Department of Information Science and Technology

A data-driven approach to predict the success of IT banking projects

Rui Alexandrino Calção Mendes

A Dissertation presented in partial fulfilment of the requirements for the Degree of Master in Computer Science and Business Management

Supervisors:

Sérgio Moro, PhD, Assistant Professor,
Department of Information Science and Technology, ISCTE-IUL

Joaquim Esmerado, PhD, Assistant Professor,
Department of Information Science and Technology, ISCTE-IUL

September, 2018

[This page was intentionally left blank]

Acknowledgements

I owe an enormous debt of gratitude to all the people that supported me in many ways in the last two years.

To my supervisors Sérgio Moro and Joaquim Esmerado that had the patience and responses to all my needs, and for all attention and full support in this intense path.

To my wife Patricia that provide all the familiar support, patience and understanding. Without you it wouldn't be possible.

To my friends, colleagues and managers that always believed in me.

I would like to express my gratitude for the company that provide all the data that made this work possible, especially Project Officers for their swiftly responses to my questions and demands.

For my parents and grandparents, a very special word. Wherever you are, I hope that you are proud for what I become and what I have achieved in the last 20 years.

Most of all, I dedicate my dissertation to my son Viriato, that with only five years old understood that my thesis was very important to me. Hope that one day you can be proud of your father's hard path.

[This page was intentionally left blank]

Abstract

Financial institutions have recurrent Information Technology needs. These needs respond, among others, to innovation features, maintenance, customer needs, cost reductions, process simplifications, and operational risk minimizations. To do such implementations, financial institution X, as many others, has project management processes implemented to respond efficiently to internal departments' requirements and needs, following standard project management practices.

Some of these projects are concluded on time and on budget while others, for several reasons, have project changes in their lifecycle. They are not concluded within scope, time and cost that were indicated after the planning phase, at contract acceptance time.

Although these reasons are known at execution time, the purpose of this study is to verify a relation between initial project variables (data features known at planning phase) and project variations that occur during execution phase. To understand this relation, a data analysis based research was conducted using data delivered by PMO's Institution. Then, the best classifiers were scrutinized and the results were discussed in detail. Specifically, the most relevant input variables for project success were further analysed. The findings unveiled that it is possible by initial variables to predict project changes. It was also concluded that, for this specific organization, variables that do not depend from project management, such as the type of the project, the department that requested the project or the project's objective, have a strong weight in the influence of project success and project changes.

Keywords: Project Management; Information Technology; Information Systems; Financial Information Systems; Data Mining

[This page was intentionally left blank]

Resumo

As instituições financeiras têm necessidades tecnológicas recorrentes. Estas necessidades, entre outras, surgem como resposta a necessidades de inovação, manutenções, necessidades de clientes internos e externos, redução de custos, simplificação de processos e mitigação de risco operacional. Para avançar com estas implementações, a Instituição financeira X, tem definido processos de gestão de projetos de forma a responder de forma eficiente às necessidades das diversas direções e departamentos, adotando ao longo dos anos as melhores práticas nesta área.

Alguns destes projetos são concluídos dentro do tempo e custo estimados, assim como e o âmbito que fora definido, implementado na sua plenitude. Outros, por diversas razões, não são concluídos no âmbito, tempo e custo indicados em altura de aceitação de contrato, após a fase de planeamento.

Embora as razões para existirem alterações sejam conhecidas na altura de execução, é objetivo deste estudo validar a existência de relação direta entre variáveis iniciais dos projetos (que sejam conhecidas na fase de planeamento) e a indicação de alterações que possam vir a ocorrer. Foi conduzida uma pesquisa baseada em análise de dados provenientes do PMO da instituição. De seguida, foram discutidos os resultados, detalhando-se quais as variáveis com maior impacto no sucesso dos projetos, tendo sido concluído que é possível, através dos dados iniciais dos projetos, prever alterações. Também se conclui que, para esta organização, as variáveis que não dependem da gestão do projeto, como o tipo de projetos, o departamento que pediu o projeto, ou o objetivo, têm um grande peso na existência de alterações.

Palavras-Chave: Gestão de Projetos; Tecnologias de Informação; Sistemas de Informação; Sistemas de informação financeiros; Mineração de dados.

[This page was intentionally left blank]

List of Contents

[This page was intentionally left blank]

List of Tables

[This page was intentionally left blank]

List of Figures

Abbreviations

ALIFT	Area of the LIFT Cumulative Curve
APM	Association for Project Management
AUC	Area under the ROC curve
ATM	Automated Teller Machines
BI	Business Intelligence
CCB	Change Control Board
CSF	Critical Success Factors
CRISP-DM	CRoss Industry Standard Process for Data Mining
DM	Data Mining
DSA	Databased Sensitivity Analysis
DT	Decision Tree
ERP	Enterprise Resource Planning
ICB	IPMA Competence Baseline
IPMA	International Project Management Association
IS	Information Systems
IT	Information Technology
LR	Logistic Regression
MLP	Multi Layer Perceptron
MLPE	Multi Layer Perceptron Ensemble
NN	Neural Network
PM	Project Manager
PMBOK®	Project Management Body of Knowledge
PMI®	Project Management Institute®
PMO	Project Management Office
Prince2	Projects in Controlled Environment, version 2
RF	Random Forests
RFOREST	Random Forest
ROC	Receiver Operating Characteristic
SA	Sensitivity Analysis
SVM	Support Vector Machine
VEC	Vector Effect Characteristics

[This page was intentionally left blank]

1 Introduction

1.1. Scope

Financial institutions have permanent Informational Technology's needs. Among other reasons, these needs emerge from new products implementation, innovation features, maintenance, cost reduction, processes simplifications, operational risk minimization or mandatory features implementation (Shu *et al.* 2005; Moro *et al.* 2015).

As in other industries, to support such implementations, financial institutions implement project management processes and methodologies. To respond efficiently to internal departments and maximize project success, IT departments follow known project management practices, gaining maturity on the process with experience, since project efficiency and probability of success are directly related with project methodologies and process implementations (Joslin and Müller 2015).

Yet, some of these projects for several reasons are not concluded within the scope, time and cost that were indicated in project baseline at contract acceptance time (after the project's planning phase).

Although the reasons are known and identified at project execution and conclusion time, is proposed in this study to identify if there is a direct relation between initial project variables and time delays, cost increases or scope variations.

The main purpose of this study is to unveil the intrinsic interactions between the several variables involved in each project and how these can influence the overall project performance. Thus, this research attempts to provide insights that can ultimately lead to better management practices of future projects by identifying those variables that influence more project deviation from the expected deadline, cost, and scope. To achieve such goal, a data mining project based on real data will model if the project at its conclusion has deviated from the initial contracted parameters.

This case addresses a European financial institution. This company can be classified as a medium sized company in retail banking and have implemented a Project Management methodology for IT development implementations and application maintenance, following PMI®'s best practices. The company has a Project Management Office (PMO) and project management tools that helps projects to be managed by

project managers, controls organization project portfolio and supplies report information to middle and top management. One of the tools used is a software tool where project managers can register all project data since the beginning of the project until its closure. Dates, costs, scope changes, budget control and risk management information are registered in this tool. This tool is also used for resource time entry and to manage the entire project workflow defined in the project lifecycle.

The project data used in this case was extracted from this tool by IT's PMO department where is possible to verify that some projects had changes in time, cost and/or scope across their lifecycles, ending in a different date, different cost and/or with different business requirements that where defined at project contract acceptance time (after project plan phase and before execution phase).

1.2. Motivation

This research aims to answer the following question: Are there initial project variables and characteristics that can be related with scope, time and cost changes? Can these changes be predicted and anticipated at project initiation and planning phase? If there are relations between project variables and project changes, their prediction can be useful for improve scope definitions for some kind of projects, accurate estimations and to complement project risk management, creating project reserves at planning phase as a response. Furthermore, the conclusions drawn can be used to change department behaviours, aligning and implementing methods and processes that can reduce project changes.

This research can be useful for both scholars and practitioners, and results can be compared with other financial institutions project data in order to understand change reasons across organizations, adding other variables for future studies as organization size, type, and culture or maturity in project management and processes.

1.3.Contributions and objectives

This research has its main purpose in the identification and the prediction of projects' changes that may occur in project's execution and that can be identified with the project information that is available until the end of the planning phase. If a relation is

identified, Company's project managers can predict them and do better planning, and make some risk reserves based on these variables. On the other hand, the institution can try to identify the origin of these changes, modifying (if possible) behaviours and processes in some departments in order to mitigate project changes at project time. Thus, the developed model can be a useful tool for the PMO in order to better understand project slippage reasons and act preventively.

To the best of our knowledge, this approach has no precedents in current literature. Specifically, empirical project management research within organizational context is traditionally conducted through article analysis questionnaires to the main stakeholders (e.g., Cooke-Davies 2002).

1.4. Methodological approach

The data used for this research was obtained from the Financial Institution's Project Management Office. The PMO delivered a raw dataset with 1361 records that correspond to projects deployed into production since May 2012 until October 2017. From this dataset it was possible to obtain 36 relevant variables (also known as features). Some of these variables were treated or cleaned according to PMO´s guidance. From these variables only those known at project planning phase were chosen. Data analysis and data mining techniques were used to extract from the dataset information about variables importance and to identify if the models could predict project changes. For the best model, results were discussed in light of current state-of-the-art literature on project management, and conclusions were drawn.

1.5.Document structure

This dissertation is divided into five chapters. In the introduction chapter, the dissertation's scope, motivations and contributes are explained. On the second chapter the literature is reviewed. On the third chapter the methodology adopted is exposed, detailing knowledge extraction techniques by data modelling and data analysis. Results and data analysis are detailed and explained on fourth chapter (results and discussion chapter). Finally, on the fifth chapter, the main conclusions are presented.

[This page was intentionally left blank]

2 Literature Review

2.1. Project Management

2.1.1 Project Definition

There are several project definitions, but all of them indicate that a project is limited in time and scope and has a specific purpose for creating a unique product or service or to achieve a specific objective. In the words of Turner (1999) cited by Prabhakar (2008) , "*a project is an endeavour in which human, financial and material resources are organized in a novel way to undertake a unique scope of work, of given specification, within constraints of cost and time, so as to achieve beneficial change defined by quantitative and qualitative objectives*" (Prabhakar 2008, p.3). For the Project Management Institute® (PMI®), a "*project is a temporary endeavour undertaken to create a unique product or result*". It is limited in time, it has a beginning and an end and its purpose is to make something unique, meeting project requirements (PMI® 2017, p. 43). For Kezner (2003), a project must have specific objectives to be achieved under certain specifications, to have delimited start and end dates, and to have funding limits when applicable. Projects consume resources and are multifunctional (projects cross multiple functional lines).

Project requirements have its origin in someone's needs that have intention to see them implemented. These persons are influent in project and they are called Customers or Sponsors (Lock 2007; PMI® 2017). The most important attribute of a project is its importance to senior management and to the organization. Only this importance can justify the cost and investment of work in a project (Meredith and Mantel 2009). People, groups and organizations that impact or are impacted by a project result or execution are called project stakeholders (PMI® 2017).

2.1.2 Project Management

Projects are managed since the beginning of humankind (Rosen 2004; Lock 2007). Its importance has been growing in the last years, specifically in the light of project management organizations that advocate best practices to achieve a better result (Lock 2007). Project management is a specialized branch of management that evolve in order to fulfil complex activities and requirements on today's industry. Today's projects have

to respond to the ever-changing dynamics of situations requiring it to become much more adaptive than ever (Lock 2007). Kerzner (2003) defines project management as the planning, organizing, directing, and controlling of company resources for a relatively short term objective that has been established to complete specific goals and objectives. It is of great importance to project management to meet requirements and objectives, satisfy stakeholder's expectations, predict delays, respond to risks, do better planning and optimize resources (PMI® 2017). Lack or poor project management may result in implementation delays, cost overruns, unsatisfied stakeholders, requirement implementation failure, time-to-market failure, and also in a lower project return on investment (Meredith and Mantel 2009).

2.1.3 **Project Management Associations**

The interest and relevance of project management and its inherent complexity have raised the need to identify common guidelines that could help in projects' success. As a result, frameworks have been promoted by associations specifically devoted to project management.

IPMA is a federation with Member Associations that develop project management competences. APM (Association for Project Management) is the United Kingdom member association, including a Journal publication encompassing also a Body of Knowledge, which is the IPMA Competence Baseline (ICB) that compiles the project management best practices. APM has a certification for project managers (Lock 2007).

Prince2 is a methodology that is used by the UK Government, but practised also by private sector (especially in the UK and Australia). It is not an association, but it also promotes standards, methodologies and a certification widely used and recognized (Rehacek 2017).

PMI® is the world's leading not-for-profit organization for individuals who work or are interested in project management (Lock 2007; Rehacek 2017). PMI® develops recognized standards and publishes them in project management body of knowledge guide (PMBOK®). PMI® publications include a professional magazine and a project journal. PMI® is also dedicated to maintain a rigorous professional certification program to advance the project management profession and to recognize the individuals that own

this certification. The most relevant is Project Management Professional® (PMP®) Certification (Lock 2007).

In spite of ICB and PMBOK® having similar competences, ICB is the base for APM's professional certification, while PMBOK® guide's main purpose is to be a guideline for managing projects. PMI® gives more depth and detail about project management. Because of its flexibility in some matters, IPMA's ICB need to get more information elsewhere, routing knowledge to other literature. For instance, ICB do not provide project tools and techniques (Eberle *et al.* 2011; Rehacek 2017). For those reasons, and since PMI® is used worldwide (including in the studied institution) covering project management in great detail, PMBOK® was chosen as a reference for explaining concepts and definitions in this study.

2.1.4 **Project Manager**

Project manager is the person that is responsible for taking a project to its end, applying all necessary tools and techniques to achieve project's main goal (PMI® 2017). He/she must be the "*man/woman in between*" management and technology (Industry), and must know how to communicate with both (Gaddis 1959). Posner (1987) refers what it takes to be a good project manager (aggressiveness, confidence, poise, decisiveness, resolution, entrepreneurship, toughness, integrity, versatility, multidisciplinary, and quick thinking). Posner (1987) refers that a good project manager must have the project skills showed on table 1.

Table 1 - Project Management Skills (source: Posner 1987)

Communication	Organizational	Team Building	Leadership	Coping	Technological
Listening	Planning	Empathy	Sets Example	Flexibility	Experience
Persuading	Goal Setting	Motivation	Energetic	Creativity	Project Knowledge
	Analysing	*Espirit de corps*	Vision	Patience	
			Delegates	Persistence	
			Positive		

Turner and Muller (2005) reference seven personal characteristics of effective project managers (problem-solving ability, result orientation, energy and initiative, self-confidence, perspective, communication and negotiating ability). Choosing the proper project manager is a decision that affects the project's success, since persons are unique, and they have soft and hard skills that are suitable for each kind of project (Belassi and Tukel 1996).

2.1.5 **Project Life Cycle**

All projects have a beginning and an end (Pinto and Slevin 1988). The phases in between are called Project Life Cycles (PMI® 2017). The number of phases in a project life cycle can vary with the industry type, the type of deliverables, and the project size, among others. Kerzner (2003) has developed a sequence of phases that may be identified with most of the projects: (1) Conceptual Phase, (2) Planning Phase, (3) Testing Phase, (4) Implementation or Execution Phase, (5) Closure Phase.

PMI® (2017) defines an example of project phases in a simpler way as showed on figure 1 (Phase1 – Starting the project, Phase2 – Organizing and Planning, Phase3 – Carrying out the work, Phase4 – Ending the project).

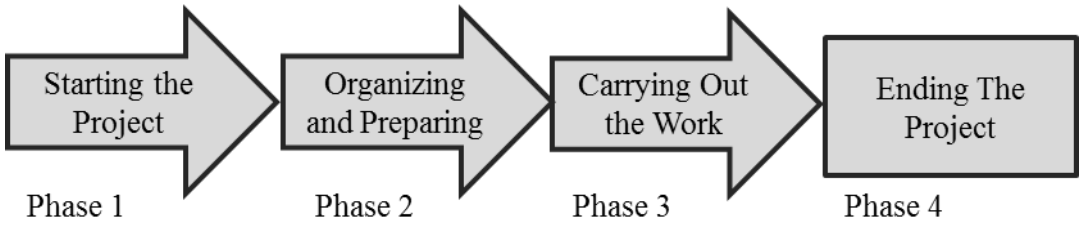

Figure 1 - Project Life Cycle (adapted from PMI® 2017)

Other IT life cycle's phases examples of a project are showed on table 2.

Table 2 - IT Life cycle (adapted from Rosen 2004)

Phase	Description
1 – Project Concept	New projects are evaluated and created and presented to the company. It is evaluated its importance and alignment with company strategy. It can be constructed a business plan for better decision. At this time, the project is identified (can be a legal and mandatory project, reduce cost effective project, among others).
2 – Project Design	At this phase project is designed in order to satisfy project requirements. Project is planned with deliverable dates, a cost is indicated, and scope is closed. The project team is defined. At the end, there is a project baseline that is presented in a customer contract.
3 – Project Development	At this phase project is developed, creating the deliverables that satisfy the business requirements.
Phase 4 – Quality Assurance	The purpose for this phase is test the requirements that were implemented to make sure that meets with the initial specifications.
Phase 5 – Beta	After Quality assurance, the project can go live for a group of users called beta testers. They can identify some errors that weren't identified in latest phases for being solved before deliverables dissemination.
Phase 6 – Release	After beta phase is concluded and approved, and the project deliverables are disseminated to all end users.
Phase 7 – General Availability	At this phase project is in use. The deliverables enter in maintenance mode, and maintenance teams fix bugs and make some small project evolution
Phase 8 – End of Life	At this time the project is discontinued since they become obsolete or expensive to support.

IT managers realize that project life cycle process can help them to identify, create, release, manage and discontinue a project. This tool helps to complete projects on time and under budget (Rosen 2004).

2.1.6 **Project Management Process Groups**

Process groups are an aggregation of inputs, tools and techniques, and outputs are generally grouped by Initiation, Planning, Executing, Monitoring and Controlling and Closing (PMI® 2017) as described on table 3.

Table 3 - Process groups (adapted from PMI® 2017)

Process Group	Description
Initiation Process group	Processes to define and approve a project begin.
Planning Process group	Processes that establish and require the strategy for achieve success. It is in this process that the scope, budget and plan are defined, creating a baseline that project customers / Sponsors approve.
Execution Process Group	Processes to achieve the plan that was defined in Planning Process Group.
Monitoring and Controlling Process Group	Processes that track and regulate the progress of a project. In this group, project manager identifies plan changes and start making actions to manage them.
Closing Process Group	Necessary actions for project completion.

All process groups are important to manage, but the planning group is critical in project success (Zwikael and Globerson 2004). When planning processes improve, project success probability gets higher (Zwikael and Globerson 2004). It is at planning phase that a project baseline is created, identifying the scope and the cost and estimation such as delivering dates for achieve it. Risk responses, the communication plans and quality measures are also included in this phase. This contract, known as Project Management Plan, is the project baseline and needs to be approved by the customer (PMI® 2017).

2.1.7 **Project Management knowledge areas**

PMI® (2017) defines Project Management Knowledge Areas as groups of understanding that a project manager must own and apply during project lifecycle. There are ten knowledge areas in project management: integration, scope, schedule, cost, quality, resource, communication, risk, procurement and stakeholder management. Each of them is applied in more than one process group, but not all are applied in all process groups (only integration management is present in Initiation, planning,

execution, monitoring and control and closing groups). Time, risk, scope and Human resources are the areas with great impact on project success (Zwikael 2009) and all have impact on project outcomes.

2.1.8 **Project Constraints**

The main challenge for project management is to achieve project success within the given constraints. The primary constraints are scope, time, quality and budget. If a primary constraint is affected, it will affect one or more of the other ones. If there is required a time reduction, for instance, it can be necessary a higher budget, a scope cut or to have less quality in the deliver (PMI® 2004). There are more known factors that are appearing as main constraints such as sustainability (Ebbesen and Hope 2013) or business oriented and customer satisfaction (Garret 2008). After 2009, Project Management Institution introduced more constraints giving more visibility to other project constraints such as risk or resources. Their balance is one of the main purposes of Project management and leads to its success (PMI® 2017).

A project achieves its success if it is completed with the scope initially defined, on the plan indicated and within the budget initially indicated (Prabhakar 2009), creating the necessary balance between them (Morris and Sember 2008). Time, cost and scope are not all the reasons for a project success but researchers still assume that they are the main reasons to determine a project success (Uluocak 2013). Some authors (Meredith and Mantel 2009) refer these constraints as project goals as showed on figure 2. Scope is defined as performance (*"Performance is what a system does")* (Meredith and Mantel 2009, p. 250).

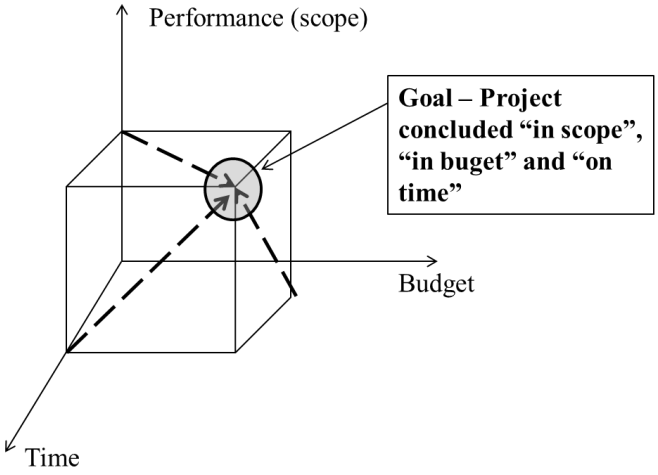

Figure 2 - Direct Project Goals (adapted from: Meredith and Mantel 2009)

2.1.9 **Project Reserves**

Projects at execution time have issues, risks known or unknown that might occur, problems that need to be solved that weren't estimated, among others. To answer them, the project manager will create a reserve at project plan. This type of reserve is identified as contingency reserve (PMI® 2007). Other type of reserve is the management reserve that usually can be applied to undefined work inside of project scope. This type of reserve is applied for unknown risk from the project, and it is not identified in the baseline, when contingency reserve is applied for identified risks and estimations identified in the project. If rework is necessary, then these reserves are used (PMI® 2017).

2.2.IT Projects in financial Institutions

2.2.1 **Projects in Organizations**

The accomplishments of important organization's goals are being achieved by adopting project management methodologies (Meredith and Mantel 2009). For PMI® (2017, p.49), *"projects are a key way to create value and benefits in organizations"*. In today's competitive environment, organizations need to be able to manage tighter budgets, shorter timelines, stricter deadlines, resource limitations, and a swiftly change of technology. Such competition results in dynamic business environments in constant change. To remain competitive, companies are adopting project management best

practices to consistently deliver business value. Effective and efficient project management should be considered as a strategic competency within organizations. It enables organizations to align project results to business goals, to compete effectively in markets and to sustain the organization, responding to business environment's changes adjusting project plans (PMI® 2017).

2.2.2 **Project Portfolios**

Project portfolio is a group of projects that share the same resources under the sponsorship of an organization, having a dynamic decision process and its projects are periodically changed and revised (Martinsuo and Lehtonen 2007). A portfolio can be defined as all projects and programs that belong to an institution or group that aim to achieve strategic objectives (figure 3). A project portfolio is used to manage its projects effectively. Therefore, portfolio management encompasses the activities devoted to managing of one or more portfolios from an institution in order to achieve project objectives in a coordinated manner (PMI® 2017).

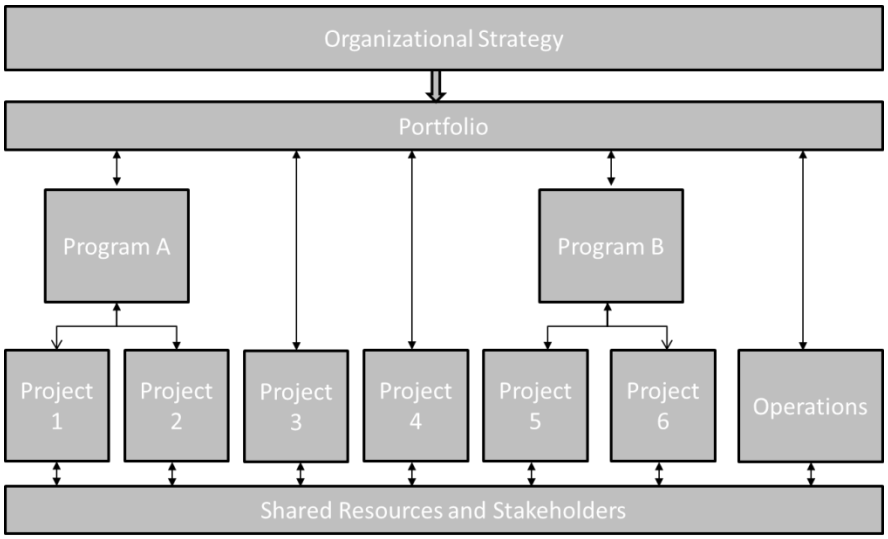

Figure 3 - Portfolio management (adapted from PMI® 2017)

It is very important that the project portfolio is aligned with enterprise's strategy, identification and prioritization of projects, in order to ensure that the best benefits and resource optimization are achieved (Petit 2012). The main focus of portfolio management is to select and prioritize projects to ensure that risks, complexity, incomes and resource management are balanced in an optimal way. Project portfolio can change when there are changes in the institution's strategy, or when there are changes in projects that belongs to organization's portfolio (Petit 2012) thus, portfolios must be

often reviewed (McDonough and Spital 2003). Delays in one project can cause delays in other projects because projects dependencies and shared resources (Meredith and Mantel 2009). The projects in portfolio selection are based in priorities. According to Meredith and Mantel (2009), usually organizations have three priority levels: (1) high priority projects (2) low priority projects and (3) urgent and mandate projects that must be started and executed immediately.

2.2.3 **Project Manager Office**

A project manager office (PMO) is an organizational structure that supports project managers implementing and disseminating project standards and processes among the organization or department. This office has the responsibility of managing project portfolio plan and support management on progress report (PMI® 2017).

According to PMI® (2017), there are three types of PMO's in the organizations (supportive, controlling and directive) as shown on table 4.

Table 4 - PMO types in organizations adapted from PMI® 2017)

PMO Type	Main Roles	Degree of Control
Supportive	Supply templates, best practices, training and lessons learned Used as a project repository	Low
Controlling	Adoption of project management framework and methodologies Conform to governance Frameworks Supply templates, best practices, training and lessons learned	Moderate
Directive	Project managers are assigned and respond to PMO	High

According to Millhollan (2008), the objectives of a PMO are to develop and to implement processes for project requests, to ensure that projects in portfolio are aligned with business strategy, to develop project management skills to project managers providing training and mentoring, to adopt best practices of project management, to report key performance indicators to stakeholders and project reports, to manage strategic initiatives and to ensure benefit realization and measurement.

2.2.4 **IT Projects**

Today's capacity of organizations to compete is supported on application software and Information Technology (IT). The continuing development in IT allows new business strategies and vice-versa. Organizations must move forward and keep software developments in various disciplines in order to succeed (Aerts *et al.* 2004).

IT projects are known to be complex, difficult and with a low rate of prediction when compared with other types of projects (Rodriquez-Repiso *et al*. 2007; Uluocak 2013). Because of that, usually these projects exceed proposed budgets, schedules and do not achieve their initial objectives. Uluocak (2013) also refers that The Standish Group in 2003 collected data concerning the factors of greatest impact on IT project success. The top 10 success factors were: user involvement, executive support, experienced project manager, clear business objectives, minimized scope, agile processes, standard infrastructure, formal methodology, reliable estimates, and skilled staff (Uluocak 2013). The success rate for information technology (IT) implementation is low. For better results, models must include stakeholder involvements, social factors attention, and integrate better change management practices. Organizations decide to invest in IT because they expect to improve their performance, cost reductions, gain productivity and quality (Legris and Collerette 2006).

2.2.5 **Banking Technologies: IT in Financial Institutions**

Banking Technologies involve the IT management from banks core business, customer relationships and risks associated with financial institutions and customers. Banks use their technology to better serve their customers, to gain profits, to adopt regulatory requirements, and to mitigate risks (credit risk, market risk or operational risk). Banks also use IT to uniform and implement standards that facilitate the communication between institutions improving service levels in banking transactions and services. Other use of technology in banking is based in data mining in order to identify customers' behaviours, customers' segmentation, scoring, and target marketing, among others. Successful use of data mining is related with profits (Moro *et al.* 2015). "*From a theoretical perspective, banking technology is not a single, stand-alone discipline, but a confluence of several disparate fields such as finance (subsuming risk management), information technology, communication technology, computer science,*

and marketing science" (Ravi 2007, p.2). The benefits for technology adoption are productivity, profitability, efficiency, faster service, customer satisfaction, convenience and flexibility, operations 24x7, space and cost savings (Ravi 2007).

Banking's competition is based on price (interest rate), quantity (deposit and loan size) and quality (reputation and relationship). Innovation in new solutions improves productivity (Chang 2003). IT solutions in banking, reduce operational costs, can get more customers and can improve reputation by innovation. It also allows 24 hours per day 7 days per week service, with the objective of the service to be consumed in different locations (Joseph and Joseph 1999).

Information Technology was found to be an important differentiator of banks that were having higher revenues in the mid-1980s, when compared to those that were less profitable (Bharadwaj 2000). A flexible IT infrastructure that allows applications to be launch in a short time, IT skills and strong customer orientation, makes that Financial Institutions can implement new products in a short time, achieving time-to-market requirements (Bharadwaj 2000). IT also reduces costs in operational processes, providing banking services in a more efficient way. IT has a considerable effect in the bank's internal processes and in its structures (Schmidt 2001). Product and service differentiation are a reality in a banking changing industry as the one that we have today. The usage of technology in an effective manner is the key to success in the banking system (Anyasi and Otubu 2009). Banks are continuously bringing innovation and state-of-the-art technology to customer services, changing from a product-driven to a customer-driven strategy. On the other hand, customers require as well that banks "become a one-stop-shop" for all their needs. Information Technology (IT) has been changing with technology evolution. For its customers, banks have delivered new or adapted services by technological solutions and new channels, such as ATM, credit cards, phone banking, or Internet banking (Ravi 2007).

2.3.Project Success Factors

Since the 1960's, project manager's researchers are trying to identify which factors lead to project success (Cooke-Davies 2002). There are many discussions about the notion of project success. There are projects that are delivered on time and under budget but considered as failure. For instance, an ERP that was delivered on time and on

budget but did not return benefits to the company (Coombs 2015). On the other hand, there are projects that are considered as a failure immediately after implementation but which can be considered as successful after some time. Due to this ambiguity, project manager and companies are forced to see project success with the measures that are quantitative, as scope, time and budget. Some organizations also include client satisfaction for measure client success. Yet, this is also a nebulous concept (Pinto and Slevin 1988). Kerzner (2003) refers that for successful project management a project must finish in time, within cost, at the desired performance and technological level, with an efficient and effective assignment of resources and with client acceptance. Project success must require its completion in time and in budget, at the proper specification level, with customer acceptance, with mutually agreement on scope changes, without disturbing the organizations flow and without changing the corporate culture.

Classic metrics referred as "the iron triangle" (figure 4) – Scope time and cost are the traditional measures in many studies of project success (Pinto and Slevin 1988).

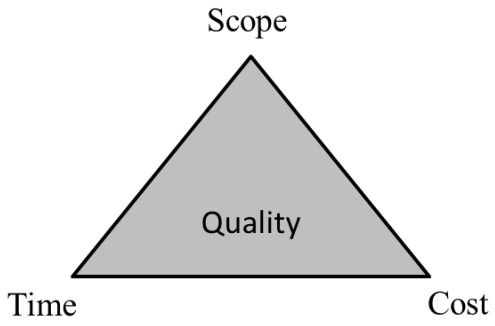

Figure 4 - The Iron Triangle

Some authors have included customer/stakeholder satisfaction as a project success factor (Millhollan and Kaarst-Brown 2016), although client satisfaction depends of the parameters of clients in measuring success, and that project success or failure is a personal perception and vary between different stakeholders (Lock 2007). Table 5 shows the outcome success factors and their change across time. Good users' requirements definition, a high involvement of stakeholders with top management support, having the right resources, and limitation of scope are examples of project success factors since the beginning of the 1980's, and they are still considered in today's project management.

Table 5 - Summary of project success factors (adapted from Millhollan and Kaarst-Brown 2016)

Author	Success Factors
Baker *et al.*, 1983	Clear requirements, proper planning and estimation, adequate staff, vision and objectives (scope definition and business case), leadership, adequate resources, absence or bureaucracy and politics.
Pinto & Slevin, 1989	User involvement, executive management support, staff, deliver on time, budget and within specifications, project members and stakeholders satisfaction, leadership, communication, absence or bureaucracy and politics.
Freeman & Beale, 1992	Technical Performance, efficiency in execution, stakeholder satisfaction, team member growth assurance, termination completeness, problem identification and solution, deliverable usability and performance.
Chaos Report, 1994	User involvement, executive management support, clear requirements, proper planning, realistic expectations, ownership, small project milestones, staff, clear vision and objectives.
Wateridge, 1995	Objective achievement, provide benefits, satisfy the users' needs, owners and stakeholders, meets pre-stated objectives, deliver on time, budget and within specifications, project team satisfaction.
Shenhar & Levy, 1997	Meet quality, time and budget goals, customer satisfaction, benefits to organization.
Whitaker, 1999	Project planning, business case, top management involvement, estimation accuracy; finish in budget and in schedule, requirement definition and meet requirements.
Shenhar *et al.*, 2001	Project efficiency, customer impact, business impact, future preparation.
Boehm, 2002	Complete requirements, user involvement, resources, executive support, no scope extension
Jugdev & Muller, 2005	Divide project success factors by eras.

To Lock (2007), at execution period, some factors are relevant for project management success: (1) good project definition and business case, (2) project strategy choice, (3) support from higher management, (4) available funds and resources, (5) firm control of changes, (6) technical competence, (7) organization's quality culture, (8) safety and health of project team, (9) project communication, (10) motivation, (11) resolution of conflicts.

Project's cost, time and scope are factors that can be measured during and after the lifetime of a project, when others might not be seen until long time passed (Millhollan and Kaarst-Brown 2016). Project managers' skills are also a success factor (Millhollan and Kaarst-Brown 2016). Cooke-Davies (2002) says that project success and project management success are distinct. Project success is measured against the main objectives of the project. Project management success is measured against traditional measures of performance, as time, quality and cost. The other distinction is between

success criteria (the measures by which success or failure of a project will be judged) and success factors (inputs to the management system that leads directly or indirectly to project success) (Cooke-Davies 2002).

Millhollan and Kaarst-Brown (2016), distinguish different types of success according to different definitions as showed on table 6.

Table 6 - Relations between definitions of success (source: Millhollan and Kaarst-Brown 2016)

Term	Definition	Hard Skills	Soft Skills
Project Success	Meet planned and desired business objectives	Clear Requirements and objectives	Change management, communication, conflict resolution, decision making, expectation management, negotiation, stakeholder analysis and management
Project Management Success	Accomplishment of cost, time and quality objectives	Knowledge of project management tools and technical knowledge	Decision Making and leadership
Project Manager Success	Successful application of project management methodology	Industry, organizational and disciplinary knowledge	Motivation, leadership, negotiation, communication, conflict resolution, team development

According to Lock (2007), project failure depends directly from (1) scope not clear or understood, (2) vague technical requirements (3) cost, timescale or benefits too optimistic, (4) incomplete task assessment, (5) inappropriate project strategy, (6) insufficient regard in cash-flow or funds payments and provisions, (7) stakeholder concerns not take to account, (8) motivation and behaviour of project team and (9) project approval without concerning business plans.

In addition, Posner in 1987 referred the main problems in project management were (1) inadequate resources, (2) unrealistic deadlines, (3) unclear goals, (4) team members not bounded, (5) insufficient planning, (6) communications breakdown, (7) changes in goals and (8) resources and conflicts between departments or functions.

According to Meredith and Mantel (2009), scope (performance), time and cost are the primary objectives to be met in a project. These are direct Project Objectives.

A project to be succeeded must take in concern four dimensions: (1) project's efficiency, that is related with primary objectives, (2) project's impact on the customer, related to customers' satisfaction and to the project acceptance, (3) business's impact in the organization relying if the project's business case was successful and (4) opening new opportunities for the future, where project benefits and opportunities can be measured.

The project's success factors are not universal for all type of projects (Dvir *et al*. 1998). Different types of projects are affected by different success factors (Hyvari 2006a). Not all success factors are directly related with project management or with the organization. External environment is also a critical success factor to be considered (Belassi and Tukel 1996). The identification of critical factors leads to better project evaluation and performance improvement (Uluocak 2013).

2.4.Project Changes

A risk is defined by PMI® (2017) as an uncertain event or condition that, if it occurs, it will have an effect in one or more project objectives.

Instead of studying risks, some authors prefer to study changes and deviations (Hällgren and Maaninen-Olsson 2005; Petit 2012). While a risk is known but did not yet occurred, a change refers to situations that occurred or will occur and that will imply a divergence in project plan. Changes are reactive, because they are unknown at planning time (Hällgren and Maaninen-Olsson 2005; Petit 2012).

Project changes usually imply scope, time and/or cost deviations from initial baseline. The most common changes are explained as a natural tendency of the client and/or the project team members to try to improve the product or service. New demands and performance requirements are discovered at execution time, and not before, at project plan where project management plan can be adapted before client approval. New

technologies, ideas or solutions can be better identified at project execution when stakeholders can have a better glimpse of the product to be delivered. As later these changes are discovered and implemented in the project, the more difficult and costly they are to complete. Without control of changes implementation, small changes can imply a significant impact on project´s schedule and cost (Meredith and Mantel 2009).

It is possible to minimize project changes by making them affect only part of the project and not all its deliverables. The solution is to break large projects into smaller projects (this is one of the most important tasks in developing projects). This will ensure spread responsibility across a greater number of people as an easier way to manage it (Clarke 1999).

Millhollan (2008) refers that it is important that any change affecting expectations must follow a formalized change request process. Impacts in expectations are impacts in scope, time delivery and in cost of the project. A formal change control system is responsible to integrate and coordinate changes through the project life cycle, controlling changes in a proper way. The purpose of the formal change control system is to review all requested changes in project, identify impacts, and translate them to project scope (performance), cost and schedule, evaluate benefits and costs of requested changes, accept or reject changes, identify synergies between changes, communicate changes to all stakeholders, ensure correct change implementation and prepare change report recurrently (Meredith and Mantel 2009).

2.5. Project variables that imply in project changes and project success factors

In this subchapter it is explained and detailed how some factors can affect project changes and project success. The information chosen is aligned with the features (variables) of the dataset used for this study.

2.5.1 Scope, time and cost changes

Scope is one of the primary constrains of the iron triangle and can be defined as the work that will be done in the project to fulfil the project requirements. "*Scope change is inevitable and natural. It is important to understand that our job is not to stop scope change, but to successfully manage that change*" (Millhollan 2008, p.2). Also for Kerzner (2003, p.6), "*scope changes are inevitable and have the potential to destroy not only the morale of a project, but the entire project*".

There are three main causes for scope changes: (1) when there is an error in planning the solution for the project's requirements; (2) when clients or end-users change the purpose from the deliverable or the deliverable specification itself; (3) when it appears mandatory and legal requirements external to the project (Meredith and Mantel 2009). Scope changes (change requests) need to be managed carefully. The ones that affects plan cost or conceal great differences in specifications, must be approved by a Change Control Board (CCB), constituted by project sponsors and stakeholders that can measure the costs and the plan changing that the change will imply (Lock 2007; PMI® 2017). Medium and small projects might not have a CCB. Yet, it must be always defined a process for the changes approval to each project (Meredith and Mantel 2009). Allowing scope changes by a mature scope change control process is positively correlated to cost performance (Cooke-Davies 2002).

There are changes that are not related to scope. Some of them only affect time and budget. Usually these changes are omissions and errors in planning that are discovered during execution process (Millhollan 2008). Time changes, or delays, may happen because of lack of production capacity, impractical planning or resources multitasking (Ayal and Globerson 2002).

As referred by Ayal (2005), as later a change occurs on a project, as higher the impact it will have on its cost. Complex or innovating projects can encompass more uncertainty and it may be required more time to solve problems or getting information. This results in a higher project cost (Yu 2017).

There are projects that do not have a favourable business case, not having a profitable reason (scientific research, for instance). Yet, even in these cases, project cost management and project control are also vital for project management success, and therefore for project success, since if costs overrun, the project can become incomplete not reaching its purpose (Lock 2007).

2.5.2 **Duration Estimation**

According to Ayal and Globerson (2002), duration's estimation is critical for project planning. It predicts how much time is needed to complete the project tasks. There is a psychological tendency to underestimate the amount of time needed to complete a project or overestimate the amount of work that can be done in a time period. This is called *planning fallacy* (Buehler *et al.* 1994). An alternative approach is to use mathematical models instead of knowledge basis estimation.

Estimation is one of the critical steps at planning phase of a project, since it involves resource allocation and time/effort/cost allocation. It is the estimation that defines most of the final project cost. A poor estimation implies that resources aren't optimized, delaying and affecting the overall project portfolio plan (PMI® 2017) A positive estimation can lead to project's budget and deadlines slippage, when a pessimistic estimation can lead to lack of budget or person resources that could be allocated to other projects. It can also involve that the project sponsor gives up from project implementation. Estimation also considers uncertain, that leads to risk analysis (PMI® 2017). In order to achieve an accurate estimation at planning phase several techniques are used. In a more classic way, expert-judgement, and bottom-up analysis are user. Yet, these estimations have limitations, since they depend of the persons that are estimating their risk adversity and their state of mind (Pospieszny 2017). In order to respond to these limitations, parametric techniques can also be used (e.g., function points, lines of code), but these ones, sometimes shows a lot of complexity and they also can be fallacious for some kind of projects as technological change projects (Pospieszny 2017). Literature also indicates the use of data mining and knowledge extraction techniques to identify real estimations (Villanueva-Balsera *et al.* 2009), but no real applications were found in literature analysis.

According to Cooke-Davies (2002), project duration must be kept below three years (one year is better) in order to achieve project success.

2.5.3 **Project Duration**

Project Duration is the time between the project initiation and its closure. Cooke-Davies (2002) identified duration as a critical factor for project success. The project duration can be defined as the sum of all task durations that have a mandatory dependency between them and do not have any lags or leads. This tasks or activities make a path, being this activity path a representation of the longest path in projects. This path is designated by PMI® (2017) as the critical path, and it is a way to determine the shortest project duration. A variation in the activities that are on the critical path (that do not have any float), will affect the project duration. Resource availability is also an important factor for project duration. Some projects can be delayed because specific resources were affected to other projects or tasks. A good project resource optimization will affect positively project duration (PMI® 2017).

2.5.4 **Client Importance**

According to several authors as shown on table 7, client consultation and client acceptance are quite relevant as critical project success factors (Uluocak 2013).

Table 7 - Project Implementation Profile CSF importance rankings (source: Hyvari 2006b)

	Hyvary, 2006	**Finch, 2003**	**Delisle & Thomas, 2002**	**Pinto & Prescott, 1988**	**Pinto & Slevin, 1987**
Project Mission	6	7	1	1	1
Top Management Support	4	6	9	7	2
Project Schedule	5	5	5	9	9
Client Consultation	**2**	**1**	**2**	**2**	**4**
Personnel	9	10	10	10	5
Technical Task	7	9	4	3	6
Client Acceptance	**3**	**4**	**6**	**4**	**7**
Monitoring and Feedback	10	3	3	5	8
Communication	1	2	8	6	9
Trouble-shooting	7	8	7	8	10

A client consultation ensures quality in the project and involves users in the project process (Pinto 1986; Slevin and Pinto 1986) and, according to Yu (2017), customers' participation and involvement in software projects makes easier to achieve information about customer's needs, facilitating implementation, raising efficiency, and saving costs and time. The impact of customer participation is greater for projects with higher levels of complexity.

2.5.5 **Project Team Size**

According to Uluocak (2013), the project team size (number of personnel affected to a team) has a direct relation with the coordination complexity, that can put in risk project's success. As projects become more complex, more areas of expertise (internal or external to the organizations) are required. Managing multidisciplinary teams becomes more problematic and the project risk increases (uncertainty increases). When the various individuals and groups working in the project aren´t integrated, they tend to work as separate and distinct parts, not communicating between them. Multi-

disciplinary teams involve conflicts and project manager must be able to solve and intermediate these same conflicts (Meredith and Mantel 2009).

2.5.6 **Organization Size and Communication**

Most of the project manager's time is spent communicating. With the project team and with stakeholders, the project manager must sell, re-sell, explain, and report project issues recurrently (Meredith and Mantel 2009). According to PMI® (2017), project managers spend 90% of their time communicating. Communication has been gaining importance across time. As it can be seen in Table7, communication is the highest ranked success factor in the latest studies.

Organization size (Hyvari, 2006b) and communication (Pinto 1986), have a direct impact in project success. Organization size and structure can difficult communication considering that large organizations have more stakeholders, making projects more complex to manage.

2.5.7 **Project types**

There are several types of projects. Dvir *et al.* (1998) identify ideal project types for general classes of projects and concluded that project success measures vary with the type of the project. Belassi and Tukel (1996) refer that an urgency of a project (e.g., mandatory projects with deadlines, time to market projects or legal requirements implementations) can affect project success, since the plan is already constrained with a due date. According to Martinsuo and Lehtonen (2007), differentiated management style applied to different project types is relevant to project efficiency.

2.5.8 **Outsourcing**

As a definition, "*outsourcing is the use of external agents to perform an organizational activity*" (King and Malhotra 2000, p.1). For companies, IT became one of the functions to be outsourced so that the focus can be in core business and not in IT. Internal IT functions are considered more ineffective and inefficient compared with specialized IT companies. On the other hand, IT outsourcing can lead to loss of control of IT assets, opportunism from suppliers, reduced IT expertise and reduced corporate

memory, as well as a decline of morale on internal employees (King and Malhotra 2000). Today, more and more companies are relying on outside third parties to provide critical IT products and services, and this phenomenon is likely to continue into the future (Lee and Kim 2005; Davis *et al.* 2006).

Organization and outsourcing partnerships are being used in Informational Systems and they are based in mutual trust. Cost savings, quality of IS services and user satisfaction are known outcomes for outsourcing services acquisition. Other reason for clients to engage in IT outsourcing is to reduce and transfer risk to vendors that are expected to be able to address this risks because of their competence and experience, since developing IT projects are their core business (Taylor 2006).

To mitigate the problems of coordination, opportunistic behaviour, and to facilitate learning/knowledge transfer, it has been given importance to assessment of vendor's abilities, such as building and sustaining a cooperative project environment and establishing problem-solving and conflict-resolution arrangements. Sourcing agreement needs to include high trust, commitment and outcome-oriented for vendors and contractors, making and sustaining a win-win relation. Information exchange, legal bonds, relationship-specific adaptations by the client and the vendor are project's success key factors for companies (Haried and Ramamurthy 2009). The conflict resolution is also an important aspect, since *"no relationship exists without an occurrence of a conflict"* (Haried and Ramamurthy 2009, p.57).

The advantages of offshoring (provision of products and services from locations in other countries) and the outsourcing of Information systems are the cost savings and the allowance for companies to focus in their core activities. The dangers are the loss of important business skills and reliance on suppliers who face risks unfamiliar to the client. According to the outsourcing principle, an organization should not do tasks that are not in its core business competences or aren't critical success factors, since other companies can do it better with less cost and faster (Davis *et al.* 2006).

It is quite often that large projects have more than one outsourcing team and sometimes in different locations. This makes more difficult the communication between all the involved teams. Since teams are not in the same physical location, can also difficult project alignment. The communication and teamwork capability issues are the

same than in virtual projects (projects where team members rarely meet as a face to face team, being distributed in different locations) (Meredith and Mantel 2009).

2.5.9 **Number of Stakeholders**

Not all stakeholders affect a project in the same way. Lack of clarity and agreement among the stakeholders may turn the project into failure (Legris and Collerette 2006). Identifying and rating stakeholders is a very important task for project management. The more stakeholders there are in a project, the more difficult it is to satisfy expectations since not every stakeholder have the same perception and interest in the project (Lock 2017). For instance, in a dam construction we can have stakeholders that agree with its construction and other ones that disagree (e.g., archaeologists and environmental organizations).

Differences in cultures are seen not only in different geographies but also in different internal functional departments. Different divisions have different cultures, and some can be more "project oriented" than others (Meredith and Mantel 2009). Also, the difference between these divisions when they are relevant stakeholders in a project can be a difficult issue. If there are two different divisions as stakeholders with opposite cultures can affect project decisions and therefore project success. Other kinds of stakeholders are the departments from organizations, or internal stakeholders. Mostly an organization project can have many departments as stakeholders. Jealousy, mistrust and conflict between departments lead to uncontrolled scope creep and to delays and budget overruns (Meredith and Mantel 2009)

2.5.10 **Customer Satisfaction**

Customer's satisfaction is an important variable for knowing if a project has succeeded (Garret 2008). A project may end on defined scope, within time and on approved budget, but end-users may not understand its value or project implementation may not face the business expectations (for instance, projects that do not bring the expected benefits) (Coombs 2015). In other hand, each stakeholder has its own way to measure project success, thus customer's satisfaction can be different from stakeholder to stakeholder (Lock 2007). Managing customer's expectations is a key success factor

for any project and to solve problems that occur at project time. It is essential to customer satisfaction that a project manager gives realistic expectations since the project beginning, avoiding, for instance, define unrealistic milestones. It is more difficult to change and clarify a false expectation in a later phase of the project (Taylor 2006). Client expectations and project team designs should be aligned, but rarely are (Meredith and Mantel 2009).

2.5.11 **Number of applications**

The term applications in this research is referred to an independent informational system that has a set of functions and objectives well defined to achieve a main purpose, in a group of interdependent programs related and with a well-defined context. Some examples are customer management application, finance reporting application, site application, and payments application, among others.

Increasing the number of applications increases the project density (ratio between the total number of precedencies and total number of activities) (Belassi and Tukel 1996), which makes projects more complex to manage.

2.5.12 **Mandatory projects**

One type of scope changes is mandatory changes. These changes usually have external organization factors based on legal requirements, new laws or directives. These changes can be only by themselves a project, changing the prioritization of other projects in the portfolio (Meredith and Mantel 2009), since usually its implementation is classified as urgent.

2.5.13 **Project dimension**

One of the structural factors that affect project success is the size of the project (Dvir *et al.* 1998). Belassi and Tukel (1996) indicate that project size is one of the factors related with project success or failure and that many large-sized projects (megaprojects) exceed their deadlines. For instance, megaprojects are projects that usually have more contracts or applications and they tend to be more risker than projects with only regional and few contractors (PMI® 2017).

2.6.Data Mining

Today, information is very important to organizations. The extraction of information is helping companies to be competitive, to achieve a faster time-to-market products implementation, and to replace knowledge extraction. Data mining helps to retrieve from large amounts of data information and knowledge that can be useful within Organizations. It uses multidisciplinary techniques from statistics, pattern recognition, artificial intelligence, and machine learning in order to extract useful information (Pospieszny 2017).

Data Mining (DM) can be used to obtain knowledge from data, and it can improve decision making (Cortez and Embrechts 2013). There are two types of Data Mining approaches. Classification that aims to obtain knowledge by giving a categorical response, and regression that responds with a numeric/real value (Chapman *et al.* 2000).

There are several models that can be applied to obtain knowledge. Some can be directly interpreted (such as logistic regression or decision trees) also known as white-box modules, while others cannot, but are capable of more accurate predictions (Neural Networks, Support Vector Machines or Random Forests). These models are known as black-box models. To understand these models, it is necessary to extract rules or to use visualization techniques (Cortez and Embrechts 2013).

Logistic regression represents occurrence probability of a dichotomy event (categorical target variable). It operates a logistic transformation over a multiple regression model allowing estimating all possibilities of classes. Although the model is easy to interpret, it is rigid and does not model complex linear relationships (Moro *et al.* 2014).

Decision trees are decision algorithms that modulate a dependent variable from a set of other independent variables that will being divided to describe better a problem. They are easy to understand and interpret because its translation to IF-THEN rules (Hastie *et al.* 2008; Moro *et al*. 2014).

As black boxes models can be used Neural Networks (NN), Support Vector Machines (SVM) and Random Forests (RFOREST) (Hastie *et al*. 2008)

Neural Networks (NN) are simplified models of the human brain. They have a parallel process composed by single process units (nodes). Knowledge is stored in node

connections. During learning, connection weights are adjusted. MLP (Multilayer Perception) is the most used Neural Network topology (Moro *et al*. 2014).

Support Vector Machines (SVM) is a supervised method used for classification problems that find optimal decision function and that has a maximal margin between support vectors (Hastie *et al*. 2008). It transforms the input space into a multi-dimensional space according the different features of the model. This model finds the best linear hyperplane related to a set of support vector points in the m-space (Moro *et al*. 2014).

Random forests are a predictive technique that consists in an ensemble of decision trees where a large number of trees are modelled and in the end the results are aggregated (Baecke and Bocca 2017).

Black box models have a high-level of accuracy, yet, they are complex to be read directly by users, unlike other models as decision trees (Cortez and Embrechts 2013). To read this black boxes models it can be used some techniques as rule extraction, where with model results it is made a prediction and an adaptation to a white box model, such as decision tree, or a sensitivity analysis where it is possible to understand which are the model features that most affects model result (Moro *et al*. 2014).

2.7.Data Mining applied to project Management

In the last years, Data Mining applications are increasing with the increase of data availability (Pospieszny 2017). Big Data is helping in taking information where it wasn't considered or possible in the past (Calderaro 2015). Thus, data mining applications are being used in industry (e.g, Construction, IT, among others) (Ertek *et al.* 2017). Data mining is also being applied to project management. Literature indicates that is being initiated its application, although there are still many gaps in literature (Ertek *et al*, 2017; Pospieszny 2017).

According to Pospieszny (2017), the application of data mining techniques in project management can improve project success rates and estimation, quality or cost identification in project planning, and PMO information can be used to extract knowledge that improves estimations, monitoring, risk and quality management. For instance, a regression model can be used to calculate project estimations and duration

instead of bottom-up analysis expert judgement, function points or other techniques that have several limitations since they are subjective or complex techniques. Models can be used to identify at project beginning how many resources will be necessary for a specific project, originating a better resource planning and management. Regression and classification techniques can be also used to monitoring the project, identifying if the project is at time and on budget or if it has delays. Other data mining techniques can be used for risk identification or issues and anomaly detections.

Data mining for project management can be useful for identify and extract knowledge that can provide information that will help to mitigate risks and projects failure. This is being applied to many industries that are identified by literature (Pospieszny 2017), but it was not found studies from its application for banking IT development project management and its use for predict project changes at execution time.

[This page was intentionally left blank]

3 Materials and Methods

3.1.Case Description

This case addresses a European financial institution. This company can be classified as a medium sized company in retail banking and has implemented a Project Management methodology for IT development implementations and application maintenance based on PMI® processes and best practices.

The Company has a Project Management Office (PMO) and project management tools that helps projects to be managed by project managers, controls and manage organization project portfolio, and supplies report information to top and middle management.

In the organization's IT development department, IT development department units are divided into functional areas to better support and respond to business divisions' requirements (e.g., credit services, payments, finance, branches, channels and internet, treasury, informational systems, customers and accounts). Each department unit has a pool of project managers and a pool of resources divided by application development teams (credit recovery team, customer's team, and cards team, among others). Each team supports the application's maintenance and supports projects managed by project managers. Projects can have several applications involved from different teams of the development department, but usually a project manager is chosen from a unit where the main application is allocated.

The IT organizational structure is shown on figure 5. Projects are required by institution's divisions (for instance, Marketing, Channels, Account, and International Division, among others). One project can have several clients, but there is only one division or department that is the primary client, where the project sponsor is chosen and identified.

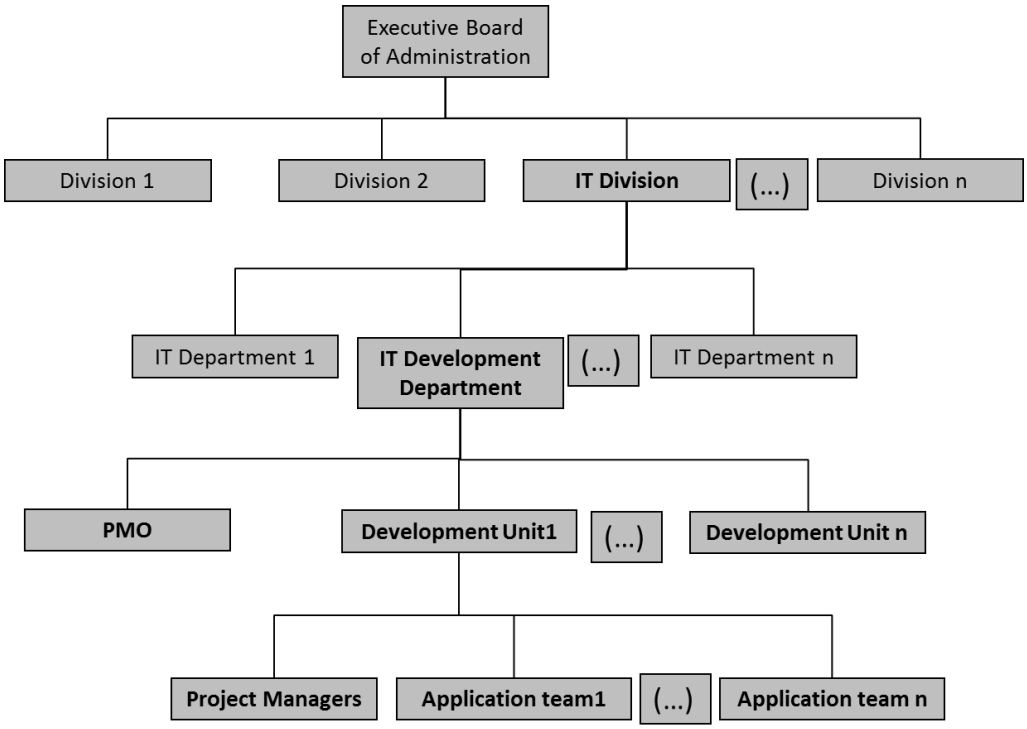

Figure 5 - Organization structure

All required projects are included in the institution's project portfolio. Projects in the portfolio are prioritized and planned in top management steering committees at the beginning of the year and can be aligned and reviewed during the year. Yet, during the year, some projects are requested with a high urgency request for being implemented as soon as possible (e.g., business extreme needs, mandatory projects, audit recommendations, among others causes.), and for several reasons they cannot wait until next committee to be panned. In these cases, top management after understand the impact in the project portfolio (which projects need to stop or be delayed for this new project start immediately), sends the approval for initiate this project and portfolio is reviewed. These projects are called interposed projects and they impact the IT portfolio project plan, especially in projects that haven't yet began.

Projects are delivered by PMO to each IT development unit manager that chooses and assigns a project manager. The project manager accepts and initiates the project and starting planning phase, allocating resources, creating functional analysis and building estimations and the project plan.

As others, this organization recurs to outsourcing resources to develop some projects. The outsourcing contracts can be "off the shelf" (Costs off the shelf or COTS), or with

resident outsourcing teams that are managed by project manager making a tailored development as internal teams do (time and materials).

After costs, estimation, time, risks and scope are defined and planned, the project manager creates a project management plan (known internally as project contract), to be approved by the internal clients. This contract defines that planning phase is concluded and a baseline is created. Plan and schedule are defined, go-live dates are set, and scope was already accepted by customers through a functional analysis document approval (that is created and approved by sponsor and stakeholders at planning phase). At this point, project costs, plan, and scope are fixed, setting the initial baseline.

After internal clients (especially the project's sponsor) approve the project management plan, the project manager starts its execution and project's team starts developing their tasks. After project team finish all tasks and make integrated tests, the solution is then promoted to a pre-production (known also as certification) environment, where internal customers and internal clients are invited to certificate the solution. Usually, it is in this phase that scope changes appear.

Users identify errors (defects) and issues (small improvement requests) which need to be implemented. If the project manager has made some project contingency reserves, they can be used to implement issues. Otherwise, a change request takes place. Another possible scenario may occur when the solution does not fit to customer expectations or when customer changes his mind (PMI® 2017) about functionalities initially required and structural changes must take place. In this case three situations may occur: (1) a change request must take place, (2) another project must be opened to implement new needs; (3) project is cancelled.

The Change request is a contract addendum made by the project manager. This addendum indicates a change in the project, by cost, plan and/or scope, and changes project's management plan initial baseline. These changes can impact on other ongoing or planned projects, since the resources and applications are the same. If so, top management must decide if the current project changes are implemented first, assuming the delays on the deliveries and in the project portfolio plan. After changes are estimated and new plan proposed, contract addendum is sent to customer approval. After his approval, changes take place and a new project baseline is defined. One project can have several change requests. Some of them only affect time (for instance, production date is delayed because is needed more time for certification), others can

affect only cost (one more resource is required to achieve project plan date). Scope usually affects time or cost. Some project changes can also affect time, cost and scope.

After the client certifies the deliverables, and if there are no impeditive errors identified, project can be deployed into production. The project team is only dismembered after one or two months of warranty period were production support is given. After project's conclusion, internal client is inquired about his project's satisfaction. Some projects because their specificity or urgency may not respect all this implementation process. These projects are known as special projects and they are exempt from project management metrics.

One of the project tools used in the company for project management is a software tool where project managers can register all project data since the beginning of the project until its closure. Dates, costs, scope changes, budget control and risk management information are registered in this tool. This tool is also used for resource time entry and to manage the entire project workflow defined in the project lifecycle, implementing PMI®'s best practices and processes.

3.2.Data Collection and Preparation

The used data for this research was obtained from a financial institution's IT PMO department. It was delivered for the study a raw dataset with 1361 records that corresponds to projects that were deployed into production between May 2012 and October 2017. From this dataset it was possible to obtain 36 variables (features) for data analysis as described on table 8. From these variables, 16 were chosen for creating a predictive model. The table also indicates if the variable is continuous or categorical, and the range of values each can assume. The purpose of the model is to predict if changes in future projects will be affected by project variables, and which variables are more important to identify project's changes during its lifecycle. The dependent variable of project changes is calculated assuming "yes" if one of the four conditions applies: (1) number of change requests is greater than zero; (2) date difference categorization is "yes"; (3) budget difference categorization is "yes"; (4) external costs difference categorization is "yes".

The project size and the duration timeframe were calculated according to project's estimation and project's duration, according to the following criteria, that is

implemented on the institution: (1) small, when project has less than 100 person-days; (2) medium, when project has less than 250 person-days; (3) large, when project has less than 500 person-days; (4) very large when project has above 500 person-days (one person-day is equal to 8 hours, according to institution's internal definition).

The project duration is calculated as the difference in days between first deployment into production date and project start date. Real duration is calculated with the difference between real production date and project start date. The first estimation and the total estimation effort are variables that are basically the same (although the source of both is different, they indicate the first estimation of the project), so only the second was considered for analysis.

Although there are other interesting variables, only the variables that can be calculated or identified at project planning phase (before project management plan approval) were chosen for being included in the model. The variables that are obtained at execution or closure phase were not considered as inputs for the model, since those features are not known at project contract elaboration.

There are two variable types in the dataset. Categorical variables that assume a limited and discrete number of values, and continuous variables that assume a numeric or a date value. Table 8 discriminates the different number of values for each categorical variable (var. type = "Cat"), while continuous variables (var. type = "Cont") assumes the value 1. Table 8 also identifies the variables that were chosen to be included in data model (last column).

Table 8 - Data variables

Variable Categories / Description	**Var. Type**	**Nbr. of values**	**Include analysis?**
Independent (Original Dataset)			
Project ID	Cat	1361	N
Project Phase	Cat	3	N
Project Type	**Cat**	**2**	**Y**
Internal Client	**Cat**	**22**	**Y**

Variable Categories / Description	Var. Type	Nbr. of values	Include analysis?
First Production Date	Cont	1	N
Real Production Date	Cont	1	N
First Estimation	Cont	1	N
Real Estimation	Cont	1	N
First External Costs	Cont	1	N
Real External Costs	Cont	1	N
Interposed	**Cat**	**2**	**Y**
Decision Type	**Cat**	**4**	**Y**
Mandatory Type	**Cat**	**7**	**Y**
Objective	**Cat**	**12**	**Y**
Number of Applications	**Cont**	**1**	**Y**
Number of Internal Clients	**Cont**	**1**	**Y**
Total Estimated Effort	**Cont**	**1**	**Y**
Internal Estimated Effort	**Cont**	**1**	**Y**
Outsourcing Estimated Effort	**Cont**	**1**	**Y**
Actual Internal Effort	Cont	1	N
Actual Outsourcing Effort	Cont	1	N
Number of Outsourcing Teams	**Cont**	**1**	**Y**
Number of Internal Team members	**Cont**	**1**	**Y**
Begin Date	Cont	1	N
Number of Change Requests	Cont	1	N

Independent (Calculated Variables)

Variable Categories / Description	Var. Type	Nbr. of values	Include analysis?
Date Difference	Cont	1	N
Date Difference Categorization	Cat	2	N
Budget Difference	Cont	1	N
Budget Difference Categorization	Cat	2	N
External Cost Difference	Cont	1	N
External Cost Difference Categorization	Cat	2	N
Estimated Duration	**Cont**	**1**	**Y**
Real Duration	Cont	1	N
Project Size	**Cat**	**4**	**Y**
Duration Size	Cat	4	N
	Dependent		
Changes in Project	**Cat**	**2**	**Y**

3.3.Data Mining Application

This study aims for a classification problem, since the purpose is to identify projects with changes according initial features (variables).

The CRISP-DM (Cross Industry Standard Process for Data Mining) was used in this study as the data mining methodology. The CRISP-DM process was adopted in this study, as identified on figure 6, CRISP-DM is divided by relevant stages or phases that were run in this study elaboration. The business understanding stage was processed when business analysis and literature review took place. Then, the data understanding stage was made when each feature of the dataset was analysed. After, the data preparation stage was achieved when relevant variables were identified and chosen to the model, and computing other ones that were relevant for model analysis and data analysis. Afterwards the modelling phase took place were modulation techniques were identified and it was chosen the more suitable model for analysis. Finally, model

evaluation stage and results were obtained and the conclusions were detailed (Chapman *et al*. 2000).

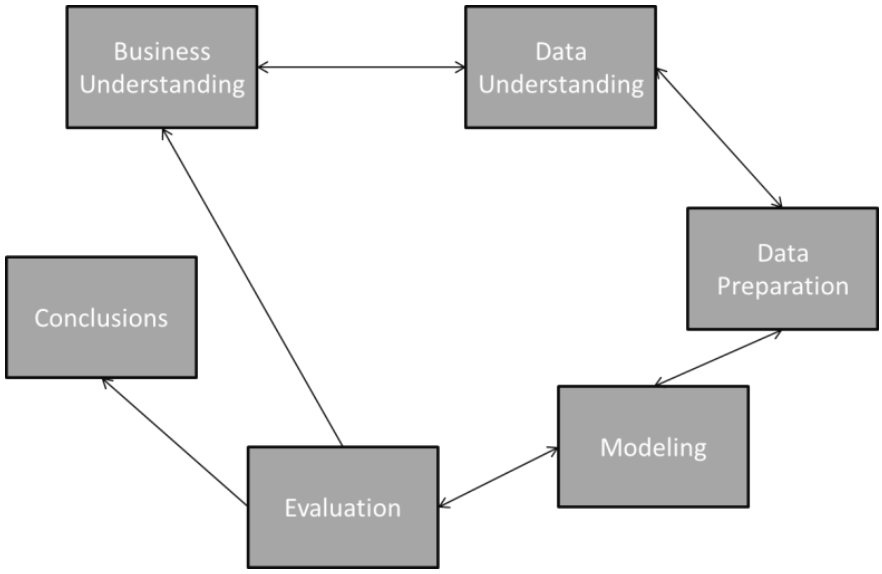

Figure 6 - Adaptation of CRISP-DM phases to this study

3.4.Model Evaluation

The performance of these models, in order to identify the one that have a better fit to the problem, is calculated using cross validation techniques. Cross validation techniques are widely accepted as a procedure for the performance estimation and model selection (Refaeilzadeh *et al*. 2009). The techniques considered for model's application were k-Fold and holdout.

In k-Fold cross validation technique, data will be divided in K parts (10 parts were used in the study, since it is the most usual number of parts) with similar size, where K-1 (in this case 9) parts are used to training sample and one part is used for validation sample. The process is repeated until each part is used for validation (Refaeilzadeh *et al*. 2009; Bengio and Grandvalet 2014).

In holdout process the approach is to divide the data in two separate parts, being one used for validation and the other one for testing the model. In the study a third of the data is used for validation, when 2 thirds are used for testing the model (Moro *et al*. 2014; Refaeilzadeh *et al*. 2009).

In comparison by these two techniques, it was decided to use k-fold instead of holdout, since it is a more robust technique, providing a better simulation of a real scenario, when compared to the more static holdout (Refaeilzadeh *et al*. 2009) although in appendix A, the results for holdout runs are presented.

Afterwards, it is calculated for each model the average "area under the curve" (AUC) for the "receiver operating characteristic" (ROC). The ROC curve shows the performance of a two class classifier on the range of a determined threshold values where x-axis determines 1 minus specificity and y-axis the sensitivity. The threshold is the value for one event is true if the probabilistic outcome of that event is bigger than the threshold. AUC is the area of ROC between 0 and 1. A random classifier has a value of 0.5, and the desirable value is 1.0 (Moro *et al*. 2014).

The area of the LIFT cumulative curve (ALIFT) is also used as a metric for measure the accuracy of each model. ALIFT is obtained by the area of the line plot of the samples of the population against the cumulative percentage of real responses captured. As AUC, as closest to 1.0 more desirable is the result. An ALIFT closest to 1 represents that the predictive model in the top deciles (division of the population in a decrease order of their predictable probability of success), while 0.5 corresponds to a random of performance (Moro *et al*. 2014).

The approach used, first identifies the best model (the one with the best value for k-fold) by using AUC and ALIFT value comparison. Then, it is applied a "data-based sensitivity analysis" (DSA) into the best model. This verified the importance and relevance of each feature in the model and how each feature influences and predicts the outcome (Cortez and Embrechts 2013). DSA was chosen for analysis since it has been successfully applied on several studies for knowledge extraction (Moro *et al*. 2017), and importance of relevant features can be visualized and analysed.

All the experiments were conducted in R Statistical Tool (https://cran.r-project.org/), since provides an open source framework with multiple resources for data analysis and data model (James *et al*. 2013). The chosen package for data mining tasks is the "rminer", since it has functions that help in knowledge extraction and in sensitivity analysis, and because it is a package already used for other studies (Moro *et al*. 2014; Moro *et al*. 2017) with interesting results, and that were validated by the scientific community.

[This page was intentionally left blank]

4 Results and Discussion

4.1.Data Analysis

In this subchapter each feature of the model is analysed. This analysis is important for the data understanding phase of CRISP-DM. Other features that exist in the dataset but were not used in the model were also analysed, since they can have relevant information to confirm initial data, and to understand business and the data.

4.1.1 Changes in projects

The project dataset has a total of 1361 projects. From these projects, 414 had changes across their lifecycle, so almost 30% of this company's projects were affected by changes as described on figure 7.

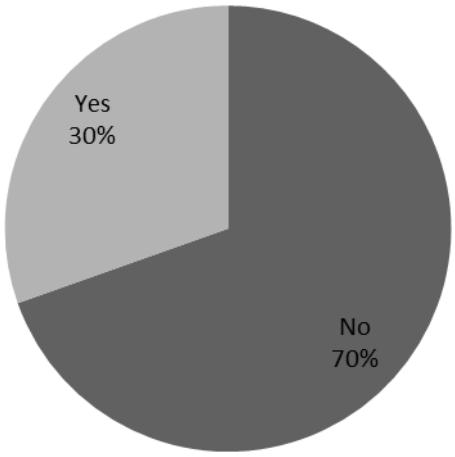

Figure 7 - Changes in projects

The "Changes in projects" feature is the dependent variable and assumes the value "yes" if one of the following four conditions apply to each project: (1) the "number of change requests" feature is greater than zero (there was at least one scope, time or budget change in the project); (2) the "date difference categorization" feature is "yes" (it means that there was a difference between the target end date and the real end date); (3) the "budget difference categorization" feature is "yes" (means that project cost in the end was different that was initially accorded at project management plan); (4) the "external costs difference categorization" is "yes" (outsourcing or time and materials costs were different than the indicated in project management plan). As shown on table

9, the number of change requests (that include scope, time and date changes) is 24.25%. There are 23.95% of projects with date differences between the initial date and real date. Budget differences affected 11.39% of the projects, when only 38 projects had external cost differences between initial budget estimation and real external costs, referring to 2.79% of the total number of projects. Yet, it is important to mention and understand that only 202 (15%) of the projects had external costs as marked on figure 8. So, 18% of projects with external costs had differences during its lifecycle.

Table 9 - Project changes Analysis

Project Changes	**No**		**Yes**	
(1) Change Requests	1031	75.75%	330	24.25%
(2) Date Difference	1035	76.05%	326	23.95%
(3) Budget Difference	1206	88.61%	155	11.39%
(4) Ext Cost Difference	1323	97.21%	38	2.79%
Changes in project	**947**	**69.58%**	**414**	**30.42%**

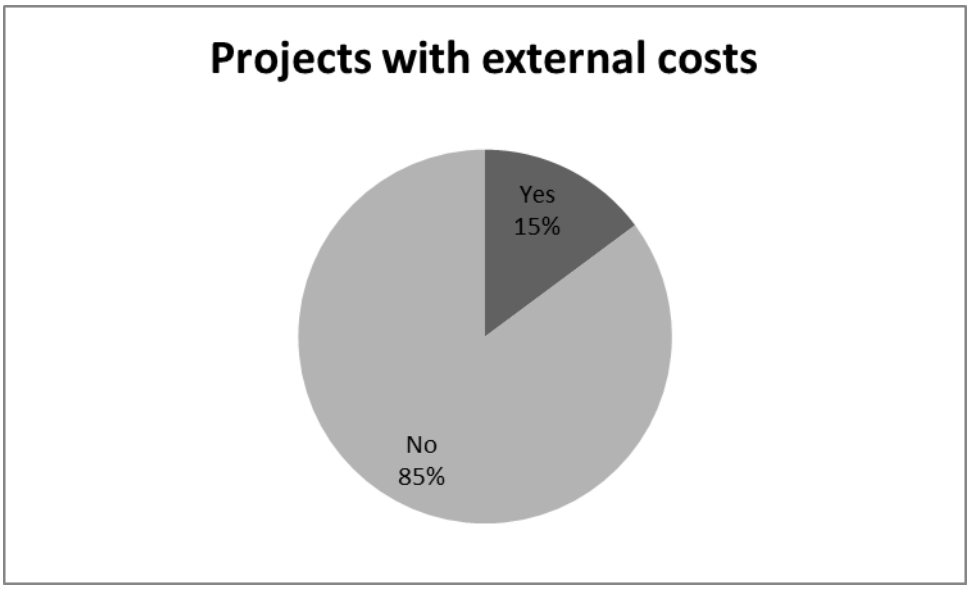

Figure 8 - Projects with external costs

4.1.2 **Project Type**

The institution's projects have two different types. Projects that are measured for internal metrics (e.g., the customer satisfaction grade, if the project has delayed, the number of errors identified in UAT testing, if the cost approved by the internal client

was not exceeded). These projects are controlled in time, scope and cost and are named by the institution as “Regular Projects”. There is other type of projects that are not measured after its conclusion. These are, for instance, urgent or critical projects, technological projects, experimental projects, or projects that for some reason do not follow the project management process. They are known as "Special Projects”, these projects are about 29% of all dataset projects as showed on figure 9. Special projects also have a final estimation and a real duration and it is possible to calculate them when they are not available (because project managers are not obligated to register this information in this type of projects). Although it is not mandatory for project managers to register initial estimation or project dates (i.e., save a baseline), it is possible to see their change requests, the times consumed by teams, its start, and the multiple production dates (when there are more than one phase of deployment). When project dates are not filled, the features before described are used to calculate duration, according to PMO’s organization definition.

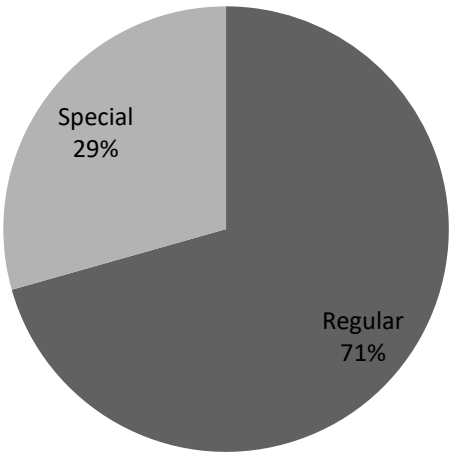

Figure 9 - Project type

When the relation between project type and project changes is drawn (figure 10), it is possible to observe that 37% of the regular projects had changes while only 14% of special projects had been subjected to changes.

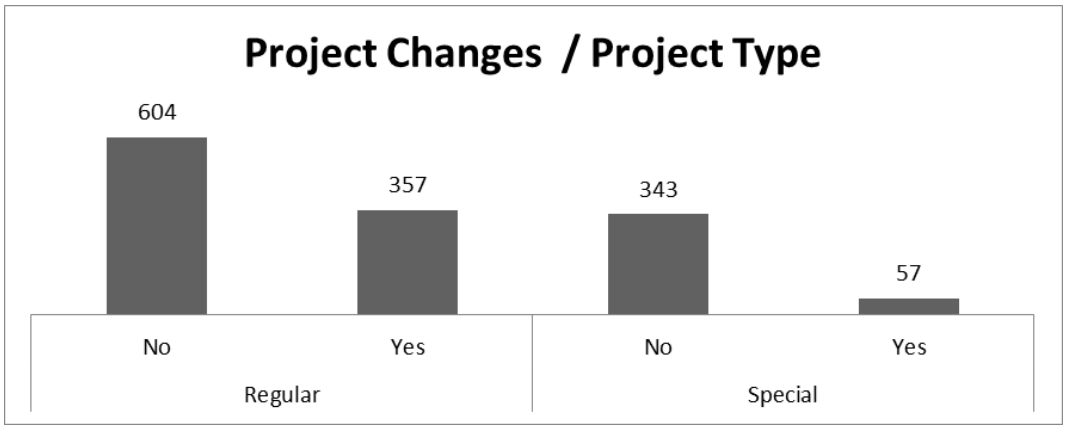

Figure 10 - Relation between Project Type and Project Changes

4.1.3 **Internal Client**

The "Internal Client" or "Internal Customer" defines the internal division or department that requested the Project as showed on table 10. Each department has a different organizational culture. By nature, some departments are more familiar with IT projects, since they have a greater dependence on IT. It is the case, for instance, of Channels and Internet related departments. Audit, compliance and rating departments are usually departments that require mandatory and legal projects, with strict deadlines. Marketing (it includes branches/commercial area requests) is the internal customer with more project requests. IT division is also client of itself. This occurs, for instance, in software technological renovations, housekeeping/maintenance projects and IT tools implementation. Departments with small project requests (less than 5) were aggregated on "others". On table 11 it is possible to understand that requests are not homogeneous between different clients.

Table 10 - Internal Customers

Internal Client	Number of projects
Retail Marketing	213
Channels and Internet	170
Operations	154
International	148
Accountant	117
IT	99
Corporate Marketing	63
Branch Support	60

Internal Client	Number of projects
Credit	54
Credit Recovery	50
Audit, Compliance and Rating	31
Risk	29
Financial Markets	28
Insurances	28
Reporting	21
Human Resources	19
Procurement and Resources	16
Segments	16
Real Estate	15
Private	13
Others	11
Communication	6

Table 11 - Internal Client and Project Changes Relation

Internal Client	Number of Projects	%
Retail Marketing	**213**	
No	142	67%
Yes	71	33%
Channels and Internet	**170**	
No	141	83%
Yes	29	17%
Operations	**154**	
No	90	58%
Yes	64	42%
International	**148**	
No	92	62%
Yes	56	38%
Accountant	**117**	
No	95	81%
Yes	22	19%
IT	**99**	
No	72	73%
Yes	27	27%
Corporate Marketing	**63**	
No	37	59%
Yes	26	41%

Internal Client	Number of Projects	%
Branch Support	**60**	
No	45	75%
Yes	15	25%
Credit	**54**	
No	33	61%
Yes	21	39%
CredITRecovery	**50**	
No	34	68%
Yes	16	32%
Audit, Compliance and Rating	**31**	
No	17	55%
Yes	14	45%
Risk	**29**	
No	23	79%
Yes	6	21%
Insurances	**28**	
No	27	96%
Yes	1	4%
Financial Markets	**28**	
No	20	71%
Yes	8	29%
Reporting	**21**	
No	15	71%
Yes	6	29%
Human Resources	**19**	
No	14	74%
Yes	5	26%
Segments	**16**	
No	15	94%
Yes	1	6%
Procurement and Resources	**16**	
No	9	56%
Yes	7	44%
Real Estate	**15**	
No	6	40%
Yes	9	60%
Private	**13**	
No	7	54%
Yes	6	46%

Internal Client	Number of Projects	%
Others	**11**	
No	8	73%
Yes	3	27%
Communication	**6**	
No	5	83%
Yes	1	17%
Grand Total	**1361**	

4.1.4 **Interposed projects**

Projects are usually planned twice a year in committees. In the first committee, an IT project plan (or portfolio) is created where projects are planned according IT capacity and resource availability. These projects are prioritized, and IT mission is to comply with this annual plan. On the second committee (when exists), the project portfolio can be reviewed and aligned, since business needs, requirements and strategies can change. Yet, some projects can appear in the middle of committees and because of this projects urgency in their implementation (time to market, legal or regulatory projects, as others), they cannot wait until next portfolio review to be prioritized and initiated. So, it is necessary to initiate these projects as soon as possible, impacting project portfolio plan baseline. These projects are called interposed projects and they affect project portfolio plan that will change after management acceptance. Almost 36% of the projects in the dataset are interposed projects (figure 11). Comparing interposed and not interposed projects, the relation between this feature and project changes is the same. There are 30% of projects with changes for interposed projects. The same number for portfolio projects as it is possible to verify in figure 12. There are 67 projects that are not identified as interposed or planned projects. 28 of these projects had changes and 39 did not have changes in their lifecycle and they were considered as “blanks”.

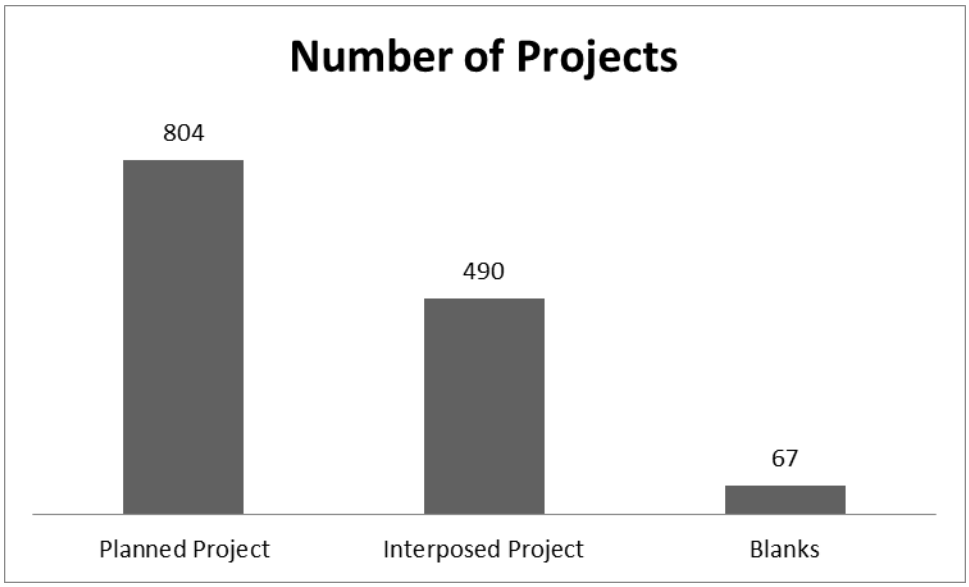

Figure 11 - Interposed projects

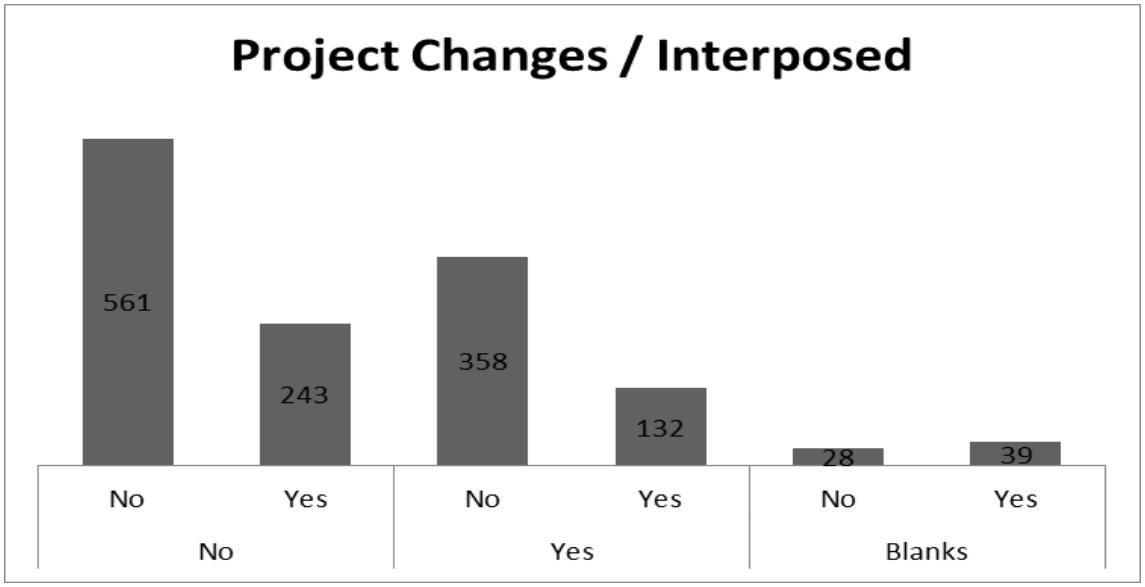

Figure 12 - Relation between interposed and Project Changes

4.1.5 **Decision and mandatory type**

Clients and Sponsors at project definition classify projects with different types. This classification groups projects in the main reason for their implementation. Projects can be accepted for implementation because they bring benefits and there is an associated profitable business case, they can be settled for implementation because they have at least one legal requirement, because they are urgent or extreme critical, because they have legal implications or respond to an audit request, or just because there are technological causes involved as referred on table 12. Almost half of projects made by

this institution are classified as mandatory projects (figure 13). At last there are projects that do not have quantitative benefits but they are, for instance, reputational projects. These have qualitative benefits and are known as “Qualitative” projects. Mandatory type variable for non-mandatory projects is always considered as “others” (Table 13). Table 12 show the division of project type. 67 projects weren’t classified and they were identified to this study as unknown in order to not impact other classifications.

Table 12 - Number of projects by decision type

Decision Type	Number of Projects
B – Benefits	399
M – Mandatory	633
Q – Qualitative	262
Unknown	67

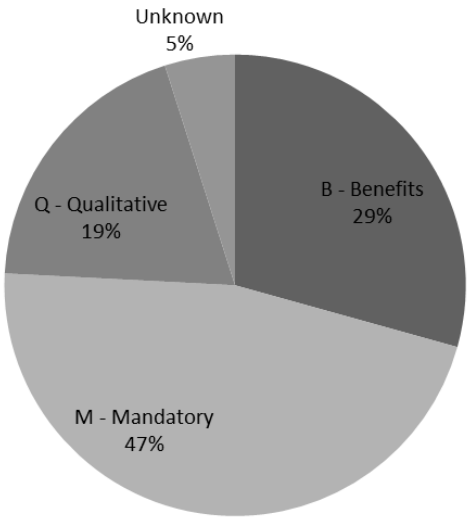

Figure 13 - Decision Type

Table 13 - Number of projects by Mandatory Type

Mandatory Type	Number of Projects
Ma - Mandatory Legal Requirement	126
Mb - Business Extreme Critical	39
Mc - Legal Necessity or Audit High Risk	84
Md - Audit Medium and Low	142
Mt - Technology Causes	62
OT – Others	841
Blanks	67

By relating decision type with project changes, it is possible to verify that this feature by itself doesn't have any visible effect. All Decision Type values have 30% of projects with changes as is shown on figure 14.

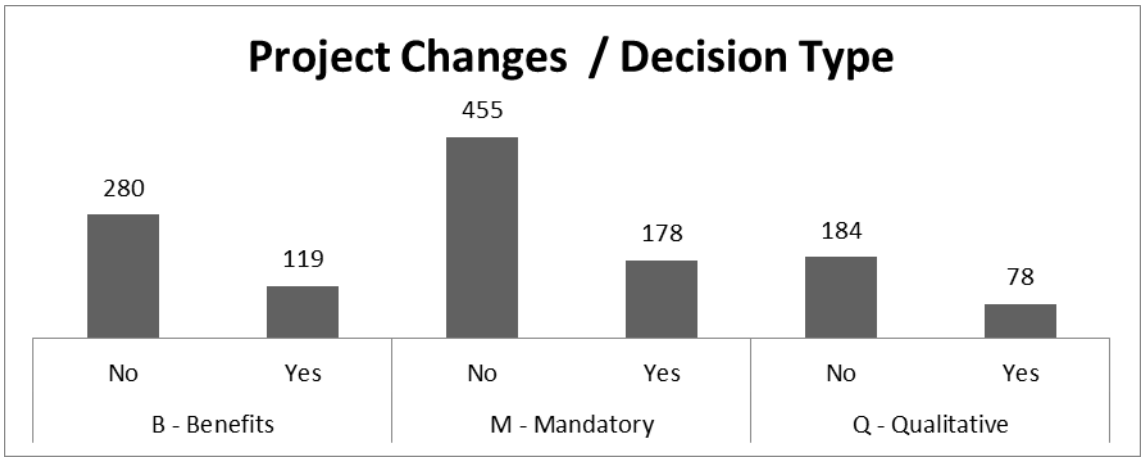

Figure 14 - Relation between Decision Type and Project Changes

For mandatory type and project type's relation showed in table 14, it is possible to understand that the values do not differ much. Projects that are classified as "business extreme critical" have a small percentage of project changes, although the number of this kind of projects is the lowest of all categories. Projects with categorization "Audit high risk non conformity" have the smallest percentage of projects with no project changes (40%).

Table 14 - Relation between Mandatory Type and Project Changes

Mandatory Type	Number of Projects	%
Ma - Mandatory Legal Requirement	**126**	
No	88	70%
Yes	38	30%
Mb - Business Extreme Critical	**39**	
No	30	77%
Yes	9	23%
Mc - Legal Necessity or AudITHigh Risk Non Conformity	**84**	
No	50	60%
Yes	34	40%
Md - AudIT Medium and Low Non Conformity or Administration Strategic Require	**142**	
No	107	75%
Yes	35	25%
Mt - Technology Causes	**62**	
No	45	73%
Yes	17	27%
OT - Others	**841**	
No	599	71%
Yes	242	29%

4.1.6 **Project Objective**

At projects' initial phase, internal clients fulfil project objective in order to identify the main reason behind project implementation. All project objectives are detailed in table 15. Since this is not a mandatory field, there are 319 projects that do not have objective classification inserted, and so, they are still waiting classification. There were 67 records not filled in (with no value). Since these are unknown, the "N/A" value is assumed.

Table 15 - Project Objectives

Project Objective	Number of Projects
0 - Waiting Classification	319
1 - Audit Committee	15
1 - Increase Income	125
1 - Management Control Information	58
2 - Cost Reduction	135
2 - Information for Customer	49
2 - Internal Recommendation	178
3 - Audit Recommendation	2
3 - Legal Requirement	257

Project Objective	Number of Projects
3 – Others	80
4 – Technology	76
Unknown (N/A)	67

4.1.7 **Project Size**

The studied financial institution considers small projects as the ones with estimation below 100 person-days of effort. These are 61% of all IT company projects. Projects between 100 and 250 person-days are considered medium-sized projects (24%). 10% are large sized projects (projects between 250 and 500 person-days of effort). Only 5% of projects are above 500 person-days as it can be seen in figure 15 (one person-day consists in 8 hours of work).

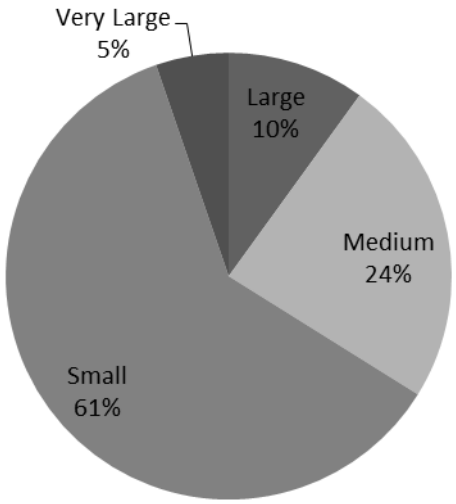

Figure 15 - Project Size

By attempting to establish a relation between project size and project changes, it can be concluded that as the estimation gets bigger, the percentage of projects with changes also grows. This confirms the statement that large sized projects have more delays and more tendencies to fail (Belassi and Tukel 1996). For large and very large projects there are more projects with changes than the ones that don't as is it showed on figure 16.

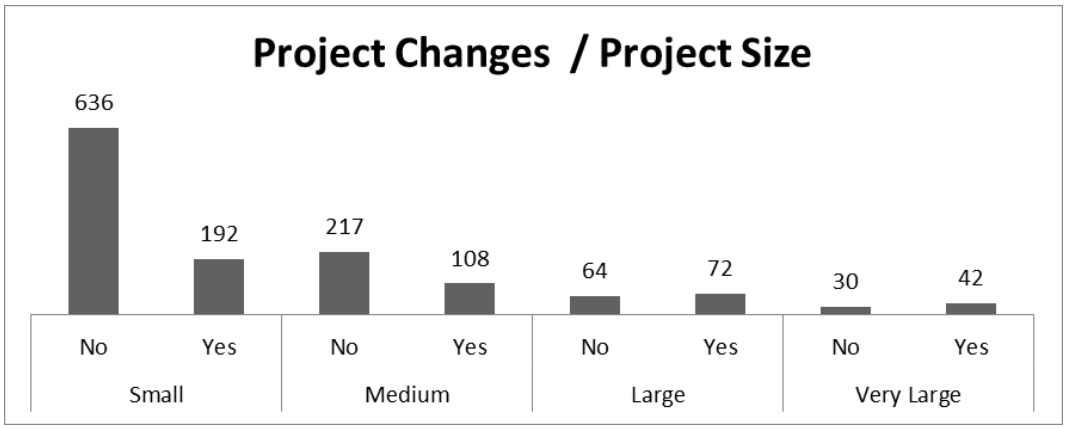

Figure 16 - Relation between project Size and project changes

4.1.8 **Duration Size**

The duration size variable was also grouped with the same rule than project size. The organization considers and defines small projects as the ones with duration below 100 days, medium projects with duration between 100 and 250 days, large projects as projects between 250 and 500 days of calendar, and very large projects with duration greater than 500 days.

The duration size is calculated with the rules above, with the number of calendar days between projects start and projects last date. Usually projects last date is bigger than last production date, since there is a warranty period for production support (usually 1 to 2 months). There are 55 projects in the dataset that do not have last production date filled, so it wasn't possible to calculate project duration. This projects assume "#N/A" in this field.

There is not a relation between project size and duration size (table 16), since duration size is the duration in days between the project start and the last production date. For instance, 26 small projects were finished in more than 500 calendar days. The duration size feature was not chosen for be a part of a model, since it is only known at project end, but is an important feature for data and business understanding. As it is possible to see in figure 17, 32% of projects are concluded below 100 days. 39% below 250 days, 18% below 500 days and only 7% are concluded above 16 months.

Table 16 - Project and Duration size relation

Very Large Duration	**90**
Small Project	26
Medium Project	15
Large Project	24
Very Large Project	25
Large Duration	**242**
Small Project	86
Medium Project	80
Large Project	55
Very Large Project	21
Medium Duration	**538**
Small	306
Medium	176
Large	45
Very Large	11
Small Duration	**436**
Small	377
Medium	45
Large	8
Very Large	6
#N/A	**55**
Small	33
Medium	9
Large	4
Very Large	9

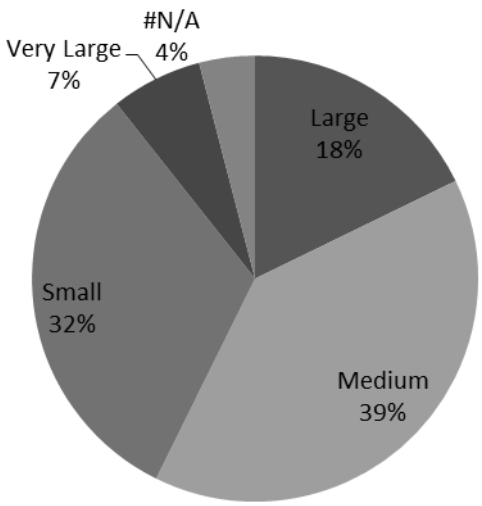

Figure 17 - Duration Size

Duration size also indicates that almost 50% of large and very large projects have project changes, when the difference for smaller duration projects is more significant. This can be seen on figure 18.

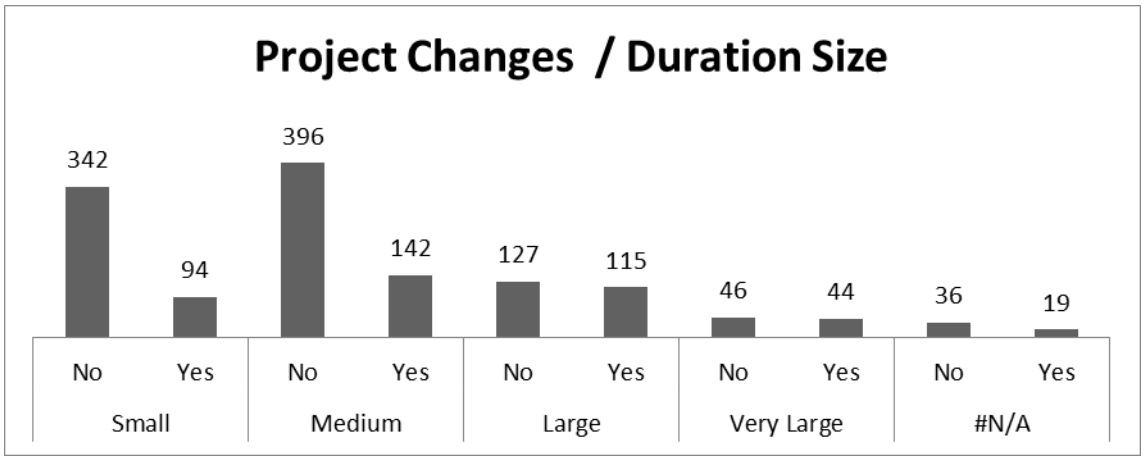

Figure 18 - Project changes by duration size

4.1.9 **Numeric Features**

Table 12 shows the statistical measures for the numeric variables. The project estimation is created by the different development teams involved in project with the support of the project manager. After a good understand of scope and requirements, and after the solution elaborated is approved with the client acceptance of functional

analysis document, estimation is created. Estimations include time effort for management (e.g., includes time for project management work, coordination, monitoring and task control), functional analysis elaboration (includes the time needed for business analysts to conceive the functional solution for the project client approve), organic analysis elaboration (e.g., it includes the time for table creation, metadata information, pseudo-coding specifications), development and unit testing (includes the measurement of the work of programmers that will conceive the solution), certification support (includes the effort for the client support in the certification environment) and production support (work measurement for supporting users in the first days after the project is deployed in live environment). The project estimation feature has a mean value of 844 person-hours (around 105 person-days of effort estimation). There are some projects that weren't initially estimated, and they are presented with zero hours of estimation. Real estimation value is the actual number of hours in the project (internal and outsourcing) at project closure time. The values for external costs cannot be disclosed due to confidentiality constraints.

The "number of applications" feature is registered by project manager at the beginning of the project, since those are identified during scope definition. Projects have at least one application, with most of the projects having 2 applications involved. The project with most applications had 41 applications involved.

The internal clients (organization's divisions or departments) usually are identified before the beginning of the project, but in some cases they can also be added after when one more client or stakeholder is involved at scope definition in the planning phase. These are the internal departments that sponsor the project. There is always a main client (from where the formal sponsor is identified). Yet, some projects can have more than one internal client. From table 17, it is possible to see that most of the projects have one internal client, although it varies from 1 to 9 different clients.

The outsourcing teams are identified at the time of project plan and project estimation. Teams that are considered for this study, are teams that usually maintain applications and resources are used as project team members, outside of "out of the shelf" type of procurement. These teams also register the hours spent in the project management tool for control. The number of outsourcing teams in the project is in median one team per project. Estimation effort for outsourcing teams rounds the 341 hours. It is not possible to understand by the data, how many outsourcer resources were

involved in the project. It is only possible to understand the number of different outsourcing teams. Internal team members are the number of internal members that are involved in the project. Internal members are all IT development employees from the institution. The median is 4 individuals per project.

The project duration rounds 191 days (6 months), although the estimated duration is 3.6 months. Number of Change requests (that formalizes one project change) is 0.37 per project.

"Estimated internal effort" is bigger than the "actual internal effort". Yet, the numbers between estimated and real outsourcing effort aren't so different from each other. Comparing first project estimation and real estimation the numbers also do not differ much.

For model purposes, only the features that are known at project planning time were assumed, since is at this phase that the study tries to predict if there will be changes in the project.

Table 17 - Numeric Variables - Statistics

Variable	Min	Median	Mean	Max	Model Inclusion?
First Estimation	0	440	843.9	37737	N
First External Costs	*	*	*	*	N
Real_Estimation	0	444	920.4	59539.1	N
Number of Applications	1	2	3,88	41	Y
Number of internal Clients	1	1	1.24	9	Y
Total_Estimated_Effort	0	580	1197	37737	Y
Internal_Estimated_Effort	0	160	416.5	13208	Y
Outsourcing_Estimated_Effort	0	28	341	20493	Y
Actual_Internal_Effort	0	319	709.8	22185	N
Actual_Outsourcing_Effort	0	88	376.5	20493	N
Number of outsoucing teams	0	1	1.66	25	Y
Number of Internal Team members	1	4	6.48	113	Y
Estimated Project Duration (days)	0	81	109.3	977	Y
Real Project Duration (days)	0	137	191.1	1480	N
Number of change requests	0	0	0.37	9	N

4.2.Modelling Analysis

The dataset contains 1361 records. Each record belongs to a different project from the institution. Only projects concluded in the last three years were extracted to dataset. There are 36 initial variables. One is the dependent variable – "Changes in projects" that can assume the values "yes" or "no". Since the dependent variable categorical, this is a classification problem. From the other 35 variables, it was chosen the ones that are known at project plan phase, i.e., until project management plan I sent to customer approval. There are 15 independent variables that were chosen for modelling analysis, and they can be identified on table 18.

Table 18 - Dataset variables for model

ID	Name	Type	Short Description
1	Project_Type	Categorical	Project Type – Regular or Special
2	Internal_client	Categorical	Project Sponsor's name of department
3	Interposed	Categorical	Identifies if the project was planned or if it is an interposed project
4	Decision_Type	Categorical	Type of decision that took place to defend the project acceptance
5	Mandatory_Type	Categorical	When a project is mandatory, the reason for being it
6	Objective	Categorical	Client perception for moving forward with project
7	Number_of_Applications	Integer	Number of different applications involved in the project
8	Number_of_Internal_Clients	Integer	Number of different customer departments involved in the project
9	Total_Estimated_Effort	Integer	Total of effort estimated in the project
10	Internal_Estimated_Effort	Integer	Internal effort estimated before project execution
11	Outsourcing_Estimated_Effort	Integer	Outsourcing effort estimated before project execution
12	Number_of_Outsourcing_Teams	Integer	Number of outsourcing teams involved on the project
13	Number_of_internal_resources	Integer	Number of internal development resources assigned to the project.
14	Estimated_Duration	Integer	Duration of the project estimated before project execution
15	Project_Size	Categorical	Project Size according institution point of view
16	Changes_in_Project	Categorical	Identifies if there was changes in project

For the modelling analysis, there were seven techniques that were adopted, in order to understand which one is the best that fits the data: (1) Logistic Regression (LR); (2, 3) Decision Trees (RPART and DT); (4, 5) Neural Networks (MLPE and MLP); (6) Support Vector Machines (SVM), (7) Random Forest (RFOREST).

All models were run 20 times for the performance method that was chosen (K-fold with K= 10). AUC and ALIFT metrics were calculated for each model with the average of the 20 runs. The results are showed in table 19 and it is possible to understand that Random Forest has the best metric for Area under the curve for AUC (0.7747) and also for ALIFT (0.6912).

This area under the curve value can be accepted as a good value since there are studies with interesting conclusions supported by values of AUC below the one achieved in this study (e.g., Baecke and Bocca 2017, who achieved 0.6135). By analysing the ROC curve in figure 19 for several models, it is also possible to understand that Random Forest has the best values.

As such, Random Forest was chosen as the best model for extracting knowledge from it, in order to understand each feature's contribution.

Table 19 - Comparison between the DM models (in bold the best value)

	AUC	**ALIFT**
	K-Fold	**K-Fold**
LR	0.731165	0.660826
RPART	0.687260	0.630179
DT	0.685774	0.629188
MLPE	0.726803	0.657768
MLP	0.678670	0.624372
SVM	0.750326	0.674174
RFOREST	**0.774745**	**0.691157**

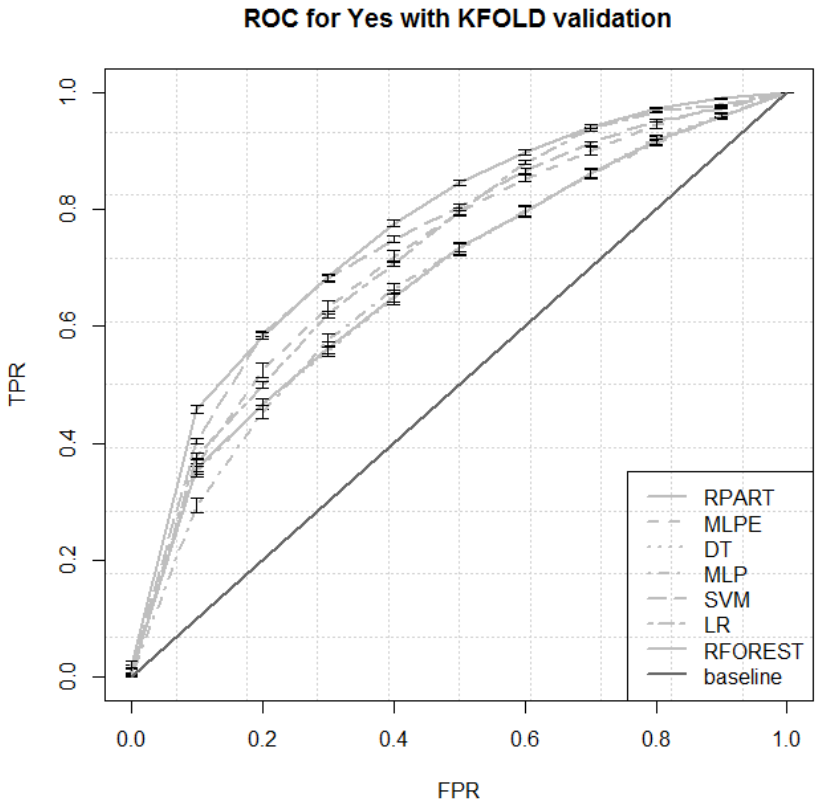

Figure 19 - ROC with KFOLD validation

In order to identify information in model extraction, DSA was used as showed in figure 20. DSA provides means that makes possible to show the percentage of importance that each variable has in the model (in this case Random Forests). DSA computation is based on a random sample selection, so the process was executed 20 times, showing the relevance of the average of these 20 executions for each variable. This approach was already use in other studies (e.g., Moro *et.al* 2017). The first 6 variables are the most relevant ones with a total relevance 73.86 % of the model.

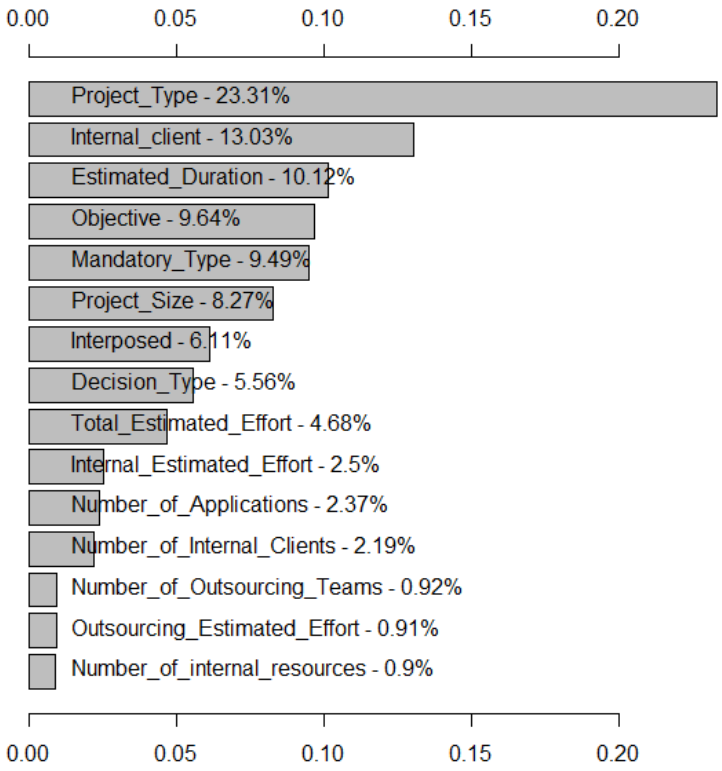

Figure 20 - DSA results

4.2.1 Project Type

The most relevant variable is the project type and it has 23.31% of relative importance in modelling project changes. The data importance on the VEC plot (figure 21) demonstrates project changes in project by each feature type. It is drawn that special projects only have 25% of possibility to have project changes while regular projects have near 40%. As already explained, special projects are projects that are considered important and critical for the institution. These projects usually are interposed and affect project portfolio plan, affecting also other projects that can be already in course. Since these projects are critical projects, and urgency affects project success (Belassi and Tukel 1996), project managers attempt to achieve dates and cost proposed at all cost. Important projects are also sponsored at a high-level management (top management) and have more absence of bureaucracy. This are pointed as success factors for project success (Millhollan and Kaarst-Brown 2016).

Other factor that can explain this difference between regular and special projects is that special projects have metric project management process exemption. For this

type of projects, project manager is not obligated to do and report change requests or to indicate dates and estimations at planning phase. This factor also reduces bureaucracy and politics that are presented as project success factors (Millhollan and Kaarst-Brown 2016).

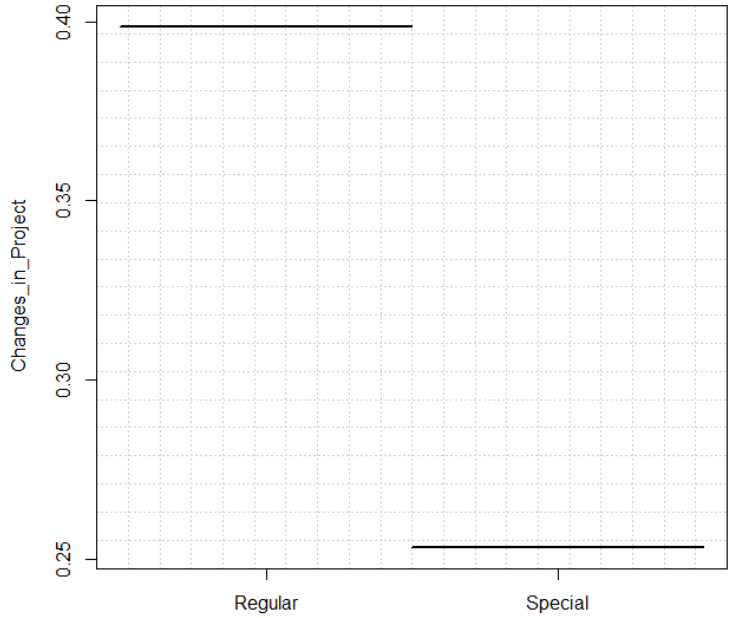

Figure 21 - Project Type VEC plot

4.2.2 Internal Client

The “Internal Client” feature is the second most relevant feature in the model. As is possible to understand from figure 22, the changes vary almost 25pp from the client with less project changes (Channels and Internet) to the ones with most projects with changes (Real Estate and IT departments). Client and project teams that support the most applications from this client projects can be relevant factors for project success (Hyvari 2006b; Meredith and Mantel 2009). Some departments are more projects oriented. It is the case of Channels and Internet department. This department needs to have continuous improvement on applications, since they are mostly end-user applications (e.g., internet sites and mobile applications). By other hand, the next two departments, accountant and risk have several mandatory projects that are also a factor for projects to be finished in scope, time and cost (Meredith and Mantel 2009).

Real Estate department has almost 50% of projects with changes, the same as IT when is his own internal Client. This can be a justification by itself. IT projects are

postponed and delayed when is necessary to respond to business needs that are considered of higher importance than internal projects (Chang 2003).

Different stakeholders have different ways to see project achievement so they have different types of satisfaction with end product deliverable (Lock 2007). This can be another reason for department differences in project changes.

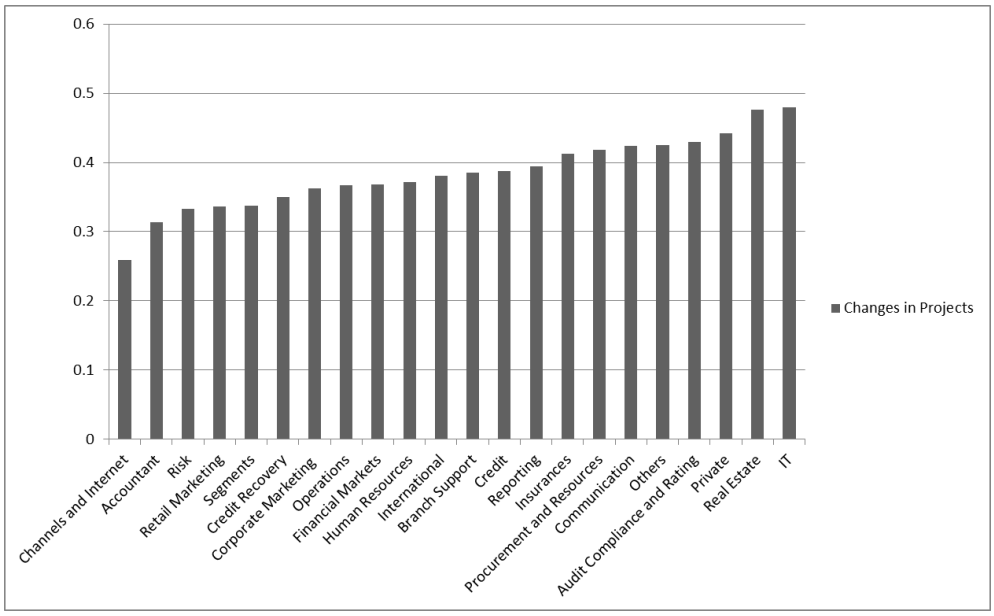

Figure 22 - Internal Client VEC plot

4.2.3 Estimated Duration

Estimated Duration has 10.12% of importance in the model. It is possible to conclude, according to figure 23, that projects with small duration have less project changes than others. Projects until 160 days are below 35 pp. Projects below 245 days are below 40 pp, when larger estimated duration projects are above. A longer duration project has more chances to fail in time, cost or scope than shorter projects (Cooke-Davies 2002).

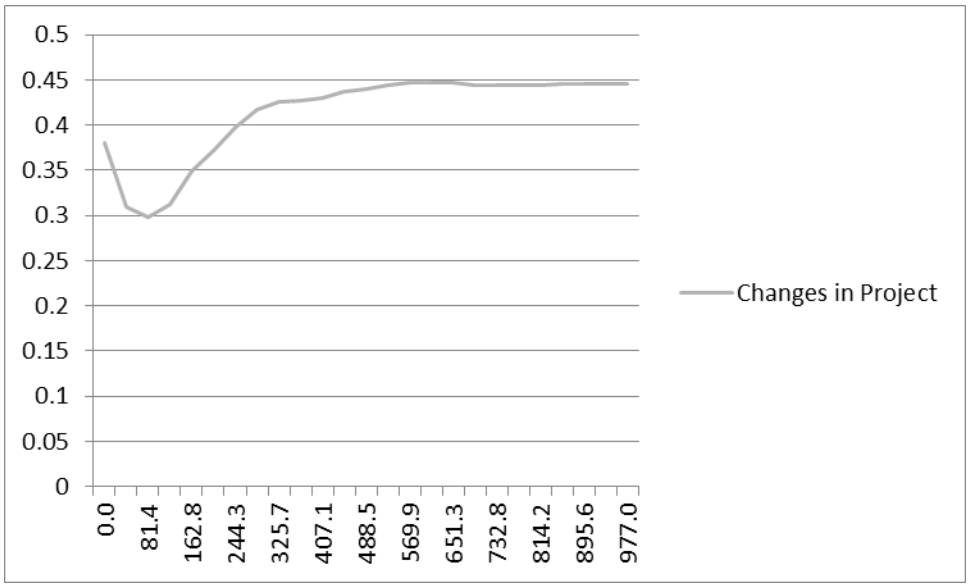

Figure 23 - Estimated duration VEC plot

4.2.4 Project Objective and Mandatory Type

Project Objective and mandatory type features are the fourth and the fifth more important features of the model. Figure 24 shows that projects associated to legal requirements implementation have less project changes than others. This makes sense since projects that have an external mandate time to be implemented have a greater sponsorship to be implemented at the proper time (Meredith and Mantel 2009).

On the other hand, there are some values in the project objectives that were not possible to take a conclusion from them. For instance, there are concluded projects that never were classified (they are still waiting internal client's classification). Others, do not have any value fulfilled ("N/A" was assumed for this cases). It is possible to see that "Audit recommendations" are more proper to fail scope, time and/or costs than projects with "internal recommendations" or "technological purposes" as objective. Business purposes (Increase income, reporting, or others) are closely aligned with the average of project changes (35pp). The values obtained for project objectives, could not be confirmed by literature.

The mandatory types should have some connection with the project objectives. Yet, there isn't none as it is possible to conclude from table 20 values. Therefore, for mandatory type, it is not possible to understand why the best values are for the "medium audit non conformities" projects when the "technological causes" projects have the worst value (figure 25), when "technological project" objective has a good value. Values "Others" and N/A were not possible to identify since they make part of every objective.

Other relevant factor is the number of project mandatory types. The “Other” classification has much more values than remain classification values, as showed on figure 26.

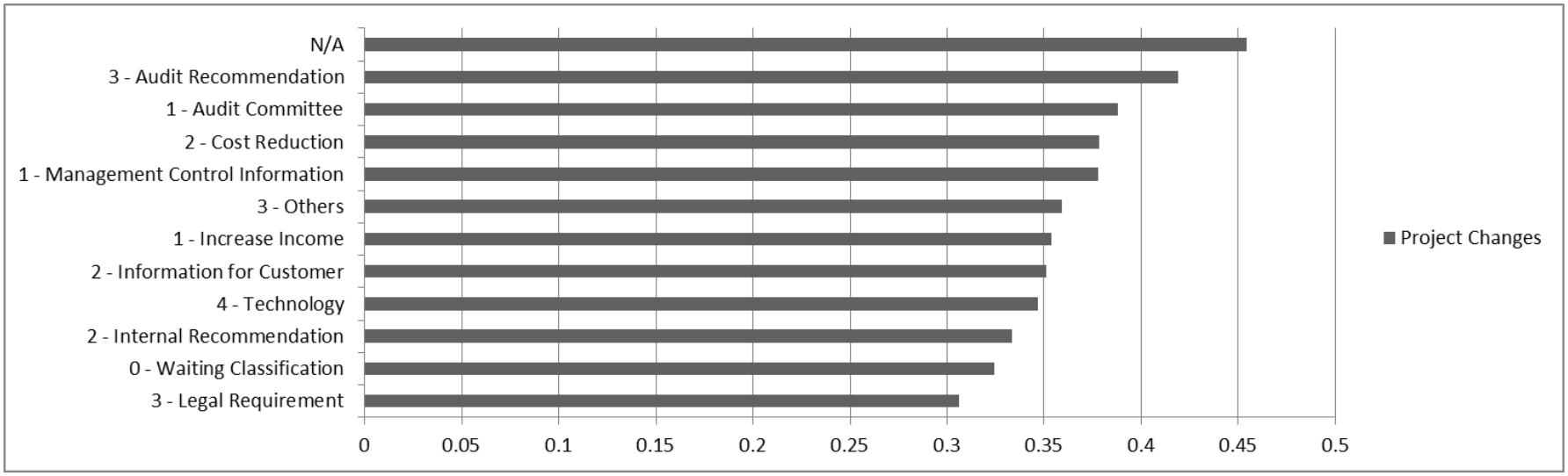

Figure 24 - Objective Importance VEC plot

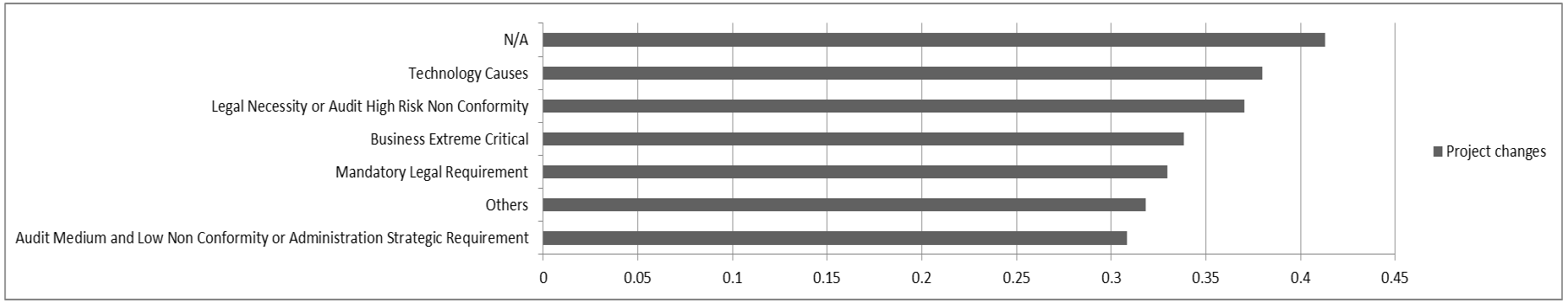

Figure 25 - Mandatory Type Importance VEC Plot

Table 20 - Relation between Objectives and Mandatory Types

	Ma	Mb	Mc	Md	Mt	OT	NA
0 - Waiting Classification	4	9	8	27	3	268	
1 - Audit Committee	1		8	2		4	
1 - Increase Income	2	8			4	111	
1 - Management Control Information				2	2	54	
2 - Cost Reduction	1			1	18	115	
2 - Information for Customer		6				43	
2 - Internal Recommendation	2	7	20	97	8	44	
3 - Audit Recommendation						2	
3 - Legal Requirement	114	5	47	9		82	
3 - Others			1		1	78	
4 - Technology	2	4		4	26	40	
NA							67

Legend:

Ma - Mandatory Legal Requirement

Mb - Business Extreme Critical

Mc - Legal Necessity or Audit High Risk Non Conformity

Md - Audit Medium and Low Non Conformity or Administration Strategic Requirement

Mt - Technology Causes

OT - Others

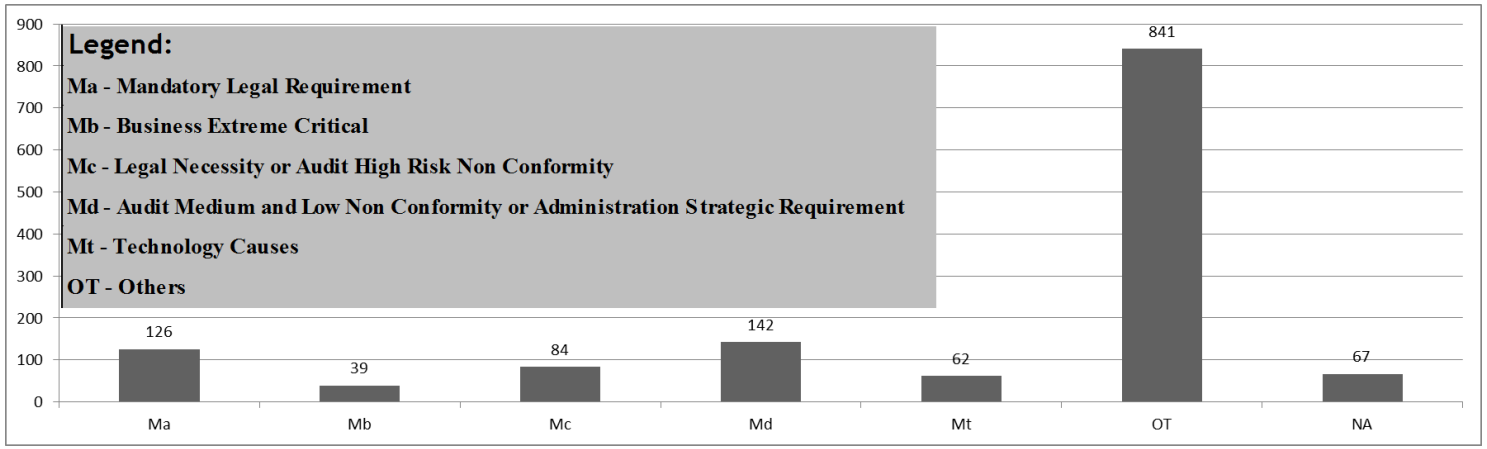

Figure 26 - Number of projects by Mandatory Type

4.2.5 Project Size

The “project size” feature has some relevance in project changes. The larger is the project, more is the project change relevance (as showed in figure 27). Project size is one of the most important factors for project success or failure (Dvir *et al.* 1998), and many large projects exceed their deadlines (Belassi and Tukel 1996). Projects should be kept below one year (Cooke-Davies 2002).

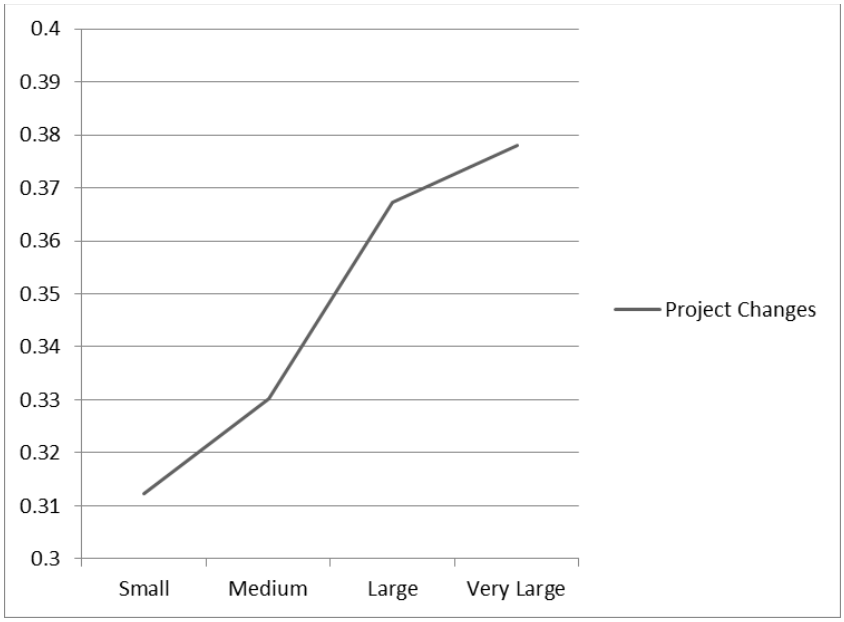

Figure 27 - Project Size importance VEC plot

4.2.6 Analysis Resume

"Project type" is the most relevant feature from the analysis. Special projects had fewer projects with changes when compared to regular and planned type of projects. Special projects are usually required to be implemented urgently, and they need to be deployed in a defined or imposed time with fixed and strict milestones. Also, most of Special projects are interposed projects. Interposed projects change "on going" (and regular) project plan. Mandatory projects are also a factor to be considered. Business extremely critical projects and audit projects have fewer changes than others, as confirmed by Dvir *et al*. (1998) and Belassi and Tukel (1996) that refer that project urgency affects project success. In this case, these projects, that are considered urgent and important to the institution, have a different type of management and control that explains the changes in projects Martinsuo and Lehtonen (2007). The internal client also influences project changes tendency. This is related with the department's culture, that also affects the requirement definition, the rigor in project certification and in project acceptance, and making some departments more "projects oriented" than others (Meredith and Mantel 2009). Another finding is related with estimation and duration. Small projects have fewer changes than projects with a bigger dimension. This reason is related with the fact that in smaller projects, it is easier to project manager to have more control in scope and smaller projects are better to manage and achieve project success, or, in this case, not having changes during the project lifecycle (Ayal and Globerson 2002; Cooke-Davies 2002).

Aggregating features and its importance by nature, as showed on table 21, it is possible to verify that features that are external to project management have the most importance on project changes (with 67.14%) than the ones that depends of project manager and project team. This means that more than 67% of relevant factors that affects project changes in the institution do not depend of project management and are not controlled by project manager. Cost and plan have 16.36% and 10.12% of relevance in changes, when the project complexity (number of persons involved) only has 6.38% of importance.

Table 21 - Feature importance aggregation

Feature	Importance
Estimation - Cost	**16.36%**
Total Estimated Effort	4.68%
Internal Estimated Effort	2.50%
Outsourcing Estimated Effort	0.91%
Project Size	8.27%
Duration - Plan	**10.12%**
Estimated Duration	10.12%
Project complexity- Stakeholder and HR management	**6.38%**
Number of Outsourcing teams	0.92%
Number of Applications	2.37%
Number of Internal Clients	2.19%
Number of Internal Resources	0.90%
Project features - external	**67.14%**
Project Type	23.31%
Objective	9.64%
Mandatory Type	9.49%
Decision Type	5.56%
Interposed	6.11%
Internal Client	13.03%
Total	**100.00%**

5 Conclusions

5.1. Main Conclusions

The project characteristics that do not depend from project management have a strong weight in the influence of project's management success (in not having project changes across project lifecycle). Projects considered by the organization as "special projects" have less tendencies to have project changes when compared to regular projects. Special projects usually are considered critical projects for the institution, and regular projects can be delayed if necessary for implementing the special ones on time and on defined scope. These projects usually do not have a well-defined scope, but requesters show a high urgency level in having them implemented. For this reason, special projects are exempted from metrics and project managers aren't obligated to insert estimations and project plan at planning phase in order to speed-up the implementation process. Usually, special projects' information only is inserted at the end of project, and that can clear the trace of project changes that may have occurred.

Organization's departments and divisions that assume the role of project clients are also a relevant factor for changes that occur in projects. Some departments are more "project oriented" than others, having more facility in the requirement identification (presenting them in a more detailed manner) (Meredith and Mantel 2009). This factor, leads projects toward success, since their scope is better defined at the beginning of the project. Each department's culture may also affect projects involvement and change requirements across the project execution (Meredith and Mantel 2009). Although these variables do not depend of project manager, doesn't mean that they cannot be mitigated. Knowing these indicators, a project manager can assume it like risk entries and apply project reserves in order to mitigate and reduce project changes. A different approach to each department according to its culture, it is also a known success factor to the project's success (Martinsuo and Lehtonen 2007).

As projects get bigger, they are also more difficult to manage. The scope is sparser and the number of applications also increases. This implies that project development teams are bigger (internal or external). With a large number of applications, the number of stakeholder's and internal clients also increases, that makes more difficult to manage projects. Simple projects, with a low estimation and a reduced number of clients are more suitable to finish on time and on budget (Cooke-Davies 2002; Uluocak, 2013; Lock 2017).

5.2.Contributes for academic and business communities

5.2.1. Academic contribution

It was possible to prove that for this specific institution, initial variables can be valuable predictors of project changes and enable to identify its impact on projects. Project size and estimation size have a direct correlation with changes in time, cost and scope. The number of teams, the resources and the stakeholder's culture has also a direct relation with changes on project baseline. External project factors, that project manager do not control, can be identified as variables that can affect the project management's success.

Data mining models can be used to study and identify project management indicators that can help project management, changing the classical risk management and experience-based estimations. The use of data mining and modulation techniques to analyse project changes in financial institutions has no precedent in current state-of-the-art project management literature, and there are still very few studies of data mining applied to project management.

5.2.2. Business recommendations

It is strongly recommended that institutions continue registering project information. For the studied institution, special projects should disappear or project managers should be obligated to insert project dates and estimations at project plan phase as they do for regular projects, even being exempt of projects metrics and measures. This measure can accurate the data quality and improve, in the future, the knowledge extraction of initial features (variables) that affects project changes.

The features "project objective", "project mandatory type" and "project decision type" are important variables for model knowledge extraction. Yet, their quality is not very good, since it is possible to see that in some cases they do not make sense. It is proposed a stricter and controlled input in this features' insertion by the institution. If possible, the projects already closed should be reviewed, since there are very projects in this stage that are still waiting for classification.

When possible, large projects should be divided to smaller projects and implemented in phases. Smaller projects are easier to control and more accurate in scope plan and time identification. Large projects can be defined as programs and divided into smaller

projects. Scope add-ons can be related with another project (another phase of the program).

Data mining models can be implemented in the institution, helping project managers and PMO in making risk management, in estimation predictions and in portfolio definition. Project reserves and lags can be predicted with a regression model that can help to better identify how features that usually are not taken in consideration affects project management. Institution's PMO can also take advantage from knowledge extraction from predictive models in order to create a more accurate project portfolio plan and to manage impacts when interposed projects affect the portfolio baseline.

5.3.Limitations

The study was made with one specific European financial institution, so its conclusions cannot be generalized for other institutions. Also, the scope of this study was on IT development projects, and cannot be assumed for other kind of projects in banking (e.g., Marketing projects, product implementation projects, rebranding, among others).

Although the data that was analysed and used in this work shows good data quality and consistence, there are some variables that can be improved when they are being filled in. For instance, there are project objectives, project mandatory types and project decision types that are still waiting for classification in projects that are already closed. For special projects, dates and estimations sometimes aren't filled in, since it is not mandatory its fulfilment to this kind of projects.

5.4.Future research opportunities

The study was conducted with data from one specific financial institution. It can be relevant to make the same study in other financial companies, obtain results and compare them. This will validate and compare initial features' importance between different institutions. Also, after a 3-year period it is relevant to know if the results of this institution are similar to the actual ones, or if initial features differ from the study results and, if so, try to understand what changed in this period that explain the new results.

[This page was intentionally left blank]

References

AERTS, A.T.M., GOOSSENAERTS, Jan B.M., HAMMER, Dieter K., and WORTMANN, Johan C., 2004, Architectures in context: on the evolution of business, application software, and ICT platform architectures. *Information & Management*. 2004. Vol. 41, no. 6, p. 781-794.

ANYASI, F. I., and OTUBU, P. A., 2009, Mobile phone technology in banking system: Its economic effect. *Research Journal of Information Technology*. 2009. Vol. 1, no.1, p. 1-5.

AYAL, Moshe and GLOBERSON, Shlomo, 2002, Forecasting project duration as a function of scope changes and delay. *Project Management Institute Annual Seminars & Symposium, San Antonio, TX. Newtown Square*. 2002. PA: Project Management Institute.

AYAL, Moshe, 2005. *Effect of Scope Changes on Project Duration Extensions*. Tel Aviv University.

BAECKE, Philippe and BOCCA, Lorenzo, 2017, The value of vehicle telematics data in insurance risk selection processes. *Decision Support Systems*. 2017. Vol. 98, p. 69-79. DOI 10.1016/j.dss.2017.04.009.

BELASSI, Walid and TUKEL, Oya Iemeli, 1996, A new framework for determining critical success/failure factors in projects. *International Journal of Project Management*. 1996. Vol. 14, no. 3, p. 141-151. DOI 10.1016/0263-7863(95)00064-x.

BENGIO, Yoshua, and GRANDVALET, Yves, 2004. No unbiased estimator of the variance of k-fold cross-validation. Journal of machine learning research. 2004. P.1089-1105.

BHARADWAJ, Anandhi S., 2000, A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*. 2000. Vol. 24, no. 1, p. 169. DOI 10.2307/3250983. JSTOR

BUEHLER, Roger, GRIFFIN, Dale and ROSS, Michael, 1994, Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*. 1994. Vol. 67, no. 3, p. 366-381. DOI 10.1037//0022-3514.67.3.366. American Psychological Association (APA)

CALDERARO, Andrea, 2015, Book Review: Big Data: A Revolution That Will Transform How We Live, Work, and Think Mayer-SchönbergerViktorCukierKenneth, Big Data: A Revolution That Will Transform How We Live, Work, and Think. Boston, MA: Houghton Mifflin Harcourt, 2013. *Media, Culture & Society*. 2015. Vol. 37, no. 7, p. 1113-1115. DOI 10.1177/0163443715596318b.

CHANG, Yoonhee Tina, 2003, Dynamics of banking technology adoption: an application to internet banking, *Royal Economic Society Annual Conference*. 2003. No. 41

CHAPMAN, Pete, CLINTON, Julian, KERBER, Randy, KHABAZA, Thomas, REINARTZ, Thomas, SHEARER, Colin, and WIRTH, Rudiger, 2000, *CRISP-DM 1.0: step-by-step data mining guide*. SPSS.

CLARKE, Angela, 1999, A practical use of key success factors to improve the effectiveness of project management. *International Journal of Project Management*. 1999. Vol. 17, no. 3, p. 139-145. DOI 10.1016/s0263-7863(98)00031-3.

COOKE-DAVIES, Terry, 2002, The "real" success factors on projects. *International Journal of Project Management*. 2002. Vol. 20, no. 3, p. 185-190. DOI 10.1016/s0263-7863(01)00067-9.

COOMBS, Crispin R., 2015, When planned IS/IT project benefits are not realized: a study of inhibitors and facilitators to benefits realization. *International Journal of Project Management*. 2015. Vol. 33, no. 2, p. 363-379. DOI 10.1016/j.ijproman.2014.06.012.

CORTEZ, Paulo and EMBRECHTS, Mark J., 2013, Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*. 2013. Vol. 225, p. 1-17. DOI 10.1016/j.ins.2012.10.039.

DAVIS, Gordon, EIN-DOR, Phillip, KING, William R., and TORKZADEH, Reza, 2006, IT offshoring: History, prospects and challenges. *Journal of the Association for Information Systems*. 2006. Vol. 7, p. 770–796.

DVIR, Dov, LIPOVETSKY, Stan, SHENHAR, Aaron, and TISHLER, Asher, 1998, In search of project classification: a nonuniversal approach to project success factors. *Research Policy.*1998. Vol. 27, no. 9, p. 915–935

EBBESEN, Jonas Balshoej and HOPE, Alex, 2013, Re-imagining the iron triangle: embedding sustainability into project constraints. *PM World Journal*. 2013. Vol. 2, no. 3.

EBERLE, Alexander, MEYER, Helga, and ROSEN, Drem, 2011. A comparison of PMI and IPMA approaches. Analysis to support the project management standard and certification system selection. *Aktuel Projekt Management*. 2011. Vol. 4, p. 31-34.

ERTEK, Gurdal, TUNC, Murat Mustafa, ZHANG, Allan Nengsheng, TANRIKULU, Omer, and ASIAN, Sobhan, 2017, Data Mining of Project Management Data: An Analysis of Applied Research Studies. *In Proceedings of the 2017 International Conference on Information Technology*. 2017. P. 35-41. ACM.

GADDIS, Paul O., 1959. *The project manager*. Boston, MA : Harvard University.

GARRETT, Dave, 2008, Is the Triple Constraint the WRONG way to Define Success? *Project Management 2.0 Blog*. 2008.

HÄLLGREN, M. and MAANINEN-OLSSON, E., 2009. Deviations, ambiguity and uncertainty in a project-intensive organization. *Project Management Journal*. 2009. Vol. 36, no. 3, p. 17-26.

HARIED, Peter and RAMAMURTHY, K., 2009. Evaluating the Success in International Sourcing of Information Technology Projects: The Need for a Relational Client-Vendor Approach. *Project Management Journal*. 2009. Vol. 40, no. 3, p. 56-71. DOI 10.1002/pmj.20113.

HASTIE, Trevor, TIBSHIRANI, Robert and FRIEDMAN, J. H., 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. New York : Springer.

HYVÄRI, Irja, 2006a. Project management effectiveness in project-oriented business organizations. *International Journal of Project Management*. 2006. Vol. 24, no. 3, p. 216-225. DOI 10.1016/j.ijproman.2005.09.001.

HYVÄRI, Irja, 2006b. Success of Projects in Different Organizational Conditions. *Project Management Journal*. 2006. Vol. 37, no. 4, p. 31-41. DOI 10.1177/875697280603700404.

JAMES, Gareth, WITTEN, Daniela, HASTIE, Trevor and TIBSHIRANI, Robert, 2013. *An introduction to statistical learning*. New York: Springer.

JOSEPH, Mathew, MCCLURE, Cindy and JOSEPH, Beatriz, 1999, Service quality in the banking sector: the impact of technology on service delivery. *International Journal of Bank Marketing*. 1999. Vol. 17, no. 4, p. 182-193. DOI 10.1108/02652329910278879. Emerald.

JOSLIN, Robert and MÜLLER, Ralf, 2015, Relationships between a project management methodology and project success in different project governance contexts. *International Journal of Project Management*. 2015. Vol. 33, no. 6, p. 1377-1392. DOI 10.1016/j.ijproman.2015.03.005.

KERZNER, Harold, 2003. *Project management workbook to accompany Project management - a systems approach to planning, scheduling and controlling, eighth edition*. New York : Wiley and Sons.

KING, William R. and MALHOTRA, Yogesh, 2000, Developing a framework for analyzing IS sourcing. *Information & Management*. 2000. Vol. 37, no. 6, p. 323-334. DOI 10.1016/s0378-7206(00)00046-x.

KOPMANN, Julian, KOCK, Alexander, KILLEN, Catherine P., AND GEMÜNDEN, H. G., 2014, Business case control: The key to project portfolio success or merely a matter of form? *Paper presented at EURAM European Academy of Management Conference, 4–7 June*. 2014. Valencia, Spain.

KUEN, Chan Wai and ZAILANI, Suhaiza, 2008, Critical factors influencing the project success amongst manufacturing companies in Malaysia. *African Journal of Business Management*. 2008. Vol. 3, no. 1, p. 16–27.

LEE, J.-N. and KIM, Y.-G., 2005, Understanding Outsourcing Partnership: A Comparison of Three Theoretical Perspectives. *IEEE Transactions on Engineering Management*. 2005. Vol. 52, no. 1, p. 43-58. DOI 10.1109/tem.2004.839958. Institute of Electrical and Electronics Engineers (IEEE).

LEGRIS, PAUL and COLLERETTE, PIERRE, 2006, A Roadmap for it Project Implementation: Integrating Stakeholders and Change Management Issues. *Project Management Journal*. 2006. Vol. 37, no. 5, p. 64-75. DOI 10.1177/875697280603700507.

LOCK, Dennis, 2007. *Project Management Ninth edition*. 2007. UK: Gower Publishing.

MARTINSUO, Miia and KILLEN, Catherine P., 2014, Value Management in Project Portfolios: Identifying and Assessing Strategic Value. *Project Management Journal*. 2014. Vol. 45, no. 5p. 56–70. DOI 10.1002/pmj.21452.

MARTINSUO, Miia and LEHTONEN, Päivi, 2007, Role of single-project management in achieving portfolio management efficiency. *International Journal of Project Management*. 2007. Vol. 25, no. 1p. 56–65. DOI 10.1016/j.ijproman.2006.04.002.

MCDONOUGH, Edward F. and SPITAL, Francis C., 2003, Managing Project Portfolios. *Research-Technology Management*. 2003. Vol. 46, no. 3p. 40–46. DOI 10.1080/08956308.2003.11671565.

MEREDITH, Jack R, MANTEL, Samuel J, 2009, *Project management: a managerial approach.* New York : Wiley and Sons.

MILLHOLLAN, Chuck, 2008, Scope change control: control your projects or your projects will control you. *PMI® Global Congress Proceedings.* 2008. Newtown Square, PA: Project Management Institute.

MILLHOLLAN, Chuck and KAARST-BROWN, Michelle, 2016. Lessons for IT Project Manager Efficacy: A Review of the Literature Associated with Project Success. *Project Management Journal*. 2016. Vol. 47, no. 5, p. 89-106. DOI 10.1177/875697281604700507.

MORO, Sérgio, CORTEZ, Paulo and RITA, Paulo, 2014, A data-driven approach to predict the success of bank telemarketing. Decision Support Systems. 2014. Vol. 62, p. 22-31.

MORO, Sérgio, CORTEZ, Paulo and RITA, Paulo, 2015, Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*. 2015. Vol. 42, no. 3, p. 1314–1324. DOI 10.1016/j.eswa.2014.09.024.

MORO, Sérgio, RITA, Paulo and COELHO, Joana, 2017, Stripping customers feedback on hotels through data mining: The case of Las Vegas Strip. *Tourism Management Perspectives*. 2017. Vol. 23, p. 41–52. DOI 10.1016/j.tmp.2017.04.003.

MORRIS, Rick A. and SEMBER, Brette Mcwhorter, 2008, *Project management that works*. New York: American Management Association.

PETIT, Yvan, 2012, Project portfolios in dynamic environments: Organizing for uncertainty. *International Journal of Project Management*. 2012. Vol. 30, no. 5, p. 539-553. DOI 10.1016/j.ijproman.2011.11.007.

PINTO, Jeffrey K. and SLEVIN, Dennis P., 1988, Project Success: Definitions and Measurement Techniques. *Project Management Journal*. 1988. Vol. 19, no 1, p. 67–72.

POSNER, Barry Z., 1987, What it takes to be a good project manager. *Project Management Journal*. 2010. Vol. 18, no.1, p. 51–54.

POSPIESZNY, Przemysław, 2017, Application of Data Mining Techniques in Project Management – an Overview. *Collegium of Economic Analysis Annals.* 2017. Vol 43, p. 199-220.

PRABHAKAR, Guru Prakash, 2009, Projects and Their Management: A Literature Review. *International Journal of Business and Management*. 2009. Vol. 3, no. 8. DOI 10.5539/ijbm.v3n8p3. Canadian Center of Science and Education.

PROJECT MANAGEMENT INSTITUTE, 2004, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)-Third Edition (ENGLISH)*. Newtown Square, PA : Project Management Institute.

PROJECT MANAGEMENT INSTITUTE, 2017, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)-Sixth Edition (ENGLISH)*. Newtown Square, PA : Project Management Institute.

RAVI, Vladlamani, 2007, *Advances in banking technology and management: impacts of ICT and CRM*. IGI Global.

REFAEILZADEH, Payam, TANG, Lei and LIU, Huan, 2009, Cross-Validation. *Encyclopedia of Database Systems*. 2009. P. 532–538. DOI 10.1007/978-0-387-39940-9_565.

REHACEK, Petr, 2017, Application and Usage of the Standards for Project Management and their Comparison. *Journal of Engineering and Applied Sciences*. 2017. Vol. 12, no. 4, p. 994-1002.

RODRIGUEZ-REPISO, Luis, SETCHI, Rossitza and SALMERON, Jose L., 2007, Modelling IT projects success: Emerging methodologies reviewed. *Technovation*. 2007. Vol. 27, no. 10, p. 582-594. DOI 10.1016/j.technovation.2006.12.006.

ROSEN, ANITA, 2004, *Effective IT project management: Using teams to get projects completed on time and under budget*. New York : Amacom, American Management Association.

SCHMIDT, Reinhard H., 2001, The future of banking in Europe. *Financial Markets and Portfolio Management*. 2001. Vol. 15, no. 4, p. 429-449. DOI 10.1007/s11408-001-0402-x.

SHU, Wesley and STRASSMANN, Paul A., 2005, Does information technology provide banks with profit?. *Information & Management*. 2005. Vol. 42, no. 5, p. 781-787. DOI 10.1016/j.im.2003.06.007.

TAYLOR, Hazel, 2006, Risk Management and Problem Resolution Strategies for it Projects: Prescription and Practice. *Project Management Journal*. 2006. Vol. 37, no. 5, p. 49-63. DOI 10.1177/875697280603700506.

TURNER, J. Rodney and MÜLLER, Ralf, 2005, The Project Manager's Leadership Style as a Success Factor on Projects: A Literature Review. *Project Management Journal*. 2005. Vol. 36, no. 2, p. 49-61. DOI 10.1177/875697280503600206.

ULUOCAK, Burak, 2013, Critical success factors (CSFs) affecting project performance in Turkish IT sector. *PMI® Global Congress 2013—EMEA, Istanbul, Turkey*. 2013. Newtown Square, PA: Project Management Institute.

VILLANUEVA-BALSERA, Joaquín, ORTEGA-FERNANDEZ, Francisco, RODRÍGUEZ-MONTEQUÍN, Vicente, and CONCEPCIÓN-SUÁREZ, Ramiro, 2009, Effort estimation in information systems projects using data mining techniques. *In WSEAS International Conference. Proceedings. Recent Advances in Computer Engineering (No. 13).* 2009.WSEAS.

YU, Ming-Chuan, 2017. Customer Participation and Project Performance: A Moderated-Mediation Examination. *Project Management Journal*. 2017. Vol. 48, no. 4, p. 8-21. DOI 10.1177/875697281704800402.

ZWIKAEL, Ofer, 2009. The Relative Importance of the PMBOK® Guide's Nine Knowledge Areas during Project Planning. *Project Management Journal*. 2009. Vol. 40, no. 4, p. 94-103. DOI 10.1002/pmj.20116.

ZWIKAEL, O. and GLOBERSON, S., 2004, Evaluating the quality of project planning: a model and field results. *International Journal of Production Research*. 2004. Vol. 42, no. 8, p. 1545-1556. DOI 10.1080/00207540310001639955. Informa UK Limited.

Appendix

[This page was intentionally left blank]

Appendix A – Dataset variables - details and description

Variable 1

Name: Project ID

Description: Project's internal identification. It has a unique reference with a project.

Model inclusion: No - This value is not relevant for modulation or analysis, since do not contains any relevant factor for knowledge extraction.

Variable 2

Name: Project Phase

Number of possible values: 3

Description: The variable indicates the phase of the project in its lifecycle. Only the projects that are already implemented were included in this study, since for projects that aren't in production phase, it is not possible to understand if there are more changes in the project, since they aren't yet concluded. For instance, a project in a development phase, it isn't in his final stage and it is possible to have changes at certification's phase. Only the last three phases were assumed:

Phase "in Production" – When a project goes to live environment (production environment), project manager passes project status to in production. This assumes that clients have accepted certification phase and agreed to activate and deploy project in production.

Phase "Warranty Finished" – After project manager's decision that project's warranty period has finished, project manager closes the project. Usually warranty takes one to two months, but if there are pendent errors detected, warranty can take more time. After warranty period is concluded, team members are not capable to input more hours to the project. At this phase, a survey is sent to project sponsor in order to evaluate and classify his satisfaction with project implementation.

Phase "Project Completed" – After sponsor response to the satisfaction survey, project passes to "Project Completed". This is the terminal phase and status of a project, and there are no more changes affected to projects at this phase.

Model inclusion: No – Project phase is a temporary phase (besides "project completed" phase) and does not give any value to understand project changes in a

project. For instance, conclude that warranty finished projects have more changes than other phases are fallacious, since all projects lead to project completion.

Variable3

Name: Project Type

Number of possible values: 2

Description: Projects keep a project lifecycle and a project management process. These rules are implemented by PMO, according best project management practices, and they need to be respected in order to lead to project management's standardization and process control. These rules are transformed in metrics in order to evaluate project and project manager's performance. For instance, projects should end on time and on scope in order to mitigate portfolio impact. If a project ends later than expected, it affects project manager's Key Performance Indicators. Other applied rule is that project development (execution) cannot start before project contract (known in literature as project management plan) is approved by the project's sponsor. These projects are called Regular projects. Yet, other projects, by its nature, are exempt of these metrics and rules since they don't use the regular project process flow. Technology projects, extreme critical projects, mandate projects that development needs to start as soon as possible, or projects with new methodologies are usually assumed as special projects. For these cases, project manager usually does not use project management tool in the same way, and can only insert project plan and estimations at project end. These projects usually have less bureaucracy than regular projects.

Model inclusion: Yes – This feature can be relevant in project change indication.

Variable 4

Name: Internal Client

Number of possible values: 22 different departments or divisions

Description: The **"**internal client" variable indicates the principal client of a project. Projects can have more than one client. Yet, there is a department or division that has more interest in the project success. Usually, it is the department or division that has created the request and this department is responsible for sponsoring the project. This department approves project management plan (contract) and facilitates network and connection with other clients and stakeholders.

Model inclusion: Yes. Each department is different from another, since they can have different ways to understand projects. Their involvement also can be different. So, this is a relevant feature for analysis.

Variable 5

Name: First Production Date

Description: This variable indicates the first date inserted by project manager as the live date for implementation. It is the first production date indicated to client in customer contract as production date.

Model inclusion: No. This date will be used to identify changes in projects, but by itself it is not relevant for the projects' change study.

Variable 6

Name: Real Production Date

Description: The real production date indicates the last production date indicated by project manager. It is known after production deploy. If this date is equal to first production date, it means that the project didn't have delays in plan.

Model inclusion: No. It will be used for understand and calculate project plan changes, but it was not included directly for model analysis.

Variable 7

Name: First Estimation

Description: The first estimation reveals the number of hours indicated to client as the effort required for making the project. This value includes internal and outsourcing resource effort, but do not assumes external costs (COTS effort, for instance). This estimation is presented to customer in the project management plan (contract), and so, is the first estimation officially given to the internal customer.

Model inclusion: No. It was decided to use total estimated effort feature for model inclusion, since their values are very similar to this one, and using both will be redundant.

Variable 8

Name: Real Estimation

Description: The real estimation variable contains the project's actual effort at project end. It is the last effort indication of project manager. If is different from first estimation, it means there was a cost variation in project.

Model inclusion: No. This feature only is known at the end of the project and not at the beginning/planning phase. Yet, it is important to calculate if there were cost changes in the project.

Variable 9

Name: First External Costs

Description: This feature indicates external project costs. External project costs are all costs that aren't effort costs (COTS software, Licence acquisitions, specific components, travels, among others). These costs are included in customer contract for approval, and they cannot be adjudicated before client approval.

Model inclusion: No. These costs could be a relevant factor for change identification, but by confidence reasons they weren't revealed. Yet, they will be used for understand if projects had cost differences between baseline and the costs that were really spent.

Variable 10

Name: Real External Costs

Description: the "real external costs" feature indicates the real value of external costs that were spent after project execution is finished. If there are changes between this feature and the first external costs indicated, then there was a cost difference in the project.

Model inclusion: No. The same reasons of first external cost variable.

Variable 11

Name: Interposed

Number of possible values: 3 – Yes, No, Blanks

Description: Projects are requested by each department to IT. These projects go to a project portfolio without any implementation plan. Usually, twice a year (and at least

once a year), project portfolio plan is made or reviewed and these projects are prioritized by high management in committees. Normal projects are prioritized and they are included in project portfolio plan, but, some other projects, because business or mandatory urgent needs need to start as soon as possible. For instance, legal requirements with mandate dates of implementation may not be able to wait for a next committee to be prioritized, planned and implemented. These projects, after having top management approval (that also approve changes in project portfolio plan), start immediately and they usually affect projects in project portfolio plan, impacting dates and costs in other project implementations. These projects are called interposed projects.

Model inclusion: Yes. This is a variable known at the project's beginning. Many of these projects usually have a great sponsorship since they have aggressive deadlines or its impact is quite relevant in the institution. It is an important variable to be analysed.

Variable 12

Name: Decision Type

Number of possible values: 4

Description: Clients and Sponsors at project definition classify projects with different decision types. Projects can be decided to implement because they bring benefits and there is a profitable business case for making them, projects can be accepted to begin because they have at least one legal requirement, urgent or extreme critical requirements, legal requirements, they may implement audit responses, for technological causes, among others. These projects are classified as mandatory projects. At last, there are projects that do not have quantitative benefits, but they are, for instance, reputational projects. These have qualitative benefits and are known as "Qualitative" projects.

Model inclusion: Yes. It is a variable at the initial phase and can be relevant for the model.

Variable 13

Name: Mandatory Type

Number of possible values: 7 different mandatory types

Description: When a project is classified as mandatory, it is required by PMO the reason for its mandate. The reason can be a legal requirement that need to be implemented, an audit non conformity, a risk mitigation, a technological cause (discontinued software support, for instance), or an extreme critical business need. Mandatory type for non-mandatory projects is always considered as "others"

Model inclusion: Yes. Mandatory type can be related with project importance and project stakeholder's involvement, which can lead to a higher management sponsorship.

Variable 14

Name: Objective

Number of possible values: 12 different objectives

Description: At project beginning, internal clients identify the project objective. This is not a mandatory field, and there are several projects that are still waiting a classification and they are already implemented. Project objectives vary by audit purposes, income increases, cost reductions, recommendations, reporting or technology reason.

Model inclusion: Yes. Although there are many records not filled, it is a known variable at project begin and can have some relevant indicators.

Variable 15

Name: Number of Applications

Description: This variable indicates the number of applications affected by the project. After project starts, the project manager identifies all applications that are involved in the project to satisfy business requirements.

Model inclusion: Yes. It is a known variable at project begin and can be related with project complexity, since as more applications are involved, as hardest the project becomes to manage.

Variable 16

Name: Number of Internal Clients

Description: This feature represents the number of internal clients involved in the project. Although only one client (a client is a division or a department that belongs to the organization) is the primary client (usually the project sponsor), a project can have more than one client, dividing "costs" between them, since the requirements affect them all, and all of them have a positive interest in the project implementation.

Model inclusion: Yes – The number of internal clients can have a direct influence in project changes, since as more clients and stakeholders a project has, more difficult and complex can be to manage.

Variable 17

Name: Total Estimated Effort

Description: Total estimated effort indicates the effort estimated by each application team for internal and resources members. This value is known at project management plan, being part of the baseline after project estimation.

Model inclusion: Yes. Project total estimated effort is known before project execution, and can be used in analysis to understand if is related with project changes.

Variable 18

Name: Internal Estimated Effort

Description: The variable "internal estimated effort" indicates total effort estimated at planning phase by internal resources.

Model inclusion: Yes. It can be relevant to understand if this feature has impact on project changes.

Variable 19

Name: Outsourcing Estimated Effort

Description: Outsourcing estimated effort feature indicates the total effort estimated by outsourcing teams that work in time and materials (in person-days) mode at planning phase.

Model inclusion: Yes. It can be relevant to understand if this feature has impact on project changes.

Variable 20

Name: Actual Internal Effort

Description: It refers to the real number of imputed hours by internal team members on projects. For some special projects that don't have estimations inserted (this input is not mandatory for this kind of projects) it is the way to calculate and assume project estimation, project size and project duration.

Model inclusion: No. It is a variable that is only known at the project's end, so it has no relevance for project changes relation in the model. Yet, this feature is used for project estimation calculation, project size calculation and project duration calculation for the special projects that do not have estimations fulfilled.

Variable 21

Name: Actual Outsourcing Effort

Description: Actual imputed hours by outsourcing team members on the project. For some special projects that do not have estimations inserted (this input is not mandatory for this kind of projects), it is the way to calculate and assume project estimation calculation, project size calculation and project duration calculation.

Model inclusion: No. It is a variable that is only known at project end, so it has no relevance for project changes relation. Yet, this feature is used for project estimation calculation, project size calculation and project duration calculation for the special projects that do not have estimations fulfilled.

Variable 22

Name: Number of Outsourcing Teams

Description: It is not possible to understand how many outsourcing resources were involved in project, since the PMO tool do not register outsourcing hours for each resource. Yet, it is possible to understand how many different outsourcing teams are involved in the project.

Model inclusion: Yes. As indicated on literature, more team members of different outsourcing teams can lead to project complexity and have negative impact on project management success, since it raises project management complexity.

Variable 23

Name: Number of Internal Team members

Description: This feature indicates the number of project team members that are internal to institution. Outsourcers in time and materials, although are being managed directly by the project manager, are not included in this feature, since their imputation is made by a unique virtual resource.

Model inclusion: Yes. As indicated on literature, more team members can lead to project management complexity and it can have a negative impact on project management success.

Variable 24

Name: Begin Date

Description: The begin date consists in the date when project manager has received the project and has started the project initiation phase.

Model inclusion: No. For a predictive or categorical analysis, it has no relevance if the project has started one year or one month ago. Yet, this variable will be useful to calculate project duration.

Variable 25

Name: Number of Change Requests

Description: When a project has a scope, time, or cost change, project manager must make a contract addendum in order to notify project's clients for the reasons of project's changes, showing their impact in the project (and sometimes on other projects). This changes can affect only project's dates (when a project date changes because a longer certification, for instance), can affect project's cost (e.g., when more resources must be included in the project for achieve target date), or can affect scope (when there is, for instance, a change in initial requirements and project must be aligned with them). Usually scope changes also affect time, cost or both. After the project clients approve a contract addendum (a new project management plan that includes the new project changes), a new baseline is set with a new plan defined. A project can have more than one project change across its lifecycle.

Model inclusion: No. This is a very important variable to understand project scope changes, since they aren't showed in estimation or in dates, and there isn't a feature that represents it. Yet, it makes no sense to include this feature in analysis, since its

relation with project changes is direct and usually only after project execution project changes are revealed (they are not identified before project management plan approval).

Variable 26

Name: Date Difference

Description: The feature "date difference" is calculated with the difference in calendar days between the real production date and first production date. If this variable assumes the value zero, there are no date changes in the project. If it is grater zero, then the project had a delay according to the first production date defined by the project manager. If the difference is negative, project has been completed before expected.

Model inclusion: No. This variable will be important to understand if project had changes, but by itself does not make sense to include in model, since its relation with project changes is direct.

Variable 27

Name: Date Difference Categorization

Number of different values: 2 - Yes or No

Description: If the date difference is greater than zero, then assumes variable "Yes" (there is a date change in the project that made as the project delayed), else assumes "No" (there aren't project date changes, or the project has finished before the indicated date).

Model inclusion: No, but this variable is essential to calculate if the project had changes.

Variable 28

Name: Budget Difference

Description: This variable is calculated with the difference between the real project estimation and the first project estimation variables. If it is greater than zero, then there were cost changes in project, and the project has consumed more effort than the originally expected.

Model inclusion: No. This variable will be important to understand if project had changes, but by itself does not make sense to include in model, since its relation with project changes is direct.

Variable 29

Name: Budget Difference Categorization

Number of different values: 2 - Yes or No

Description: If budget difference feature is greater than zero, then this variable assumes the value "Yes" (there is a cost change in the project), else assumes "No" (there aren't project cost changes, or the effort was less than the original expected, if the value is negative).

Model inclusion: No, but this variable is essential to calculate if the project had changes.

Variable 30

Name: External Cost Difference

Description: This variable is the difference between the real external cost and the first external cost indicated in project management plan. If this cost is greater than zero, then there was a higher external cost in the project than the one that was expected.

Model inclusion: No. This variable is important to understand if the project had changes, but by itself does not make sense to include in model, since its relation with project changes is direct.

Variable 31

Name: External Cost Difference Categorization

Number of different values: 2 - Yes or No

Description: If external cost difference is greater than zero, then this feature assumes the value "Yes" (there is a cost change in the project that imply a higher project cost), else assumes "No" (did not exist any cost differences or it was spent less budget than the originally expected).

Model inclusion: No, but this variable is essential to calculate if the project had changes.

Variable 32

Name: Estimated Duration

Description: Estimated duration is a calculated variable. By definition, project duration is the time since project start until its end. For this study, project duration was considered as the time between the project start date and the production date. For estimated duration was considered the time in days between project start date and estimated production date.

Model inclusion: Yes. Literature says that project duration may have impact on the project, so this is a variable to consider.

Variable 33

Name: Real Duration

Description: Real Duration is the time in days between the project start date and the real production date

Model inclusion: No. Since project production date is not known at project management plan approval time, it makes no sense to include this variable as a model feature.

Variable 34

Name: Project Size

Number of different values: 4 (Small Medium, Large and Very Large)

Description: The institution's PMO has the following size order defined: Small sized projects are projects below 100 person-days (each person-day consists in 8 labour hours) of total estimation. Medium sized projects are projects above 100 and below 250 person-days. Large sized projects are projects above 250 and below 500 person-days. Very large projects are projects above 500 person-days of estimation.

Model inclusion: Yes. This is a variable known at the end of project plan phase and can be relevant for data analysis.

Variable 35

Name: Duration Size

Number of values: 5 (Small, Medium, Large and Very large and '#N/A')

Description: for this feature definition, institution's PMO has the following duration size classification: Small sized projects are projects below 100 calendar days of duration. Medium sized projects are projects above 100 and below 250 calendar days. Large projects are projects above 250 and below 500 calendar days. Very large projects have a duration that is longer than 500 days.

Model inclusion: No. This is a variable that is calculated with real duration date. Also, a project can have a larger duration because other project external features (lack of resource availability, project suspension because an interposed project, among others). For this reasons, this feature wasn't included for model analysis.

Variable 36

Name: Changes in Project

Number of different values: 2 – Yes or No.

Description: Changes in projects feature assumes the value "yes" if there were changes in project. For calculate this value one of the four conditions applies: (1) number of change requests is greater than zero (there was at least a scope, time or budget change in the project); (2) date difference categorization is "yes" (there was a positive difference between the target end date and the real end date); (3) budget difference categorization is "yes"(project cost in the end was higher than the one that was initially accorded); (4) external costs difference is "yes"(outsourcing or time and materials costs is higher than the initially contracted). If none of this conditions apply, then there are no project changes and the value assumed is "no".

Model inclusion: Yes. It is the model's dependent variable.

[This page was intentionally left blank]

Appendix B – Model's metric results

Below are detailed and explained the model results for each model applied to the dataset.

RPART results:

Table 22 - RPART metric results

RPART			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.675886	0.6964204	0.6222705	0.6381884
0.6920162	0.6994588	0.6335024	0.639058
0.705916	0.6423133	0.6427536	0.5938406
0.6796622	0.6754151	0.6245169	0.6239855
0.706282	0.7157288	0.6435507	0.6402899
0.6777543	0.7069116	0.6238406	0.6365217
0.6717412	0.6012085	0.619372	0.5648551
0.6829308	0.6201958	0.6270773	0.587971
0.7053969	0.6708173	0.6431159	0.6149275
0.6965972	0.7105004	0.6364976	0.6468841
0.6799542	0.6989543	0.6255556	0.6319565
0.6890256	0.7111539	0.6311111	0.6489855
0.678795	0.6763094	0.6244203	0.6197101
0.6751884	0.6531026	0.6219324	0.6022464
0.6973547	0.6652449	0.6371981	0.6163043
0.6773513	0.6963516	0.6233333	0.6350725
0.6884925	0.7026922	0.6308696	0.6296377
0.6852507	0.747638	0.6289614	0.6689855
0.6969568	0.7475005	0.63657	0.6726087
0.6826464	0.7455627	0.6271256	0.6673188
Average			
0.687260	0.689174	0.630178745	0.62896739

MLPE results:

Table 23 - MLPE metric results

MLPE			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.7209915	0.708769	0.6536715	0.6450725
0.727715	0.7116584	0.6584541	0.6474638
0.7175903	0.710959	0.6515217	0.6468116
0.7380706	0.7257613	0.6655797	0.6572464
0.7207403	0.728559	0.6535024	0.658913
0.7206357	0.7014195	0.6536473	0.6399275
0.7230155	0.7413319	0.6550725	0.6678986
0.7273758	0.7336957	0.6581643	0.6627536
0.7267522	0.7448977	0.6577295	0.6704348
0.7318356	0.758886	0.6614734	0.6802899
0.7413699	0.7196042	0.6679227	0.6531159
0.7075917	0.7025431	0.6445169	0.6405072
0.7396163	0.7087232	0.6665459	0.6450725
0.7215195	0.7312878	0.653744	0.661087
0.7425266	0.673936	0.6686715	0.621087
0.7249922	0.7192602	0.6566425	0.6523188
0.7070089	0.6754609	0.6438889	0.622029
0.730207	0.7549762	0.6600966	0.6774638
0.7335675	0.7013163	0.6625121	0.6401449
0.73294	0.720166	0.6620048	0.6535507
Average			
0.726803105	0.71866057	0.657768115	0.652159425

Decision Tree Results:

Table 24 - DT metric results

DT			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.6978904	0.6627912	0.6375845	0.6210145
0.6884683	0.6607274	0.6311594	0.6173913
0.7119189	0.6444001	0.6473913	0.6065942
0.6728163	0.7037241	0.6200966	0.6407246
0.676377	0.726713	0.6226812	0.6600725
0.6760696	0.6862961	0.6219324	0.6304348
0.6729017	0.6272588	0.6199275	0.5872464
0.6713701	0.6512796	0.6192029	0.5976087
0.6864699	0.6985759	0.6294686	0.6376087
0.6708191	0.6750367	0.618599	0.6221739
0.6802259	0.6850693	0.6255072	0.6266667
0.6972081	0.7179646	0.6372947	0.6497826
0.6950987	0.6829366	0.6358696	0.6223913
0.6841743	0.6617822	0.6282609	0.6128261
0.6935364	0.6455582	0.6344928	0.6013768
0.6720587	0.6601312	0.6195652	0.6084058
0.6842534	0.6258599	0.6280435	0.5871739
0.6878383	0.6886351	0.6307971	0.6299275
0.6928057	0.6306297	0.6344203	0.5935507
0.7031702	0.6739704	0.6414734	0.6236232
Average			
0.68577355	0.670467005	0.629188405	0.61882971

MLP Results:

Table 25 - MLP metric results

MLP			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.683478	0.6365231	0.6277778	0.5957246
0.6686128	0.7234108	0.6174879	0.6555797
0.6894069	0.6741653	0.6316184	0.6215217
0.6725217	0.6622065	0.6200966	0.612971
0.6912064	0.6686159	0.6329227	0.6175362
0.6886864	0.6111837	0.631256	0.5773913
0.667414	0.6907219	0.6168116	0.6332609
0.6792107	0.6865827	0.6247585	0.6292754
0.6495939	0.6534581	0.6041546	0.6068116
0.6836603	0.6589961	0.6279469	0.6105797
0.6831145	0.737445	0.6274155	0.6662319
0.661318	0.6901486	0.6124396	0.6321739
0.6885678	0.653653	0.6310628	0.6086232
0.6715665	0.6652908	0.6191063	0.6147101
0.6715116	0.6729958	0.6194203	0.6203623
0.6768998	0.6632155	0.6234541	0.6149275
0.6788689	0.6355371	0.6244203	0.5945652
0.6925786	0.6158388	0.6340097	0.5808696
0.6880411	0.6929577	0.6311111	0.6344203
0.687133	0.7004219	0.6301691	0.6396377
Average			
0.678669545	0.669668415	0.62437199	0.61835869

SVM Results:

Table 26 - SVM metric results

SVM			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.7535263	0.7247065	0.6763043	0.6567391
0.7540313	0.7468584	0.6766667	0.6718116
0.7488675	0.7164855	0.6733575	0.6505797
0.7474047	0.7324115	0.6720531	0.6613768
0.7568217	0.733902	0.6788164	0.6627536
0.7474047	0.732572	0.6720048	0.662029
0.7489887	0.7586452	0.6732609	0.6797826
0.7489224	0.750344	0.673285	0.6743478
0.7488356	0.7509631	0.6730918	0.6747101
0.7554762	0.7347276	0.6778986	0.6633333
0.7539178	0.7850853	0.6766667	0.6985507
0.7519984	0.7355302	0.6755072	0.6639855
0.7490524	0.7628417	0.6734058	0.6826087
0.7519653	0.7232389	0.6753623	0.655
0.7482745	0.7377545	0.6725845	0.6652899
0.7451142	0.7173225	0.6704831	0.6513768
0.7571125	0.7463768	0.6788164	0.6715942
0.7449765	0.7537149	0.670628	0.6764493
0.7403037	0.7651348	0.667029	0.6846377
0.7535288	0.7758783	0.676256	0.6918841
Average			
0.75032616	0.744224685	0.674173905	0.669942025

RFOREST Results:

Table 27 - RF metric results

RFOREST			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.7704893	0.7666369	0.6881159	0.6852899
0.7712532	0.7681159	0.6886715	0.6867391
0.7761568	0.7596427	0.6922222	0.6807971
0.770715	0.7778045	0.6883575	0.6934058
0.7744938	0.7638163	0.691087	0.6836232
0.7679119	0.7652266	0.686401	0.6845652
0.7787151	0.7569483	0.6938647	0.6791304
0.7770381	0.7903137	0.6927536	0.702029
0.7741673	0.7858535	0.6908937	0.6992754
0.7714662	0.7780338	0.6887681	0.6935507
0.7773021	0.7820698	0.6928019	0.6966667
0.7740475	0.7627844	0.6907246	0.6826087
0.7815693	0.7822189	0.6959903	0.6964493
0.7819404	0.7484292	0.696087	0.6728986
0.7677359	0.7451385	0.6863285	0.6706522
0.774615	0.7610759	0.6909662	0.6816667
0.7790544	0.7294304	0.694058	0.6596377
0.7766721	0.7878371	0.6925604	0.7004348
0.7729239	0.7982595	0.6898309	0.7072464
0.7766236	0.7834572	0.692657	0.6973188
Average			
0.774744545	0.769654655	0.691157	0.687699285

Logistic Regression Results:

Table 28 - LR metric results

LR			
AUC		ALIFT	
Kfold	Holdout	Kfold	Holdout
0.7243074	0.738775	0.6558937	0.6662319
0.7281652	0.7218171	0.6586957	0.6542754
0.732764	0.7039763	0.6619082	0.642029
0.7324962	0.7481884	0.6618116	0.6727536
0.7341273	0.7363672	0.6629227	0.6648551
0.7273273	0.7296941	0.6580676	0.6601449
0.7304978	0.7121285	0.6603382	0.6476087
0.730624	0.7678178	0.6604106	0.6863043
0.7279484	0.7245918	0.6586715	0.6563043
0.7322233	0.7211177	0.66157	0.654058
0.7317438	0.7626582	0.6613285	0.6826087
0.7384583	0.7433269	0.6658213	0.6694203
0.7325778	0.7473858	0.6618116	0.6726087
0.7294686	0.7462736	0.6599275	0.6713768
0.7291906	0.7152128	0.6596135	0.6497826
0.7355634	0.7560195	0.6639855	0.6782609
0.7322182	0.7573954	0.6615459	0.6789855
0.7292926	0.7192258	0.6595652	0.6525362
0.7348364	0.7444964	0.6632609	0.6702899
0.7294609	0.7159925	0.659372	0.6500725
Average			
0.731164575	0.73562304	0.660826085	0.664025365

[This page was intentionally left blank]

Appendix C - R coding samples

Libraries used:

```
# library to do the scientific part of analysis which contains
classification methods
library(rminer)

# library that plots classification graphics
library(rpart.plot)
```

Original Dataset:

```
# Atributes names class
ColNames=c("Project ID", "Project_Phase", "Project_Type",
"Internal_client","First_Prod_Date","Real_Prod_Date", "DateDiff",
"DatediffCat", "First_Estimation","Real_Estimation",
"BudgetDiff", "BudgetDiffCat",
"First_Budget_External_Costs","Real_Budget_External_Costs",
"External_Cost_difference","External_Cost_Difference_CAT",
"Interposed", "Decision_Type", "Mandatory_Type", "Objective",
"Number_of_Applications","Number_of_Internal_Clients",
"Total_Estimated_Effort", "Internal_Estimated_Effort",
"Outsourcing_Estimated_Effort","Actual_Internal_Effort",
"Actual_Outsourcing_Effort", "Number_of_Outsourcing_Teams",
"Number_of_internal_resources","Begin_Date", "Last_Date",
"Number_of_Change_Requests", "Estimated_Duration",
"Assumed_Estimation_duration","Duration", "Real_Duration",
"Project_Size", "Duration_Size", "Changes_in_Project" )
```

Dataset Used to model knowledge extraction:

```
df2 <- df[,c("Project_Type","Internal_client",
             "Interposed","Decision_Type","Mandatory_Type",
"Objective","Number_of_Applications","Number_of_Internal_Clients","Tot
al_Estimated_Effort", "Internal_Estimated_Effort",
"Outsourcing_Estimated_Effort","Number_of_Outsourcing_Teams","Number_o
f_internal_resources", "Estimated_Duration", "Project_Size",
"Changes_in_Project")]
```

Metric obtain example (for SVM):

```
KSVM <- mining(Changes_in_Project~., df2,Runs=20,
               model="svm",
               method=c("kfold",10))

HSVM <- mining(Changes_in_Project~., df2,Runs=20,
               model="svm",
               method=c("holdout"))

mmetric(KSVM, metric="AUC", TC=2)
mmetric(HSVM, metric="AUC", TC=2)

mmetric(KSVM, metric="ALIFT", TC=2)
```

```
mmetric(HSVM, metric="ALIFT", TC=2)
```

ROC Curve generation:

```
AllCWithKfold <- vector("list",7);
AllCWithKfold[[1]]=KRPART;
AllCWithKfold[[2]]=KMLPE;
AllCWithKfold[[3]]=KDT;
AllCWithKfold[[4]]=KMLP;
AllCWithKfold[[5]]=KSVM;
AllCWithKfold[[6]]=Klr;
AllCWithKfold[[7]]=KrandomForest;

mgraph(AllCWithKfold,graph="ROC",TC=2,leg =
list(pos="bottomright",leg=c("RPART","MLPE","DT","MLP","SVM","LR","RFO
REST")),col = c("GREY","GREY","GREY","GREY","GREY","GREY","GREY"),
baseline=TRUE,Grid=10,
       main="ROC for Yes with KFOLD validation")
```

Weight predict on classification models example:

```
RPART <- fit(Changes_in_Project~.,df2,model="rpart") #decision tree
algorithm
DT <- fit(Changes_in_Project~.,df2,model="dt") # fit a Decision Tree
with tr
MLPE <- fit(Changes_in_Project~.,df2,model="mlpe") #neuronal network
MLP <- fit(Changes_in_Project~.,df2,model="mlp") #neuronal network
SVM <- fit(Changes_in_Project~.,df2,model="svm") # fit a SVM with
Algorithm
RFOREST <- fit(Changes_in_Project~.,df2,model="randomForest") # fit a
RF with Algorithm
```

Sensibility analysis example:

```
Imprf <- Importance(RFOREST,df2,method="DSA",RealL=51)
Listrf <- list(runs=20,sen=t(Imprf$imp))
```

VEC Plot example for Total_Estimation_Effort variable

```
xval <- c(which(names(df2)== "Project_Type"))
vecplot(Imprf,graph="VEC",xval=xval,TC=2,
        pch=c(16,4),Grid=20, xlab="",ylab="Changes_in_Project",
        leg=list(pos="bottomright",leg="ProjectType"))
res<-data.frame(Type=Imprf$sresponses[[xval]]$x,
                No=Imprf$sresponses[[xval]]$y[,1],
                Yes=Imprf$sresponses[[xval]]$y[,2])
```

Appendix D – Importance results for all dataset features – RAW Data

1. Project Type

Importance results:

Table 29 - Project type importance results

Type	No	Yes
Regular	0.6014195	0.3985805
Special	0.7466422	0.2533578

VEC Plot:

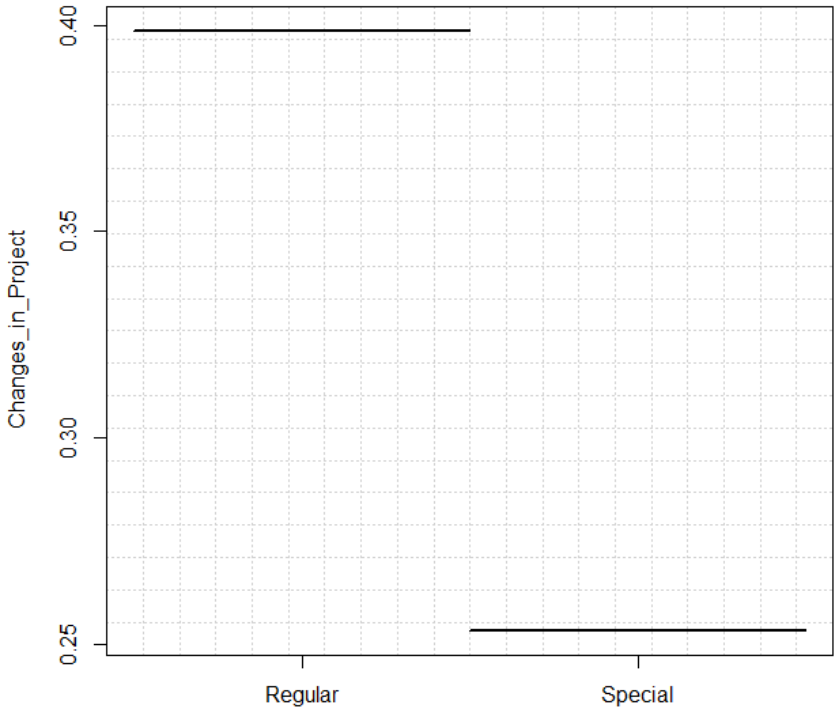

Figure 28 - Project type importance VEC plot

2. Internal_client

Importance results:

Table 30 - Internal client importance results

Internal Client	No	Yes
Accountant	0.685809	0.314191
Audit Compliance and Rating	0.569935	0.430065
Branch Support	0.615161	0.384839
Channels and Internet	0.741348	0.258653
Communication	0.576231	0.423769
Credit	0.612927	0.387073
Credit Recovery	0.649652	0.350348
Financial Markets	0.631954	0.368046
Human Resources	0.628626	0.371374
Insurances	0.587254	0.412747
International	0.618885	0.381115
IT	0.520511	0.479489
Marketing Empresas	0.638035	0.361965
Operations	0.632595	0.367405
Others	0.574677	0.425323
Private	0.558066	0.441934
Procurement and Resources	0.581968	0.418032
Real Estate	0.523774	0.476226
Reporting	0.605705	0.394295
Retail Marketing	0.66414	0.33586
Risk	0.666839	0.333161
Segments	0.662745	0.337255

Importance plot

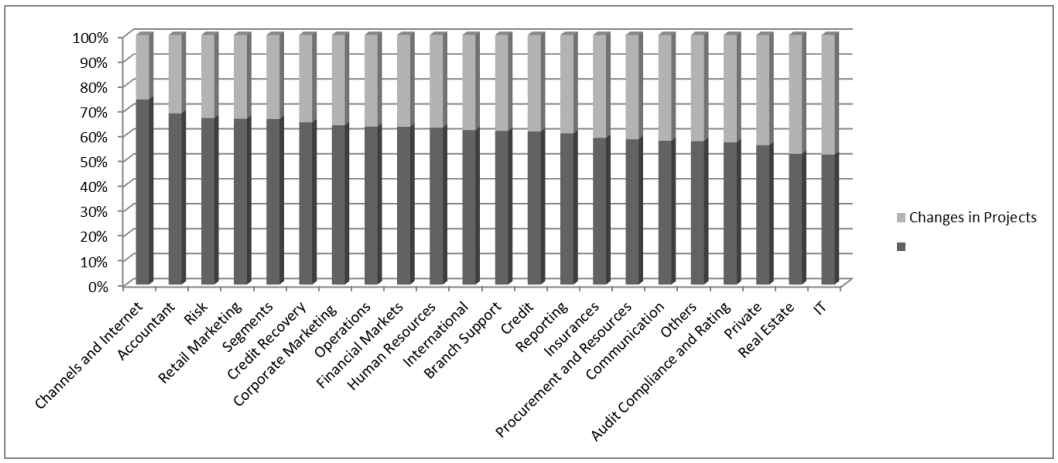

Figure 29 - Internal client importance plot

3. Estimated Duration

Importance results:

Table 31 - Estimation duration importance results

Estimated Duration	No	Yes
0	0.619421	0.380579
40.70833	0.690849	0.309151
81.41667	0.701687	0.298313
122.125	0.687533	0.312467
162.83333	0.651193	0.348807
203.54167	0.628814	0.371186
244.25	0.602976	0.397024
284.95833	0.582623	0.417377
325.66667	0.574423	0.425577
366.375	0.573013	0.426988
407.08333	0.570513	0.429487
447.79167	0.562851	0.437149
488.5	0.559543	0.440457
529.20833	0.555295	0.444705
569.91667	0.553628	0.446372
610.625	0.552679	0.447321
651.33333	0.552927	0.447073
692.04167	0.555718	0.444282
732.75	0.555803	0.444197
773.45833	0.555624	0.444376
814.16667	0.556256	0.443744
854.875	0.554861	0.445139
895.58333	0.554861	0.445139
936.29167	0.554861	0.445139
977	0.554861	0.445139

Importance VEC Plots:

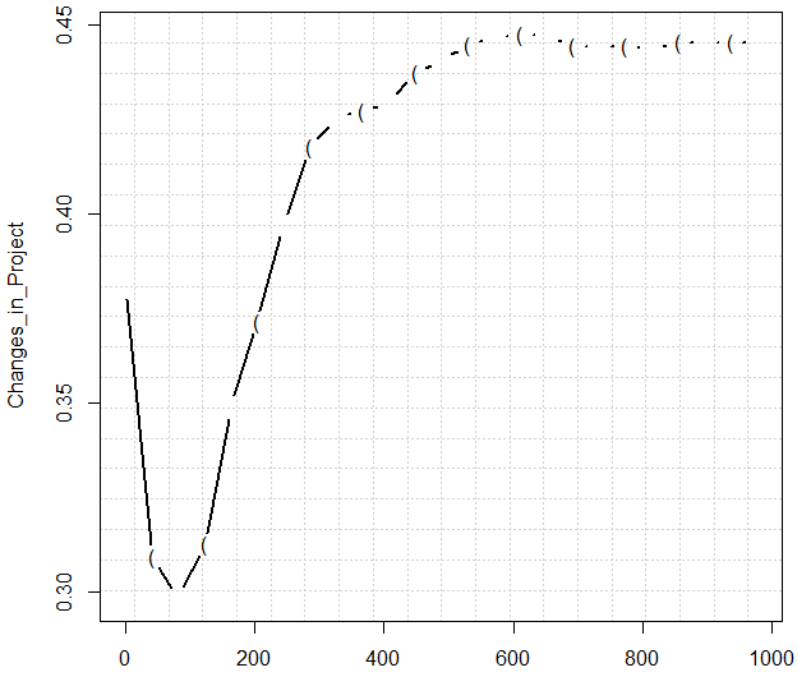

Figure 30 - Estimated duration importance plot 1

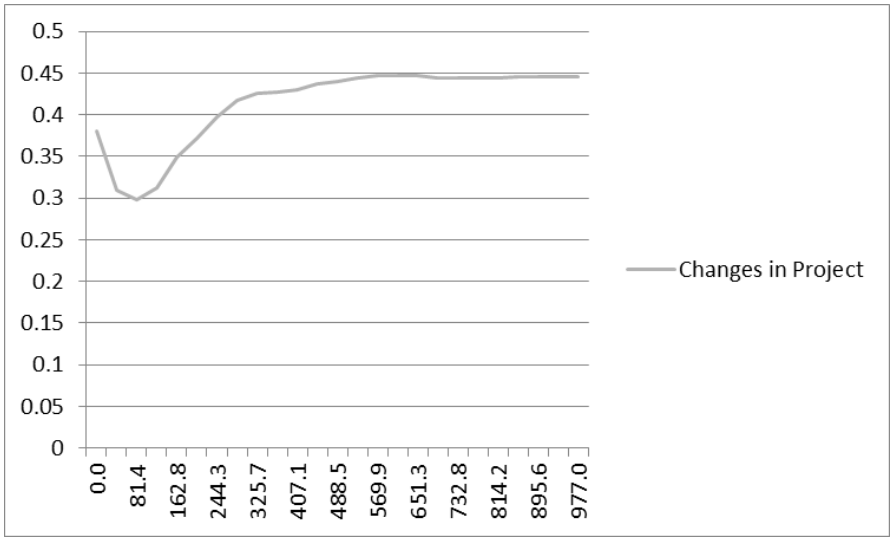

Figure 31 - Estimated duration importance plot 2

4. Objective

Importance results:

Table 32 - Objective importance results

Objective	No	Yes
3 - Legal Requirement	0.693831	0.306169
0 - Waiting Classification	0.675906	0.324094
2 - Internal Recommendation	0.666359	0.333641
4 - Technology	0.652943	0.347057
2 - Information for Customer	0.648952	0.351048
1 - Increase Income	0.646414	0.353586
3 - Others	0.641136	0.358864
1 - Management Control Information	0.622347	0.377653
2 - Cost Reduction	0.621815	0.378185
1 - Audit Committee	0.611985	0.388015
3 - Audit Recommendation	0.580823	0.419177
N/A	0.545761	0.45424

Objective importance Plot:

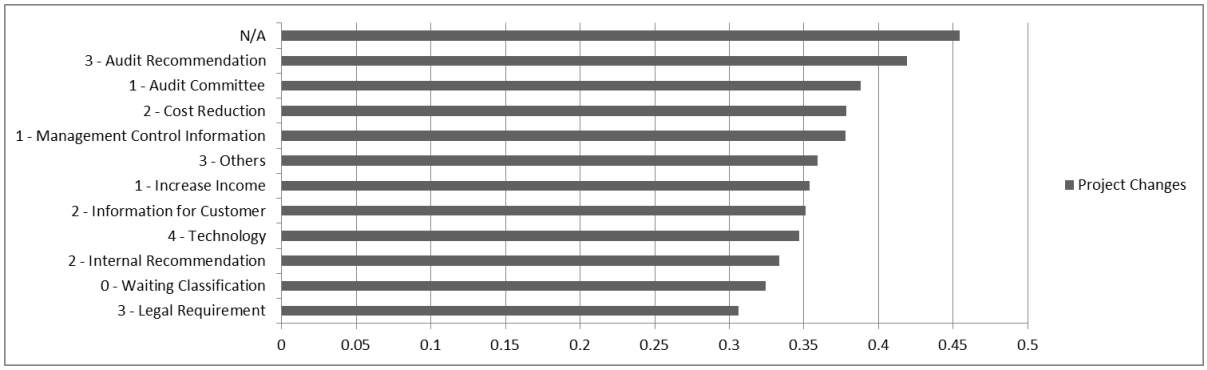

Figure 32 - Objective importance plot

5. Mandatory Type

Importance results:

Table 33 - Mandatory type importance results

Mandatory Type	No	Yes
Audit Medium and Low Non Conformity or Administration Strategic Requirement	0.691408	0.308592
Others	0.681843	0.318157
Mandatory Legal Requirement	0.67024	0.329761
Business Extreme Critical	0.661771	0.338229
Legal Necessity or Audit High Risk Non Conformity	0.629608	0.370392
Technology Causes	0.620129	0.379871
N/A	0.586871	0.413129

Importance plot:

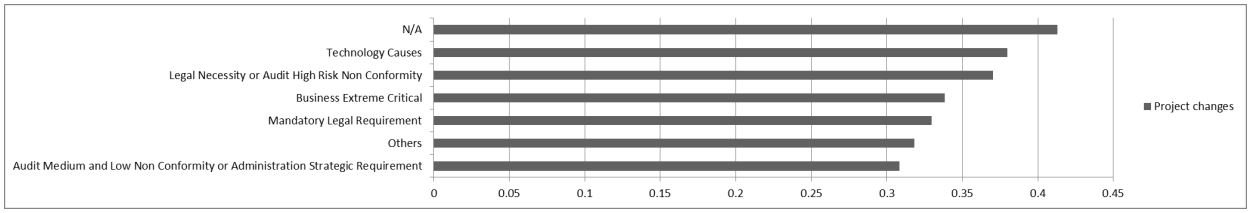

Figure 33 - Mandatory type importance plot

6. Project_Size

Importance results:

Table 34 - Project size importance results

Project Size	No	Yes
Small	0.6877399	0.3122601
Medium	0.6698898	0.3301102
Large	0.6326392	0.3673608
Very Large	0.6218854	0.3781146

Importance VEC plots:

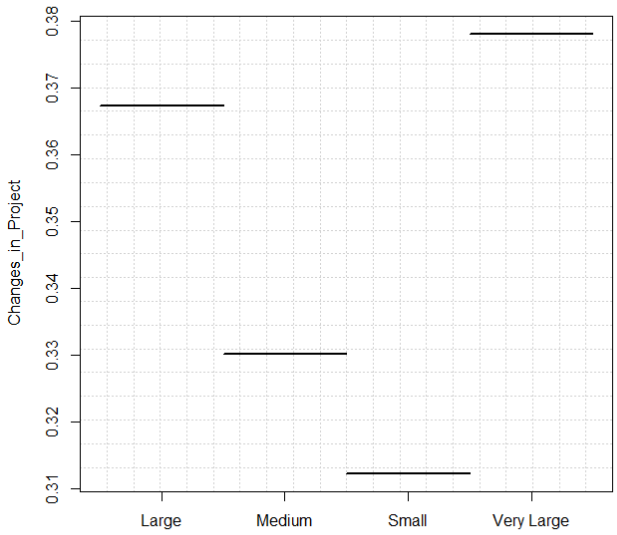

Figure 34 - Project size importance plot 1

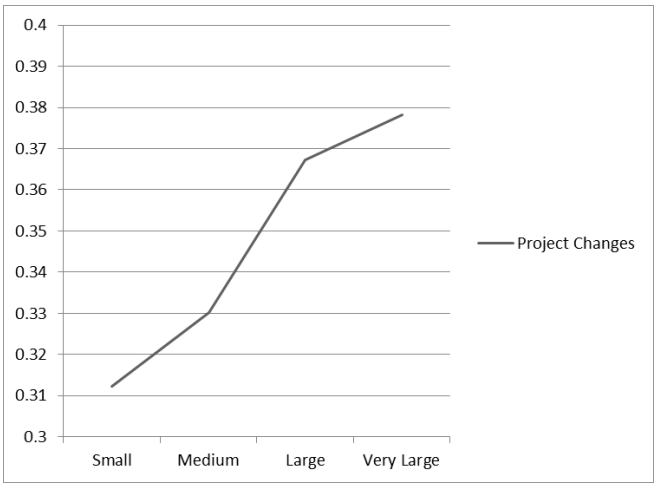

Figure 35 - Project size importance plot 2

7. Interposed

Importance Results

Table 35 - Interposed importance results

Interposed	No	Yes
Blanks	0.631849	0.368151
No	0.679968	0.320032
Yes	0.688926	0.311074

Importance VEC Plot:

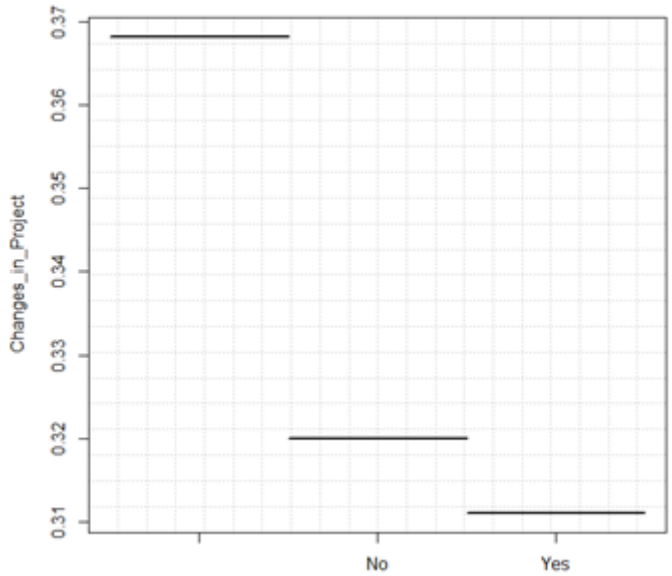

Figure 36 - Interposed importance plot 1

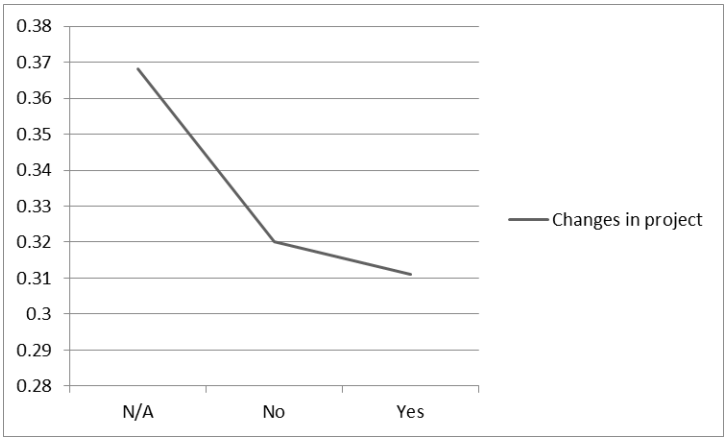

Figure 37 - Interposed importance plot 2

8. Decision Type

Importance results:

Table 36 - Decision type importance results

Decision Type	No	Yes
N/A	0.6272417	0.3727583
B - Benefits	0.6769287	0.3230713
M - Mandatory	0.6837781	0.3162219
Q- Qualitative	0.664147	0.335853

Importance VEC Plots:

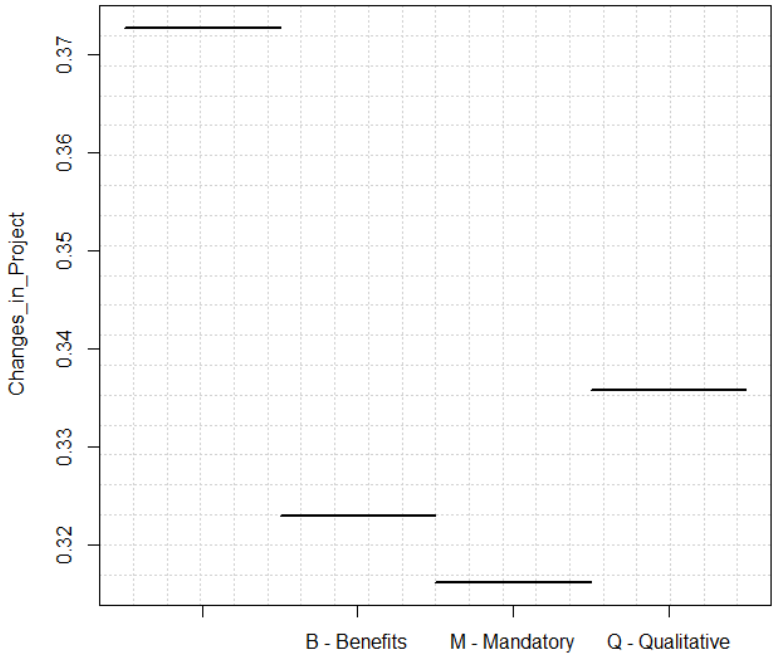

Figure 38 - Decision type importance plot 1

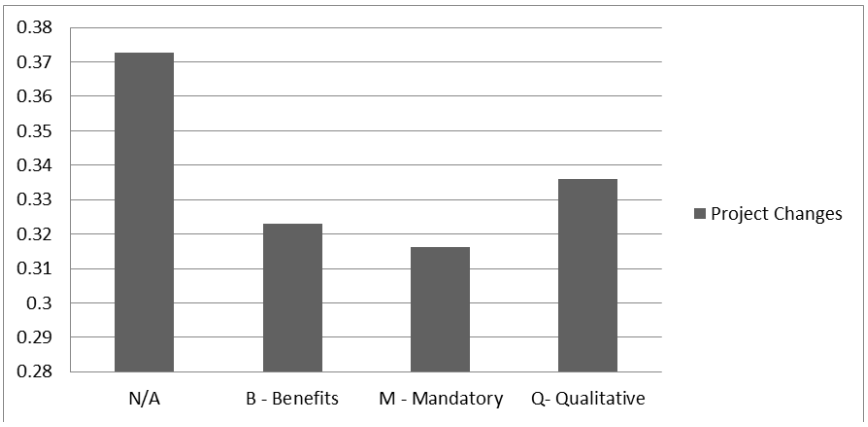

Figure 39 - Decision type importance plot 2

9. Total estimation Effort

Importance results

Table 37 - Total estimation effort importance results

Total Estimation Effort	No	Yes
0.00	0.711916	0.288084
1572.38	0.618364	0.381636
3144.75	0.518935	0.481065
4717.13	0.509162	0.490838
6289.50	0.504735	0.495265
7861.88	0.504131	0.495869
9434.25	0.503351	0.49665
11006.63	0.501035	0.498966
12579.00	0.501035	0.498966
14151.38	0.501035	0.498966
15723.75	0.501481	0.498519
17296.13	0.501481	0.498519
18868.50	0.501481	0.498519
20440.88	0.501459	0.498541
22013.25	0.501618	0.498382
23585.63	0.501117	0.498883
25158.00	0.50114	0.49886

Importance VEC Plots

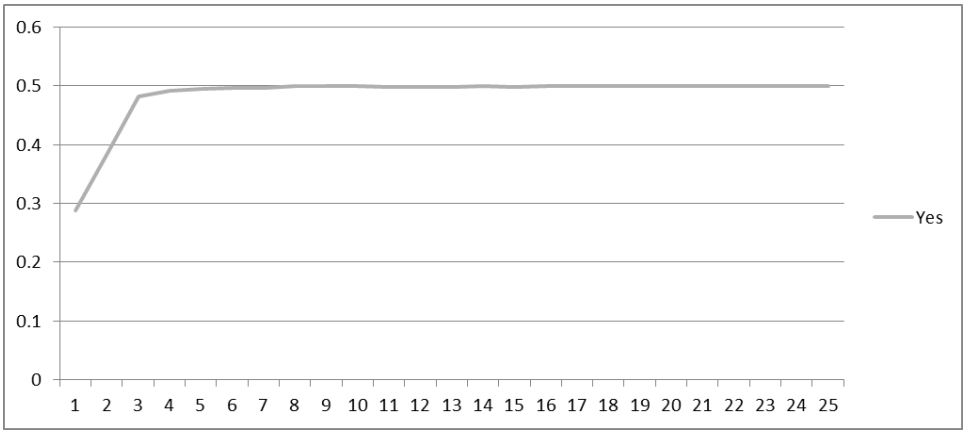

Figure 40 - Total estimation effort importance plot 1

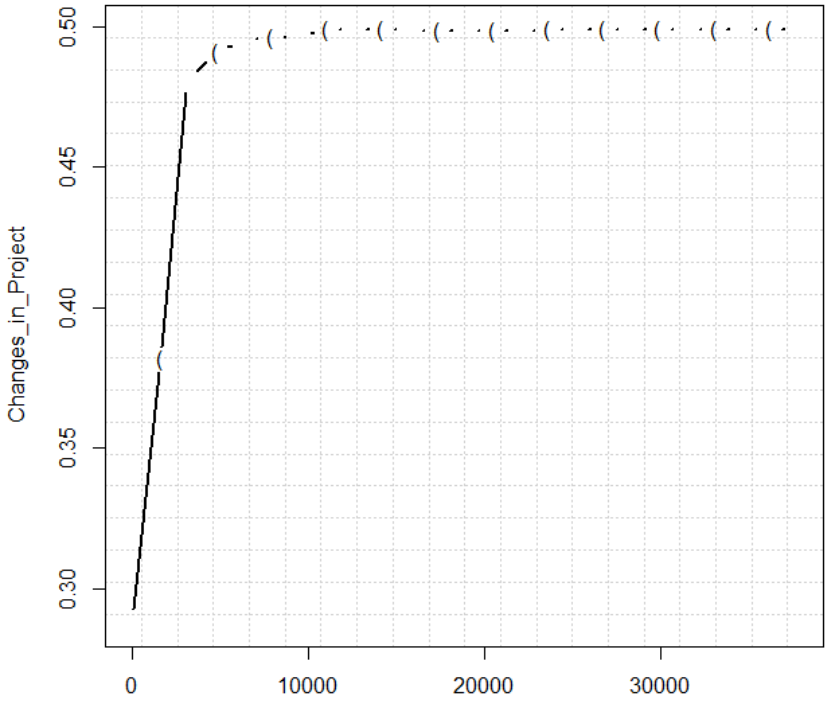

Figure 41 - Total estimation effort importance plot 2

10.Internal Estimated Effort

Importance results

Table 38 - Internal estimation effort importance results

Internal Estimated Effort	No	Yes
0	0.63552	0.364481
550	0.682578	0.317423
1101	0.648078	0.351922
1651	0.627271	0.372729
2201	0.622489	0.377511
2752	0.624354	0.375646
3302	0.618319	0.381681
3852	0.611997	0.388003
4403	0.610146	0.389855
4953	0.608924	0.391076
5503	0.609259	0.390741
6054	0.609306	0.390694
6604	0.609306	0.390694
7154	0.609284	0.390716
7705	0.609934	0.390066
8255	0.609915	0.390085
8805	0.611567	0.388434
9356	0.611567	0.388434
9906	0.611567	0.388434
10456	0.611567	0.388434
11007	0.611567	0.388434
11557	0.611567	0.388434
12107	0.611567	0.388434
12658	0.611567	0.388434
13208	0.611567	0.388434

Importance VEC Plots

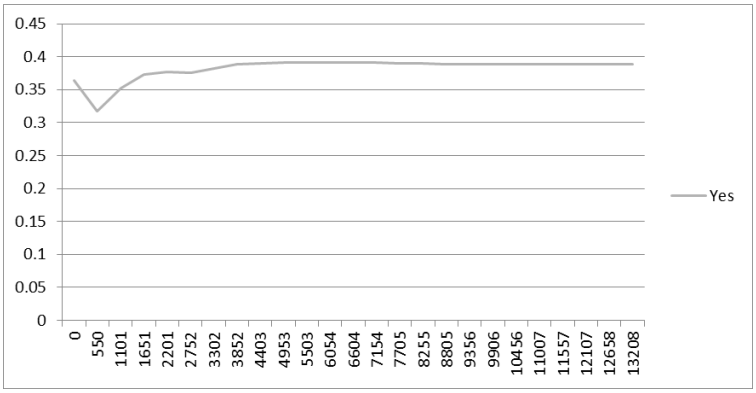

Figure 42 - Internal estimated effort importance plot 1

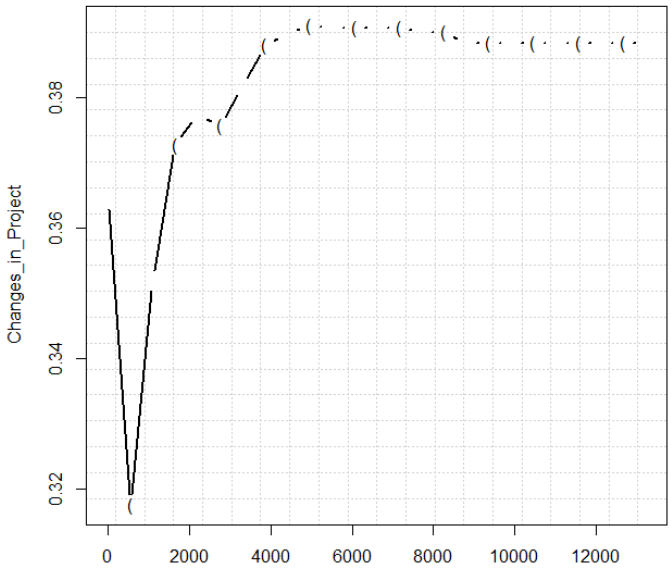

Figure 43 - Internal estimated effort importance plot 2

11.Number of applications

Importance results:

Table 39 - Number of applications importance results

Number of applications	No	Yes
1	0.68145	0.31855
3	0.68035	0.31965
4	0.682532	0.317468
6	0.675621	0.324379
8	0.67005	0.32995
9	0.67001	0.32999
11	0.673212	0.326788
13	0.673606	0.326394
14	0.669212	0.330788
16	0.669298	0.330702
18	0.656676	0.343324
19	0.656372	0.343628
21	0.65534	0.34466
23	0.655168	0.344832
24	0.655759	0.344241
26	0.655921	0.344079
28	0.656472	0.343528
29	0.656498	0.343502
31	0.656542	0.343458
33	0.656669	0.343331
34	0.656669	0.343331
36	0.656432	0.343568
38	0.65688	0.34312
39	0.65688	0.34312
41	0.65683	0.34317

Importance VEC Plots

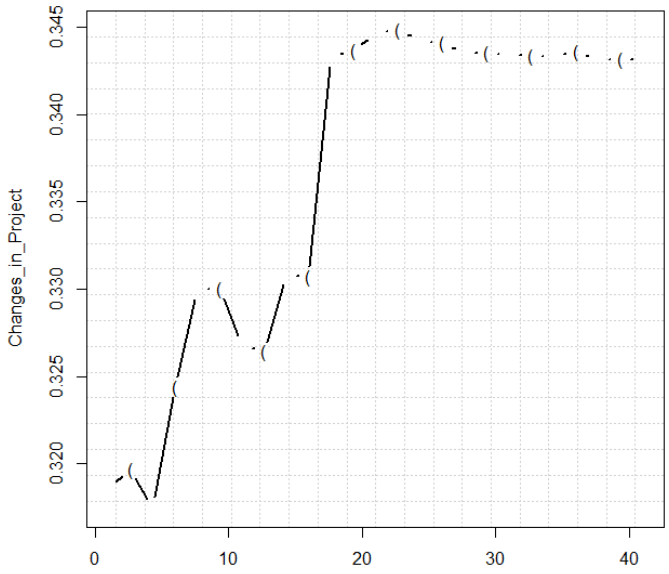

Figure 44 - Number of applications importance plot 1

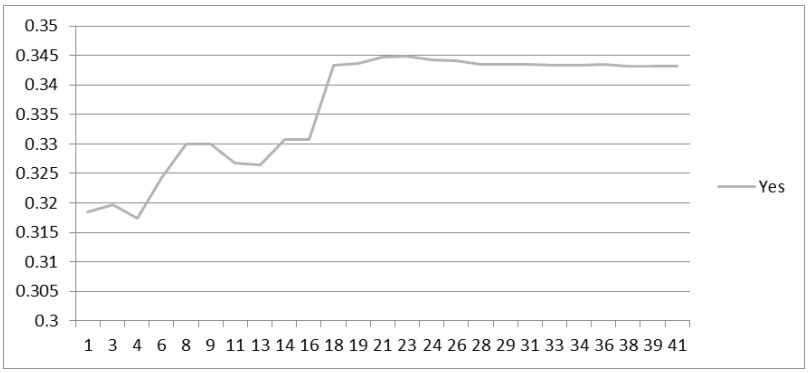

Figure 45 - Number of applications importance plot 2

12.Number of internal clients

Importance Results

Table 40 - Number of internal clients importance results

Number of internal clients	No	Yes
1.0	0.685024	0.314976
1.3	0.685024	0.314976
1.7	0.666996	0.333004
2.0	0.666996	0.333004
2.3	0.666558	0.333442
2.7	0.663252	0.336748
3.0	0.663252	0.336748
3.3	0.66415	0.33585
3.7	0.667528	0.332472
4.0	0.667528	0.332472
4.3	0.668031	0.331969
4.7	0.669082	0.330918
5.0	0.669082	0.330918
5.3	0.669304	0.330697
5.7	0.663055	0.336945
6.0	0.663055	0.336945
6.3	0.657713	0.342287
6.7	0.654564	0.345436
7.0	0.654564	0.345436
7.3	0.654564	0.345436
7.7	0.654338	0.345662
8.0	0.654338	0.345662
8.3	0.653092	0.346908
8.7	0.650921	0.349079
9.0	0.650921	0.349079

Importance VEC plots

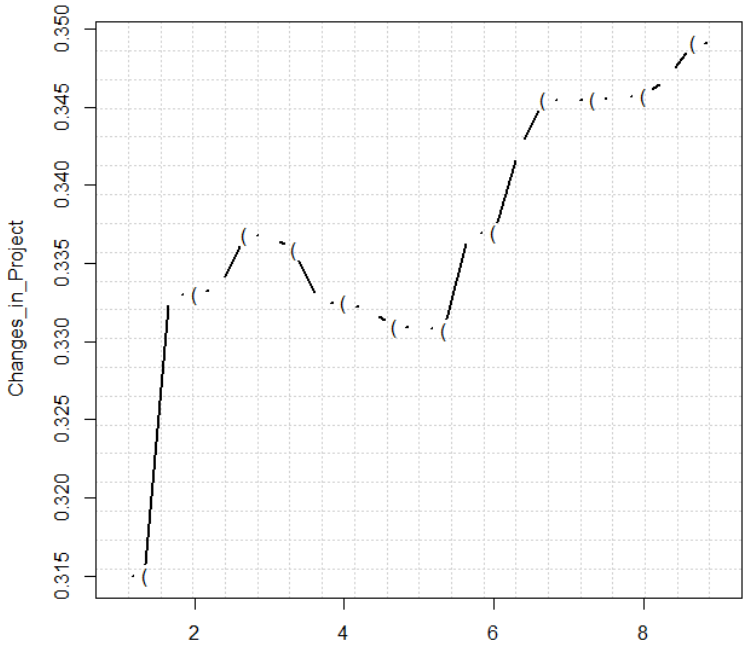

Figure 46 - Number of internal clients importance plot 1

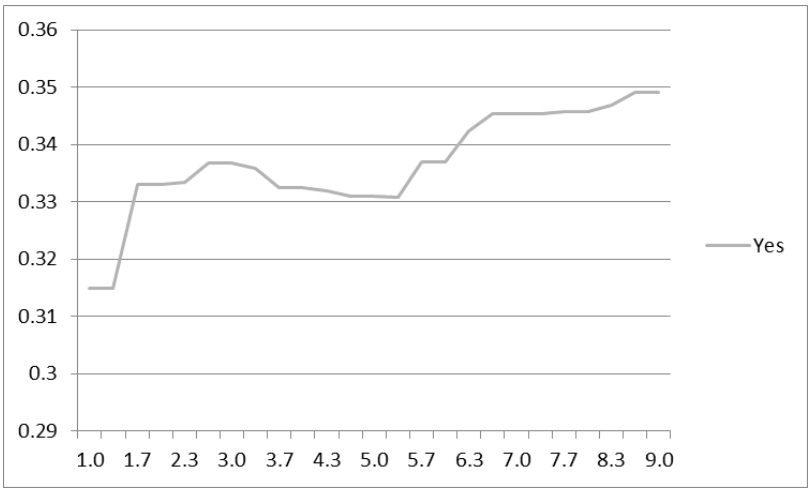

Figure 47 - Number of internal clients importance plot 2

13.Number of outsourcing teams

Importance results

Table 41 - Number of outsourcing teams importance results

Number of outsourcing teams	No	Yes
0	0.66902	0.33098
1.041667	0.675712	0.324288
2.083333	0.686866	0.313135
3.125	0.688106	0.311894
4.166667	0.687872	0.312128
5.208333	0.688184	0.311816
6.25	0.678065	0.321935
7.291667	0.684018	0.315982
8.333333	0.682513	0.317487
9.375	0.684535	0.315465
10.41667	0.684588	0.315412
11.45833	0.68455	0.31545
12.5	0.684363	0.315637
13.54167	0.679865	0.320135
14.58333	0.680107	0.319893
15.625	0.680096	0.319905
16.66667	0.680254	0.319746
17.70833	0.680354	0.319646
18.75	0.680295	0.319705
19.79167	0.680295	0.319705
20.83333	0.680295	0.319705
21.875	0.680295	0.319705
22.91667	0.680295	0.319705
23.95833	0.680257	0.319743
25	0.680257	0.319743

Importance VEC plots

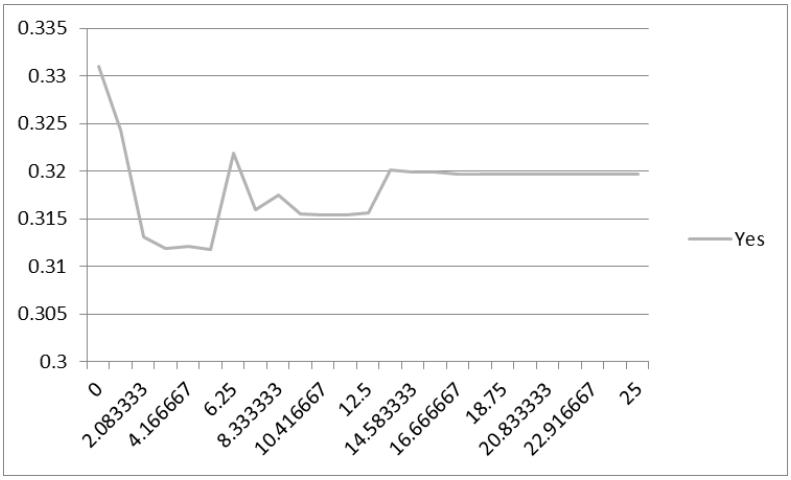

Figure 48 - Number of outsourcing teams importance plot 1

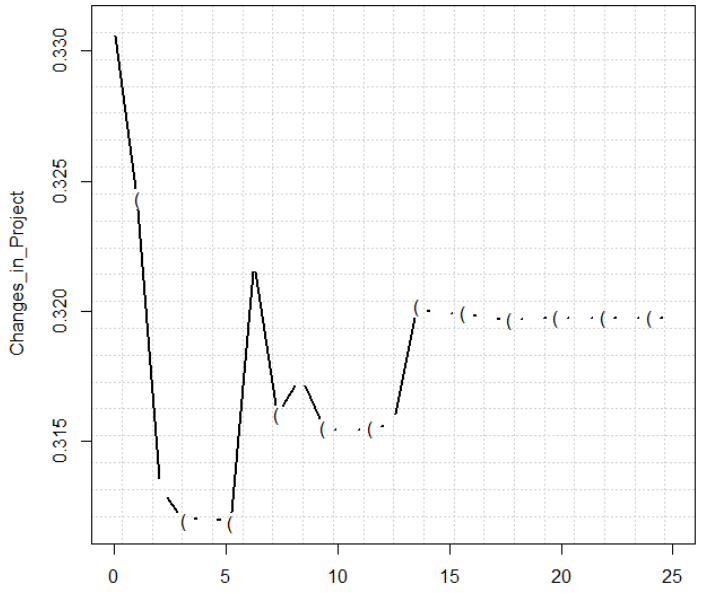

Figure 49 - Number of outsourcing teams importance plot 2

14.Outsourcing estimated effort

Importance results

Table 42 - Outsourcing estimated effort importance results

Outsourcing estimated effort	No	Yes
0	0.684466	0.315534
853.875	0.617987	0.382013
1707.75	0.622081	0.377919
2561.625	0.624855	0.375145
3415.5	0.625299	0.374701
4269.375	0.624608	0.375392
5123.25	0.624636	0.375364
5977.125	0.62416	0.37584
6831	0.624231	0.375769
7684.875	0.624316	0.375684
8538.75	0.624316	0.375684
9392.625	0.624316	0.375684
10246.5	0.624316	0.375684
11100.38	0.624316	0.375684
11954.25	0.624316	0.375684
12808.13	0.624316	0.375684
13662	0.624316	0.375684
14515.88	0.624316	0.375684
15369.75	0.624316	0.375684
16223.63	0.624316	0.375684
17077.5	0.624316	0.375684
17931.38	0.624316	0.375684
18785.25	0.624316	0.375684
19639.13	0.624316	0.375684
20493	0.624316	0.375684

Importance VEC Plots

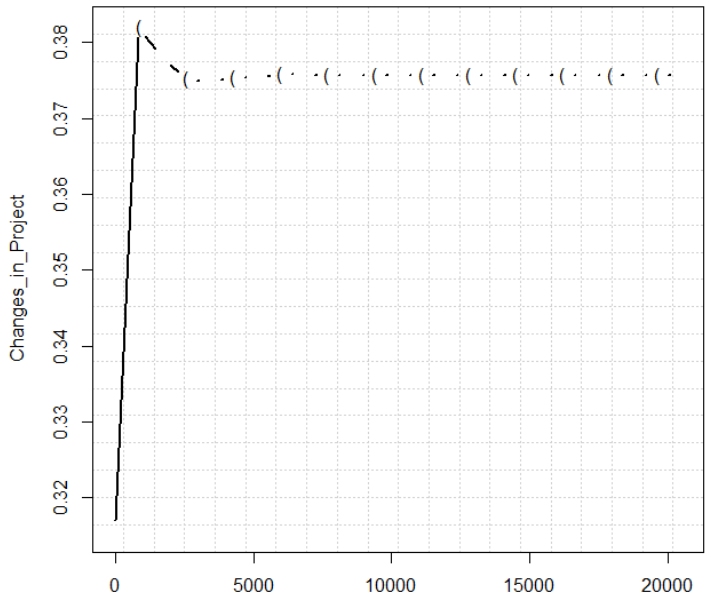

Figure 50 - Outsourcing estimated effort importance plot 1

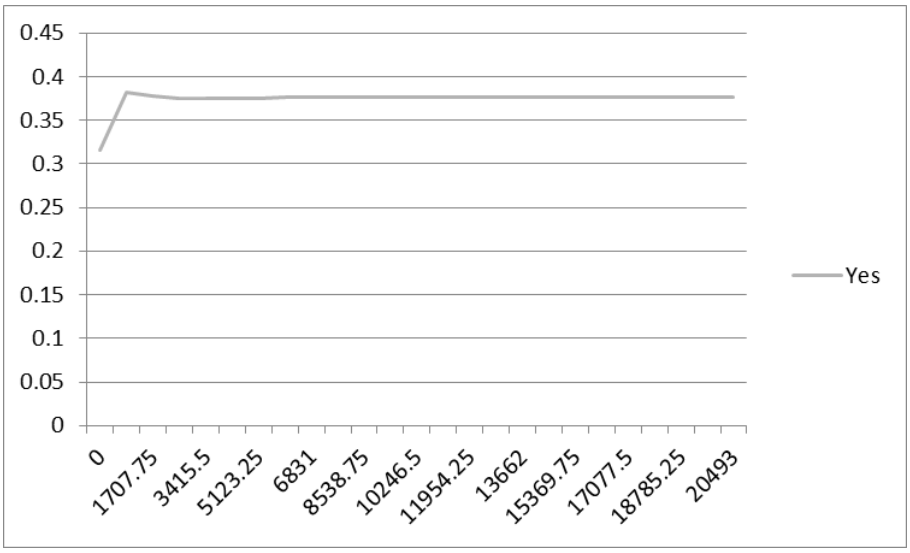

Figure 51 - Outsourcing estimated effort importance plot 2

15.Number of internal resources

Importance results

Table 43 - Number of internal resources importance results

Number of internal resources	No	Yes
1	0.66647	0.33353
5.666667	0.678237	0.321763
10.33333	0.678948	0.321052
15	0.671539	0.328461
19.66667	0.669994	0.330006
24.33333	0.665568	0.334432
29	0.667261	0.332739
33.66667	0.666044	0.333956
38.33333	0.666536	0.333464
43	0.66652	0.33348
47.66667	0.666335	0.333665
52.33333	0.66635	0.33365
57	0.66635	0.33365
61.66667	0.666325	0.333675
66.33333	0.66461	0.33539
71	0.664053	0.335947
75.66667	0.664053	0.335947
80.33333	0.664034	0.335966
85	0.663919	0.336081
89.66667	0.662442	0.337558
94.33333	0.662442	0.337558
99	0.662442	0.337558
103.6667	0.662442	0.337558
108.3333	0.662442	0.337558
113	0.662442	0.337558

Importance VEC Plots

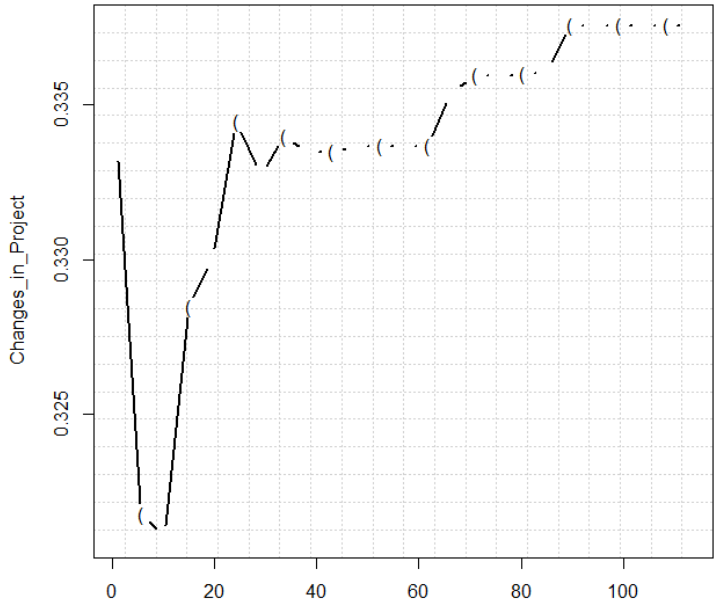

Figure 52 - Number of internal resources importance plot 1

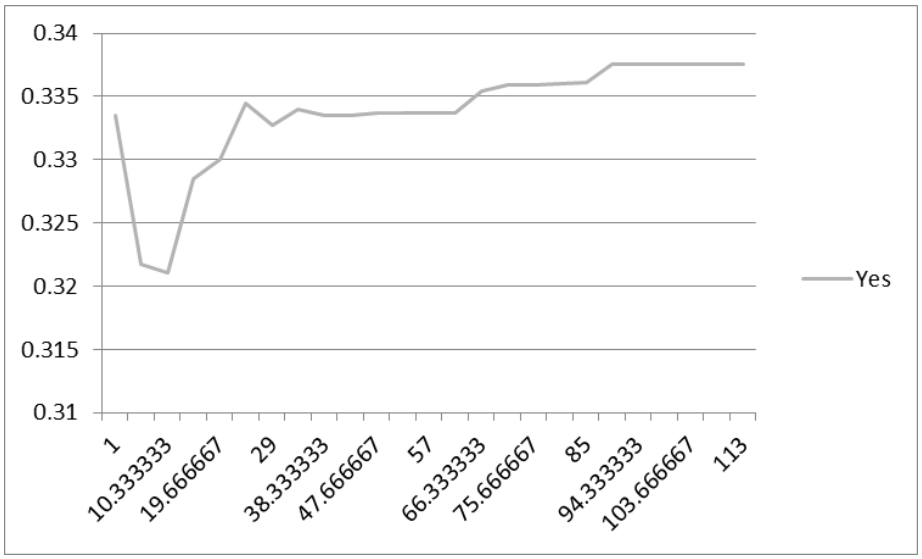

Figure 53 - Number of internal resources importance plot 2