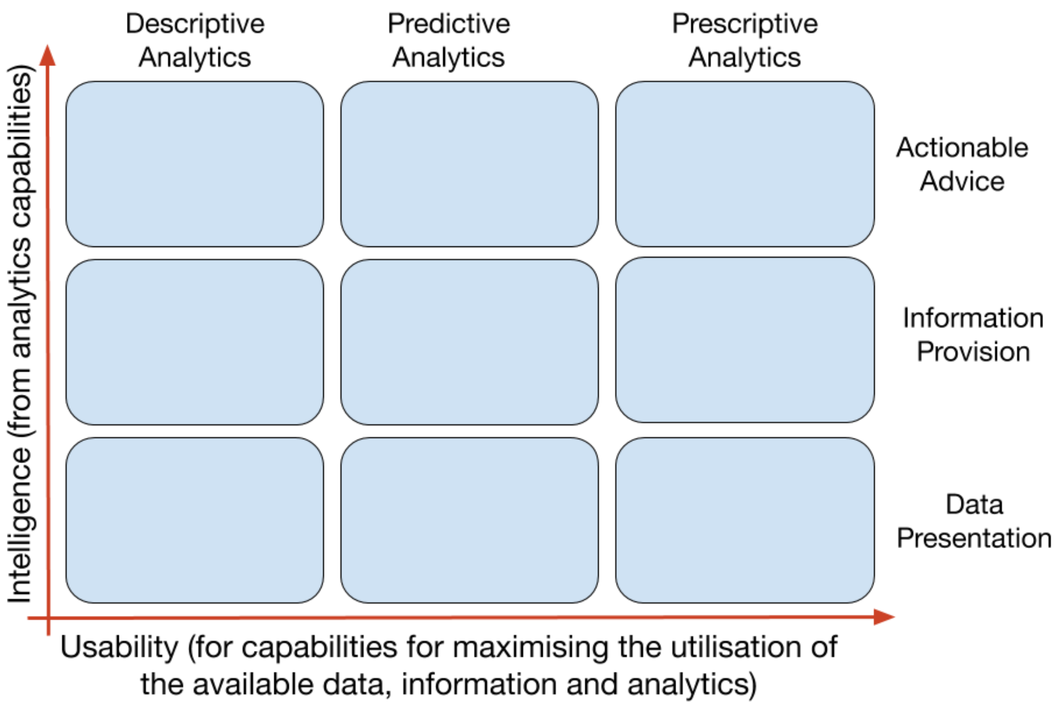


Figure 3.3: *Framework for analysing tools for Project Portfolio Management*

3.3 Data Mining method for Project Evaluation

In order to facilitate the support of decision-making in Project Portfolio Management, we will focus on introducing the element of Predictive Analytics in the part of project evaluation and selection. The intention is to develop a Prediction Machine or Classifier that estimates at a certain level of accuracy the outcome of projects. Our predictions are intended to act as input to the portfolio manager, human actor that will make the final decision. The benefits of our Prediction Machine are that it facilitates the more rational evaluation of projects based on their available data. Therefore, data-driven decision-making will be possible. Moreover, the use of models from our Prediction Machine (or classifier) provides a significant economic benefit. Specifically, such evaluations of projects at the scale of 378,661 projects or 7,200 mobile applications is a highly expensive and demanding task to be executed manually. The cost of just running the models to get a preliminary analysis are virtually none since they can be run at a simple computer machine. Therefore, there is a significant improvement on the costs of analysing and estimating the outcome of projects and thus, the performance of the portfolio.

As indicated in figure 3.4, our approach is what Janssen et al. (2017) describe as the process for data-driven decision-making in their study. First, we collect data that are representative to the decision at hand. Specifically, the collected data are from the digital footprint of independent projects that are in their initial phase and they request funding or budgeting in order to proceed and evolve. Then, we prepare the collected data in order to suit our analysis. The raw, collected data need to be cleaned and prepared accordingly in order to have a quantification of their features and outcome. In addition, in order for the analysis to be reliable, it is required to include only data that would normally be available during the phase of project evaluation and not in later stages of the evolution of a project. Further, the data are analysed with the use of a set of Machine Learning models, that lead to the development of the prediction machine or classifier. Finally, we provide the most suitable interpretations of the predictions generated, in order to achieve data-driven decision-making.

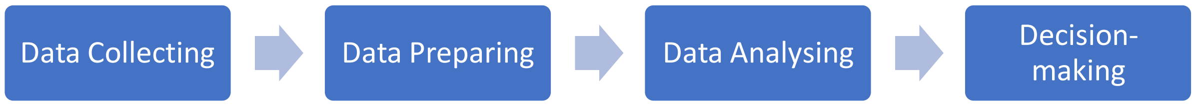


Figure 3.4: *Research Methodology for data-driven decision-making (Janssen et al., 2017)*

3.4 Data Gathering of cases

In order to achieve the quantitative project evaluation, operational data are required to be in place of projects and portfolios. Such data need to cover a set of attributes that describe the projects and our objective is to find patterns and correlations between such attributes and the outcome of projects. The attributes of the project in the data can be descriptive data of the projects' details or more in-depth representations of the project management and operations of the team. Descriptive data include information like category or type of each project, project objective, duration, requested budget or funding, customer satisfaction and so on. Operational data might be more detailed metrics of the way that project teams operate, and information gathered from process mining.

In our analysis, we intend to analyse different types of data that represent projects of organisations that operate with Agile Portfolio Management. Agile Portfolio Management predisposes organisational structures of autonomous independent teams that create value-networks as well as innovation and entrepreneurship within the organisation. Consequently, we considered that innovation projects that take place within the same innovation ecosystem provide a representative sample of how value-networks work within organisations. As a result, we gathered and analysed open source data that are provided in open innovation ecosystems like Kaggle and specifically, data of projects from Kickstarter and App Store. The first dataset used in our analysis is published in Kaggle (www.kaggle.com/datasets) by (Mouillé, 2017) under the licence "Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)" as described in (Creative Commons, n.d.). Therefore, our analysis is also publicly available under the same licence and this research contains our interpretation in relation to project portfolio management. The data are used as a metaphor to how an agile organisation of autonomous teams operates. The second dataset we used is also published in Kaggle (www.kaggle.com/datasets) by (Perumal, 2017) under the GNU General Public licence as described in (Free Software Foundation, 1991). The data describe mobile applications from iOS Apple App Store, and they are mined in 2017.

3.5 Testing and evaluation of Prediction Machines

Moreover, in order to increase the reliability of our analysis, we evaluate the developed machine learning models with the ten-fold cross-validation technique. This enables the testing of the models in the entire data-set. Specifically, the dataset is divided into a training set and a testing set ten times and the Machine Learning models are developed iteratively, as shown in figure 3.5. Therefore, the models are built and tested in ten different combinations of training and testing. As a result, their performance calculated is more reliable and the accuracy is closer to real-world situation. This way, it is possible to calculate the probability of success of a project with a standard deviation.

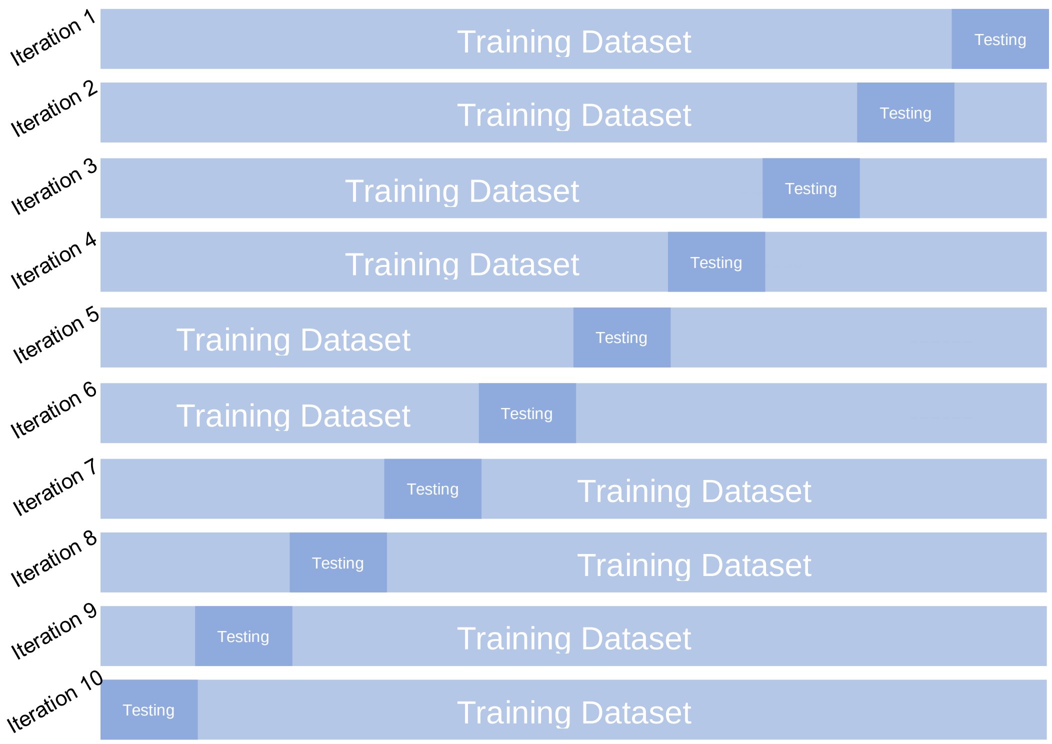


Figure 3.5: *Ten-fold cross-validation*

4 Analysis of Portfolio Management Decision Support Systems

First, we analyse the current state-of-the art in portfolio management tools based on our analysis described in chapter 3. The objective is to classify the tools that assist decision-making of portfolio managers based on their analytical capabilities and usability. This will allow us to practically verify our argument that current decision-support systems in portfolio management operate mainly in the descriptive spectrum of analytical capabilities. Therefore, this chapter provides a description and analysis of tools that are used by portfolio managers for supporting their decision-making.

4.1 Project Portfolio Management Dashboards and Metrics: FlightMap

The first decision-support system that we cover is through a summary of FlightMap and its dashboards. As a Decision Support tool for Project Portfolio Management, FlightMap intends to assist the creation of value through better decisions in a project portfolio. The web-based application provides several dashboards and views with insights for managing and understanding the different projects of a portfolio. Our analysis provides a summary of some dashboards available in FlightMap with a description, purpose, potential metrics that can be included and quality. Next, we describe our analysis based on the described theoretical framework with some more details on the application, dashboards and analysis.

The information that FlightMap provides and in combination with other data and information that can be included in the dashboards, results in a very rich tool in terms of content. There are many features that help visualise different types of information and the potential combinations of different types of content are many.

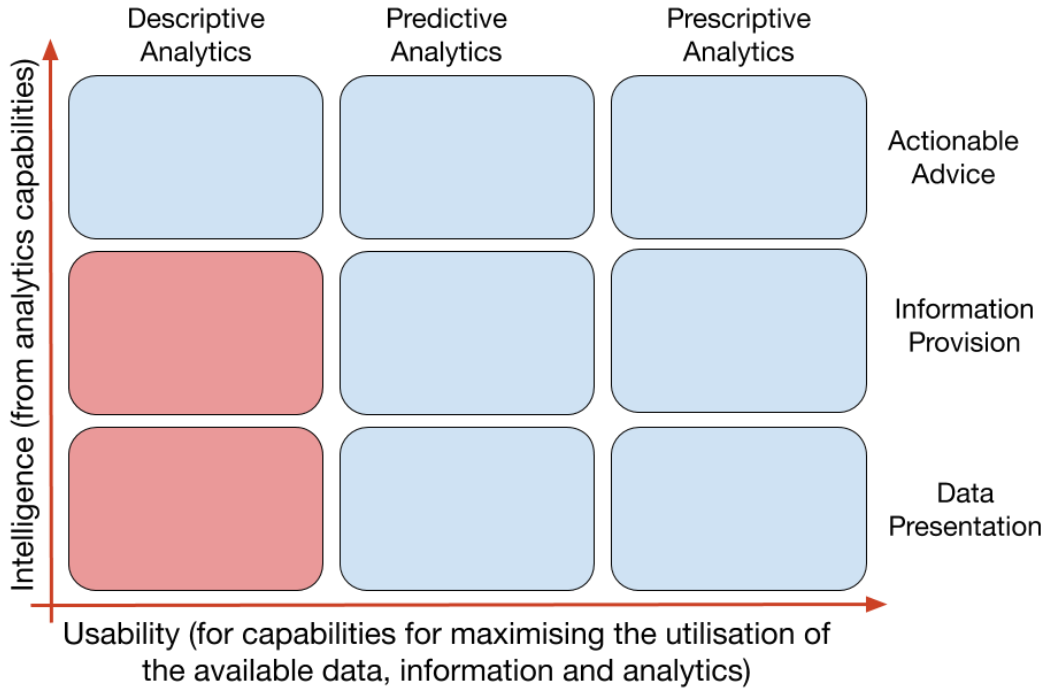
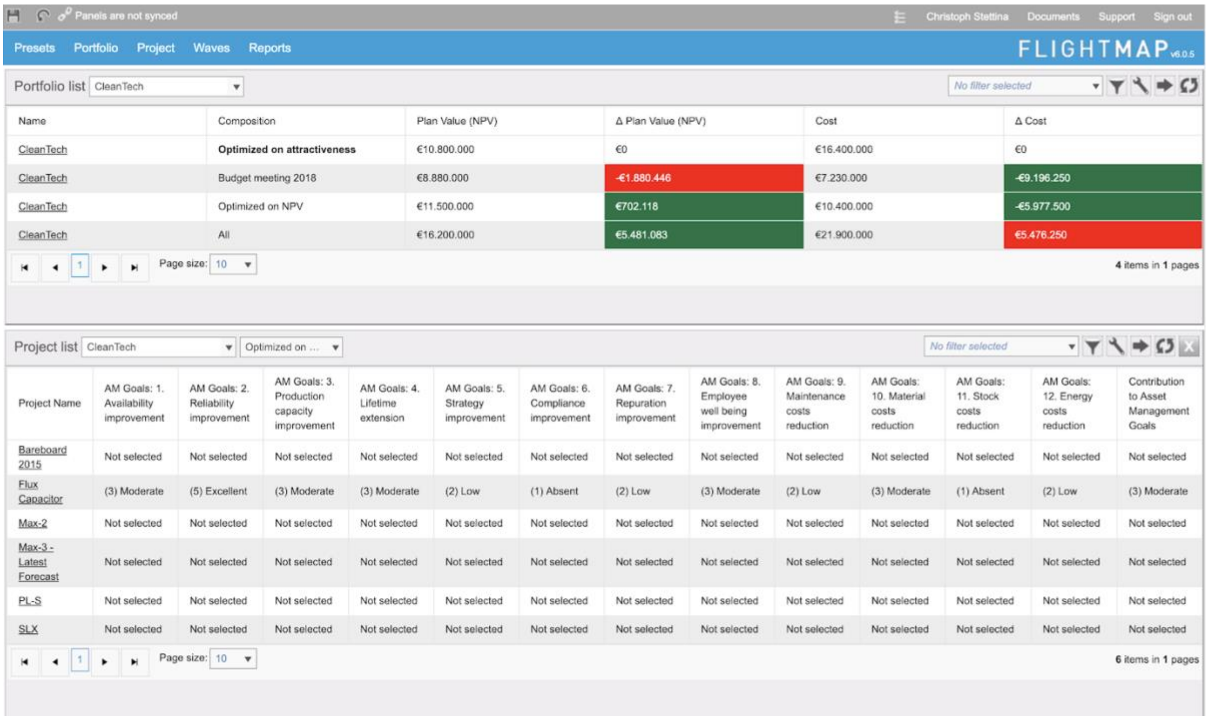


Figure 4.1: *Analysis of FlightMap*

In addition, the tool visualises data of the past and current states of a portfolio and its projects. However, it seems that there are limited predictive capabilities based on the aggregation and statistical analysis of large amounts of data. Therefore, there is not available any prescriptive capability of providing actionable recommendations. Furthermore, the information provided is many times static with limited interactions on the composition of information from data. However, that is not always the case and therefore, it is usually performing better than just for data presentation. Even though there are not many options for collaborating in the tool.

The first dashboard we analysed is that of *Portfolio List & Project List* as showed in the figure below. The purpose of the dashboard is to provide an overview list of the portfolio with its different compositions and their financial performance. In addition, it views an overview list of the projects in a specific composition with their impact/contribution in the business goals. The objective of such a dashboard is to take judgement on portfolio performance and project contribution



Name	Composition	Plan Value (NPV)	Δ Plan Value (NPV)	Cost	Δ Cost
CleanTech	**Optimized on attractiveness**	€10.800.000	€0	€16.400.000	€0
CleanTech	Budget meeting 2018	€8.880.000	-€1.880.446	€7.230.000	-€9.196.250
CleanTech	Optimized on NPV	€11.500.000	€702.118	€10.400.000	-€5.977.500
CleanTech	All	€16.200.000	€5.481.083	€21.900.000	€5.476.250

Project Name	AM Goals: 1. Availability improvement	AM Goals: 2. Reliability improvement	AM Goals: 3. Production capacity improvement	AM Goals: 4. Lifetime extension	AM Goals: 5. Strategy improvement	AM Goals: 6. Compliance improvement	AM Goals: 7. Repuration improvement	AM Goals: 8. Employee well being improvement	AM Goals: 9. Maintenance costs reduction	AM Goals: 10. Material costs reduction	AM Goals: 11. Stock costs reduction	AM Goals: 12. Energy costs reduction	Contribution to Asset Management Goals
Bareboard 2015	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
Flux Capacitor	(3) Moderate	(5) Excellent	(3) Moderate	(3) Moderate	(2) Low	(1) Absent	(2) Low	(3) Moderate	(2) Low	(3) Moderate	(1) Absent	(2) Low	(3) Moderate
Max-2	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
Max-3 - Latest Forecast	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
PL-S	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected
SLX	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected	Not selected

Figure 4.2: *Portfolio List & Project List of FlightMap application*

Another dashboard that FlightMap has is that of the Portfolio Bubble. Specifically, this dashboard provides a representation of the impact that each project has in a portfolio composition based on specific criteria, like *Attractiveness* and *Fit* to the Business goals. This allows the comparison of different projects and the specific example shows how the metric of Innovation Cost relates to contribution.

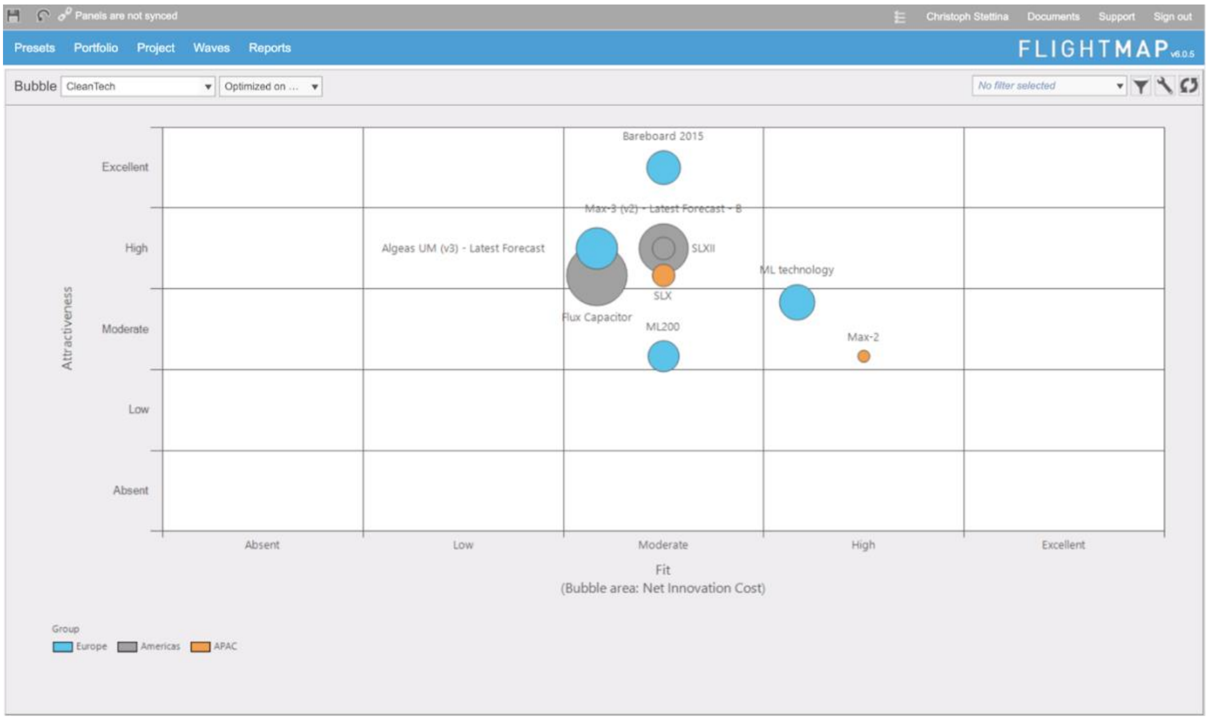


Figure 4.3: Portfolio Bubble dashboard from FlightMap

Decisions of managers can be influenced depending on the progress that projects are undertaking, and the dashboard of Portfolio Funnel provides an overview of the undergoing projects. This dashboard shows a representation of the progress' state and the risk for each project in a portfolio composition. This allows a portfolio manager to compare the state of different projects in relation to allocated resources or costs.

Figure 4.4: Portfolio Funnel dashboard from FlightMap

Another type of content is provided through a dashboard of Portfolio over time and Resource Planning. Specifically, this dashboard provides the relation of projects with Innovation Costs and allocated resources (in FTEs) over time. This allows the portfolio manager to relate Innovation Costs of each project with allocated resources.

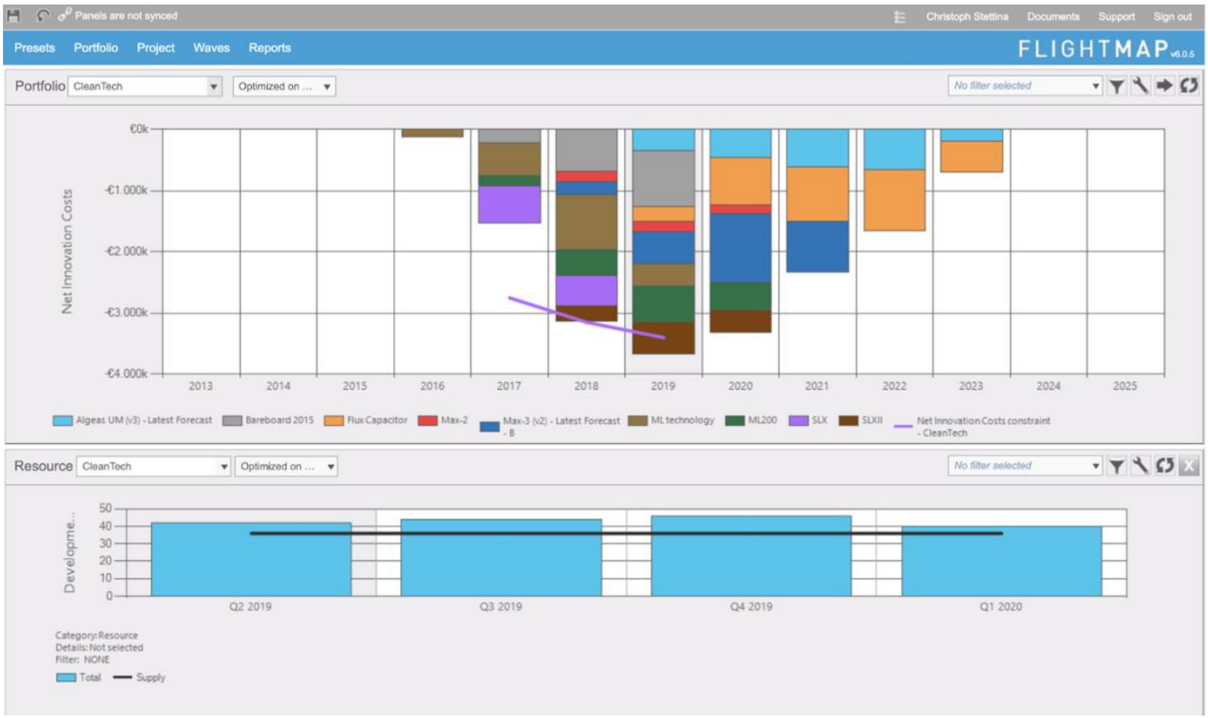


Figure 4.5: Portfolio over time and Resource Planning dashboards from FlightMap

Finally, in the dashboards of Balance and Roadmap a portfolio manager can view the balance of the value (to-be) delivered in each stage of the portfolio and the roadmap with the projects of each stage. Therefore, this can help to make a judgement for the value (to-be) delivered over time.

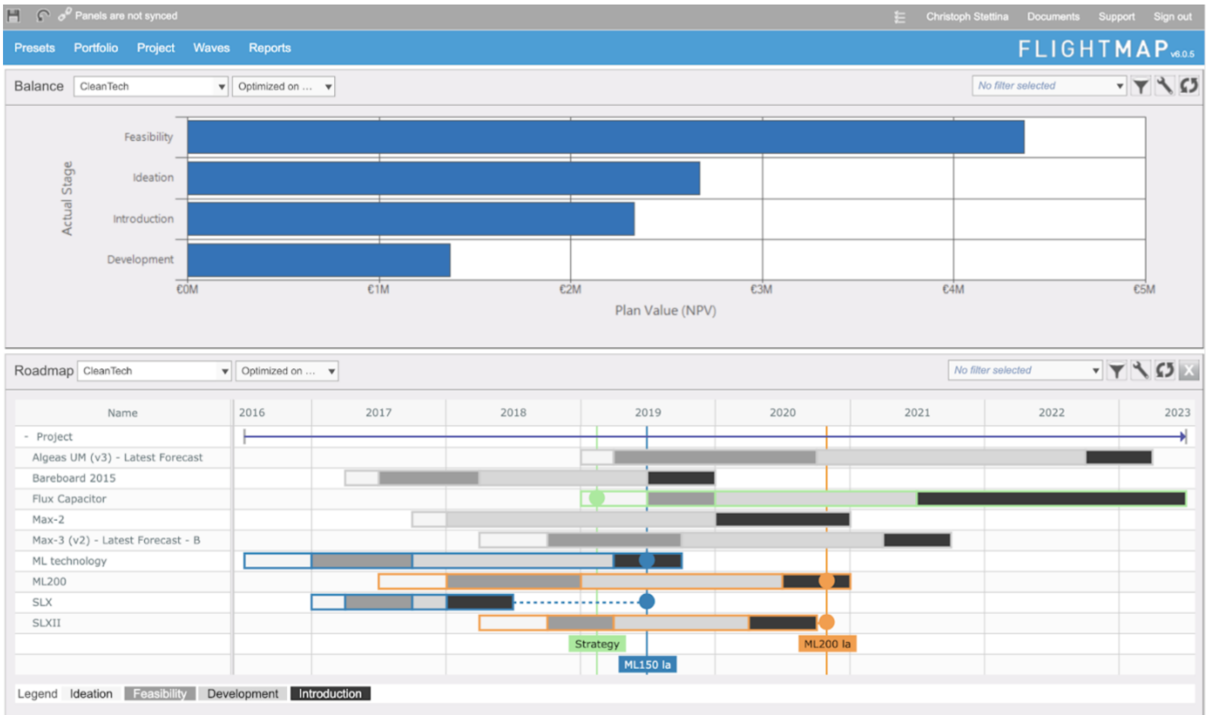


Figure 4.6: Balance and Roadmap dashboards from FlightMap

4.2 Agile Portfolio Management Dashboards and Metrics: TargetProcess

The second decision support system that we cover is Target Process which is in fact also a collaboration tool. Target Process is a relatively focused application with a deep specialisation in Agile Project Management and Agile Portfolio Management. Specifically, it provides configurations that can support the application of agile methodologies in large scale, with implementations of frameworks like SAFe and LeSS, Agile Portfolio Management configuration and Scrum of Scrums. To do so, it provides typical features that are required to manage in an agile way a project, a program and/or a portfolio. These features are keeping track of the different initiatives and teams with backlogs and cards that represent activities, tasks and projects as epics (depending on the level required to view). For the management of such activities it provides not only collaborative features but also reports, overviews and several ways to list the cards. Followingly, a set of such reports and overviews are described in an analysis on the axels of Intelligence and Usability.

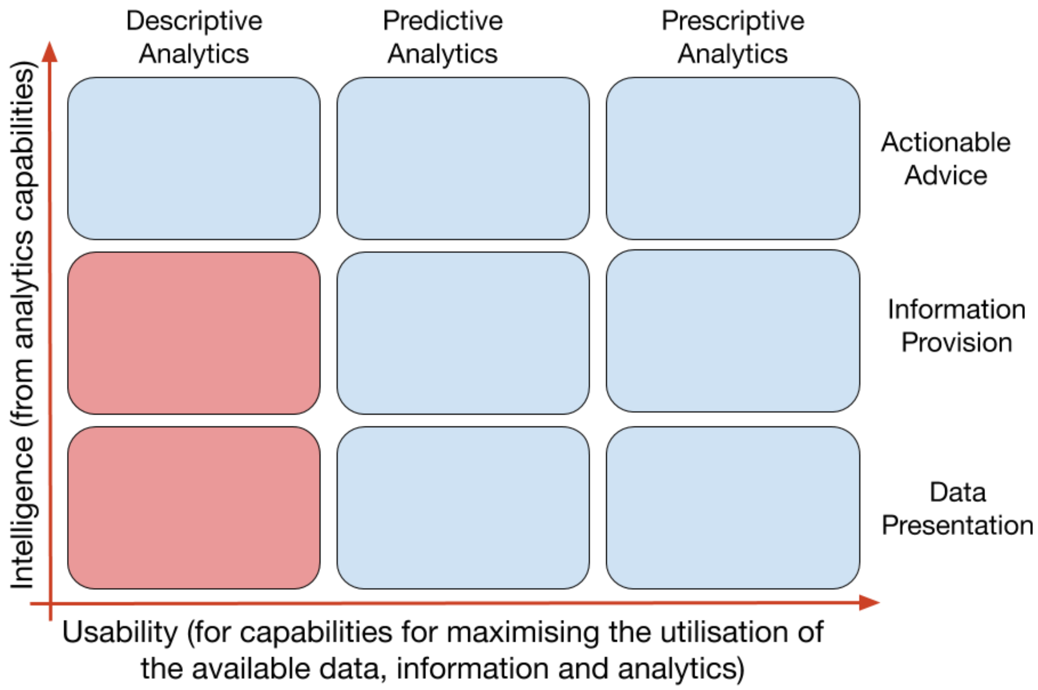


Figure 4.7: *Analysis of TargetProcess*

Firstly, one of the core dashboards of Target Process is the overview listing of cards and because we focus on the portfolio level, in Figure 2.10 we show an example of the Epic Backlog. This provides an overview of the projects that a portfolio has as epics, along with their relevant information. Specifically, the relevant information that are included in every epic are the Features that its project contains and within each Feature the User Stories or Bugs tasks that belong to it. For example, in Figure 2.10 projects are represented from epics like *List*, *Process Workflow* and so on. The project *Detailed View (aka Entity View)* contains features like *Basic Views Re-work, Views: Re-design Attachments area* and so on. Each of this Features has a set of activities in the form of User Stories or Bugs. The Feature *Make detailed view components' layout independent* will be realised by User Stories such as *Make detailed view title layout independent* and has a Bug that needs work which is *Incorrect styles for Requester title.*

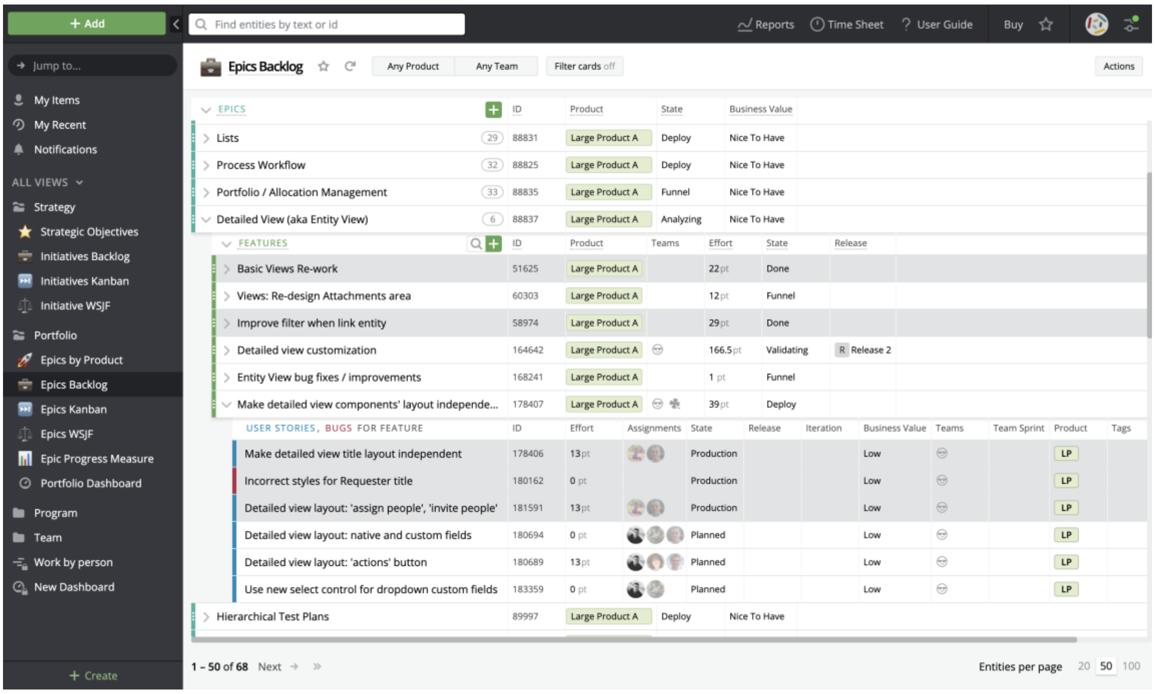


Figure 4.8: Target Process Epics Backlog for projects in a portfolio

Another dashboard of Target Process is the Epics Kanban shown in Figure 4.9. Specifically, a Portfolio manager is able to collaboratively manage the different epics of projects and see an overview of the portfolio's progress. The different projects are represented in epics that are in phases (e.g. project funnel, analysing etc). For example, in figure 4.9, we can see the project *Detailed View (or Entity View)*, which is the same project we have seen before in the stage of *Analysis*. This view can be shared among different team members and therefore, it is a strong tool for alignment between stakeholders in regard to the progress of the portfolio.

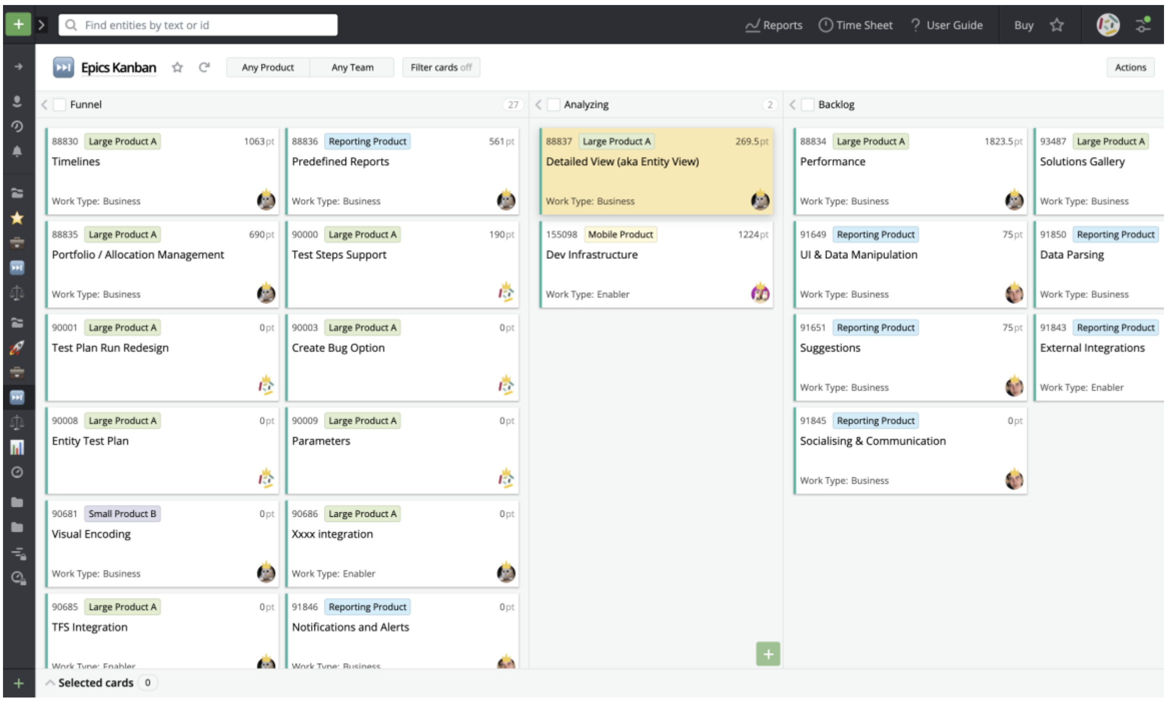


Figure 4.9: Target Process Epics Kanban for projects in a portfolio

In order to make educated judgements on the projects of a portfolio, the tool provides to portfolio managers measurements about them and one of them is the measurement of resources assigned to the portfolio. Figure 4.10 shows the dashboard in which a project manager can see measurements on the progress of projects based on Initial Estimates, Effort Completed, Effort To-Do and Cumulative Effort. This can indicate the state of the portfolio in terms of expectations about projects and projections of how these expectations can be met.

Figure 4.10: *Target Process Progress Measure of projects in a portfolio*

5 Analysis for Project Evaluation

Having verified that decision-support tools of project portfolio management are mainly descriptive, we intend to provide a way to achieve the next analytical capability which is predictive analytics. To do so, we analyse two datasets and create a set of Machine Learning models that are able to predict the outcome of the projects. The first dataset we analyse has 378,661 samples of projects that are published in KickStarter and the second dataset describes 7,200 mobile applications. Chapter 5 describes the analysis process along with the data gathered and the created Machine Learning models.

5.1 Preliminary Analysis of Data

The first dataset has 378,661 entries about projects that were published in Kickstarter and requested crowdfunding in the site. The attributes of the data are as follows (Mouillé, 2017):

- *ID* which is the internal kickstarter id,
- *Name* which is the name of the project (defined as a finite work with a clear goal that somebody would like to bring to life),
- *Category,* which is the type of the project,
- *Main_category* which is the category of the project's campaign,
- *Currency* used to support the project with funding,
- *Deadline* for crowdfunding,
- *Goal* of fundraising which is the funding goal is the amount of money that creators need in order to complete their project,
- Date *Launched,*
- Amount *Pledged* by "crowd",
- *State* of every project that describes its current condition (successful, failed, cancelled, undefined, suspended, live),
- Number of *Backers* which is the amount of people invested in the project
- *Country* pledged from,
- *USD* pledged (amount of money pledged),
- *Goal amount* in USD (amount of funding requested).

The second dataset describes 7,200 mobile applications from the Apple iOS app store and its samples were collected in July 2017. The attributes of the data are as follows (Perumal, 2017):

- *id* which is the Application's ID
- *track_name* which is the name of the Application as published
- *size_bytes* which is the size of the Application (in Bytes)
- *currency* used to sell or purchase the application
- *price* which is the amount requested for purchase
- *rating_count_tot* which is the user rating counts for all versions
- *rating_count_ver* which is the user rating counts for the most recent version
- *user_rating* which is the average user rating value for all versions

- *user_rating_ver* which is the average user rating value for the most recent version
- *ver* which corresponds to the latest version of the Application
- *cont_rating* which is the rating of the content
- *prime_genre* which is the main category or genre that the Application falls into
- *sup_devices.num* which is the number of supporting devices
- *ipadSc_urls.num* that correspond to the number of screenshots showed for display in the App Store
- *lang.num* which is the number of languages that the Application supports
- *vpp_lic* that indicates if VPP (Apple Volume Purchase Program) Device Based Licensing is enabled.

The attributes of the datasets cover partially a subset from all the predictors identified in literature, as shown in table 5.1. An important note is that the mapping of dataset and predictors is only relevant for the time frame that the specific decision at hand takes place. Therefore, we only consider predictors that are relevant for project analysis and evaluation.

PREDICTORS IDENTIFIED IN LITERATURE	KICKSTARTER DATASET COLUMNS	APP STORE DATASET CLUMNS
PROJECT VISION, MISSION AND GOALS	Category, Main_Category, Goal	prime_genre, track_name
SYSTEMATIC DECISION-MAKING AND SUPPORT FROM TOP MANAGEMENT		
EFFORT IN (AGILE) PROJECT SCHEDULING AND PLANNING		
CLIENT CONSULTATION		rating_count_tot, rating_count_ver, user_rating, user_rating_ver
PERSONNEL AND TEAM EXPERIENCE		
PROJECT COMPLEXITY / TECHNICAL TASKS	Category, Deadline, Date Launched	cont_rating, sup_devices.num, ipadSc_urls.num, lang.num, vpp_lic
CLIENT ACCEPTANCE		
AVAILABILITY OF INFORMATION FOR MONITORING & FEEDBACK	Currency, Number of Backers, Country	size_bytes, price
COMMUNICATION		
TROUBLESHOOTING		

Table 5.1: *Datasets and their coverage*

Our objective is to develop a set of models based on the available data that will predict if a project will succeed or not. Therefore, it is initially important to understand how the potential outcomes of projects distribute across the dataset, as it is shown in figure 5.1. That will enable us to judge if the dataset is useful and representative. For example, if most of the projects are in an undefined state then the dataset might not be useful and if all projects are considered successful then the dataset is not representative. A big benefit in using the Kickstarter dataset is that the data describe real-life projects and therefore, they are within the uncertainty context that typically projects go through. In addition, in the agile settings that organisations operate, with intra-company entrepreneurship the dynamics are similar to those of the crowdfunding community.

From the distribution of states, we observe that for most of the projects the outcome is known and specifically, 35.4% is known as successful and 52.2% is known as failed. Additionally, there is a small percentage of the projects that their state is unknown. Specifically, 10.2% is labelled as cancelled, 0.94% as undefined, 0.74 % as live and the rest as suspended. The unknown states cannot be used in our models because the reasons and rationale of projects being cancelled or suspended do not indicate if the outcome would be successful or not. Also, projects that are still live might manage to become successful and for those that are undefined is simply impossible to know their outcome. As a result, we consider only the outcome of successfully pledging the targeted amount and failure to doing so.

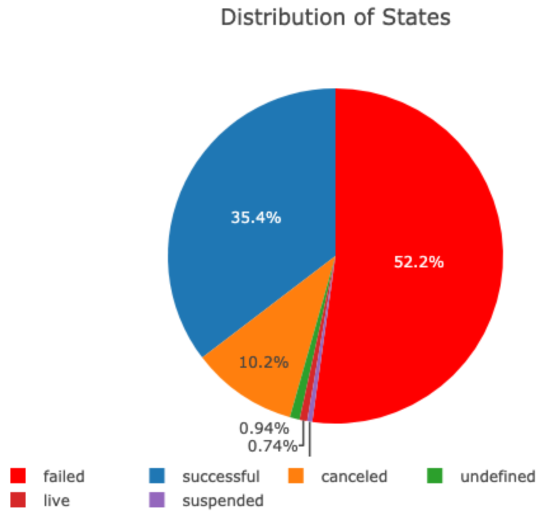


Figure 5.1: *Distribution of projects' states*

5.2 Prediction Machines

Based on the defined data that we describe above; we developed a set of Machine Learning models that try to predict the outcome of projects. The process to do so, follows the steps of data pre-processing, data cleansing, data normalisation and finally knowledge extraction from data, as indicated in figure 5.2. The final step of knowledge extraction using data mining with Machine Learning, is basically the applicability of classifiers. These learn how to classify projects that are likely to succeed on getting funding (or budget) from projects that are not likely to do so on the first dataset. Therefore, our classifier-based prediction machines imitate the decisions that are previously taken for providing or not providing funding and replicate them. For the second dataset, the classifiers learn how to predict if a project will achieve a high rating from users or a low rating. However, in order to achieve this, the data needed to be processed and transformed accordingly.

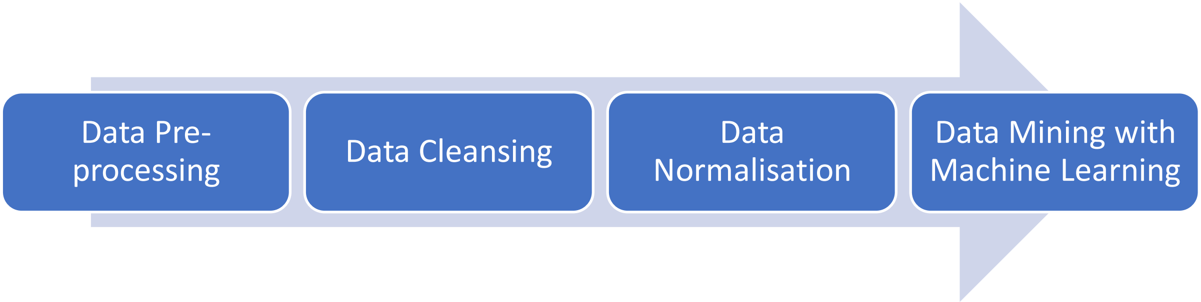


Figure 5.2: *Prediction Machine development process*

5.2.1 Data Pre-processing

First of all, the columns that represent unnecessary attributes were deleted. Specifically, some columns from the dataset do not add significant value for predicting a project's outcome. For example, the version of the project or application is weakly correlated with the outcome since all projects have different versions and there are no significant patterns that can be extracted. Moreover, some of the attributes need to be processed so that their use is better defined. Specifically, the attribute of the date launched by itself does not provide any insight and similarly, the deadline of a project does not provide any significant knowledge that can predict the outcome. However, the combination of these two attributes, can result to knowing the duration of the project, which is important for the outcome.

5.2.2 Data Cleansing

Furthermore, some projects performed in a very different way than the way in which the majority of the projects performed. This abnormal performance can characterise these projects as outliers. Therefore, in order to achieve a realistic modelling of the data they need to be discarded, since they do not represent the typical behaviour of the projects. Another type of data that is cleaned are whitespaces that appear in the text. Text pieces of the data, like descriptions, need to be cleaned from special characters and clarify the boundaries between each piece of text.

5.2.3 Data Normalisation

This leads us to the need of encoding the data in order to be able to quantify them. Specifically, parts of the processed data-set can be transformed from text into a form in which our models would understand it and consequently, into machine language. This includes the shaping of the data and of course, their split into training and test data. The part of the training data is used to develop the intelligence in our Machine Learning models and the part that is testing data is used to evaluate the trained models and estimate their accuracy. Finally, most of the attributes were vectorised in order to establish a quantification of their values without creating distance between different values. For example, the attribute "Category" was transformed into one attribute for every possible entry that a category can have with value 0 or 1. This took place in all columns that can have a finite number of different options.

5.2.4 Machine Learning models that predict project success and project failure

Finally, after the preparation of the data, a set of classifiers is developed that provide the required predictions of success or failure. These classifiers learn from the dataset and classify the projects into two potential categories. Therefore, they are all binary classification techniques.

5.2.4.1 Logistic Regression

The first classifier developed is Logistic Regression, which is a machine learning classification method that acts as a linear model and estimates the probabilities of the possible outcomes. In this model, we approach the decision of project selection as a binary classification problem that has two potential outcomes, success and failure. Specifically, our prediction using Logistic Regression is the probability of the outcome of success and the probability of the outcome of failure. The probabilities are calculated using a logistic function and their evaluation using a cost function that calculates the cost of the error, as shown in equation 5.1. In addition, logistic regression is a parametric algorithm which means that the data need to be regularised in a certain form in order to be processed by the algorithm. This is described by the Penalty parameter of the model.

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \log(\exp(-y_i (X_i^T w + c)) + 1).$$

Equation 5.1: *Logistic Regression Cost Function* (Pedregosa et al., 2011)

The nature of the Kickstarter dataset guided the configuration of the model towards the best approach for binary classification of data that do not have multiple dimensions. Specifically, since our data are mainly descriptive, there are not many attributes that would lead to a high number of dimensions and the number of samples is much larger than the number of dimensions (378,661 samples versus 15 attributes). Consequently, the following configuration was used, as it is indicated in the table 5.2.

Penalty	l2
Formulation	Primal
Tolerance	1e-4
Inverse of regularization strength	1
Bias / Intercept	1
Weights Adjustment	None
Solver	liblinear
Iterations	100
CPU cores	1

Table 5.2: *Configuration of Logistic Regression model*

5.2.4.2 k Nearest Neighbours Classification

With the k Nearest Neighbours Classification algorithm we built a model that is non-parametric, as opposed from logistic regression and therefore, there is no need to regularise the data. Furthermore, the nature of the algorithm considers similarity of the samples in different features and therefore, it does not generalise to a model that is mathematically strict. That means that the model estimates the closeness of samples in the different features and from the distance between samples it classifies the data-points. Figure 5.3 shows a plot on how data points can be classified based on their k-closest neighbours.

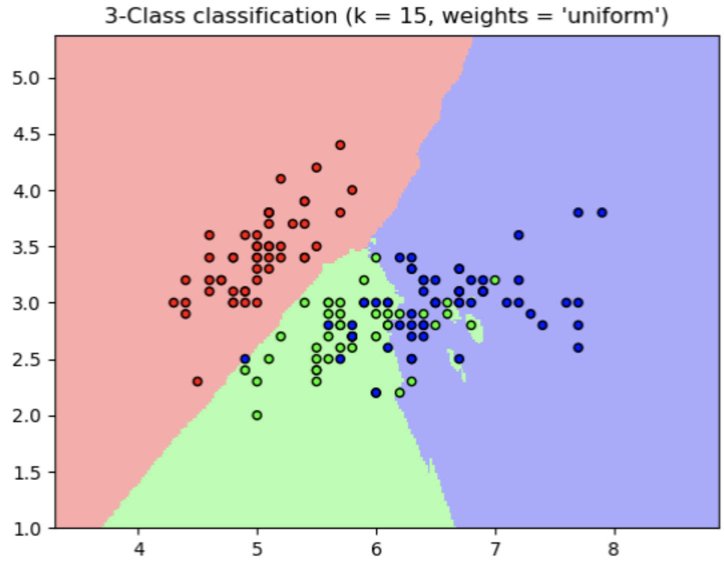


Figure 5.3: *Example plot of k-nearest classifier* (Pedregosa et al., 2011)

Since the projects need to be classified into 2 possible categories, the data-set guided to the parameters shown in table 5.3.

Number of Neighbours	5
Weights	All points are weighted equally
Leaf size	30 in order to execute within the available computational resources
Distance Calculation	Euclidean Distance
Metric	Minkowski
CPU cores	1

Table 5.3: *Configuration of K-Nearest Neighbours model*

5.2.4.3 Decision Tree

Another Machine Learning technique that was used to address the prediction of projects' outcome and help portfolio managers with project selection is that of Decision Trees. Specifically, a Machine Learning model has been built that is not parametric, just like KNN and that classifies through supervised learning the outcome of projects.

However, the way Decision Trees do that is with a set of decision rules that provide a set of choices to go through. Such a set of decision rules are what make a decision tree (a simple example is shown in figure 5.4) and it can be easily interpreted schematically in similar ways. Therefore, decision trees are simple to understand and visualise and very accurate in certain types of data and decisions, especially in types that the options are finite and can be deterministically described. On the other hand, decision trees are prone to overfitting and their use of heuristics might predispose the development of biased trees. Thus, generalisation of decisions is sometimes difficult, and this can influence the accuracy with data that are different from those that trained the model.

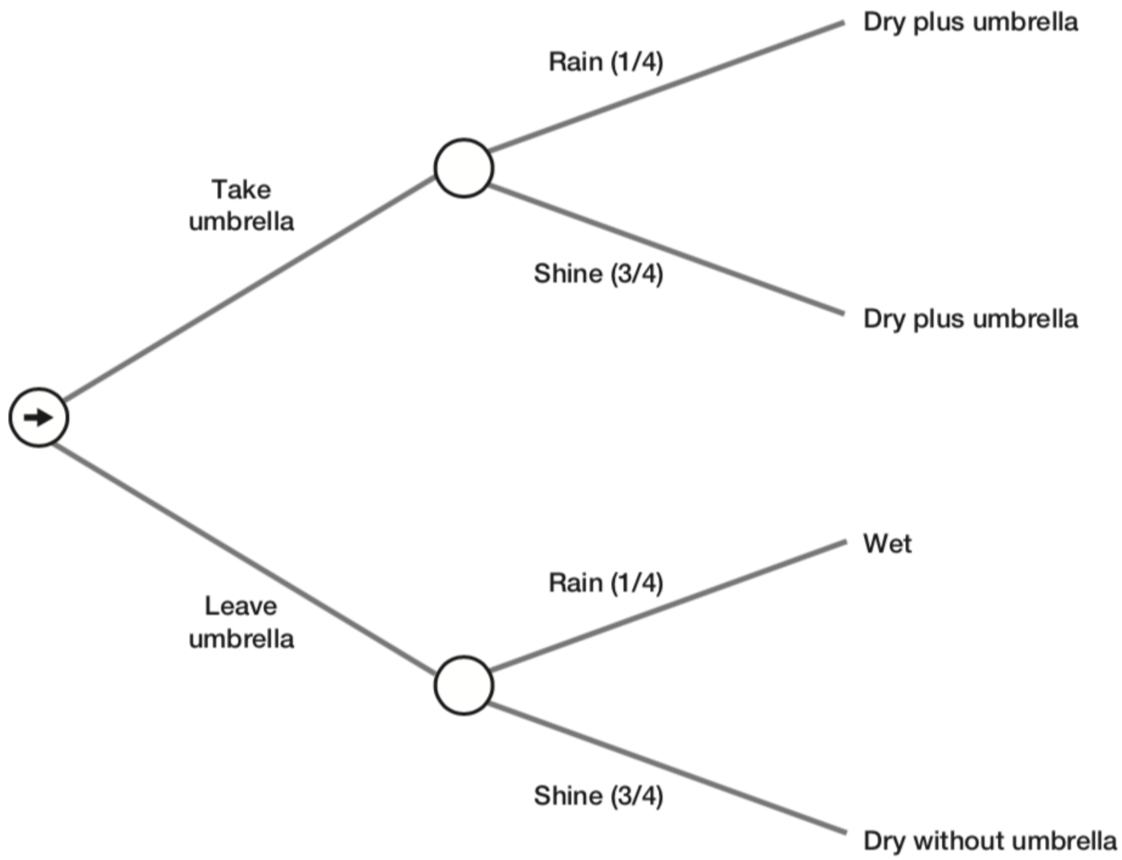


Figure 5.4: *Decision Tree for the decision of taking an umbrella* (Agrawal et al., 2018)

5.2.4.4 Random Forest

The next technique that was used is Random Forest which is a set of Decision Trees that are generated in random subsets from the data. When the decision trees are generated, they are aggregated, and the final classification is calculated. Specifically, in our Random Forest there are ten decision trees that work in the same way as the Decision Tree described in section 5.3.3.3.

5.2.4.5 ADA Boost Classifier

Another technique that was used is ADA Boost Classifier, also known as Adaptive Boost Classifier. This classifier works similarly to Random Forest which means that it develops a classifier for every random subset from the dataset and then aggregates their results in order to achieve a more accurate classification. However, one important difference from Random Forest is that during the iterative training, the different classifiers are weighted, and their weights are adjusted in every iteration. Therefore, ADA Boost Classifiers typically perform even better than Random Forests. In our model, the type of classifier used is again a Decision Tree, but in contrast to Random Forest, we used 50 weighted Decision Trees.

5.2.4.6 Gradient Boosting Classifier

In a similar iterative and aggregative manner of Adaptive Boost Classifier, Gradient Boosting it creates a set of classifiers and gradually improves the accuracy that their combination can achieve. However, the difference in Gradient Boosting Classifier is that instead of adjusting the weights of the classifiers, the improvements take place with the calculation of a loss function. This leads to a more accurate estimation of the difference of each classifier from the correct predictions. Therefore, the classifiers' combination is better trained and thus, achieving better accuracy. Our models use Regression Tree which means that Decision Tree classifiers are used in order to classify the dataset and the logistic regression is used to calculate the loss function.

5.2.4.7 LGBM Classifier

The final model used is LGBM classifier which is a Light Gradient Boosting Classifier based on gradually improving a set of classifiers in an iterative manner. The main benefit of the LGBM Classifier is its ability to train very efficiently and effectively. Specifically, it requires low computation resources of processing power as well as memory. Therefore, it is possible to build a model with marginally more Decision Trees to be aggregated and consequently, better accuracy can be achieved. Our model uses 100 Regression Trees that are trained on the dataset and these classifiers are gradually improved using a regression loss function.

6 Results

6.1 Prediction Accuracy

Our dataset was split and tested with the k-fold cross validation technique into training data and test data. Specifically, every Machine Learning model, run ten times the 378,661 samples and 7,200 applications between training and testing. Every time, the model was tested in a different subset from the data in order to estimate a more accurate metric of performance. The training data allowed the models to develop and create the necessary experience in order to make predictions. On the other hand, the test data are those that we use to evaluate and test our developed and trained Machine Learning models.

After the models were trained, the testing of the models indicates accuracy and the performance of each model can be represented from the average mean and standard deviation of all the testing instances. For the dataset with projects from KickStarter, the performance of the models is in the level of 61.01% for Logistic Regression, 62.47% for K-Nearest Neighbour, 63.72% for Decision Tree and 65.02% for Random Forest, 66.53% for ADA Boost Classifier, 65.21% for Gradient Boosting Classifier and 68.45% for LGBM Classifier. For the dataset with mobile applications from iOS App Store, the performance of the models is in the level of 66.43% for Logistic Regression, 62.43% for K-Nearest Neighbour, 56.88% for Decision Tree and 70.08% for Random Forest, 64.76% for ADA Boost Classifier, for 74.77% Gradient Boosting Classifier and 42.3% for LGBM Classifier.

In such levels of accuracy, even though they are not very high (60 to 70 %), they allow the models to be considered as good predictors of the projects' outcomes, since they are marginally better than the naïve approach of following the majority. However, the used dataset did not contain dimensions and variables that create strong causation relationships with the outcome. For example, there were no attributes that described critical success factors or other predictors of project success like team coordination and leadership. Consequently, our models can act as a good evaluator of projects based on descriptive information and provide insights to Project Portfolio Managers.

Followingly, we provide an analysis of the models' results, in order to understand their strengths and limitations. Table 6.1 shows how every model performs in the first case using the Dataset with the KickStarter projects. Table 6.2 shows how every model performs in the second case using the dataset that describes the mobile applications from iOS. In the next sections of this chapter detailed descriptions are included for the Machine Learning models applied. Specifically, the accuracy of every iteration is provided for every model in the two datasets of the case studies. Furthermore, a selection of calibration plots is presented that show how the main machine learning models perform against the perfect prediction plot. The calibration plots are of k Nearest Neighbours, Decision Tree and Random Forest. Along with the calibration plots, the distribution of the predicted values is presented.

MACHINE LEARNING MODEL	ACCURACY MEAN	IMPROVEMENT ON NAÏVE	STANDARD DEVIATION
NAÏVE (ALWAYS WITH MAJORITY)	0,5959	-	-
LOGISTIC REGRESSION	0,6101	0,01	0,013
KNN	0,6247	0,03	0,006
DECISION TREE	0,6372	0,04	0,005
RANDOM FOREST	0,6502	0,05	0,005
ADA BOOST CLASSIFIER	0,6653	0,07	0,006
GRADIENT BOOSTING CLASSIFIER	0,6521	0,06	0,004
LGBM CLASSIFIER	0,6845	0,09	0,003

Table 6.1: *Machine Learning Models Accuracy and standard deviation for the first case*

MACHINE LEARNING MODEL	ACCURACY MEAN	IMPROVEMENT ON NAÏVE	STANDARD DEVIATION
NAÏVE (ALWAYS WITH MAJORITY)	0,6643	-	-
LOGISTIC REGRESSION	0,6643	0	0.001
KNN	0,6243	-0,04	0,043
DECISION TREE	0,5688	-0,1	0,204
RANDOM FOREST	0,7008	0,04	0,242
ADA BOOST CLASSIFIER	0,6476	-0,02	0,239
GRADIENT BOOSTING CLASSIFIER	0,7477	0,08	0,246
LGBM CLASSIFIER	0,4230	-0,24	0,356

Table 6.2: *Machine Learning Models Accuracy and standard deviation for the second case*

6.2 Analysis of Predictions from Classifiers

For the first dataset, the Logistic Regression model achieves the lowest levels of accuracy, in comparison to the rest of the models. In addition, the accuracy is very close to the naïve approach of following the majority. Specifically, the probability of predicting correctly using the Logistic Regression model is better than the naïve approach at a factor of less than 2 percent and therefore, there is no significant improvement. Consequently, the analysis of the model evolution does not add any significant value. With ten-fold cross-validation, the model was trained and tested ten times in a different permutation of the dataset as training and testing data. Each iteration resulted to a probability of successful prediction as shown in table 6.3 and all the probabilities together were used to calculate the standard deviation and potential error of the prediction.

Iteration	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
Accuracy	0,608	0,608	0,608	0,608	0,608	0,608	0,609	0,608	0,608	0,629

Table 6.3: *Predictions for each training iteration of Logistic Regression for the first case*

For the second dataset, the Logistic Regression model achieves the same accuracy as the naïve approach and it is by far the most reliable, in comparison to the rest of the models. Specifically, the accuracy does not improve on the naïve approach of following the majority. With ten-fold cross-validation, the model was trained and tested ten times in a different permutation of the dataset as training and testing data. Each iteration resulted to a probability of successful prediction as shown in table 6.4 and all the probabilities together were used to calculate the standard deviation and potential error of the prediction.

Iteration	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
Accuracy	0,664	0,664	0,664	0,664	0,664	0,664	0,665	0,665	0,665	0,665

Table 6.4: *Predictions for each training iteration of Logistic Regression for the second dataset*

On the other hand, the k Nearest Neighbours classifier improves prediction by almost 3 percent for the first and second dataset which is a valuable improvement. The learning process of the k Nearest Neighbours classification algorithm can be described as stable and linear. The more data and the more diverse data are learnt from the model, the more similarities can be found with the testing data. This is a natural consequence from the nature of the algorithm which identifies similarities between data through the distance between samples. Figure 6.1 shows how the k Nearest Neighbours model calibrates and improves in accuracy.

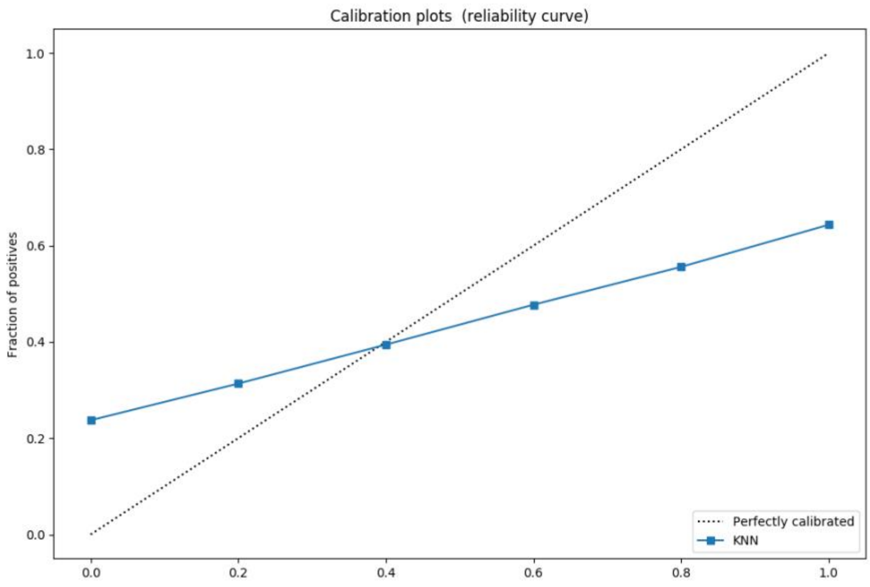


Figure 6.1: *k Nearest Neighbours Calibration*

With ten-fold cross-validation, the model was trained and tested ten times in a different permutation of the dataset as training and testing data. Each iteration resulted to a probability of successful prediction as shown in table 6.5. All the probabilities together were used to calculate the standard deviation and potential error of the prediction.

Iteration	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
Dataset #1	0,623	0,627	0,621	0,622	0,623	0,629	0,626	0,621	0,627	0,627
Dataset #2	0,603	0,618	0,651	0,625	0,613	0,610	0,656	0,644	0,638	0,584

Table 6.5: *Probability of successful prediction for each iteration of KNN*

The next model, which is the Decision Tree, achieves improved accuracy in comparison to logistic regression and k Nearest Neighbours for the first dataset and performs purely for the second dataset. Specifically, even the naïve approach of following the majority performs better than the Decision Tree. The development of the model shows a curve that reaches a plateau and even decreases in accuracy while approaching the extreme values. This is an indication of overfitting based on specific data samples and inability to generalise. In addition, the decision tree identifies specific patterns that classify samples into distinct failure or distinct success as indicated in the distribution of the predicted values. Specifically, most of the samples fall into the category of predicted near-zero values or predicted near-one values. Figure 6.2 shows how the Decision Tree model calibrates and improves in accuracy, as well as how the predicted values distribute across the samples (the predicted values are the model's estimated outcome of the project – success or failure).

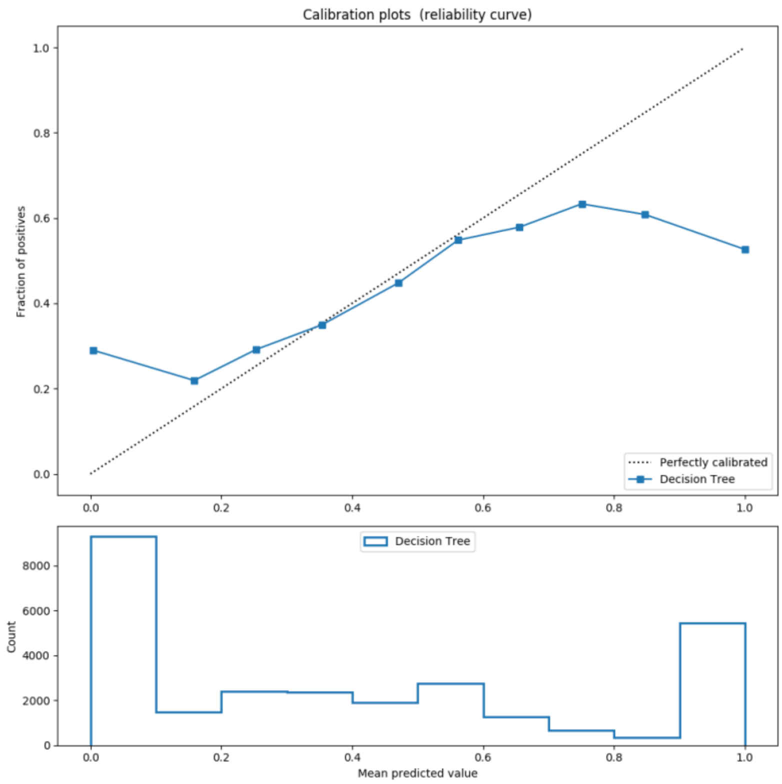


Figure 6.2: *Decision Tree Calibration*

With ten-fold cross-validation, the model was trained and tested ten times in a different permutation of the dataset as training and testing data. Each iteration resulted to a probability of successful prediction as shown in table 6.6. All the probabilities together were used to calculate the standard deviation and potential error of the prediction.

Iteration	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**
Dataset #1	0,636	0,637	0,635	0,640	0,634	0,639	0,638	0,635	0,642	0,635
Dataset #2	0,610	0,738	0,665	0,565	0,551	0,499	0,524	0,567	0,630	0,338

Table 6.6: *Probability of successful prediction for each iteration of Decision Tree*

The Random Forest classifier seems to deal better with the generalisation since the curve in the calibration plot does not show such a steam decrease and the predicted values are more equally distributed. Therefore, learning is increasingly improved as indicated in figure 6.3. In addition, the higher accuracy is evident also in the distribution of the mean predicted value. Specifically, the mean predicted value of different samples distributes in a wider range than the previous models. Therefore, the samples are classified more accurately.

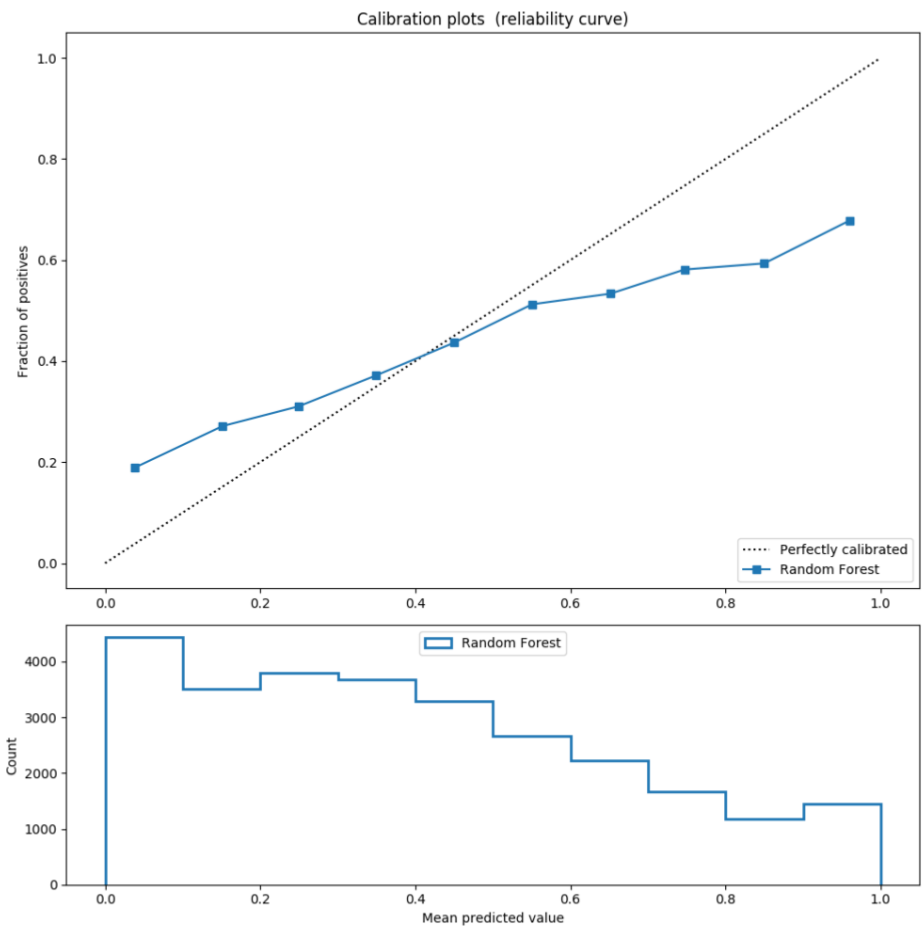


Figure 6.3: *Random Forest Calibration*

With ten-fold cross-validation, the model was trained and tested ten times in a different permutation of the dataset as training and testing data. Each iteration resulted to a probability of successful prediction as shown in table 6.7. All the probabilities together were used to calculate the standard deviation and potential error of the prediction.

Iteration	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10
Dataset #1	0,652	0,651	0,646	0,650	0,647	0,651	0,649	0,648	0,655	0,650
Dataset #2	0,691	0,751	0,781	0,749	0,757	0,738	0,732	0,709	0,757	0,345

Table 6.7: *Probability of successful prediction for every iteration of Random Forest*

Similarly, the same took place for the rest of the models that gradually improve a set of classifiers. Their accuracy of each iteration in the ten-fold cross validation is shown in table 6.8. Interestingly, LGBM classifier struggles to predict in the second dataset due to the smaller data set. In addition, in the second data set the tenth fold systematically achieves lower prediction rates in relation to the rest.

ADA Boost Classifier										
Iteration	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**
Dataset #1	0,664	0,667	0,665	0,662	0,664	0,670	0,669	0,659	0,665	0,668
Dataset #2	0,682	0,736	0,743	0,643	0,618	0,696	0,780	0,573	0,669	0,335
Gradient Boost Classifier										
Iteration	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**
Dataset #1	0,652	0,652	0,650	0,651	0,658	0,653	0,654	0,648	0,653	0,654
Dataset #2	0,652	0,740	0,771	0,761	0,814	0,825	0,851	0,825	0,821	0,417
LGBM Classifier										
Iteration	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**	**# 6**	**# 7**	**# 8**	**# 9**	**# 10**
Dataset #1	0,685	0,685	0,684	0,685	0,685	0,683	0,683	0,682	0,687	0,686
Dataset #2	0,713	0,743	0,574	0,203	0,311	0,300	0,362	0,268	0,394	0,363

Table 6.8: *Probability of successful prediction for every iteration of ADA Boost, Gradient Boost and LGBM Classifiers*

7 Discussion & Limitations

This chapter includes the main interpretations and findings from the performed research and the results from the analysis. Also, these findings and interpretations are related with similar studies and reflected upon similar contributions. This enables the identification of how this study contributes further in Decision-Support Systems for Agile Project Portfolio Management. The research question of this study is also answered, along with some of the guiding questions that were not fully addressed in the previous chapters. In addition, after answering the research question and some of the guiding questions, the main limitations of the performed analysis are indicated.

Specifically, in section 7.1 the possibilities of Predictive Analytics in the context of Agile Project Portfolio Management are reflected upon with the criteria of viability, integration with the agile process, utilisation of operational data and enrichment of toolset. Section 7.2 describes how can one enable his/her own prediction machines for agile portfolio management through Machine Learning, reflected upon the evolution of analytical capabilities than one can have and the type of data that one can utilise for each type of analytics. In section 7.3, according to the case studies, prediction machines can integrate with the agile portfolio management process by decomposing decision-making milestones and automating them, which enables the iterative nature of the agile process. Section 7.4 reflects on the case studies and the findings against the external forces that lead to the need of Agility and dynamism in organisations in the first place.

7.1 Using Predictive Analytics for Agile Portfolio Management

The results from the performed analysis show the viability of introducing predictive analytics in project evaluation and project selection, during agile project portfolio management. Specifically, the results provide an indication that it is possible to analyse and predict if a project will achieve pre-set targets and goals. This finding is in line with what other similar studies observe. Studies that apply Machine Learning for project selection in portfolio management and attempt to predict project success (Costantino et al., 2015; de Oliveira, Valentina, & Possamai, 2012). Therefore, with this study the analytical capabilities of Decision-Support Systems for Agile Project Portfolio Management are evolved and enhanced with the element of Prediction. This is done with multiple Machine Learning classifiers, in the specific context under investigation.

The context under investigation includes the project portfolio management process, in which there are certain milestones that such a tool can be used by providing predictions. These milestones are before every major phase that projects go through. In that way, such a tool can be embedded in the process. Specifically, every time an epic that represents a project moves from one phase to another, the models can run and predict the likelihood of successfully achieving pre-set targets of the projects in the phase. An important contribution of this study's findings is in achieving a decision-making process in agile project portfolio management that is data-driven and evidence-based, using existing operational data (Gandomi & Haider, 2015). Specifically, our analysis contributes in enriching the set of methods that process, analyse and transform operational data of projects into meaningful comprehensions in a specific context. Which is the action required according to Gandomi & Haider (2015) in order to make a decision-making process data-driven and enhance it with analytics. Also, this study validates that similar findings that are observed in data-driven decision-making for portfolio management (Costantino et al., 2015) apply in data-driven decision-making for agile project portfolio management as well.

Most importantly, the analysis performed in this study takes place with operational, real-world data that describe projects. This is a difference in relation to most of the currently available studies that use Machine Learning models and prediction techniques with data obtained through surveys. Studies that use operational data are mainly in process mining and lack context specificity in decision-making for agile portfolio management. On the other hand, studies that apply Machine Learning techniques on data from surveys are limited to datasets that are not large or diverse enough to sufficiently train a model. In addition, operational data are more representative to the real situation and they are not affected by the shortcomings of surveys, like respondents' bias. Therefore, the findings are based on recorded evidence that describe projects in their real-time situation.

Moreover, with the proposed methodologies and the tools applied in this study, the ability to systematically analyse projects in an agile and iterative manner is enhanced. This takes place through predicting if a project will achieve its pre-set targets and goals in an efficient way. In this study we are seeing how decision support systems have potential for more intelligent support to portfolio managers. Specifically, the applied models for analytics show the ways in which technological developments enable the viability of digitalising analysis and prediction during decision-making (Agrawal et al., 2018). Consequently, this shows the newly introduced ways in which computer-aided decision-making develops data-driven management and business models (Mayer-Schönberger & Ramge, 2018; McAfee, 2002). Using such prediction machines enables the analysis of large numbers of projects in an instant manner and application of better educated judgement.

7.2 Machine Learning support in decision-making of Agile Project Portfolio Management

This study shows how prediction can be used a single time during the initial phases of portfolio management, but the developed methodology can be used also for other phases in the iterative process of Agile Project Portfolio Management. According to industry reports and validated from our analysis of the available tools in chapter 4, (Agile) Project Portfolio Management is supported by tools that are mainly characterised by descriptive analytics. Therefore, decision-support systems for Agile Portfolio Management do not utilise the full potentials that lower cost predictions can provide with predictive and prescriptive analytics (Chen et al., 2012; Sivarajah et al., 2017). In order to achieve the required support, it is needed to have in place the needed analytical abilities to make sense of available data. Ultimately, the aim is to achieve prescriptive analytics that provide advice of solutions and recommendations to portfolio managers. Currently, decision-support systems in agile portfolio management are mainly descriptive and before being able to achieve prescriptive analytics and solid actionable recommendations, they first need to evolve in prediction.

Therefore, portfolio managers supported by decision-support systems can first obtain from systems solid overviews of the current and past state of their portfolio, then obtain predictions of potential future outcomes and then obtain recommendations of actions to take. For example, information that are descriptive can provide performance overviews on different projects and teams in order to be able and decide on strategy. In addition, seeing different milestones of decision-making in portfolio management is important in order to plan accordingly related activities for the future. Thus, they can be useful for both predictive and descriptive analytics. Another characteristic of descriptive analytics is to see which features and user stories are in high priority in order to align with strategy. Predictive analytics can help to influence prioritization of features in order to decrease risks during evaluation and selection.

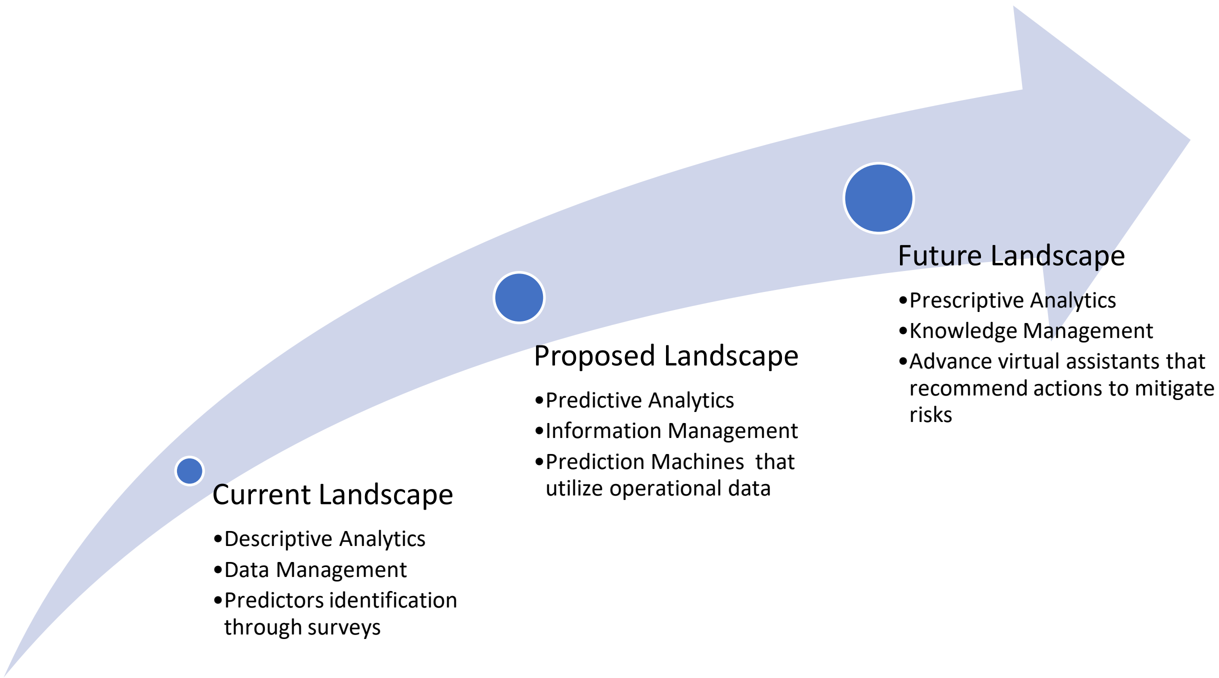


Figure 7.1: *Evolution of analytics for Agile Project Portfolio Management*

Decisions of project selection and project evaluation that are supported by Machine Learning are improved with relatively high confidence. The improvement of accuracy against the naïve approach of following the majority is not huge but big enough to achieve better results in the analysis of projects. Specifically, the best performing model on the first dataset achieves an improvement at a difference of 9% against the naïve approach and on the second dataset achieves a difference of 8% improvement in deciding whether a project will achieve its goal. This is a large number of projects in the long term and the confidence on these models is relatively high due to the low standard deviation. Specifically, on the first dataset the standard deviation of the models' performance ranges between 0,003 and 0,013 which is very low and thus, the confidence of systematically achieving that rate of right predictions is high. On the other hand, the standard deviation of the models' performance in the second dataset is higher and therefore, the confidence of systematically keeping their prediction rates is lower.

Consequently, we come to an answer for our research question of *“How can Machine Learning support decision-making processes in Agile Project Portfolio Management?”*. Specifically, our answer is that Machine Learning can be used to predict the outcome of projects and the predictions can be used to (at least) evaluate and select which projects to invest on. Evaluating the performance of agile projects helps the decision-making process on a management and operational level via the provision of a preliminary evaluation of a large number of projects, that otherwise it would be costly and time-consuming to analyse manually.

Data that simply describe projects can be successfully used to predict achievement of certain targets with confidence. The larger size of the dataset helps to predict with more confidence. Targeted values that describe behaviours, like user satisfaction, are more challenging to predict with descriptive data. Our models identify patterns about behaviour but are not able to explicitly explain the exact causal variables of behaviour. This has similarities with other prediction machines or automated processes that imitate human behaviour (Agrawal et al., 2018; McAfee & Brynjolfsson, 2017). Additionally, the applicability of Machine Learning in research enhances findings that show how Artificial Intelligence can represent alternatives in research to conventional analysis (de Oliveira et al., 2012). The reasons that the models in this study predict in such high levels of accuracy is the imitation of investing behaviour of funders / investors and of rating behaviour of end-users of mobile applications. There are patterns in the descriptive data that lead to predicting factors that describe success in projects. Finally, there is an uncertainty element which means that there are external factors that are not included in the used datasets, that have an impact on the outcome. A solution to this would be to include more detailed data of the projects that were not available during the time of the execution, as discussed further in section 7.5.

7.3 Accelerating parts of the decision-making process with prediction

Decision Support Systems with predictive analytics can be valuable for project portfolio management and this can be seen by answering the guiding question of *“How can parts of the decision-making process in Agile Portfolio Management be accelerated with prediction?”*. Our developed models showed improvements in the phase of project selection from the portfolio management process. Project selection takes place based on likelihood of achieving pre-set targets of initial budgeting/funding and of user satisfaction. The same analysis can be replicated in other targets that describe success factors and other phases as well, with similar improvements. Also, the approach of imitating the investment behaviour of people that fund or allocate budget to projects is an efficient way to develop models with advanced analytics. Many examples are available that follow a similar approach (McAfee & Brynjolfsson, 2017) and our study shows that a similar approach can be applied in Agile Portfolio Management.

Consequently, the followed approach in this study and the resulted models, can act as a framework for introducing predictive analytics in different phases of Agile Project Portfolio Management. This framework is tested in the phase of project evaluation and selection and it can be used to predict project success in other phases as well. The applicability of Machine Learning techniques during decision-making in Agile Portfolio Management can improve the efficiency of analysis in decision-making and the quality of decisions. As a result, in order to introduce predictive analytics in Portfolio Management with the intention to help decision making, we try to quantify data that describe projects and mobile applications. The decision-making process for Portfolio Management can be separated into the judgement element of managers and the prediction element of machines. The basic elements we identified can be replicated in different stages of Agile Portfolio Management in order to automate other decisions as well. These elements include the decomposition of decisions that take place in portfolio management, identifying which of them are better to be automated, understand the analytical capabilities of current Decision-Support tools and evolve them to the next level of analytics.

The main benefits of the developed Machine Learning models are their contribution to a better adoption of Agile principles during project portfolio management. Specifically, this can take place with the cost-effective preliminary analysis of large and diverse portfolios in a nearly instant manner and hence, enabling a more iterative process. Initial evaluation of projects at that scale is usually impossible to happen due to the high effort that requires, and the relatively low returns on that effort. Our developed models can change this, since they provide a way to quickly pre-analyse a large number of projects and evaluate them through predicting their outcome. This can afterwards help portfolio managers to filter projects and have a smaller set to judge upon for detailed project evaluation and selection. The analysis methodology of this study shows that projects that before were not possible to analyse during decision-making, due to the large amount of data, now can be considered and analysed. The predictive analytics we introduced seem to improve marginally the process and prescriptive analytics can be even more influential (Chen et al., 2012).

The way to accelerate decision-making in Agile Portfolio Management is by evolving the analytical capabilities of tools that support the entire process (Gandomi & Haider, 2015). First, factors that correspond to project success need to be identified (Serrador & Pinto, 2015; Stettina & Hörz, 2015). Then, it is required to identify predictors for each factor from the attributes available in the data (Costantino et al., 2015). Finally, Machine Learning models can be developed that evaluate effectively and efficiently projects through prediction and help in selecting those that are more likely to succeed (Elgendy & Elragal, 2016). In the cases analysed in this study, the attributes available in the data are mainly descriptive and not particularly dense with many dimensions, but they still achieve relatively high prediction ratings. Some of the characteristics that influence the analysis are size and type of the data, what they describe and how frequently they are updated with the changes that take place in business. Furthermore, we identify that for small datasets the Gradient Boosting Classifier performs better than the rest of the tested algorithms and in larger datasets LGBM performs even better and also more efficient.

7.4 How do decision support systems assist Agile Portfolio Management of software-intensive projects?

According to literature, software-intensive projects shape entire organisations and industries. Specifically, software is evolving into an integral part of our lives and businesses with the evolution of technologies like the Internet of Things (Atzori, Iera, & Morabito, 2010b; Gubbi et al., 2013). Moore's law results to increased business connectivity at lower costs and this enables the development of Ecosystems and Digital Marketplaces around physical projects within organisations, that contain information goods and digitalised offerings (McAfee & Brynjolfsson, 2017). This shared, digital market is usually facilitated in common technological platforms and common industrial cyber-physical spaces that come to life with the exchange of information and other resources between actors (Colombo, Karnouskos, Kaynak, Shi, & Yin, 2017). Software intensive initiatives that participate in such marketplaces are the 7,200 mobile applications analysed in this study. And the machine learning models developed were able to predict based on descriptive data if an application is likely to achieve high user ratings or not.

Moreover, the project landscapes in organisations assemble internal innovation hubs that are close to networked structures like those of digital marketplaces that appear in software ecosystems (Schroeder, 2016) or innovation hubs like Kickstarter. Hence, they are complex systems in which actors interact with each other or interact as units in a shared market for software and services (Mayer-Schönberger & Ramge, 2018). Therefore, organisational structures that group people in functional teams based on technical skills give their place to networks of autonomous teams or self-managed work groups that take development responsibility of software products or services (Moe et al., 2010). Specifically, the simultaneous development on a variety of software systems might predispose change in operational models towards organisations that are project-oriented (Gemünden, Lehner, & Kock, 2018).

According to literature, this has several challenges that can be addressed with taking from managers some analytical tasks and allowing them to focus in applying their judgement. For example, strategy is developed from organisations and it usually acts as a reference point of the organisation's potential and future aspirations (Romano, 2014). Also, there is a tendency of managers to plan more than execute, passing crucial responsibilities of implementation to lower management levels and not-sufficiently coping with the complexity of implementation, due to longer time and more people involved (Hrebiniak, 2006). Through that, it is becoming increasingly challenging to accommodate end-to-end processes of customers, many times across their respective industries (Mayer-Schönberger & Ramge, 2018).

When managing sets of projects, complexity increases significantly, and the right methodologies are needed to support this task. Such methodologies need to support decision-making with data-driven capabilities and provide insightful predictions and recommendations on coordinating independent and autonomous teams of projects. Having this in mind, it is necessary to acknowledge both the technical challenges as well as behaviours in developing products or services. The systems that support the development of business value, besides the often-complicated technology, are comprising of social contexts, environments, people that are an integral part of these environments and finally organizational structures and business processes (Baxter & Sommerville, 2011).

7.5 Limitations

This study illustrates how the decision-making process in agile project portfolio management can become more data-driven, through predicting pre-set targets and goals in two case studies. The first case study is described through a dataset of Kickstarter projects and the second case study through a dataset of mobile applications from iOS. To achieve this data-driven decision-making process, it is needed to define targets from the used data. In the cases analysed in this study, the data are descriptive of projects and mobile applications. On one hand, the predicted targets in the first case are about achieving budgeting or funding targets. On the other hand, the predicted targets of the second case are about achieving user satisfaction ratings. An interesting finding is that even though the two datasets are different in volume and descriptive characteristics, the resulted prediction accuracy is on a similar range.

Even though the followed methodology achieves the prediction of projects' success factors, there is room for further improvement. Specifically, better predictions are possible with the analysis of more comprehensive datasets. For example, an improvement would be if the data of the mobile applications also include attributes that describe the interactions of users with the software applications. Such data could describe the number of different screens that the users have access to and the navigation options of the applications. More importantly, better explanation of the predictions is possible to be achieved with data that have a better coverage on the predictors identified in literature. The datasets used in this study have a partial coverage and thus, it is expected to achieve better performance in predictions with a more complete coverage. Specifically, the analysis of operational data that come from project portfolio management tools like Target Process have bigger potential for better prediction accuracy and more comprehensive explanations on predictions. Hence, we are suggesting for the future to replicate this analysis in data that are more detailed and describe operations in more detail. An example of that would be derived from historical data of different epics to predict if an epic or project has the state that it was planned to have. Then, it is possible to predict which projects will end-up in unplanned states with Machine Learning and using attributes like effort planned, type, job size, related strategic objective, progress, related project, release, start date, related workflow, related role, related process and so on. Finally, at a later stage the predictions can be complemented with recommendations on potential predicted scenarios.

8 Conclusion

This study provides a set of methodologies and tools that introduce prediction in decision-making for Agile Portfolio Management. This can enable the achievement of the full potential that computer-aided decision-making with prediction machines can provide. Data that describe initiatives can be valuable if analysed and can provide insights about projects (Michael Bloch, Blumberg, & Laartz, 2013). In addition, the data that are generated from agile teams can come in big volumes and many times they are not used sufficiently to support decision-making. Most importantly, the tools available mostly support descriptive analytics and do not utilise the full potentials that lower cost predictions provide with predictive and prescriptive analytics (Chen et al., 2012; Sivarajah et al., 2017). As a result, we identify and address the gap of introducing analytical capabilities in Project Portfolio Management that go beyond descriptive analytics and with the intention to help decision making. The methodology proposed can lead to the development of a system that accelerates the evaluation of projects and enables a data-driven project selection.

Moreover, in this study we analysed a set of Decision-Support Systems in Agile Project Portfolio Management and we attempted to introduce predictive analytics in the project evaluation and selection. In the different stages of Portfolio Management there is an elaborate decision-making process that we analysed. Then, we identified which parts of it can be better automated in order to optimize decision-making. In addition, we analysed a set of Decision-Support systems and identified that they are currently suitable for Descriptive Analytics but not Predictive or Prescriptive Analytics. Therefore, we contribute in filling this gap with the implementation of a set of Machine Learning models that can predict the outcome of projects with accuracy in the ranks of 60 to 70 percent. This was done with two case studies in which we analyse a dataset of projects that represent independent initiatives. The datasets describe 378,661 projects that are in their initial phase of funding request and 7,200 mobile applications. Therefore, our models basically imitate the decisions that humans make to provide or not provide the requested budget, based on information that are available only on the initial phase of projects. This enables to take place a preliminary performance evaluation of a large number of projects. Last but not least, the followed approach and developed models can be used as a framework and be replicated in other stages of Agile Project Portfolio Management, beyond project evaluation and project selection.

8.1 Practical Recommendations

Our current analysis takes place with data of projects from KickStarter and mobile applications from iOS and provides an approach for doing a similar analysis with more detailed data that are also more specific to software-intensive projects. Specifically, our approach defines success factors from data, that describe targets that projects have and then identify attributes from the data that are predictors to the targets. For example, in the first case study we identify the target of achieving the requested funding/budget and we predict if a project is likely to secure the requested amount. Therefore, our models try to imitate the decisions that investors of projects made, based on the available information. Decision-making in managing portfolios of autonomous teams can be improved when there is a large number of projects to be evaluated and filtered. The form of existing project evaluation in large scale agile is mostly in the form of principles and self-assessment checklists which rely on individuals to consider several factors and teams to assess their situation and adopt specific practices (Baxter & Sommerville, 2011). Such methodologies possibly work in many contexts but in a dynamic, highly paced and extremely practical environment, they have shortcomings that a system that automates the approach can overcome.

8.2 Future Directions

Once the prediction element is addressed and covered, we identify a future gap which is the coverage of the automatic prescription of solutions that can be based on predictions. Specifically, models can be formulated that enhance these decision-support tools with recommendations for large scale, organisational agility. In addition, the underlying reasons of large-scale agile maturity in the organisation based on certain challenges needs further investigation in order to successfully estimate the risk that these challenges can create on success criteria for organisational agility. To do so, it is possible to advance the mapping of patterns of challenges like multi-team alignment, commitment, resource allocation with project success. In addition, the imposed risk from these challenges can be calculated through success criteria like involvement and engagement of stakeholders, multi-team coordination and management of dependencies. Furthermore, with the optimisation of prediction, recommendations might be possible to be given for more specific actions regarding ways to mitigating risks.

Finally, this study does not cover the operational performance of agile portfolio management at its full, but only covers the stage of project evaluation and project selection. This means that it is possible to further define specific changing circumstances that organisations typically come across in other phases of Agile Project Portfolio Management. The definition of these circumstances of change can be the result of further analysing more types of operational data that can be extracted from collaboration and project management tools. There is an increasing amount of collaboration and portfolio configuration tools that act as a starting point to deal with the portfolio management challenges. These tools are the place that stakeholders go in order to manage their work and configure portfolios and as a result, they generate data about development processes. Such data can be the input of a similar system that uses the same approach as the one developed in this study, that gives recommendations to the people responsible for coordinating multiple projects. Other data of the digital footprint that teams leave in decision-support tools for portfolio management can be analysed and lead to more ways of understanding and optimising the decision-making process.

9 References

Agrawal, A., Joshua, G., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Agyapong, L., Guitton, P. C., Fairdoon, L., & Lasar, I. (2016). Portfolio strategy panel: successful strategy execution through project portfolio management. In *Paper presented at PMI® Global Congress 2016—EMEA, Barcelona, Spain. Newtown Square, PA: Project Management Institute.*

Anttiroiko, A. V., Valkama, P., & Bailey, S. J. (2014). Smart cities in the new service economy: Building platforms for smart services. *AI and Society*, *29*(3), 323–334. https://doi.org/10.1007/s00146-013-0464-0

Anyosa Soca, V. (2009). Linking portfolio, program, and projects to business strategy: a way to gain competitive advantage in this turbulent time. In *Paper presented at PMI® Global Congress 2009—North America, Orlando, FL. Newtown Square, PA: Project Management Institute.*

Archer, N. P., & Ghasemzadeh, F. (1999). An integrated framework for project portfolio selection. *International Journal of Project Management*, *17*(4), 207–216. https://doi.org/10.1016/S0263-7863(98)00032-5

Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79–80*, 3–15. https://doi.org/10.1016/j.jpdc.2014.08.003

Atzori, L., Iera, A., & Morabito, G. (2010a). The Internet of Things: A survey. https://doi.org/10.1016/j.comnet.2010.05.010

Atzori, L., Iera, A., & Morabito, G. (2010b). The Internet of Things: A survey. *Computer Networks*, *54*, 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data Analytics: Hyped Up Aspirations or True Potential? *Vikalpa*, *38*(4), 1–12. https://doi.org/10.1177/0256090920130401

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Bloch, M., Blumberg, S., & Laartz, J. (2012). Delivering large-scale IT projects ontime, on budget, and on valuE. *WWW Document, McKinsey Q*.

Bloch, Michael, Blumberg, S., & Laartz, J. (2013). Big data: new tools for mitigating project complexity. In Newtown Square (Ed.), *Paper presented at PMI® Global Congress 2013*. EMEA, Istanbul, Turkey: PA: Project Management Institute.

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Brown, T. (2008). Design Thinking. *Harvard Business Review*, *86*(6), 84.

Brown, T., & Wyatt, J. (2001). Design Thinking for Social Innovation IDEO, 29–32.

Caglio, A. (2003). *Enterprise Resource Planning systems and accountants : towards hybridization?* (Vol. 8180). European Accounting Review, 12:1, 123-153,. https://doi.org/10.1080/0963818031000087853

Chaphalkar, N. B., Iyer, K. C., & Patil, S. K. (2015). Prediction of outcome of construction dispute claims using multilayer perceptron neural network model. *International Journal of Project Management*. https://doi.org/10.1016/j.ijproman.2015.09.002

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Imppact. *MIS Quarterly, SPECIAL ISSUE: BUSINESS INTELLIGENCE RESEARCH*, *36*(4), 1165–1188. Retrieved from https://pdfs.semanticscholar.org/f5fe/b79e04b2e7b61d17a6df79a44faf358e60cd.pdf%3E.

Christensen, C. M., & Overdorf, M. (2000). Meeting the challenge of disruptive change. *Harvard Business Review, 78(2)*, 66–77.

Colombo, A. W., Karnouskos, S., Kaynak, O., Shi, Y., & Yin, S. (2017). Industrial Cyberphysical Systems : A Backbone of the Fourth Industrial Revolution, (March). https://doi.org/10.1109/MIE.2017.2648857

Conforto, E. C., Amaral, D. C., da Silva, S. L., Di Felippo, A., & Kamikawachi, D. S. L. (2016). The agility construct on project management theory. *JPMA*, *34*(4), 660–674. https://doi.org/10.1016/j.ijproman.2016.01.007

Cooper, R. G. (2008). Perspective: The Stage-Gate Idea-to-Launch Process—Update, What's New, and NexGen Systems. *The Journal of Product Innovation Management*, *25*, 213–232.

Costantino, F., Di Gravio, G., & Nonino, F. (2015). Project selection in project portfolio management: An artificial neural network model based on critical success factors. *International Journal of Project Management*, *33*(8), 1744–1754. https://doi.org/10.1016/j.ijproman.2015.07.003

Creative Commons. (n.d.). Attribution-NonCommercial-ShareAlike 4.0 International. Retrieved from https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode

de Bruijn, H., & ten Heuvelhof, E. (2002). Policy analysis and decision making in a network: How to improve the quality of analysis and the impact on decision making. *Impact Assessment and Project Appraisal*, *20*(4), 232–242. https://doi.org/10.3152/147154602781766627

de Oliveira, M. A., Valentina, L. V. O. D., & Possamai, O. (2012). Forecasting project performance considering the influence of leadership style on organizational agility. *International Journal of Productivity and Performance Management*, *61*(6), 653–671. https://doi.org/10.1108/17410401211249201

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems*, *55*(1), 359–363. https://doi.org/10.1016/j.dss.2012.05.044

Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and success factors for large-scale agile transformations : A systematic literature review. *The Journal* *of Systems & Software*, *119*, 87–108. https://doi.org/10.1016/j.jss.2016.06.013

Dingsøyr, T., Nerur, S., Balijepally, V., & Moe, N. B. (2012). A decade of agile methodologies : Towards explaining agile software development, *85*, 1213–1221. https://doi.org/10.1016/j.jss.2012.02.033

Elgendy, N., & Elragal, A. (2016). Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science*, *100*, 1071–1084. https://doi.org/10.1016/j.procs.2016.09.251

Esteves, J. (2001). Enterprise Resource Planning Systems Research : An Annotated Bibliography, *7*. https://doi.org/10.17705/1CAIS.00708

Free Software Foundation, I. (1991). GNU General Public License, version 2. Retrieved from http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gemünden, H. G., Lehner, P., & Kock, A. (2018). The project-oriented organization and its contribution to innovation. *International Journal of Project Management*, *36*(1), 147–160. https://doi.org/10.1016/j.ijproman.2017.07.009

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2013.01.010

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*. Harper Business, 1993.

Hrebiniak, L. G. (2006). Obstacles to Effective Strategy Implementation. *Organizational Dynamics*, *35*(1), 12–31. https://doi.org/10.1016/j.orgdyn.2005.12.001

Hunton, J. E., Lippincott, B., & Reck, J. L. (2003). Enterprise resource planning systems : comparing firm performance of adopters and nonadopters, *4*, 165–184. https://doi.org/10.1016/S1467-0895(03)00008-3

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kuusinen, K., Mikkonen, T., Pakarinen, S., Kuusinen, K., Mikkonen, T., Pakarinen, S., … Pakarinen, S. (2017). Agile User Experience Development in a Large Software Organization : Good Expertise but Limited Impact.

Lindsjørn, Y., Sjøberg, D. I. K., Dingsøyr, T., Bergersen, G. R., & Dybå, T. (2016). Teamwork quality and project success in software development : A survey of agile development teams, *122*, 274–286. https://doi.org/10.1016/j.jss.2016.09.028

Martinsuo, M., Korhonen, T., & Laine, T. (2014). Identifying, framing and managing uncertainties in project portfolios. *International Journal of Project Management*, *32*(5), 732–746. https://doi.org/10.1016/j.ijproman.2014.01.014

Martinsuo, M., & Lehtonen, P. (2007). Role of single-project management in achieving portfolio management efficiency. *International Journal of Project Management*, *25*(1), 56–65. https://doi.org/10.1016/j.ijproman.2006.04.002

Mayer-Schönberger, V., & Ramge, T. (2018). *Reinventing Capitalism in the Age of Big Data*. Basic Books.

McAfee, A. (2002). The impact of enterprise information technology adoption on operational performance: An empirical investigation. *Production and Operations Management Society*, *11*(1), 33–53.

McAfee, A., & Brynjolfsson, E. (2012). Big Data : The Management Review. *Harvard Business Review*, (October), 1–12. https://doi.org/10.1007/978-3-319-05029-4

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future.* WW Norton & Company.

McGinnis, J. O., & Pearce, R. G. (2014). The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services. *Fordham L. Rev.*, *82*, 3041–3066.

McLeod, L., & MacDonell, S. G. (2011). Factors that affect software systems development project outcomes. *ACM Computing Surveys*, *43*(4), 1–56. https://doi.org/10.1145/1978802.1978803

Meskendahl, S. (2010). The influence of business strategy on project portfolio management and its success - A conceptual framework. *International Journal of Project Management*, *28*(8), 807–817. https://doi.org/10.1016/j.ijproman.2010.06.007

Meyer, C., & Schwager, A. A. (2007). Understanding customer experience. *Harvard Business Review*. https://doi.org/10.1108/00242539410067746

Mintzberg, H. (1997). The Manager's Job: Folklore and Fact. In *Leadership: Understanding the dynamics of power and influence in organizations* (pp. 35–53).

Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team : A case study of a Scrum project. *Information and Software Technology*, *52*(5), 480–491. https://doi.org/10.1016/j.infsof.2009.11.004

Moe, N. B., Dingsøyr, T., & Moe, N. B. (2013). Research Challenges in Large-Scale Agile Software Development. https://doi.org/10.1145/2507288.2507322

Mouillé, M. (2017). Kickstarter Projects | Kaggle. Retrieved May 1, 2019, from https://www.kaggle.com/kemical/kickstarter-projects

Neill, P. O., & Sohal, A. S. (1999). Business Process Reengineering A review of recent literature, *19*, 571–581.

Patanakul, P., Chen, J., & Lynn, G. S. (2012). Autonomous Teams and New Product Development *, *29*(5), 734–750. https://doi.org/10.1111/j.1540-

5885.2012.00934.x

Patnayakuni, R., & Ruppel, C. P. (2010). A socio-technical approach to improving the systems development process, 219–234. https://doi.org/10.1007/s10796-008-9093-4

Pedregosa et al. (2011). Scikit-learn: Machine Learning in Python. *JMLR*, *12*, 2825–2830. Retrieved from https://scikit-learn.org/stable/index.html

Perumal, R. (2017). Mobile App Store (7200 apps) | Kaggle. Retrieved July 15, 2019, from https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps

PMI. (2013). *A Guide to the Project Management Body of Knowledge*. *Project Management Institute* (5[a], Vol. 5). Pennsylvania: ©2013 Project Management Institute, Inc. https://doi.org/10.1002/pmj.20125

Porter, M. E. (2008). *Competitive advantage: Creating and sustaining superior performance.* Simon and Schuster.

Porter, Michael E, & Heppelmann, J. E. (2015). How Smart, Connected Products Are Transforming Companies. *Harvard Business Review*.

Rasnacis, A., & Berzisa, S. (2016). Method for Adaptation and Implementation of Agile Project Management Methodology. *Procedia Computer Science*, *104*(December 2016), 43–50. https://doi.org/10.1016/j.procs.2017.01.055

Razavian, M., Paech, B., & Tang, A. (2019). The Journal of Systems and Software Empirical research for software architecture decision making : An analysis. *The Journal of Systems & Software*, *149*, 360–381. https://doi.org/10.1016/j.jss.2018.12.003

Rehman, M. H. U., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*(6), 917–928. https://doi.org/10.1016/j.ijinfomgt.2016.05.013

Romano, L. (2014). Corporate strategy for project managers: why strategic alignment and awareness is so important. In *Paper presented at PMI® Global Congress 2014—EMEA, Dubai, United Arab Emirates. Newtown Square, PA: Project Management Institute.*

Schroeder, R. (2016). Big data business models: Challenges and opportunities. *Cogent Social Sciences*, *2*(1). https://doi.org/10.1080/23311886.2016.1166924

Serrador, P., & Pinto, J. K. (2015). Does Agile work? - A quantitative analysis of agile project success. *International Journal of Project Management*, *33*(5), 1040–1051. https://doi.org/10.1016/j.ijproman.2015.01.006

Simon, H. A. (2000). Bounded Rationality in Social Science: Today and Tomorrow. *Mind & Society*, *1*(1), 25–39.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Souza, E., Moreira, A., Araújo, J., Abrahão, S., Insfran, E., & Silveira, D. S. da. (2018).

Comparing business value modeling methods: A family of experiments. *Information and Software Technology*, *104*(June), 179–193. https://doi.org/10.1016/j.infsof.2018.08.001

Stettina, C. J., & Hörz, J. (2015). Agile portfolio management : An empirical perspective on the practice in use. *JPMA*, *33*(1), 140–152. https://doi.org/10.1016/j.ijproman.2014.03.008

Strode, D. E., Huff, S. L., Hope, B., & Link, S. (2012). Coordination in co-located agile software development projects. *The Journal of Systems & Software*, *85*(6), 1222–1238. https://doi.org/10.1016/j.jss.2012.02.017

Svejvig, P., & Andersen, P. (2015). Rethinking project management: A structured literature review with a critical look at the brave new world. *International Journal of Project Management*, *33*(2), 278–290. https://doi.org/10.1016/j.ijproman.2014.06.004

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, *18:7*(March), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO 2-Z

Turban, E., Sharda, R., & Delen, D. (2014). *Business intelligence and analytics: systems for decision support*. Pearson Higher Ed.

Vlaanderen, K., Jansen, S., Brinkkemper, S., & Jaspers, E. (2011). The agile requirements refinery : Applying SCRUM principles to software product management. *Information and Software Technology*, *53*(1), 58–70. https://doi.org/10.1016/j.infsof.2010.08.004

Vliet, H. Van, & Tang, A. (2016). The Journal of Systems and Software Decision making in software architecture, *117*, 638–644. https://doi.org/10.1016/j.jss.2016.01.017

Vroom, V. H., & Jaago, A. G. (2007). The role of the situation in leadership. *American Psychologist*, *62*(1), 17–24. https://doi.org/10.1037/0003-066X.62.1.17

Walker, D., & Lloyd-Walker, B. (2016). Rethinking project management: Its influence on papers published in the international journal of managing projects in business. *International Journal of Managing Projects in Business*, *9*(4), 716–743. https://doi.org/10.1108/IJMPB-08-2013-0040

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*(1), 1247–1268. https://doi.org/10.17705/1CAIS.03465

West, D. M. (2018). *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press, 2018.

Yeow, A., Soh, C., & Hansen, R. (2018). Aligning with new digital strategy : A dynamic capabilities approach. *Journal of Strategic Information Systems*, *27*(1), 43–58. https://doi.org/10.1016/j.jsis.2017.09.001

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., Faraj, S., & Griffith, T. L. (2007). Information Technology and the Changing Fabric of Organization. *Organization Science*, *18*(5), 749–762. https://doi.org/10.1287/orsc.1070.0307

Appendices

Appendix A: Detailed data form from Kickstarter

Delivered as a separate deliverable

Appendix B: Code in Python for the Data Analysis

Delivered as a separate deliverable