INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA DE INTELIGÊNCIA ARTIFICIAL




Metrix Analyzer Automatic Insights generator for teams performance

MIGUEL FRANCISCO COUTINHO PANDA
Setembro de 2024

P.PORTO

Metrix Analyzer – Automatic Insights generator for teams performance

Miguel Francisco Coutinho Panda

Aluno nº: 1190896

Dissertação para obtenção do Grau de
Mestreem Engenharia de Inteligência Artificial

Orientador: António Constantino Lopes Martins, Professor Adjunto do Instituto Superior de Engenharia do Porto do Instituto Politécnico do Porto

Co-orientador: Carlos Fernando da Silva Ramos, Professor Coordenador Principal do Instituto Superior de Engenharia do Porto do Instituto Politécnico do Porto

Júri:

Presidente:

Isabel Cecília Correia da Silva Praça Gomes Pereira, Professora Coordenadora do Instituto Superior de Engenharia do Porto do Instituto Politécnico do Porto

Vogais:

Arguente: Dalila Alves Durães, Professora Auxiliar da Universidade do Minho
Vogal: António Constantino Lopes Martins, Professor Adjunto do Instituto Superior de Engenharia do Porto do Instituto Politécnico do Porto

Porto, September 2024

Abstract

The African continent is witnessing a technological renaissance, with its market competitiveness escalating at an unprecedented pace. In this rapidly evolving landscape, software companies must swiftly adapt to not only survive but thrive. Central to this endeavour is the imperative to enhance team performance, as it is vital for sustaining profitability and driving innovation. As demand for comprehensive performance analysis grows across teams, Agile Coaches are increasingly stretched thin, balancing the urgent need for insights with their ongoing projects. This challenge underscores the critical necessity for advanced automation—a solution that can streamline the generation of actionable insights while requiring fewer human resources. It was within this context that the Metrix project was born. This dissertation examines the detailed and step-by-step process of developing a hybrid expert system solution designed to meet the growing needs of the company. It covers the technical research carried out, the careful analysis of requirements, and the practical application design that laid the groundwork for the project. Additionally, it describes the data preparation and the experimentation phase that led to accuracy results of 85%. By the conclusion of this thesis, the results and insights resultant from this journey will be presented, illuminating the transformative potential of the Metrix project in revolutionizing team performance analysis and paving the way for future advancements.

Keywords (Theme): *Process Mining*, *Recommendations module, Automate, Forecast, Predictability*

Keywords (Technologies): *Java, Python, Decision Tree, Expert System*

Resumo

A evolução tecnológica e a competitividade do mercado no continente africano estão a crescer a um ritmo sem precedentes Neste panorama de rápida evolução, as empresas de software devem adaptar-se rapidamente não apenas para sobreviver, mas para prosperar. Central a este esforço está a imperativa de melhorar o desempenho das equipas, uma vez que é vital para sustentar a rentabilidade e impulsionar a inovação. À medida que a procura por análises de desempenho cresce entre as equipas, os Agile Coaches estão cada vez mais sobrecarregados, equilibrando a necessidade urgente de gerar recomendações para as equipas com os seus projetos em curso. Este desafio sublinha a necessidade crítica de automação avançada—uma solução que pode gerar recomendações automaticamente requerendo menos recursos humanos. Foi neste contexto que nasceu o projeto Metrix. Esta dissertação examina o processo detalhado e passo a passo de desenvolvimento de uma solução de sistema pericial híbrido, concebida para atender às crescentes necessidades da empresa. Abrange também a pesquisa técnica realizada, a análise cuidadosa de requisitos e o design prático da aplicação que estabeleceu as bases do projeto. Além disso, descreve a preparação de dados e a fase de experimentação que culminou em resultados com precisão de 85%. No final, serão ainda apresentados os resultados e insights resultantes do projeto, iluminando o potencial transformador da aplicação Metrix na revolução da análise de desempenho das equipas e abrindo caminho para avanços futuros.

Palavras-chave(tema): *Process Mining*, Módulo de recomendações, Previsibilidade

Palavras-chave(tecnologias) : *Java, Python, Árvores de decisão, Sistema Pericial*

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my family, especially my parents, for their unwavering support and understanding throughout these challenging academic years. Their encouragement has been invaluable to me.

I am also immensely grateful to my girlfriend Marta Silva for her constant companionship and for being my source of inspiration during my academic journey. The belief in my abilities has been a powerful motivator, helping me stay focused through the most demanding moments.

Secondly, I would like to recognize the invaluable guidance and continuous availability of my ISEP supervisor, Professor Constantino Martins. His insightful feedback, exceptional expertise and consistent encouragement were key factors that helped me navigate the many challenges I faced.

I am also grateful to my ISEP co-supervisor, Professor Carlos Ramos, for offering me the opportunity to join this innovative master's program.

Thirdly, I would like to express my appreciation to the company I collaborated with during the project—JUMIA—for allowing me to integrate into the Engineering Experience team and align my work with my master's thesis. The constructive and supportive environment not only enabled the successful completion of this thesis but also contributed significantly to my professional growth. For this, I am profoundly thankful to the members of the Engineering Experience team: Emanuel Miranda, Alan Conceição, Junior Staudt, Fabiano Freire, Virgínia Gomes, Heba Elatfy, and Brenda Belíssimo.

I would also like to offer my heartfelt thanks to my company supervisor, Bruno Aguiar, whose invaluable insights during the second phase of the project were instrumental in its success.

Finally, I extend my warmest appreciation to the friends I made throughout the course, especially João Azevedo, Gaspar Lopes, and Diogo Risca, who, in their own unique ways, contributed to the completion of this master's degree.

Index

Figures Index

Tables Index

Code block Index

Acronyms and Symbols

Acronyms List

AI	Artificial Intelligence
AIG	African Internet Group
API	Application Programming Interface
BDD	Behaviour Driven Design
CI/CD	Continuous Integration / Continuous Deployment
CLI	Common Language Infrastructure
DORA	DevOps Research Assessment
DSR	Design Science Research
ERD	Entity Relationship Diagram
ES	Expert System
EXP	Engineering Experience
GDPR	General Data Protection Regulation
KPI	Key Performance Indicator
IDE	Integrated Development Environment
IT	Information technology
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QA	Quality Assurance
RFR	Ready for Development
TBD	Trunk Based Development
TDD	Test Driven Design
TOC	Theory of Constraints
UI	User Interface

UX	User Experience
USD	Unite States Dollar
WIP	Work In Progress

Symbols List

L	Work in Progress
Ɵ	Threshold
λ	Throughput
W	Lead Time

1 Introduction

At this initial chapter there are explored at some extent topics that will provide a better understanding of the entire project, such as the problem in hands, the company context and the technologies used for the solution development.

1.1 Context

Process mining and efficiency metrics are worldwide used at many companies. According to a process mining survey conducted by Deloitte in 2021, 83% of the companies have already in place some sort of process mining mechanism at a global scale (Deloitte, 2021).

The process of achieving data-driven recommendations is simple, although sometimes it is manual and time consuming: data needs to be gathered, organized and interpreted by someone or some team that understands Agile principles. Ultimately, recommendations to improve teams' performance are written and given to the teams so they can react to it (Davis, 2015b).

Process mining is an area undergoing constant advancements and investment. The process mining market report made in 2023 highlights that process mining related projects are predicted to increase their investment from the actual USD 1.8 billion to USD 12.1 billion in the timeframe of a year (*Process Mining Market by Offering (Software (Process Discovery Tools, Conformance Checking Tools), Services), Mining Algorithm (Deep Learning, Sequence Analysis), Data Source (ERP Systems, CRM Systems), Vertical and Region - Global Forecast to 2028*, n.d.). That means an unprecedent increase of approximately 572.22%

Pan-African company Jumia Technologies AG is an e-commerce company that runs their commercial activities at the African continent. In 2012 the Africa Internet Group (AIG) launched Jumia in Nigeria, Morocco, South Africa and Egypt.

At this moment, the African market is facing an expansion with several companies performing investments in it (Sampene et al.,2022). To level up with this demand, JUMIA invested time and resources in pursuing the best possible performance for their teams. This action was assigned to a new department – Engineering Experience (EXP)

The Engineering Experience department was created by two technical Agile Coaches with a strong technical background that established the foundations and produced the basis for different initiatives. Later, four other Agile coaches with a product background joined the department, bringing different perspectives into the work being done.

From the different programs that the team was dynamizing at the company, one that gained high priority was the indicators measurement. This program was about collecting team efficiency and Dev-Ops Research Assessment (DORA) metrics to analyze them and produce useful insights and recommendations to enhance the teams performance and redirect the team towards a better development experience (Mauvius, 2022).

1.2 Problem Statement

Even with all the advantages that efficiency Metrics analysis brings to a business, this area demands a huge amount of specialized human resources. From the beginning to the end of each team analysis some actions should take place (Davis, 2015a).

First, the data needs to be manually extracted from any tool or "team board" where the teams are tracking their work progress. At the next step, the data must be reviewed and get some special handling, such as outliers identification and duplicated data removal. The third and last phase is to perform the analysis, write and share the recommendations and insights with the directors and teams.

At this analysis, there is an additional problem that must be considered - different teams have different workflow definitions (Liu et al., 2008). This lack of standardization makes the task even more time-consuming for the agile coaches with a significant increase of human error possibility (Verbruggen et al., 2019).

With concrete calculations, using JUMIA company as example with around 15 software departments, this entire process can take more than an entire work week for a team of 8 Agile Coaches.

The software teams take high advantage of this analysis (Saltz et al., 2022). However, with the available resources it is impossible to scale this process to 30-50 teams at companies that are using this manual and time-consuming process.

There is a need for a process automation, that could provide a significant time reduction at the extraction step. Additionally, it should automatically handle the data treatment addressing the lack of teams workflow standardization and even produce analysis recommendations by itself. This would simplify all the human work described above to an easier review and adjustment task.

1.3 Objectives

The Metrix project goal is to solve the automation and insights production problem and provide answers to the below mentioned research questions by using a hybrid system with decision trees and an expert system based in heuristics.

The research question that is supporting this thesis is:

Can a hybrid expert system, integrating Agile professional's insights and decision trees, effectively address brain drain and offer practical recommendations to enhance team processes and efficiency in technological companies?

The project goals are:

1. Automate the extraction of efficiency metrics allowing it to scale to all delivery teams at the company.

2. Reduce the manual time spent on data treatment actions by 50%. Including the processing of dates format and solution for the lack of workflow standardization.

3. Automatically update the spreadsheet with the recent team's data at a defined recurrent time interval such as weekly, biweekly or monthly as intended and required to produce the best insights.

4. Provide automatic recommendations based on metrics for a given group (team, vertical)

5. Answer the scientific hypothesis about the use of a hybrid expert system integrating decision trees to solve the intended problem.

1.4 Expected contributions

The Metrix project aims to solve the stated problem by making the Flow Metrics analysis process more automatic, scalable and practical, providing teams with updated and insightful performance recommendations.

For companies relying on Agile Coaches to perform a manual process, this project offers a potential solution to maintain team performance improvements while freeing up time spent on these tasks. This would allow human resources to be allocated to other projects and initiatives. This approach can lead to cost reductions and resource management benefits for the company.

For software development teams, the Metrix project has the goal of providing accurate recommendations and pointing out problems in the team workflow. The teams can benefit from this analysis by replacing the time they would allocate to identify bottlenecks or points of their workflow where they are underperforming with discussions and development of considerable improvements on their daily work in a data-driven environment.

This project offers a potentially compelling hybrid solution for the scientific community by combining decision trees with an expert system driven by heuristics. It aims to provide a cost-effective alternative to existing tools for data extraction and organization, while also introducing an innovative automatic recommendation generator.

1.5 Research Methodology

At this thesis, the research methodology used is Design Science Research (DSR), since it suits the development and assessment of innovative projects where the existing solutions are not full established (Lawrence et al., 2010).

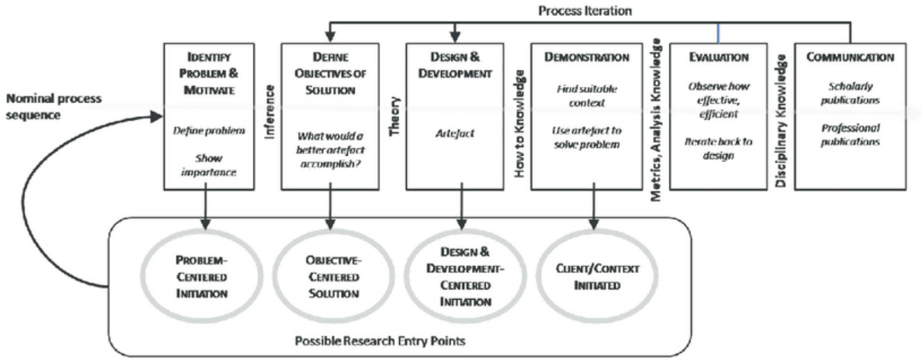


Figure 1 - Design Science Research Diagram (Peffers et al., 2007)

The DSR methodology provides a systematic approach that addresses complex problems through the creation and evaluation of innovative artifacts (Weber, 2018).

DSR has a high rate of uses in computer science projects (Storey & Baskerville, 2022) since it mixes the traditional scientific research with the practicality of design processes. It emphasizes the development of solutions to real-world problems (vom Brocke, Gau, et al., 2021).

Starting with the problem statement, at sub-chapter 1.2., where a solution gap is identified at efficiency metrics analysis. Subsequently, outlining the research explicit objectives that will guide the development process, and at the following chapters the literature review where ideas and innovative ways are explored on how to develop the possible problem solution.

As DSR demands there is a design and development stage where the intended solution, the data needed, the data gathering process, the ethical considerations regarding data privacy and the solution experimentation are documented.

At the end, regarding the DSR structure, Figure 1, a demonstration model, containing the solution results and final project considerations brings an essential perspective on the outcomes and takeaways of this innovative solution proposal.

Concurrently, at the literature review the study incorporates a simplified version of a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. This systematic review ensures a thorough and transparent exploration of existing literature, enriching the research with a broader context. Specific particularities related to the application of PRISMA guidelines are elucidated, addressing the modifications unique to the research (Page et al., 2021).

1.6 Deliverables/Outcomes

In terms of overall work management, the project was divided in different deliverables with different duration, deadlines and outcomes, table 1, correlated with the DSR structure, figure 1.

Table 1 - Metrix project work distribution

Milestone	**Initial Date**	**End Date**	**Outcome**
Milestone 0 – Project discovery and research	06/12/2022	27/12/2022	Technical Decisions (programming language, software architecture)
Milestone 1 – Technical Excellence basis and Efficiency Metrics Automation for some verticals	01/01/2023	22/03/2023	Integration with JIRA software Metrics extraction through IDE
Milestone 2 – Coaches using Metrix to extract Issues	23/03/2023	27/07/2023	Application launched to CLI
Milestone 3 - Automatize Metrix running	28/07/2023	28/10/2023	Application automatically generating csv file at spreadsheet
Milestone 4 – Recommendation module research	20/10/2023	20/02/2024	The decision trees and the expert system have good accuracy
Milestone 5 – Automatize Metrix recommendations	20/02/2024	23/07/2024	Application produces a text with correct recommendations

The Metrix project work was developed at JUMIA. Naturally, a company's scope and priorities change at a fast pace. The milestones were progressively defined accordingly to what would bring more business value at the moment.

The project was started from scratch and the first milestone focus was setting standards and a solid basis for what could come. Research was conducted to understand which programming language, software architecture and workflow would be the best fit for the project.

The following milestones, Milestone 1 and Milestone 2, were mainly about pure implementation. The integration with the JIRA software, the data handling and automatically storage at spreadsheets were achieved by a solid code basis.

After this milestones the project became a possible internal basis java project that could be used as reference for any team at the company when starting a new implementation, because of its architecture, tests practices and quality.

Prior to the achievement of this result, a ponderation and decision stage went on bringing the expected outcomes of Milestone 3 and 4 as the more rewarding objectives regarding business value to the challenging moment the company was facing.

A research on how to properly add a recommendations generator module to automate and save professionals time at the company was set for Milestone 4, and its prior implementation set to Milestone 5.

During the Metrix project timeframe, this thesis was set as objective, and its conclusion is expected to document the above-mentioned milestones, table 1.

1.7 Thesis structure

Apart from the introduction chapter, this report contains five more sections:

The next chapter focuses on the state of the art, providing context about the project domain while conducting a comparative analysis of existing work and similar technologies. Within this chapter, research questions are formulated to explore domain specific contributions in across different areas and identify possible related challenges.

The third chapter focuses on three key aspects, the data collection, the datasets formulation and the considerations regarding the data usage, privacy, security and ethical concerns that must be addressed.

In the fourth chapter, the development of the solution is explored in detail, covering the technological stack, workflow definition, software architecture, decision-making processes, and the design of the use cases, including the interactions between the different system components.

In the fifth chapter, the results from the decision tree experimentation are presented, along with the outputs generated by the application. This section also provides a comprehensive description of the evaluation process, detailing the experts contribution and the application usages.

The last chapter the overall results and conclusions are discussed including a critical view of the project limitations and of the possible future improvements.

2 State of the Art

This chapter begins with a conceptual approach of the domain where the software development technologies, process mining and efficiency metrics are contextualized and explained with detail. Their usage impact, the benefits and existent challenges are considered. At the second part, different technologies that could be used to solve the problem identified at the point 1.3 of this report are identified, presented and compared.

2.1 Conceptual Approach of the domain

This section analyzes software development methodologies, agile efficiency metrics, flow metrics, and their role in producing deliverable forecasts and recommendations, as well as their impact on team performance. It also identifies and presents various existing literature on these topics and their distributors in the market.

2.1.1 Software development methodologies

This sub-chapter focuses on four distinct software methodologies, each addressing critical facets of the development lifecycle: Behaviour Driven Design (BDD), Test-Driven Development (TDD), Continuous Integration/Continuous Deployment (CI/CD), and Trunk Based Development. These methodologies play important roles in shaping the way software is conceived, developed, tested, and delivered.

2.1.1.1 Behaviour Driven Design

Behaviour Driven Design (BDD) (Binamungu & Maro, 2023) is a software methodology that places a strong emphasis on collaboration and communication between different members of a software development lifecycle: testers, engineers and non-technical members.

BDD seeks to bridge the gap between business requirements and technical implementation by promoting the creation of executable specifications written in a natural language (Tsilionis et al., 2022) that is easily understandable by both technical and non-technical individuals. It also offers clear test scenario visualization and understanding, making it the approach of choice for quality control engineers (Bezsmertnyi et al., 2021).

BDD can be used in software design processes to support specification, transformation, and verification of behaviour designs (Lee, Huang, et al., 2014) and can play an important role in software understanding and simplification.

However, like all methodologies, BDD has its own set of challenges and limitations. Writing clear and meaningful scenarios requires time investment and effort, particularly for teams that are new to the approach.(Nascimento et al., 2020) The learning curve can vary from team to team and it may have a negative performance at the short term.(Binamungu et al., 2018)

Additionally, BDD effectiveness can be challenged when the characteristics of a project are complex or purely technical and where the behaviour is difficult to express in business terms. Despite these limitations, BDD has proven to be an invaluable tool for improving software quality, fostering collaboration, and aligning technical work with business goals (Pereira et al., 2018).

2.1.1.2 Test Driven Design

Test Driven-Design (TDD) software development methodology that flips the traditional development process by placing testing at the forefront of the workflow. (Janzen & Saiedian, 2006).

This iterative cycle involves three main steps: writing a failing test, implementing the minimum code necessary to make the test pass, and then refactoring the code while ensuring the tests still pass (Farcic & Garcia, 2015).

TDD is a practice that aims to improve product quality and maintainability and help developers and software engineers to easily identify errors and problems at the code. (Farcic & Garcia, 2015).

The use of TDD is sometimes surrounded by criticism and debate. Some argue that using TDD can delay the team and the project deadlines due to the extensive tests. (Ünlü, 2008). On the other hand, experts argue that thinking about tests and cover all possible corner cases is an frequently overlooked preventive measure that can save valuable time later in a project.(Bhadauria et al., 2020)

Furthermore, TDD fosters a culture of collaboration between developers and testers, encouraging shared responsibility for the quality of the final product.(Owen & Varhol, n.d.) In many cases, TDD can be used in conjunction with other methodologies like the BDD, explained at the previous section, to ensure that both the technical and business requirements of the software are met(Eriksson, 2023).

2.1.1.3 Continuous Integration and Continuous Delivery

Continuous Integration / Continuous Deployment (CI/CD) principles had a huge impact at software development by automating and streamlining key processes. This continuous integration allows developers to push changes more frequently and consequently achieve faster feedback about possible problems (Hroncova & Dakic, 2022). The continuous deployment is the automation on the updates deployment to different environments. (Hroncova & Dakic, 2022).

CI/CD pipelines automate the steps involved in building, testing, and deploying software, offering numerous advantages to development teams. However, while the benefits of CI/CD are substantial, there are also security considerations that need to be addressed. The automation of deployment processes can potentially introduce vulnerabilities if not managed correctly.(Poth et al., 2018) Development teams must ensure that security practices are integrated into the CI/CD pipeline to mitigate risks such as unauthorized access, data breaches, and infrastructure damage.(Loriot & Lebossé, 2021).

On the other hand, by applying continuous integration with the correct appropriate preventive measures, possible errors are reported with more frequency what can lead the team to perform adjustments or discover a solution in a timely manner before the incident could scale. (Jarmusch et al., 2024)

2.1.1.4 Trunk Based Development

Trunk Based Development (TBD) is a software development methodology that promotes the speed and simplicity at the version control process by addressing the use of a single branch at the development process (Shaked & Reich Phd, 2019).

This methodology is the opposite as the traditional feature branches that in complex projects can become a development dependency. This strategy can be used together with Continuous delivery what can maximize their use advantages (Schneid, 2017). In order to successfully use them together it is recommended that feature flags and some control mechanisms are implemented as prevention on the deployment side (Schneid, 2017).

TBD also is distinguished from other methodologies by the collaborative environment that it generates, promoting the development team to work together and not individually divided into feature branches (Malohlava et al., 2013). This collaboration leads to more effective communication and quicker resolution of issues. The practice of merging code changes frequently promotes collective ownership of the codebase, creating a sense of shared responsibility among team members. Furthermore, by reducing the complexity associated with multiple branches, TBD allows for more straightforward debugging and testing processes. (Helsinki & Wikström, 2019)

TBD can also be combined with preventive mechanisms like the use of feature toggles to assure even more the code quality. These control mechanisms allow a feature to be turned on or off if needed. Using this sort of mechanism can be effective when different team members are constantly integrating their code at the same unique branch. (Prutchi et al., 2022)

2.1.2 Process mining

Process mining is widely used to improve processes and enhance efficiency. It is used among different topics and industries, such as education, finance, public works, software development, healthcare, e-commerce, manufacturing, between many others.

Process mining uses data science and process analytics with a common purpose of bottlenecks and process improvements identification (What Is Process Mining? | IBM, 2020).

This powerful mix between data science and process science, uncovers an entire discovery process, from the data gathering and mining phase to the workflow management and analysis (vom Brocke, van der Aalst, et al., 2021), figure 2, by implementing process mining the decision-making is performed in a data-driven environment what can lead to more reliable decisions. (vom Brocke, van der Aalst, et al., 2021)

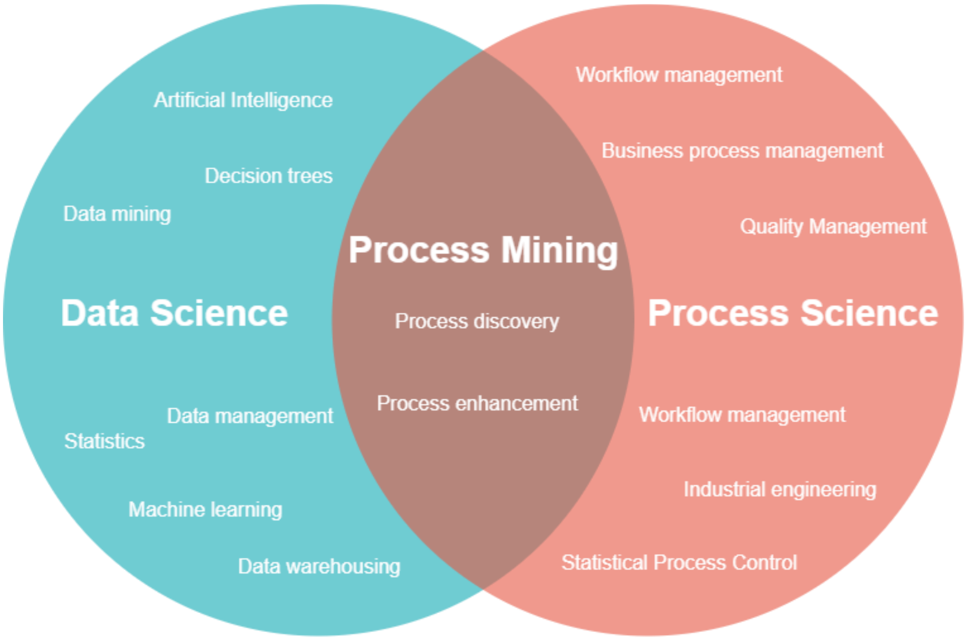


Figure 2 - Process Mining Diagram (adapted from (van der Aalst, 2022))

Software development is from the different areas where process mining is applied the one that has the strongest connection with what the project intends to solve. Software engineering Agile or Kanban methodologies, explained in detail at 2.1.3 are widely used approaches. This type of work tracking management helps the data science phase of process mining, since the teamwork data is stored in common tools such as JIRA, Trello.

In 2019 a study was conducted where different software development teams under the same conditions (same programming language and integrated development environment - IDE) were given the same task to do. By capturing the resolution activities of each team and by using process mining techniques to discover development process models and evaluate their quality. The results showed that the less complex development processes lead to the best efficiency results, highlighting once more the importance of Process Mining to discover process improvements and efficiency (Staron & Meding, 2015).

Another common problem that software development teams often face when trying to keep a good delivery rate are bottlenecks. These can be found in processes and workflows and can limit the efficiency and capacity of a team (Staron & Meding, 2015).

For example, a team using the common Kanban four stages workflow with "Backlog, In Development, Review/In Quality Assurance (QA) and Done" has five tasks that may go through the workflow to achieve a successful conclusion. However, the workflow has a bottleneck at the Review stage. At this scenario, all five tasks conclusion would be delayed, and the performance of this workflow would point downwards.

This simple explanative scenario may provide the wrong idea that bottleneck identification is a simple task, but some variants are important to consider and can disrupt the simple interpretation. The reason for a stage to take more time than others can simply be the nature of the step, a dependency from other team or an external factor.

A formal case study at Ericsson company highlighted the importance of identifying and removing these bottlenecks to improve efficiency. At the study, a measurement system with several indicators was used to identify bottlenecks at the workflow. Bottlenecks can be process killers. They can sometimes be identified through specialized agile members at a team, by doing research or even by questioning team members (Raquel & Lopes, 2023).

When dealing with a data-driven system this identification can be less painful and ways to process improvement can be found (Ram et al., 2019, 2020a, 2020b).

2.1.3 Process Science

One of primary side of process mining is Process Science. It includes relevant topics such as business process, quality and workflow management. These are only possible since most software teams use some lean methodology to track their progress. (Petersen & Wohlin, 2011). Usually, the work stages are displayed in a team board where generally the members move the work items assigned to them to provide alignment and visibility to the other colleagues regarding the work they are doing.

2.1.3.1 Kanban efficiency metrics

Kanban methodology was first developed at Toyota with the goal of improving manufacturing efficiency. From those days back in the late 40s, kanban started to be applied to different areas. At software development, Kanban boards can be used with the idea of keep track of the progress on the tasks and work to be done.

Usually, a kanban board would have some standard stages during software development process (Camara et al., 2020) being:

- Backlog: The stage where a task that is planned to be done is waiting to enter the workflow.
- In Progress: The stage where a task was picked up from backlog and it is currently being developed.
- Ready for Release: When the task was already developed but is waiting to be released. Sometimes waiting for additional testing or waiting for a specific release process.
- Done: When the task went through the workflow and is completed.

It is important to mention that even though these four are main stages on Kanban (Mahnic, 2019), there are workflows with some adaptations due to the nature of the teamwork or specific business needs. As example:

- In review / In test /In QA: This can be part of the In Progress or Ready for Release stage, but some teams prefer to divide this step. At this point the task is going to be reviewed and tested by Quality Assurance.
- Frozen / Waiting for dependencies / Cancelled: These stages can be used at teams where their tasks have high probability of being stuck waiting for dependencies on other teams' development or when teams want to reserve a stage for the tasks that

are no longer the priority or for some reason were discarded not making sense to be developed anymore.

Since one Kanban characteristic is the adaptive environment (Ahmad et al., 2013a) many more stages can be added fulfilling each team's needs. Apart from each team workflow that can vary, when it comes to Kanban methodology there are four key metrics to observe, table 2.

Table 2 - Main efficiency metrics

Lead time	Time from the moment the customer makes a request to the time they receive something.
Cycle time	Time it takes the development team to work on the request and deliver it.
WIP (Work in progress)	The total number of tasks being worked on at the same time — whether by a team or an individual.
Throughput	The number of work items exiting from the system or a given part of it. Delivery rate.

These efficiency metrics, table 2, when analysed and compared with previous registered values can bring an unmeasurable value by improving predictability (Fagarasan et al., 2023) and providing visibility for everyone about how the team is performing (Vacanti, 2016).

2.1.3.2 Dashboards and metrics correlation

The efficiency metrics can be compiled and displayed by using dashboards or even reactive views. The dashboards are used as a tool to enhance metrics organization and visibility. The dashboard's use extends to different areas and businesses and is considered to improve the data readability and even support real time throughput decision making (Franklin et al., 2017).

The lack of visibility and knowledge about Key Performance Indicators (KPI) and team performance measurement is a challenge that is commonly identified in different areas such as logistics, where the teams strive to optimize their processes (Piela, 2017). With dashboard and visually adaptable models for sharing and monitoring the measures the teams can increase the KPIs understanding what in the end leads to improvements at their team-level processes.

Even if the dashboard use in metrics analysis can be a useful tool to simplify the information, the use of reactive dashboards is still considered to face some challenges such as age of the individuals and adaptability, cost implications, dependency on IT infrastructure and support, learning curve and training requirements (Kothandapani, 2019).

The existence of these challenges makes some companies opt from the traditional dashboard techniques in which the data is manually organized and analysed, and dashboards are used as support tools and simple visualization tools.

Recently dashboards are evolving from just a tool to display and present the most relevant information to a real analysis tool that can be used at the business layer incorporating new features such as scenario analysis (Kothandapani, 2019).

If these analyses are not automated, this can lead to more human resources allocation and efficiency costs depending on the size of the companies.

2.1.3.3 Little's Law

An important law that is commonly used for teams' performance measurement and combines some efficiency metrics, displayed at table 2, together at the same formula, is Little's Law (Leskinen, 2015).

$$L = \lambda W$$

L = WIP (Work In Progress)

λ = Throughput

W = Lead Time

Little's Law states that the average wait time for service from a system equals the ratio of the average queue length divided by the average processing rate (Leskinen, 2015).

In practical terms, it can be an indicator that goes against a common misassumption that when a team is lacking efficiency and delivering slowly it can be solved by adding people to the team. In that scenario the WIP would increase what will mean that the throughput will decrease. That raises the question about how can that efficiency problem be addressed? The answer lies in searching for problems or bottlenecks that the team can be facing at their processes and on that efficiency, metrics have an essential role (BUDACU & POCATILU, 2018).

2.1.3.4 Theory of Constrains

In the context of performance measurement, particularly in team environments, the Theory of Constraints (TOC) by Goldratt provides a valuable contribution by identifying the bottleneck or limiting factor within a team's workflow(“The Theory of Constraints,” 2013).

It helps the team in prioritizing improvements that can lead to significant gains in efficiency and effectiveness. This approach not only enhances team performance but also aligns with the broader organizational goals by ensuring that efforts are concentrated on the most impactful areas (Şimşit et al., 2014).

By following the Theory of Constraints if there is more than one bottleneck at a team's workflow the team effort should lean towards solving the one that is estimated to have the highest negative impact on the team performance (Gundogar et al., 2016).

The recommendation would be to tackle that problem with the highest number of resources possible. Once it is solved the effort should be redirected to the first step again and identify the next bottleneck that is negatively impacting the workflow (Şimşit et al., 2014).

2.1.3.5 Benefits and challenges of metrics analysis

Including team metrics analysis can bring unmeasurable value to the teams and to the company business. A data-driven system that measure teams performance through agile metrics can have several advantages:

- All the team members can get visibility on the work that is being developed at a micro and macro scale what improves collaboration (Kamath, 2023) (Ahmad et al., 2013b)
- Better performance can be achieved what can lead to better delivery predictability (BUDACU & POCATILU, 2018)
- Increase delivery speed, the work quality (Kamath, 2023.)
- Improve problem solving processes by helping their early identification (Ahmad et al., 2013b).
- Business priorities and requirements changes can be added mid-flow since everyone is aware of the progress being made, the process bottlenecks and used to a data-driven environment (Ahmad et al., 2013b).
- Can enable teams to negotiate and set reliable expectations for their deliveries (Ahmad et al., 2013b)

Even with all the presented benefits of applying metrics analysis to improve software development teams there are some challenges that need to be addressed to provide the best analysis. The fact that this sort of analysis relies on real data configures the first challenge because it implies that the data quality will directly impact the outcomes reliability. The data quality can be an issue since:

- Some teams suffer from communication issues; some members lack visibility what can lead to poor milestone definition.
- Sometimes, the project can be rushed, and the tight deadlines leave insufficient time to correctly log all the data.
- Different software teams can have different workflows what can lead to confusion and difficulty in the professional's analysis (Raquel & Lopes, 2023).

Another challenge is the fact that these metrics require interpretation. Some of these metrics are simple and can easily reflect the development team's actual performance.

Although, interpretation is still a key factor being the lack of specialized skills and training one challenge for possible solutions (Ahmad et al., 2013b)(Ram et al., 2020b). That can be a challenge to possible solutions that try to use efficiency metrics data and provide insights for the teams because the interpretation is still done by specialized professionals that have the sensibility to understand the metrics and correlate them with the team's actual context what leads to reliable conclusions.

However, this sort of analysis requires time and human resources allocation what makes it a challenge to convince top level management that the usage of new kanban methods can be profitable and provide long-term results (Ahmad et al., 2013b).

2.1.4 Expert Systems and Decision Trees

The other crucial part of Process Mining is Data Science, figure 2. Data science is transforming decision-making across various industries by leveraging vast amounts of data to extract actionable insights and contribute to data-informed systems. In this section, we will explore key applications of data science, focusing on two essential tools: Expert Systems and Decision Trees (Majdzadeh, 2024).

2.1.4.1 Expert systems

Expert systems (ES) are characterized by their rule-based inference engines and knowledge representation frameworks (Eremia et al., 2016).

Over the years, expert systems have become essential in enhancing efficiency and precision across different industries. These software applications are designed to replicate human expertise, allowing high-quality decision-making in complex situations (Khan, 2015).

Organizations increasingly utilize expert systems to improve their operational efficiency, enabling them to make better-informed decisions while increasing productivity. Despite being less glamorous than deep learning models, expert systems have emerged as one of the most commercially successful branches of Artificial Intelligence (AI), demonstrating their practical value in real-world applications (Hingole, 2015).

Expert systems demonstrate an unparalleled capacity to sift through vast amounts of data, identify patterns, and offer informed recommendations or solutions (Chekushina et al., 2013).

In addition, expert systems are often used with an important educational role. The expert systems are crucial for organizations to preserve valuable knowledge that can be utilized for training new employees or informing future decision-making. (Jancura & Overbey, n.d.; Saleem & Moses, 1994). As organizations continue to recognize the importance of knowledge management and preventing brain drain, the expert systems role in replicating expertise will continue to increase (Saleem & Azad, 1994).

2.1.4.2 Decision Trees

Decision trees are one of the most powerful machine learning and artificial intelligence decision-making tools (Pathak et al., 2018) with wide applicability across diverse domains (Enayati et al., 2022; Lee, Liu, et al., 2014).

These graphical models have the capability of systematically breaking down complex decision processes into a series of simpler, hierarchical choices (Rutkowski et al., 2013).

This capability allows the conclusions to be easily interpreted (Blockeel et al., 2023) what can make the results of this Artificial Intelligence tool to be used and accessed even for naive individuals due to their tree nature and the simplicity with which they are visually explained (Pathak et al., 2018).

Originally, decision trees were conceived as simple flowcharts for decision-making but with the modern solutions, decision tree algorithms evolved to the point where there are integrated with advanced techniques and responsible Artificial Intelligence (Blockeel et al., 2023).

Decision Trees can produce high accuracy results which is a critical factor for businesses aiming to enhance their decision-making processes. They are particularly effective in handling both categorical and numerical data, making them suitable for a wide range of diverse applications, including customer segmentation, risk assessment, and predictive (Blockeel et al., 2023).

2.2 Review of Relevant Literature

At the current investigation, after the conceptual approach of the domain, a simplified version of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) has been applied (Page et al., 2021) . PRISMA stands as a well-established framework within the field of systematic reviews, offering a systematic and transparent approach to the conduct and reporting of such studies (Page et al., 2021).

The simplified PRISMA version used at this research uses a low number of search engines and repositories and a less exhaustive search model. The systematic review intends to provide answers to the research questions.

2.2.1 Research Questions

This systematic review has the following research questions:

Q1 – What is the impact of metrics analysis in different business areas and particularly in software development?

Q2 – What are the main strategies related to metrics found at different markets?

Q3 – What are the main strategies used to identify bottlenecks and improve processes and teams' performance?

Q4 – What are the existent gaps and the main domain related challenges?

2.5.2 Search Engines and Repositories

The papers included at this systematic review were carefully selected at the following repositories:

- **IEEE Xplore Repository**: This database is a reliable source of information since it provides access to a range of validated scientific papers that were published at official conferences and known scientific journals.

- **b-on repository**: This platform aggregates different repositories what provides access to a wide range of papers. The combination between software engineering, metrics, agile analysis and processes automation is a search topic in which the latest information can be challenging to be found. B-on can be even more valuable in a search context like this by its wide coverage of different scientific repositories.

Additionally, as search engine:

- **Google Scholar**: Given the wide range of the research theme, Google Scholar is a valuable search engine since it compiles different scientific repositories and important areas for this search such as product, software development and agile. The google scholar also offers an advance search customization that considering the project innovative characteristics can be a useful tool when defining more specific search queries.

2.2.3 Search Queries

To answer the research questions, the following keywords were defined:

Keywords: flow metrics & kanban & team & efficiency & automation & report & insights & generation & wip & slicing & backlog & limit & process mining

The defined search queries and results were:

Search Query 1 keywords flow metrics & kanban & team & efficiency & automation

The search query 1 produced around 9960 articles.

Search Query 2 keywords flow & metrics & kanban & team & efficiency & automation & report & insights & generation & wip & slicing & backlog & limit and Date 2021 – 2023

The search query 2 produced around 80 articles.

Search Query 3 process mining flow metrics kanban wip limit team code automation report insights and Date 2021-2023

The search query 3 produced around 25 articles.

2.2.4 Inclusion and Exclusion criteria

The articles selection is an important systematic review phase. This step ensures the inclusion of relevant studies that directly contribute to the research objectives. By defining clear and objective inclusion and exclusion criteria the research can have a better quality, because it guarantees that outdated or not totally related documentation is discarded from the research.

From the multiple articles obtained with the search queries, the following inclusion (Table 3) and exclusion criteria (Table 4) were used to filter the most valuable and relevant results.

Table 3 - Inclusion Criteria

ID	Inclusion Criteria
IC1	Abstract contains valuable information on team process improvements
IC2	Paper refers the value of applying efficiency metrics analysis in teams' performance
IC3	Paper shows the importance of data-informed systems applied in different areas
IC4	Paper identifies challenges and space to improve on metrics analysis automation topic

Table 4 - Exclusion Criteria

ID	Exclusion Criteria
EC1	Paper written before 2015
EC2	Paper not written in English
EC3	Paper that is neither a journal, article nor book
EC4	Paper that is not directly related with efficiency metrics analysis or team performance improvement topics

At this review, after the search queries produced their results, exclusion criteria were systematically applied to filter out documents that did not meet specific predefined criteria. Following this, the inclusion criteria were meticulously applied to identify and retain documents that aligned closely with the research objectives, Figure 3.

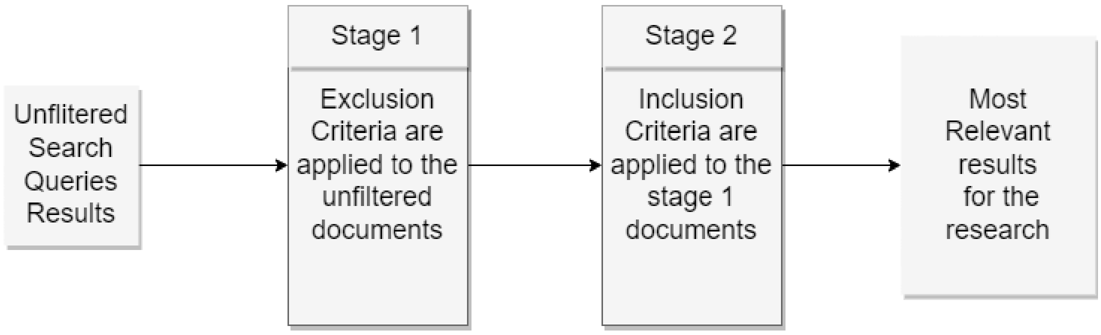


Figure 3 - Research strategy steps (Inspired (Page et al., 2021))

Proper documents selection can help avoid bias, enhance conclusion reliability and improve the topic comprehension. The exclusion criteria, table 4, helps with filtering the papers that do not have proof of scientific validation. Papers that were written before 2015 are also discarded because efficiency metrics area is growing with constant developments what can draw a line between relevant and outdated information.

On the other hand, due to the innovative nature of the research topic, the inclusion criteria, table 3, considers inclusion for sub related topics such as team process improvements, the use of data-informed systems among different business areas and papers that identify challenges or gaps at the metrics analysis automation.

2.2.5 Outcomes

The simplified version of PRISMA used at this systematic review produced the following results:

Table 5 - Research outcomes

Query	Number of Results	Included Number	Exclusion Number
Q1	9960 articles	~20	~9940
Q2	80 articles	~10	~70
Q3	25 articles	~10	~15

As visible in table 5, the research produced a significant number of results. Although by applying the inclusion and the exclusion criteria, a large number was discarded. Applying these criteria to the simplified version of PRISMA made sure that all the documents included in the research were highly relevant and related to the topic.

The use of Google Scholar as search engine for the research also contributed to a long number of results, allowing articles and papers from multiple sources and scopes to be part of the search queries results.

Due to the project's innovative nature, even with the large number of results, a small number ended up being included in the systematic review result. The uniqueness of the topic made it difficult to find literature that supports already existent products.

A vast number of Information regarding kanban methodology, efficiency metrics and bottleneck analysis was found during this literature review, but only a small number of sources applied these topics to the context of software teams. Also, the large number of challenges and gaps that were identified during the research proved that the team efficiency metrics and the analysis recommendation automation is still a field with a big prospect to grow and that in the next years is expected to receive more documentation and studies.

2.2.6 Research Outcomes

After the systematic review completion, it is important to answer the Research questions, with the knowledge obtained by the relevant documents found during the research.

2.2.6.1 What is the impact of metrics analysis in different business areas and particularly in software development?

Addressing the first research question, we navigate through the multifaceted impact of metrics analysis across various business sectors, with a particular focus on its significance in software development. (Kupiainen et al., 2015) The analysis of metrics in different business areas, particularly in software development, has proven to be unequivocally positive (Kirovska & Koceski, 2015). The influence it brings ranges from enhancing performance and decision-making support to providing strategic and reliable delivery predictability. (C. R. van der Vegt, 2021)

During this review, the use of metrics was identified in distinct areas, such as supply chain management, job shops, software engineering and automotive manufacturing. In all cases, these sectors experienced the benefits of metrics in their processes. (Aduwa et al., 2023; Braun, 2021; Saltz et al., 2022) This impact was felt either directly on profitability or by delivering high-quality products, improving the teams visibility and cooperation or fostering transparency in their goals and achievements (Ram et al., 2018).

The positive impacts extend to companies like Toyota, which utilize metrics analysis and Kanban principles, such as visualizing the workflow, limiting work in process and managing the flow (Hachemi & Bakhouche, 2024).

In the field of software engineering, metrics analysis is worldwide used and can bring significant positive impacts to teams. The primary goal of implementing efficiency metrics analysis is to identify ways to improve and adjust workflows leading to substantial impacts on the team level.(Shore, 2022; Sofia Barroso de Araújo et al., 2022; Visser, 2017)

However, there are challenges related to efficiency metrics implementation within software team specially when some members have some level of scepticism. Team members often have different perspectives and may lack trust in one another regarding to blame and merit allocation (Trainer & Redmiles, 2018). Nevertheless, this behaviour contradicts the intention behind metrics usage, which aims to promote a data-driven and a no-blame environment.

With metrics, team members can achieve better visibility of the work done. When outcomes are communicated positively, the team can benefit from the metrics usage in building of a constructive work environment and collaboration among their members (Lindstrom, 2023).

Overall, the use of metrics and its analysis revealed a positive impact across different business areas is a widely used tool used in the field of the software engineering. When correctly applied can significantly improve teams performance. (Heidenberg, 2019)

2.2.6.2 What are the main strategies related to metrics found at different markets?

The second research question aims to uncover the strategies associated with metrics in distinct markets. This exploration not only sheds light on innovative approaches but also draws connections between strategic choices and their implications across various business environments (Harrison & Lively, 2019; Kupiainen et al., 2015; Liang, 2021).

Metrics analysis reveals considerable differences between markets due to their unique characteristics, which can influence how the metrics are utilized and interpreted (Kirovska & Koceski, 2015). Despite these unquestionable differences, there some common strategies were identified.

The biggest similarities is the need to track the work progress and provide visibility to all the team members about the goals and accomplishments. This transparency fosters collaboration and promotes a sense of ownership, ensuring that everyone is aligned with the team objectives (Verbruggen et al., 2019; Καθηγητής & Λουκάς Τσιρώνης, 2024).

Additionally, the pursuit of process improvements is something that was observed across different contexts withing multiple business areas. In some domains the recruitment of agile specialists is a common strategy. The use of these professionals expertise not only aims to increase the team efficiency but also adapt the team to changing market demands. These improvements often include the metrics analysis to identify bottlenecks and discover flow enhancement opportunities (DeGrandis, 2017; Pekarcinova et al., 2021).

2.2.6.3 What are the main strategies used to identify bottlenecks and improve processes and teams performance?

The third research question is related to the metrics analysis role in identifying bottlenecks, optimizing processes and enhancing team performance. It aims to unravel the different methods utilized in diverse organizational contexts.

To effectively identify and address process bottlenecks, several key strategies emerged from the systematic review. The first is the importance of investing time to organize the team’s work and ensuring that every task is trackable. This involves setting up a clear workflow, assigning responsibilities and creating systems where each team member's progress is visible (Kadam, 2022; Viehhauser, 2023). By implementing these practices, the steps where delays occur can be pinpointed and the corrective actions can be applied (Kurkela , 2020.; Wang et al., 2024; Zalevskaya, 2019).

Another strategy is to leverage the teams work data by organizing it into dashboards or visual elements that enhance data interpretation (Raftoudi & Christina, 2023). Visual tools, such as charts or Kanban boards, simplify the teams understanding of their performance, making it easier to identify trends and spot inefficiencies (Fernando & Mattos, 2019; Forsberg et al., 2015.; Nogués & Valladares, 2017). However, this data analysis task demands a high level of expertise. A crucial factor in addressing process bottlenecks is the investment in specialized human resources who can analyse and identify inefficiencies within workflows (Greening, 2015; Van Looy, 2024).

These individuals or teams are often experts in process management, data analysis and optimization techniques (Gotz, 2021; Van Looy, 2024). By bringing in dedicated personnel to focus on these areas, organizations can benefit from a more informed and precise approach to identifying problems.

Including agile coaches can promote teams to operate with transparency and have a shared understanding of the metrics, encouraging the team members to participate and actively collaborate towards improve processes and their team performance (Kupiainen et al., 2015).

2.2.6.4 What are the existent gaps and the main domain related challenges?

Finally, the fourth research question examines the existing gaps and the challenges within the domain. This analysis provides a comprehensive understanding of the lacunes and limitations that current solutions have. All the different solutions that were found in this review have their particularities due to their own scope and business characteristics. However, some common limitations can be identified.

Most of the discovered solutions rely on a dashboard at some point of the process. However, even with the dashboard being a solid way to present and organize metrics, the analysis and the subsequent recommendations would always require the participation of a professional with specialized knowledge (C. Van Der Vegt, 2023). This kind of solutions represent an existent gap at this domain, because by depending on human resources participation, it tends to be less scalable, a higher number of teams would require a higher number of human professionals. The brain drain problem would also affect negatively this solution (Issagaliyev, 2023; Krishnan, 2023).

None of the found solutions had the capability to replicate the human behaviour in terms of interpretation by providing recommendations at the team (Weflen et al., 2022). That can be explained with the importance of teams and overall context to formulate insightful recommendations (Zeggaf, 2023).

This domain is intrinsically connected with the data quality. That can be identified as a challenge at this domain because the tracking of the work done is frequently overlooked inside some companies (Anttila, 2016). This poses a challenge to innovative solutions at this domain since solutions built on top of low-quality data could produce unreliable results (Kamath, 2023; Verbruggen et al., 2019).

2.2.7 Summary

The primary goal of this research was to identify existing projects, studies, or developments that align closely with this document scope. By using a simplified PRISMA version, a systematic review of relevant literature, technological advancements, and prior implementations in this domain was conducted to assess whether any existing solution would directly match the core aims of this project.

During the research phase, a wide range of sources was explored, including scientific papers, industry reports, and projects that were developed in business context within the field. This investigation had the goal of ensuring that existing methodologies, frameworks, or tools that could be directly used to solve the problem presented at subchapter 1.3 would be identified and thoroughly analysed.

Some interesting case studies were found that pinpointed the multiple advantages of efficiency metrics analysis not only in software engineering context but also across different areas. Better visibility about the processes, easier bottleneck identification and predictability were some of the highlighted advantages of efficiency metrics use. Also, some papers brought the interesting perspective that even with significant differences between business areas, efficiency metrics analysis can be applied to them and retrieve similar benefits in terms of process improvement.

This research also compiled some of the actions that real companies and teams are applying to identify their process bottlenecks and improve their processes. The studies highlighted the importance of investing time to organize the team's work and ensuring that every task is trackable. Organizing the work data into dashboards or visual elements that enhance data interpretation and the investment in specialized human resources, experts, who can analyse and identify inefficiencies within workflows were considered some of the most used strategies.

Considering that all the above-mentioned steps need human interaction at some point, the last part of the research focused on the identification of some applications and developments that intended to address this problem and turn the efficiency metrics analysis process more automatic and less time consuming. From the existent projects some challenges and limitations were highlighted such as the need for humans to manually perform the analysis, the lack of an automated software with capability to replicate human behaviour in terms of interpretation and the strong connection that exists between the data quality and an automated analysis that can be affected by the overlook of work tracking resulting in bad data quality.

In conclusion, after the entire research, no exact match to the specific objectives and features of the project were found. Even projects that were directly referring to the same domain, either addressed different aspects of the problem, utilized alternative approaches, or focused on distinct outcomes. None of the existing solutions fully encapsulated the unique combination of techniques and innovations proposed in this project. These results emphasize the innovative nature of the project by introducing a new perspective on how to address some of the existent gaps identified during the research.

3 Methods, Tools and Data collection

At this research the quality of the data provided is an essential part to achieve good results. At the following subchapters the data gathering and experimental processes are explained along with the ethical considerations and actions that took place regarding data privacy and protection.

3.1 Data gathering process

This research uses decision trees that receive input for its training with the research topic data. The first important step is understanding how many decision trees the analysis will be split into.

As stated on point 1.3, the research intent is to create reliable efficiency metrics insights by using an expert system powered by decision trees. The monthly insights provided by the Engineering Experience (EXP) department used five main graphics to support the analysis:

- **Work item input vs Throughput vs Average WIP** – This histogram provides a correlation between the items that enter the workflow (items in), the items that are delivered (Throughput) and the Average Work in Progress (WIP).

Higher Throughput and less items in will cause the WIP to decrease. Lower throughput and a higher number of items in will cause the WIP to increase. Work in Progress is an essential metric to a development team since having higher WIP values means that the team is probably working on multiple tasks at the same time (Lewis, 2013).

This constant changing of scope generally leads to more cognitive load to the individuals and will reduce the overall team performance and ability to meet the deadlines. On the other hand, a low WIP limit can block the team if there is any third-party dependency, because the team must wait for the dependency to not break the WIP limit (Skeie, 2015).

There is no ideal WIP value, and it will depend on team to team, from context to context, but by using this histogram, it is possible to analyse the overall progress of the team during the defined time, uncover ways to enhance the process and identify needed improvements (Sjøberg, 2018; Staats et al., 2011).

- **Average days spent in each step** – This graphic organizes in a view the average time spent in each step of the workflow. From the moment an item enters the workflow it will travel through the process until it is delivered.

- **Cycle time histogram** – This graphic organizes in a view of calendar days, the amount of time it took to deliver an item. The accumulation of this data provides the team with the capacity to predict with some certainty the time an item will take to deliver (Hemalatha et al., 2021).

This is as mentioned at state of the art, a valuable information when negotiating with stakeholders and to have a general view of the team capacity. It can be important to understand the motive behind the outliers this graphic helps identify (Ahmad et al., 2017).

The distance between the 75% and the 90% percentile can provide information about the predictability of the team - the wider the distance the less predictability has the team.

- **Cycle time 75% type** - The cycle time of type correlates the time spent with the purpose of the ticket. Software teams often deal with different types of issues that can range from bugs investigation to new features development. It is considerably important for a team to understand the distribution of these tickets and acknowledge the time that is being spent at each step. This, apart from helping in negotiations, can be valuable to better allocate the team resources when dealing with pain points (Khaled Yacoub et al., n.d.).

- **Cycle time control chart** – The cycle time control chart provides an overview of the time it takes to deliver an item but has a considerable difference from the cycle time histogram. This chart also gives a notion of the time, and the perspective of the evolution to better or worse (Alaidaros et al., 2018). Analyzing the reason behind the outliers and correlations between this chart and the historical team data can help identify pain points and patterns within the team. Identifying these correlations is essential to allow the team to take measures against them.

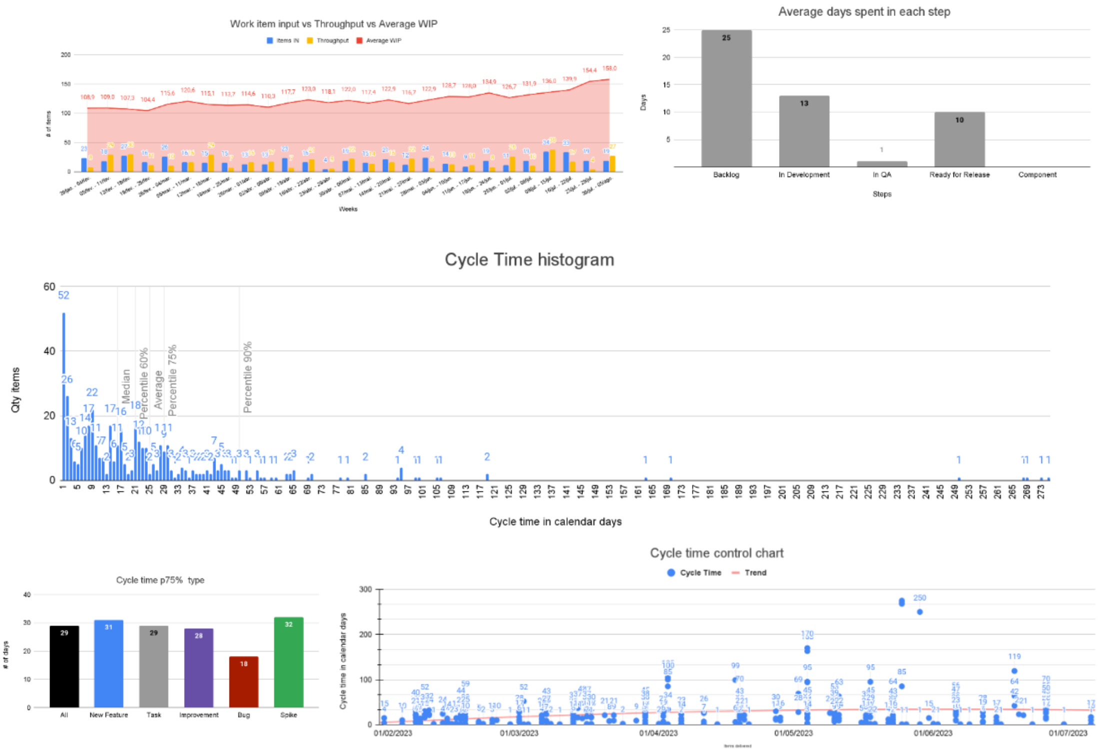


Figure 4 - Metrics visual presentation

The first step of the data-gathering process was to manually organize the raw data used to build the graphics and correlate it with the insights generated by human analysis (Figure 4).

This type of annotation faces the challenge of being limited by the available reports, which in this study had 10 as limit. Each report contains approximately 5 team performance analyses, resulting in a maximum data collection of 50 rows.

3.2 Datasets

The datasets are the base for any decision tree classification. During this project, the initial datasets were refined and supplemented with additional data from expert reviews.

The dataset for "Work Item Input vs Throughput vs Average WIP" initially included individual values for WIP, Items IN, and Throughput per week over the previous 6 months. Afterwards, calculations and tests were conducted with these values to refine the dataset using the accuracy and results obtained from the decision tree.

The final graphic dataset included the following values: the average Work in Progress, throughput, items IN, and progress calculations for each 4-week period within the previous 6 months. The progress values were calculated using the following formula:

$$\frac{(Actual\ week\ WIP\ value\ -previous\ week\ WIP\ value)}{MAX(all\ time\ WIP\ value)}$$

This calculation consists of the subtraction between the actual week WIP to the previous week, that will return a value that can be positive or negative. If the value is positive, it will mean that the WIP increased what normally is related with a negative progress between weeks. If the value is negative, the WIP decreased, resulting generally in a negative progression for that period. After the subtraction the value is divided by the maximum value for the WIP in that timeframe, resulting in a normalized index that can be included in the dataset.

The second dataset, related to the "Average Days Spent in Each Step" graphic, initially contained the number of days spent in each of the following workflow steps: Backlog, In Development, In QA, and Ready for Development (RFR). The workflow could include all or fewer steps, depending on the team. If a team does not have one of the workflow steps, it would still be included in the dataset, with a null value. Additionally, like the first dataset, some normalization was applied to the raw values to improve the accuracy of the decision-tree outcomes. These values were achieved through the following formula:

$$\frac{Days\ spent\ at\ a\ step\ at\ the\ previous\ month}{Higher\ value\ spent\ at\ that\ step\ in\ the\ previous\ 2\ months}$$

This formula will generate a normalized index, that will show if the progression at that workflow step is positive or negative from the previous months period. As an example, if the formula is applied in March to a team whose work items had passed 20 days at the Backlog step in January, and 18 days in February, then the resultant index would be:

$\frac{18}{20} = 0.9$, although if the team had 20 days or more spent in Backlog step in February the resultant index would be $\frac{20}{20} = 1$, because if the previous month value is the highest in two months, then it is divided by itself.

From a critical perspective, if the index is 1 it means that the team had zero or negative progression at that workflow step because the time spent is higher or the same as the previous period. Although, if the value is less than 1 it means that there was a positive progress at that step of the workflow. The smaller the index value, the greater the progress.

Adding this index to the dataset allows the decision tree to be trained to recognize progress at each workflow step, resulting in more accurate predictions about which step may be acting as a bottleneck and negatively impacting the teams performance.

For both dataset outcomes, all expert conclusions from each teams report were recorded and compiled into a spreadsheet. These results were used to create the "result" column in the datasets and to feed the knowledge base, which considered the frequency of each insight by comparing the number of occurrences in the reports.

Expert reviews provided invaluable qualitative insights that complemented the quantitative analysis. While the decision tree relied on historical performance data, the experts deep understanding of team dynamics and workflow processes helped identify less obvious patterns.

For instance, they highlighted situations where high Work In Progress (WIP) values coincided with changes in team structure, an aspect that might not be immediately evident solely through the data. Their feedback led to adjustments at the dataset results in order to minimize the context necessity impacts at the end solution.

In terms of limitations during the dataset generation process, several key challenges emerged. One significant issue was the low availability of historical data for certain teams; in some cases, only a few months of data were accessible. Additionally, some datasets had to be discarded due to their strong dependence on specific team contexts. The fact that experts rely on contextual factors during their analysis made it difficult to annotate values that the decision tree could use for training. If values regarding specific team events like vacation or onboarding processes are included at the datasets, it can compromise its adaptability across different teams and scenarios.

3.3 Data privacy, security and ethical considerations

The application of data anonymization techniques plays a critical role in maintaining privacy without sacrificing the utility of the data for research purposes. Techniques such as pseudonymization, randomization, and data masking ensure that sensitive information is obscured, yet meaningful patterns and trends can still be extracted for analysis. This approach not only protects individual privacy but also allows for the continued use of the dataset in future studies, as it complies with the General Data Protection Regulation (GDPR) while remaining valuable for decision-making and insights.

Data anonymization irrecoverably transforms the raw data into a protected version by eliminating direct identifiers and removing sufficient details from indirect identifiers to minimize the risk of re-identification when there is a requirement for data publishing (Senavirathne & Torra, 2020).

In accordance with the principles of the GDPR, the research methodology employed at this study prioritizes the confidentiality and protection of sensitive data. Additionally, a Non-Disclosure Agreement (NDA), was signed to ensure both parties commitment to handling sensitive information responsibly.

The data used at this research is real and from JUMIA teams. For public demonstrations and for any data sharing outside the company network, a comprehensive anonymization protocol has been implemented to ensure that all published information is entirely devoid of any personally identifiable details.

As an additional protective measure, all files containing real data are not shared outside the company network, being entirely handled in Google Spreadsheets who's only the authorized JUMIA company personally have access.

The possible concerns raised by the need to evaluate the application with experts external to the company were also addressed. A presentation was made for each expert with anonymized data that would not compromise the quality of their analysis result. Before the testing sessions the presentation was reviewed inside the company to guarantee that no relevant or compromising information was going to be accidentally shown or leaked to outsiders.

Regarding the final thesis results they will be presented with randomized data and fictional team names, making the tracking of the real values impossible for anyone that accesses the document and the research.

3.4 Summary

The main goal of this chapter was to outline the methods, tools, and data collection processes used at the project, ensuring that the data quality was sufficient to generate meaningful insights. By detailing the data gathering methods and the datasets used, this section highlights the foundational steps that led to the analysis using decision trees and expert systems, as discussed in earlier chapters.

The data collection process was conducted by using real reports that contained key performance metrics, particularly focusing on workflow analysis. The decision trees relied on those datasets that tracked essential metrics such as Work in Progress (WIP), throughput, and cycle time, enabling a structured approach to identifying bottlenecks and inefficiencies within development teams. By analyzing patterns in the data, the study provided a comprehensive view of team performance and delivered actionable insights for process improvement.

The datasets were critical to correctly train the decision tree models and were carefully refined to ensure accuracy. Initially derived from raw performance data, these datasets were normalized to reveal trends and predict future bottlenecks. A thorough methodology for calculating workflow progress and understanding team capacity was applied, ensuring that the decision-tree models had reliable inputs to produce meaningful outcomes.

In terms of data privacy and ethical considerations, rigorous protocols were implemented to anonymize sensitive information. This process complied with GDPR standards and ensured that no personally trackable information was exposed during research. The real data used from JUMIA teams was anonymized before being shared externally, while a strict access control mechanism ensured that only authorized personnel could handle sensitive data internally.

In conclusion, by following strict data safety measures, this research had a balance between the collection of sensitive data, allowing the decision trees to be trained with real data datasets, while at the same time preventing any potential threats.

4 Solution Development

The solution development is a crucial phase of the project, and it tells a story on how the different components and responsibilities are built together to provide a full operational system.

4.1 Project technical definitions

A project development is a process that requires thoughtful technical choices. These decisions made before and during the solution development can directly impact the project longevity and its short and long term use.

Code maintainability, efficiency costs and execution time are some of the parameters that can be positively affected by the definition of an adequate technological stack, high quality development practices and a clear software architecture.

4.1.1 Technological stack

The technological stack definition is a step of development that inevitably new projects go through. Considering the Metrix project goals, the technological stack should support a system that is able to connect with external applications, perform API requests and deal with large amounts of data.

From the entire universe of programming languages with these capabilities, Java and Golang became the main candidates. The choice between them was made via scientific method and using the following criteria:

- **Performance**: Different languages will imply different benchmark performances. The benchmark measures a programming language memory usage (mem), source code size (gz) and processor load (cpu). When starting a project, it is important to consider these measurements to provide the best performance and less damage possible for the infrastructure running it.

- **Community language**: The programming language acknowledgement within Jumia teams is an important factor to consider. It would make it easier for Jumia community to deal with improvements or code maintenance in the future. At the same time, thinking on the external community acknowledgement of the language will improve the access to valuable information through a bigger quality of the documentation, not excluding the fact that if needed, external help would be easier to find.

- **Application scope**: Some programming languages have a better capability to deal with limited scopes (example: The metrics extraction for Y teams, where Y is a defined number) or unlimited scopes (example: an Shop application where people can place as much orders as they want).

When compared, Golang won in terms of performance, but choosing Java would fulfil the need to develop an application in a language that most of the software engineers at the company are comfortable with. It means that the system can possibly be maintained in the future with less costs. Also, by choosing Java, the majority of JUMIA systems would be easily integrated with the application rather than if Golang was used. Having this in mind, Java was picked to be the main programming language of the Metrix project.

Additionally, considering the need to develop a Hybrid Expert System that includes the use decision trees, a research was conducted and Drools was picked to develop the expert system while Python became the choice to train the decision trees. This choice was made considering Python popularity among the Artificial Intelligence community.

Usually, a popular language has good quality documentation and community support forums what enhances the development experience and promotes code maintainability. The up-to-date technology and easy integrations with other systems are also Python characteristics that the Metrix project can benefit from.

In conclusion, after research, comparison and ponderation Java and Python were chosen to integrate the technological stack with which the Metrix project was built.

4.1.2 Development workflow

The establishment of a development workflow is a crucial project planning step in which the development practices are aligned with the team.

At the Metrix project, the development workflow was designed following Behaviour driven design (BDD), Test Driven Design (TDD), Continuous integration and deployment CI/CD and Trunk Based development. All these concepts were detailed at the second chapter of this document.

These four methodologies where combined together assuring that the project would be able to meet business requirements, automatize the extraction process and produce automatic recommendations for the teams without neglecting the need for a robust system and test covered system.

The possibility for the application to grow in the future, turned essential the need of solid tests suites, turning the adoption of BDD and TDD of extreme importance due to their characteristics.

Additionally, by being developed by a small team and to address the business needs to have an application like Metrix in the minimum possible time, CI/CD pipelines were created allowing the code to be constantly updated and integrated in production environment. This release process allowed the application to deliver frequent value, with constant updates. The use of trunk base development, once more, was picked to support a fast-paced growing application, by allowing the team to daily integrate code into production.

4.1.3 Project Software Architecture

The software architecture definition is the step where the team discusses the code structure and design a plan for the application development and growth.

The software architecture chosen for the Metrix project was the clean architecture. This architecture easily fulfils the need for API requests since it well establishes different layers. The core layer is where all the logic and business rules reside inside the application. This allows the domain to be separated and agnostic from layers that are in constant transformation like the User Interface (UI), the database or external APIs. It promotes the decoupling between the system components.

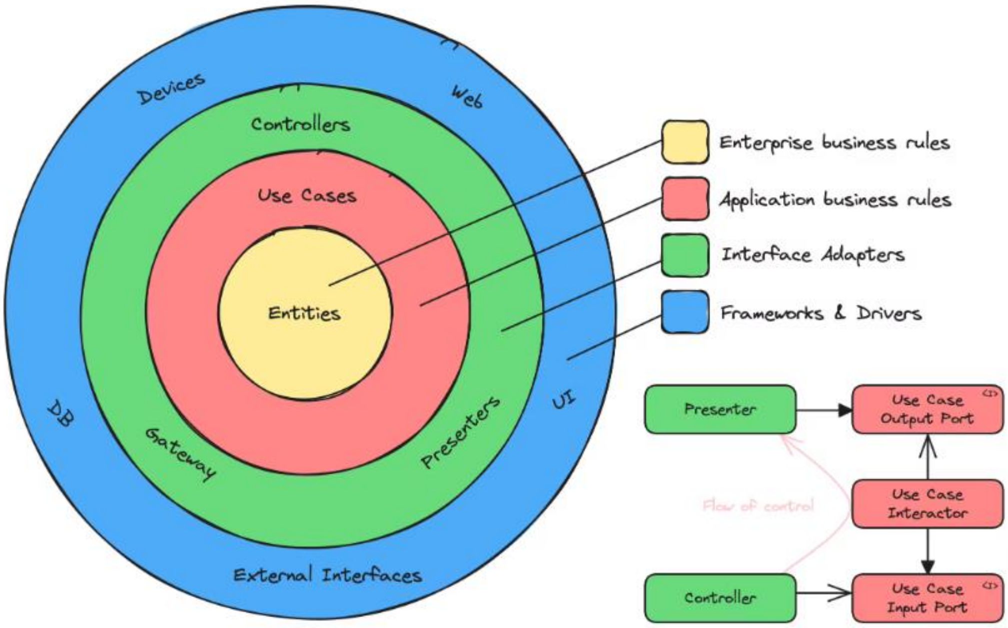


Figure 5 - Clean Architecture (Scalable Backend, 2023.)

External Metrix application connections or dependencies such as the Google Slides, Google spreadsheet and JIRA API's are placed at the system external layer, highlighted at figure 5 as Frameworks and Drivers layer. It allows the application to perform these requests while the business rules, for example the configurations regarding the teams data mapping or rules for recommendations generation are protected at the application core.

4.2 Solution overview

The project overview, Figure 5, provides an overview of the entire flow. The Metrix application has three main components:

1. Import CSV File with Issues Service
2. Import Recommendations Service
3. Expert System/Knowledge base/Decision Trees Service

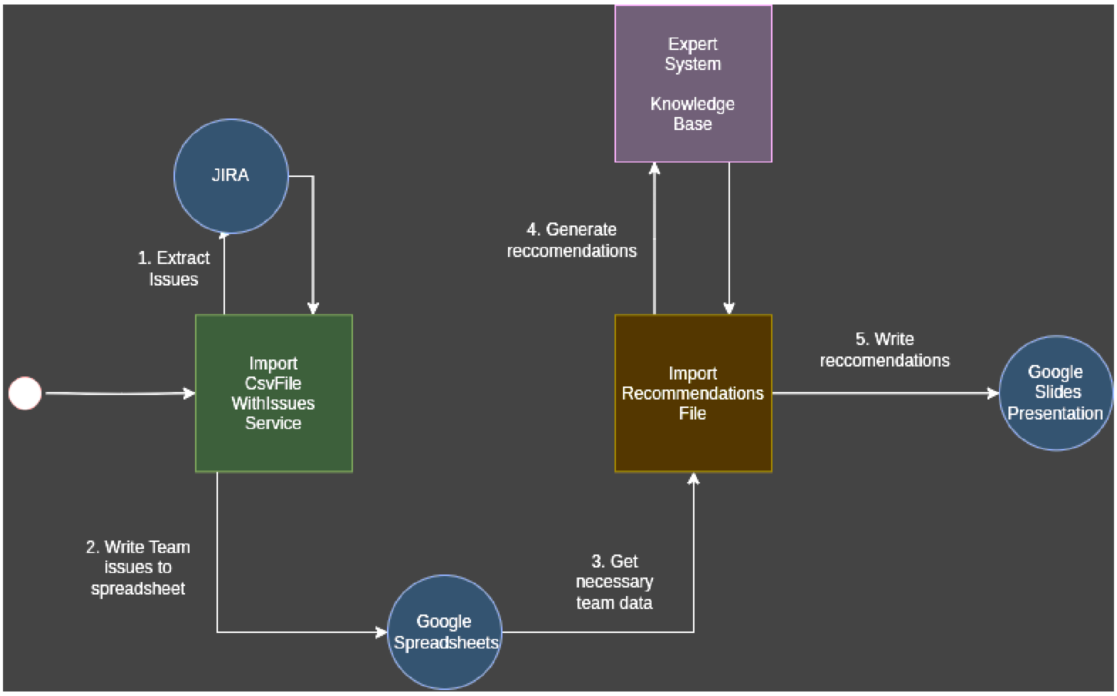


Figure 6 - Solution components interaction overview

Acknowledging these three main system components and their high-level interactions makes it possible to get a perspective of the actions that take place since the moment the user interacts with the system to the end when he can see the team efficiency metrics report at the Google Slides presentation.

The application generates two outputs during a full execution. The first output is a new google spreadsheet tab with the issues from the team, already mapped considering their unique workflow. The second is a presentation slide with auto generated recommendations regarding the team performance at the selected time period. The first output can be independently generated, however the autogenerated recommendations depend on spreadsheet data.

These outputs are part of different use cases: The import CSV file and the generate recommendations. Both use cases have their unique responsibility and set of interactions.

4.2.1 Import Csv File use case

The import CSV file use case is responsible for handling all the processes that lead to the first application output. It starts with the user providing a vertical or team name to the application. A vertical refers to a group of teams inside the company but the application supports directly a team as input if that is the user's intention.

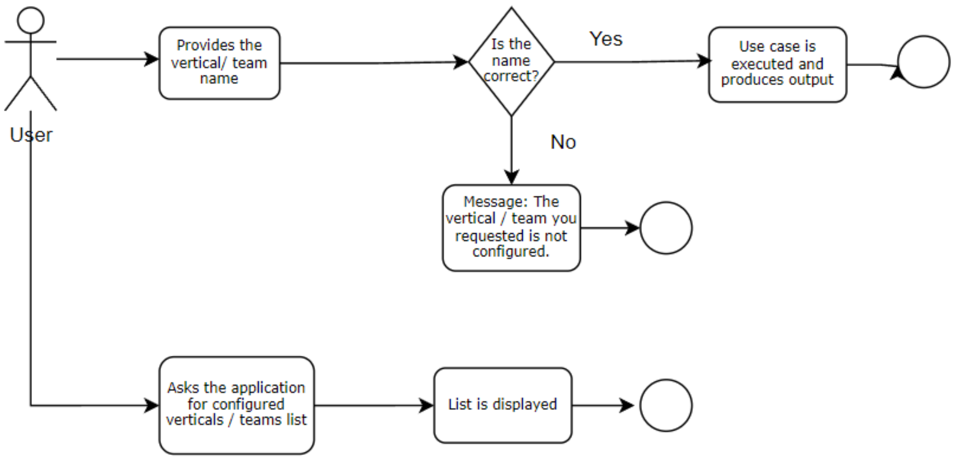


Figure 7 - Metrix user interactions

After receiving the vertical/team name information, the first step is to map that name to the configurations of the vertical, present inside a configuration file of the application, figure 7. To enhance the user experience, the user can request the application a list of the configured verticals. For the failing scenario in which the provided vertical or team name does not exist at the application, the user will be informed with an explanative message, figure 7.

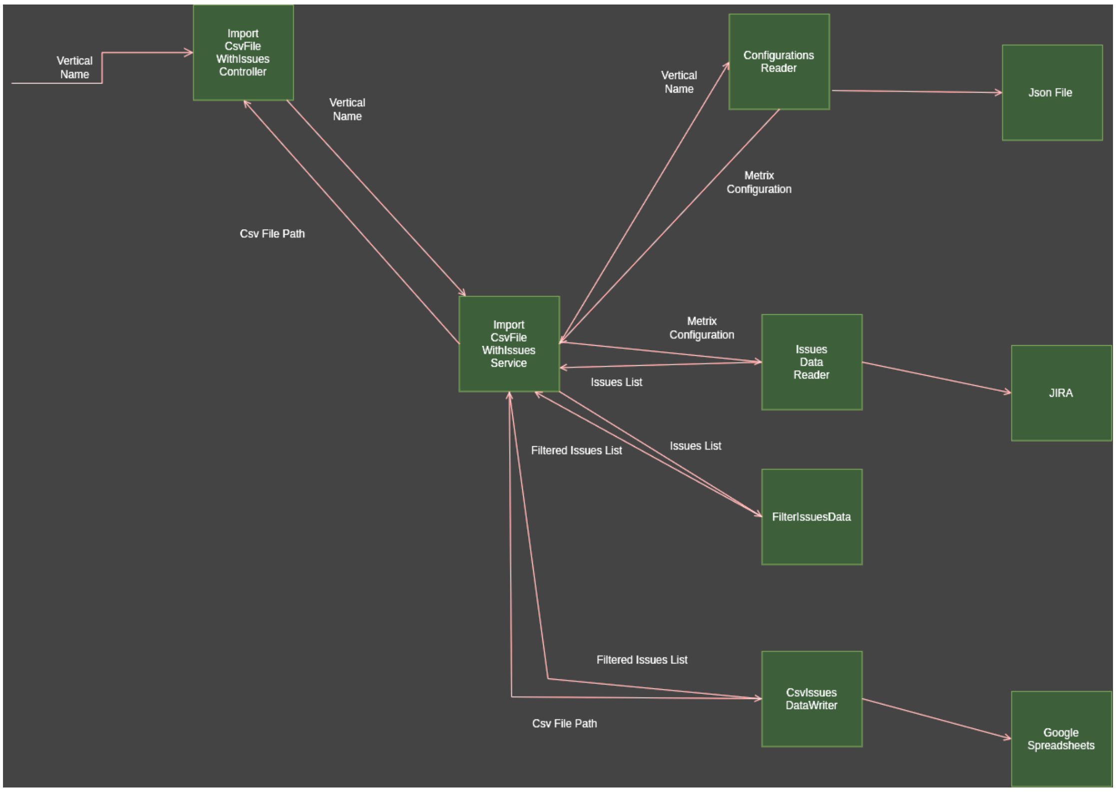


Figure 8 - Import Csv File Use Case

```
"verticalName": "vertical",
"filterId": 31424,
"statusColumnMapping": {
          "Backlog": ["Backlog", "Ready for Development"],
          "In Development": ["In Development", "Ready for Review", "In
Review"],
           "In QA": ["Ready for QA", "In QA", "Approved", "Awaiting
Integration"],
           "Ready for Release": ["Ready for Release"],
           "Done":["Closed"]
},
"flowMetricsSpreadsheetId": "1Jc1at_KJpfvqZ5c9IcDYo-jVL90kh2YjDrYRm7Y3X_c",
"extractLastDateSince": "In QA",
"isComponentColumnDisplayed": true,
"verticalSlideId": "1Jc1at_KJpfvqZ5c9IcDYo-jVL90kh2YjDrYRm7Y3X_c2"
```

Code Block 1 - Vertical / team configuration mapping

The configurations will provide the unique information related to the vertical / team, figure 7, such as:

- filterId: the identifier of the Jira filter that will return the team issues information
- statusColumnMapping: a mapper of each team unique workflow to a standard 4 or 5 stage (Backlog, In Development, In QA, Ready for Release, Done) workflow.
- flowMetricsSpreadsheetId: the identifier of the vertical google spreadsheet where the issues information is going to be imported after the data treatment and refinement.
- extractLastDateSince: the step of the development from which the transition date is going to be considered the last
- isComponentColumnDisplayed: an indication value if for this vertical a column called component that maps the data to the vertical teams is going to be displayed at the spreadsheet or not.
- verticalSlideId: the slide id to where the recommendations are going to be written.

The second step is to use this information, especially the filter id, to perform an API request to JIRA and retrieve the team work data. The raw data is transformed during a treatment process to generate the csv file.

The optimal csv file format would contain the work items that the team finished at the select time period, ordered by conclusion date and with the correct workflow mapping, such as displayed at Figure 10. To obtain that wanted format the raw data goes through a treatment process - the third step highlighted at Figure 6 as “Filter Issues Data”.

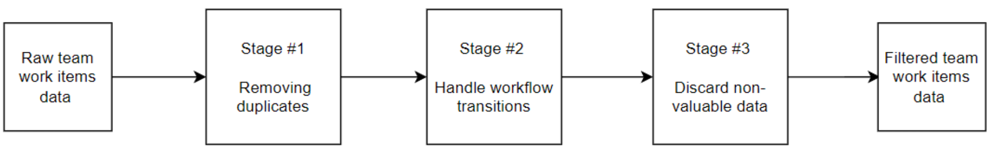


Figure 9 - Data treatment process

This step goal is to select the relevant issues and assure high data quality and organization before publishing it to the user. With that intention the data goes through a three-phase process, figure 9.

The unfiltered data enters the process and goes through the first stage that is to remove possible duplicates from the data, improving the data consistency.

After that, at the second stage, the workflow transitions are handled. A transition is the date where a work item is moved between workflow steps (e.g. the date when a ticket is moved from Backlog to In development). However, since different teams have different workflows definitions, this process guarantees that all team transitions are mapped to 4 or 5 common stages - Backlog, In Development, In Quality Assurance, Ready for Release and Done. This action allows the application to be scalable for teams with different workflows.

At the third and last stage, the work items that only have transition date for Backlog and Done workflow steps are removed from the data. This work items did not pass through the development process, so they are not considered relevant to the team efficiency analysis. The reason for the absence of "in development" data can range from tickets that were mistakenly moved into the workflow to a wrongful workflow mapping. For that reason, at this stage a list of the discarded issues is kept for tracking purposes.

The application first use case ends with the csv with issues being published into the Google Spreadsheet. By using the spreadsheet identifier value, code block 1, and connecting to the Google Spreadsheet API, the data is correctly imported to a new tab with the current timestamp at the vertical/ team spreadsheet.

At the end, the user will receive a message informing the about the operation success with a direct link to the chosen vertical/team spreadsheet and a list of the discarded work items. The link to the spreadsheet enhances the user experience by intuitively providing a direct access to the new generated tab. The list of discarded items allows the user to understand which work item information is missing from the spreadsheet and provides the necessary details to investigate these cases when needed.

4.2.2 Generate Recommendations Use Case

The generate recommendations use case, figure 10, encapsulates the second main flow of the application. The goal of the use case is to automatically generate recommendations for a vertical or team based on its spreadsheet data. Since this use case depends on the success of the new csv file generation, it can either be started in sequence with the first use case or by manual interaction with the user, with him providing a vertical or team name to the system.

The recommendations are generated based on the team work items historical data. That information is obtained by retrieving the necessary data from the Google Spreadsheet where it is written as output from the first use case.

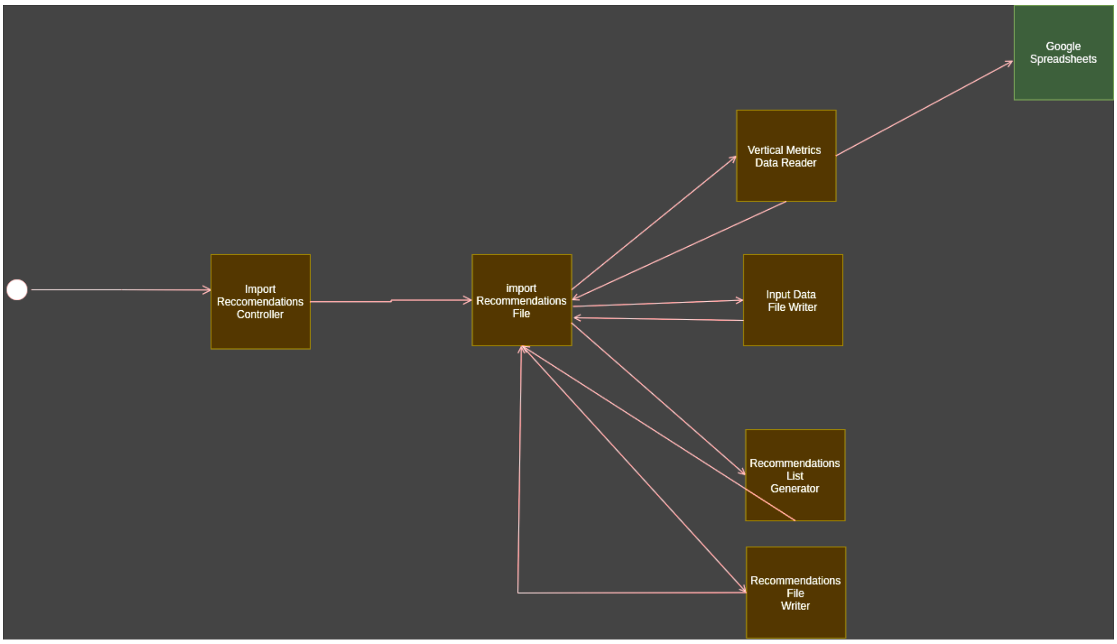


Figure 10 - Generate Recommendations Use Case

After the data collection phase, the system as the important role of formatting the input for the third component – the hybrid expert system. This component consists of two decision trees, which experimentation is explored with more depth at point 6.1, and an expert system containing a knowledge base and an inference engine, displayed at Figure 11.

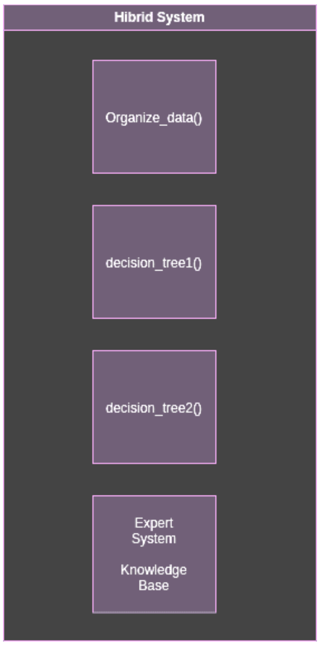


Figure 11 - Hybrid system design

At the second phase of the use case. the data should be organized and delivered to the hybrid system, figure 11. In order to achieve the best possible results, the input data should contain:

- A list with the indexes and values for the “Work item input vs Throughput vs Average WIP” team graphic. These indexes are average Work in Progress, throughput, items IN, and progress indexes calculated for each 4-week period within the previous 6 months.

- A list with the indexes from the “Average days spent in each step” graphic. This list includes the following values: the number of days spent in each of the following workflow steps: Backlog, In Development, In QA, and Ready for Development (RFR), calculated indexes on the progress of each stage considering the last month period. If a team does not have one of these steps the value is sent as null.

After the input data is generated, the first list is sent to the first decision tree. This decision tree will provide a classification about the team overall progress during that period of time. The possible classification values range from 0 to 2, table 6.

Table 6 - First decision tree classification options

Classification	Meaning
0	the team had a negative progress
1	the team had no progress
2	the team had a positive progress

The second list is sent to the second decision tree. The goal is to identify process bottlenecks, and the classification results can range from 0 to 4, table 7.

Table 7 - Second decision tree classification options

Classification	Meaning
0	There is no visible bottleneck at the workflow
1	The Backlog step can be a workflow bottleneck
2	The In Development step can be a workflow bottleneck
3	The In QA step can be a workflow bottleneck
4	The Ready for Release step can be a workflow bottleneck

.

After the production of the classification results, those are sent to the expert system along with a third list containing important team statistic values gathered at the spreadsheet by comparison with the previous month – the existence of outliers and the percentile 75%.

The expert system contains a knowledge base with an extensive set of rules assembled using the team reports annotations. The rules suffered some adjustments during the project, taking in consideration the experts feedback. The inference engine fires all rules and from that action the final results of the use case are compiled into a list of conclusions.

This list will include at the top solid information about the vertical/team progress and their possible workflow bottlenecks. The rest of the list will contain recommendations that can guide the team to solve their identified problems. For example, if the team has a possible bottleneck at the Backlog step, the conclusion list will have recommendations to tackle that specific issue such as: the suggestion to invest time in backlog management, the use of slicing techniques or applying a limit value to the backlog growth.

At the last use case step these conclusions are written at the vertical/teams slides presentation and the user will receive a direct link to the specific slide where the conclusions were automatically generated. Publishing the conclusions into Google Slides improves the user experience and data visualization.

4.2.3 User Experience

The Metrix project main goal is the quality of both csv file import and generate recommendations use cases output, however the application was designed to provide a satisfactory user experience (UX). Considering that the project end user would be technical personal, the application was designed to be accessible through the command line interface (CLI).

The use of the CLI to reach the application was a decision made based on several aspects: the application end user technical expertise; the application practicality; the development costs and the project deadlines.

To support the user experience, some specific UX features were added to the application. The first was the help command. By asking for help the application would display at the user command line a practical guide to use the application. This small guide contains a brief application description and informs the user of the parameters content, that at the existent use cases are the vertical or team name.

The second feature that supports the user experience at his first interaction with the application is the command to list the existing configurated verticals / teams. Upon requesting this information, the user will see displayed at his command line a list with the verticals or team names that are configured inside the application. As previously mentioned, at the point x of the document, the user can also take advantage from customized error messages that will warn him in case he mistypes or insert a wrong input at the application. This sort of error messages was pointed by some of the users to have a considerable impact on their application usage:

"I consider that the Metrix application has an intuitive set of commands. The experience is enjoyable, and the configured teams list feature helped me in quickly identifying that the vertical I wanted to use was already configured at the application." – Metrix user

4.3 Summary

The main goal of this chapter was to detail the solution development process for the Metrix project, explaining the decision-making process that led to the tools and methodologies combination that were used at the project development.

The chosen technological stack that includes Java for its community support and integration capabilities within Jumia existing systems, alongside with Python for its strong presence in the Artificial Intelligence domain. It not only streamlined development efforts but also ensured that the project could leverage of powerful integrations such as Drools to support the expert system.

The development workflow was established with a focus on Behaviour Driven Design (BDD), Test Driven Design (TDD), Continuous Integration/Continuous Deployment (CI/CD), and Trunk Based Development (TBD). These methodologies were combined to ensure that the application could efficiently meet business requirements, automate processes, and facilitate future scalability. The emphasis on robust testing practices highlights the importance of maintaining high-quality standards throughout the development lifecycle.

The two main application use cases were detailed during the chapter, by explaining the components interaction and providing examples of the desired outcomes. Failing scenarios were also described giving a complete overview of the solution flows.

In terms of user experience, the application was designed to be accessible via a command-line interface (CLI). Features such as help commands and detailed error messaging were implemented to enhance the overall user experience and ensure a smooth interaction process.

In conclusion, this chapter provided a detailed overview of the methods and practices used in the solution development of the Metrix project, emphasizing the selection of technologies, workflows and design.

5 Experimentation and evaluation process

After the data collection phase, the experimentation is started with tests at the datasets. This phase includes analyses, simulations, or experiments aiming to achieving the best possible outcomes.

Deeply at this chapter, the solution evaluation process will be explained. This phase consists in describing how the solution is going to be asserted, and the project real impact will be measured.

5.1 Experimentation outcomes

The first dataset was obtained with annotations from the raw data that originates the Work item input vs Throughput vs Average WIP graphic (Figure 4). Initially this dataset started with nearly raw data but suffered updates during the experimental phase ending up with indexes, variations, averages and normalized data.

The goal of this experimentation is to obtain the highest accuracy possible on the decision tree that is meant to classify the overall team progress as good or bad by analyzing historical data from the previous six months.

Table 8 - First decision tree outcomes

Dataset Work item input vs Throughput vs Avg WIP	Accuracy
Nearly raw data	40% – 60%
Nearly raw data, double of the size	40% – 60%
Combination between raw and normalized data	60% – 80%
Indexes, variations, averages and normalized data	80%

The dataset suffered some adjustments during the experimentation stage and consequently different accuracy results were obtained. The best outcome was using the dataset with indexes, variations, averages and normalized data information, table 8.

The experimentation step on the second dataset which was also obtained with the same process using the “Average days spent in each step” histogram, figure 4. The purpose of this decision tree is to identify and classify possible bottlenecks at different workflows steps (Backlog, In Development, Ready for Release, In QA).

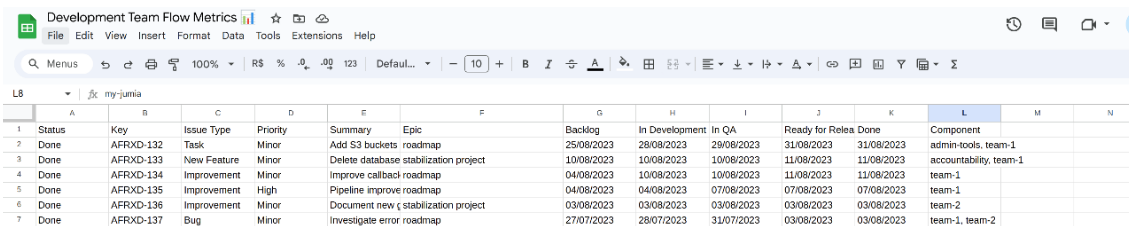
Table 9 - Second Decision Tree Outcomes

Dataset Average days spent in each step	Accuracy
Combination between raw and normalized data	80% – 90%
Indexes, variations, averages and normalized data	100%

The dataset suffered some adjustments during the experimentation phase but with only two different approaches it has experienced a high accuracy reliable value, table 9.

5.2 Solution outcomes

As previously mentioned, the application has two independent outcomes. The first one being a spreadsheet tab containing the organized team data. These data are treated to delete duplicates, clean faulty values and apply the correct transitions.


Development Team Flow Metrics

Status	Key	Issue Type	Priority	Summary	Epic	Backlog	In Development	In QA	Ready for Relea	Done	Component
Done	AFRXD-132	Task	Minor	Add S3 buckets	roadmap	25/08/2023	28/08/2023	29/08/2023	31/08/2023	31/08/2023	admin-tools, team-1
Done	AFRXD-133	New Feature	Minor	Delete database	stabilization project	10/08/2023	10/08/2023	10/08/2023	11/08/2023	11/08/2023	accountability, team-1
Done	AFRXD-134	Improvement	Minor	Improve callback	roadmap	04/08/2023	10/08/2023	10/08/2023	11/08/2023	11/08/2023	team-1
Done	AFRXD-135	Improvement	High	Pipeline improve	roadmap	04/08/2023	04/08/2023	07/08/2023	07/08/2023	07/08/2023	team-1
Done	AFRXD-136	Improvement	Minor	Document new (	stabilization project	03/08/2023	03/08/2023	03/08/2023	03/08/2023	03/08/2023	team-2
Done	AFRXD-137	Bug	Minor	Investigate error	roadmap	27/07/2023	28/07/2023	31/07/2023	03/08/2023	03/08/2023	team-1, team-2

Figure 12 - First Application output: Automatic extraction

As an outcome, the user can expect the teams tab, figure 12, to display the following information of the team: status (the status of the ticket), key (the identifier of the ticket), issue type (the issue type, that goes from task to bug), priority (the criticality of the ticket), the summary (a title or brief description on the ticket intent), epic (the project the ticket is referring to), the transitions (date where the different steps of the workflow were completed: backlog, in development, In QA, ready for release, done), the component (optional field that displays the teams the tickets are assigned to).

The second outcome is a Google Slides slide, displayed at figure 13, where that the user can access as if he was using a simple dashboard and analyze the team progress by looking into the graphics and to the insights summary that are auto generated and provide recommendations to the user on how to enhance the team efficiency and possible bottlenecks identification.

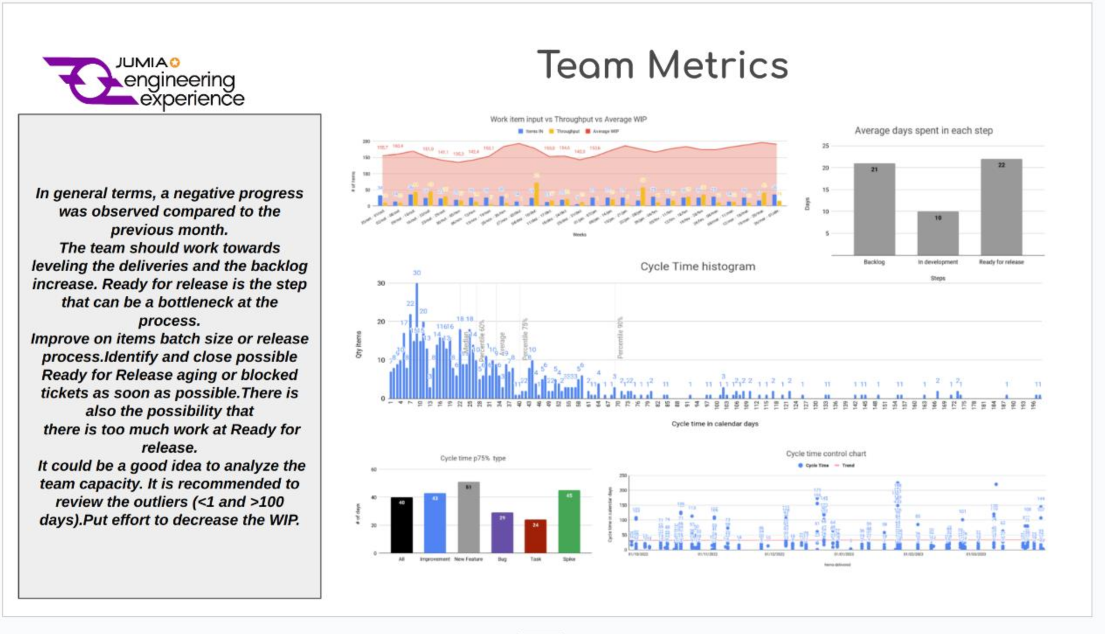


Figure 13 – Second application output: Auto-generated recommendations

5.3 Solution evaluation

Solution evaluation involves testing and assessing the proposed solution to determine its effectiveness. In this thesis, the research question seeks to explore whether a hybrid solution, as outlined in this document, can effectively prevent brain drain within a company.

5.3.1 Experts contribution at the testing process

This phase gathered different experts to be part of the evaluation process. The first expert type were Agile coaches that worked in the department at the time the project started. Since they were the ones manually producing the recommendations used to build the application their input makes an important contribution to evaluate how well the decision trees and the expert system is accurate and correctly trained.

The second type of expert is external to all the application development and has similar experience in agile positions and software teams' management. Their input can bring a different and independent perspective about the usability of the application out of the department context and how well the insights are accurate through external eyes.

The third type of expert is also external to all the application development but is a member of a team inside the company.

5.3.2 The evaluation process

The evaluation process consisted of three different phases. Each one has its own goal and is applied to the expert type that best suits the test. The experts that were part of the Engineering experience team participated in:

The first phase – three team performance analysis that existed on the Monthly reports archive were chosen and the application was given the same raw data that originated the human analysis, and an auto-generated group of recommendations was collected as output.

For each one of the three cases the experts had two options, figure 14 , the human and the machine generated analysis of whose origin the experts were not previously informed. After looking into the graphics and organizing their own thoughts, the experts are given the choice of which one was generated by a human, and which one was auto generated.

This first test idea is to evaluate how similar the application output can be from a human, while evaluating the accuracy of the recommendations as well.

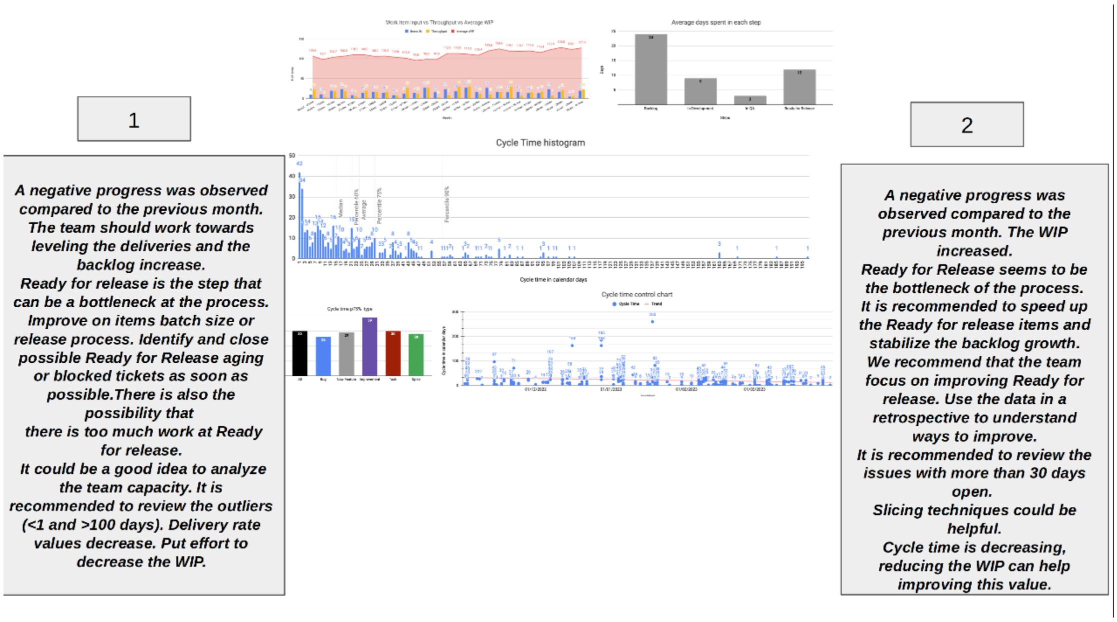


Figure 14 - Solution Evaluation Process: human vs machine comparison test

The second phase – three other real team performance raw data from random software teams at the company were extracted and auto generated using the application, figure 15. After that the graphics data was anonymized by removing any reference to a possible real team or active work item. One by one, the metrics graphics and information were displayed to the Agile coach who analyzed the team performance and wrote down some recommendations.

These expert analyses were later compared to the auto generated evaluating the hybrid solution accuracy and adaptability for real new cases.

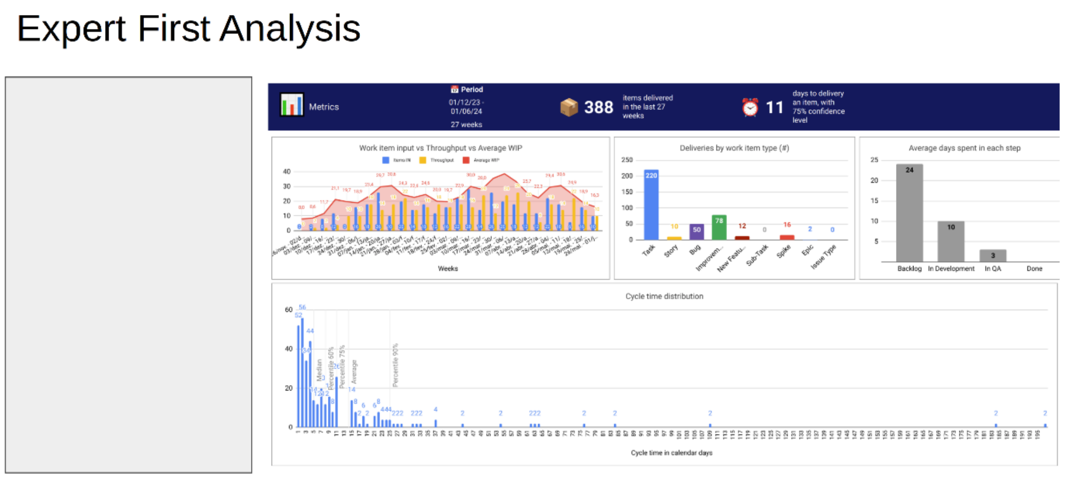


Figure 15 - Solution Evaluation process: expert vs machine test

The third phase, in which the external experts took part was similar to the second phase but with the additional plot that by being part of the team, the third expert type can have a little amount of indirect bias because. There is the possibility that a human person cannot totally detach themselves from the knowledge they have about the company teams processes the team specific context.

The idea of bringing in a totally independent expert but also a member of the team from whom the metrics are is to compare how well the application outcome will deal with these scenarios. If it is comparable to the analysis of someone that is slightly biased.

Also, some quantitative measures will be compared during the evaluation process such as the time that the human professionals took to analyze the data against the time the application took.

5.3.3. The evaluation results – Quantitative

The evaluation produced the following results:

Table 10 - Experts evaluation results

	Testing Phase 1	Testing phase 2	Would you benefit from using the application?
Expert 1	Identified the machine analysis as human.	92.31% accuracy, 12 of 13	Yes, as a basis for the analysis
Expert 2	Identified all the human analysis correctly.	75% accuracy, 12 of the 16	Yes, because the time it saves
Expert 3	Identified all the human analysis correctly.	88.235% accuracy, 15 out of 17	Yes
Expert 4	2/3 human analyses identified correctly	92.60% accuracy, 25 out of 27	Yes, using to compare manual reviews to assure that important details are not missing.

For the first type of experts, table 10, the group that contributed to the dataset with their own analysis and worked at the department at the beginning of Metrix project:

- Around 50% of the Agile Coaches chose the auto-generated result as being human generated. This result can confirm that the application produces near human written recommendations. Even with the knowledge about the data that feed the dataset for the decision trees and the heuristics for the expert system the results show that half of the experts from the group were unable to distinguish between the auto generated and the human written analysis. The fact that they have doubts about which was generated by machine and human proves that the application has a good level of accuracy at the produce recommendations.

- The auto-generated recommendations were a solid match to most of the expert conclusions, around 80% of the recommendations that the experts suggested when analyzing the graphics in real time where part of the application analysis already automatically generated. This confirms that the application can replicate with a good level of accuracy the expert's knowledge.

From a critical point of view, the fact that the first group of experts had difficulty distinguishing between the human generated and the auto generated analysis can validate the quality of the data gathering and of the assembled datasets. This difficult means that the application was able to replicate the experts conclusions and write them in a human format, that in some cases was misleading to the own experts whose conclusions were part of the data collected.

Additionally, the fact that the expert analysis had a high percentage of similarities with the auto generated proves that the decision trees were well trained, classifying new real scenarios with a accuracy comparable to a human specialist.

Table 11 - Second expert evaluation results

	Testing results
Expert 1	100% accuracy, 5 out of 5
Expert 2	80% accuracy, 4 out of 5

The second type of experts, table 11, the group external to the application development with similar experience in agile positions and software teams' management:

Had a match of 90% in which 9 from their 10 recommendations when analyzing their team's scenario was a direct match to the application auto-generated recommendations. This result confirms that the system can produce accurate results for new scenarios that will match different experts other than the participants on the dataset data, proving that the application can produce versatile recommendations.

From a critical perspective, the results of this evaluation step can lead to the conclusion that the system is able to successfully match different human experts conclusions. This can mean that the data used for the datasets and the knowledge base was either generic or had a high quality.

The third type of solution evaluators, the group of team members, when confronted with their team's performance auto generated analysis highlighted that some parts of the recommendations, they had already considered to improve their processes. They also pointed out that this analysis would help the team on a regular basis by identifying and steering the efforts of the team in the right direction.

Another quantitative evaluation of the application was the comparison of the time spent on analysis before and the time with Metrix usage.

Table 12 - Human analysis vs machine time comparison

	Manual extraction process (per team)	With Metrix project extraction (per team)	Manual analysis (per team)	With Metrix generated extractio (per team)	Total process time (per team)	Total time with Metrix
Expert 1	1 to 2 hours	3 to 5 minutes	30 minutes	5 minutes	1h30 to 2h30	10 minutes
Expert 2	2 hours	3 to 5 minutes	20 – 30 minutes	10 minutes	2h30m	15 minutes
Expert 3	1h30 hours	3 to 5 minutes	25 minutes	5 to 10 minutes	2 hours	10 to 15 minutes
Expert 4	1h30 to 2 hours	3 to 5 minutes	15 minutes	15 to 20 minutes	1h45 to 2h15	20 to 25 minutes
Expert 5	1 hour	3 to 5 minutes	30 minutes	10 minutes	1h30m	15 minutes

The values were gathered with the type 1 experts that provided their real time based on its experience. It is important to consider that the values per team differ between teams and their characteristics, number of works produced and workflow complexity.

From a critical perspective the Metrix application proved to be a valuable solution by not only replicating with success the experts conclusions, writing the results in a human-like manner, but also by reducing the time for the entire extraction process by an average value of 89.80%.

5.3.4 The evaluation results – Qualitative

Apart from the quantitative results all the experts were asked to provide overall feedback about the Metrix project. Sharing their opinions about the application usage, their ideas for improvement and if they would benefit from including the use of application in their daily work.

After compiling all the answers, the application was considered to have some ways to improve such as: adjusting the auto-generated insights to give less subjective inputs following the recommendations, for example the “Good sign!” or “bad sign!” information.

As highlighted by the experts it is natural that the application reacts like this since it was built on top of a dataset of reports where these sorts of comments were commonly made. Although these comments were felt to be dependent on each company context so they should be handled with caution considering that they are not universal. Also, the experts considered that if the application can gather even more information, it can frequently catch corner cases that the human experts cannot be aware of, contributing to a more detailed result.

Regarding the positive aspects of the project, the group of experts considered that the Metrix solution was able to accurately produce recommendations and that the use of Metrix would reduce the time they spent on analyzing team metrics to significantly more than a half.

Another interesting piece of feedback brought to the table by the experts was that they would confidently use this project not only to provide feedback to the teams but also as software to train new Agile Coaches, addressing brain drain problems at the company by helping to deal with knowledge loss and saving the remaining experts time on training new joiners.

Lastly, even if they a review is advised at the auto generated recommendations before showing to a software team the results, some of the experts agreed that in case of time or resource shortage the auto generated analysis would be sufficient to point the team in the right direction to improve their workflow and internal processes.

5.3.5 Project results summary

The solution goals were the following: automate the extraction of efficiency metrics allowing it to scale to all delivery teams at the company and reduce the manual time spent on data treatment actions by 50%. Including the processing of dates format and solution for the lack of workflow standardization, provide automatic recommendations based on metrics for a given group (team/vertical).

Overall, after with the evaluation results it is possible to conclude that the Metrix can effectively contribute with a path to metrics extraction automation. The solution does that by reducing the extraction process, which includes raw data extraction, data treatment, removing outliers, organize the data, deal with the different team's workflow and parse the correct dates format, time around 90%. This reduction can address the initial problem regarding the lack of scalability at the manual process and the lack of workflow standardization and automated data treatment. By using Metrix application, the measure of performance indicators can reach more teams at the company.

The Metrix application achieved 90% accuracy for external experts, around 80% for experts that took part in the dataset data with auto generated recommendations text which the similarities to the human style mislead some experts when choosing between both analyses. This indicates that the use of a hybrid expert system integrating decision trees produced results with high accuracy and that can lead to the solution of the intended problem.

5.3.6 Project usage at the company

The time reduction, the accuracy, the training door that Metrix opens to training new experts and the way it can prevent brain drain at the company by retaining the knowledge and replicating the expert's analysis, turns Metrix into the real solution that JUMIA is currently using for some teams.

The Metrix auto-generated recommendations are being used at some Scrum ceremonies at team level, such as retrospectives to help identifying ways to improve or planning meetings contributing to a data-informed environment that will allow the team to make informed decisions meetings.

The application usage is having good feedback from the teams. Another use that the company is giving to the application is the metrics extraction to use at company-wide meetings. The flexibility of the project allows the data to be extracted and easily configured what makes it a solution to gathering core information about the entire company team's performance.

5.4 Summary

The main goal of this chapter was to detail the experimentation and evaluation process for the Metrix project, emphasizing the methodologies and analytical frameworks employed to assess the solution effectiveness. After the data collection phase, experimentation commenced with rigorous testing on various datasets, including analyses, simulations, and experiments to achieve optimal outcomes. This chapter outlined the systematic approach to evaluating the solution's real impact, including how decision trees were trained to classify team progress based on historical data.

The experimentation outcomes demonstrated varying accuracy levels depending on the dataset configurations, with the most refined dataset yielding an accuracy of 80% to 100%. The evaluation process involved multiple expert contributions, including Agile coaches and external software management professionals, who provided valuable insights on the usability and accuracy of the generated recommendations.

The evaluation methodology was structured in three phases, each tailored to the specific expert type, allowing for a comprehensive assessment of the hybrid solution's capabilities. Quantitative results indicated that the auto-generated insights closely mirrored human analysis, validating the quality of the dataset and the decision tree training. Additionally, the chapter presented qualitative feedback from experts, highlighting both the strengths of the Metrix application and areas for improvement.

The project demonstrated a remarkable reduction in analysis time, achieving an average decrease of 89.80% in total processing time compared to manual methods. It emphasized the potential of Metrix to automate efficiency metrics extraction across various teams while addressing challenges related to knowledge retention and brain drain within the organization.

In conclusion, this chapter provided a detailed exploration of the experimentation and evaluation processes of the Metrix project, showcasing the methodologies used, the accuracy of the outcomes, and the potential for real-world application within the company. The successful integration of the hybrid expert system signifies a promising advancement in automating team performance analysis.

6 Conclusion

The realization of this work primarily focused on developing an innovative solution that enables the automation of metrics analysis in agile development teams. This system aims to enhance decision-making within companies by leveraging data analysis to improve performance, delivery predictability, and work visibility, chapter 5.

The primary objective of this work is to develop a hybrid system that combines existing metrics analysis techniques, such as expert systems and decision trees, with data visualization tools. This system not only automates time-consuming tasks, such as data extraction and cleaning, but also provides automated recommendations that assist teams in making more informed decisions.

Because of the theme domain specificity, a conceptual approach of the domain provides readers with a foundational understanding of key concepts, including development methodologies, process mining, process science and some decision-making support systems – specifically decision trees and expert systems.

The subsequent chapters build on these concepts, presenting extensive research that identified existing solutions that made use of similar approaches. Notably, several of these explored solutions demonstrate their own limitations, which is crucial for understanding the domain existing gaps and some that this thesis project aims to address.

The first phase of this research revealed that the use of metrics analysis across various domains is growing rapidly due to its substantial benefits in improving performance, enhancing delivery predictability, and increasing work visibility (Budacu & Pocatilu, 2018). However, in highly competitive markets, the application of these metrics often relies on specialized professionals, resulting in time-consuming and inefficient processes that directly impact the agility of development teams.

To address these challenges, the proposed solution is a hybrid expert system that generates automatic recommendations. This system leverages the research findings by integrating existing techniques and combines them with decision trees, expert systems and data visualization tools, to automate tasks that would otherwise be performed manually.

The use of decision trees takes advantage of an AI powerful model that is actively utilized during this project to classify teams overall performance progress and identify bottlenecks their processes and workflows (Pathak et al., 2018).

Meanwhile, the implementation of an expert system allows the valuable knowledge of experts to be translated into heuristics and integrated into a robust knowledge base. This approach not only mitigates the risk of knowledge loss but also replicates expert experience by producing reliable, human-like recommendations (Saleem & Azad, 1994).

Ultimately, the integration of these powerful tools culminates in the development of the Metrix application. This innovative solution stands to significantly enhance the efficiency and effectiveness of metrics analysis in agile development teams, fostering better decision-making and improved outcomes.

6.1 Objectives Achieved

This thesis outlines the comprehensive discovery and development process of a potential solution to the following research question: **Can a hybrid expert system, integrating Agile professionals' insights with decision trees, effectively address brain drain and offer practical recommendations to enhance team processes and efficiency in technology companies?**

In addition to this primary research question, several other goals were set. The Metrix application aims to address the analysis scalability problem - automate the extraction of efficiency metrics allowing it to scale to all delivery teams at the company. The solution should be able to solve the human time consumption issue - by reducing the manual time spent on data treatment actions by 50%. Furthermore, it should also find a possible solution to the workflow lack of standardization problem – by solving the Including the processing of dates format and solution for the lack of workflow standardization. Additionally, the application should be able to solve the experts manual review problem - by automatically generating recommendations based on metrics for a given group (team, vertical).

After the solution evaluation, it became clear that it met its objectives by reducing data extraction time by approximately 90% and generating recommendations with accuracy rate of an 85% compared to human specialists.

The application results, presented in section 5, provide a **positive answer to the research question** of this thesis. Yes, a hybrid expert system that integrates Agile professional insights with decision trees can effectively generate automatic practical recommendations to enhance team processes and efficiency in technology companies. This is supported not only by the accuracy that the solution achieved of 85 % when compared to the experts, but also by its practical usage within the company.

The research question of this dissertation questions if the use of the automatic insights generated from the solution would be able to prevent brain drain at companies. The Metrix project can answer this question with a positive note. As supported by its usage inside the company, explored on the section 5.3.6, the solution can indeed help mitigate brain drain within the company. With optimal accuracy results, displayed at the section 5, **the solution can effectively prevent brain drain at a company.** Inside JUMIA, after the agile department got dismantled, the Metrix application auto-generated conclusions accurately pinpointed team workflow deficiencies and suggested improvements. Additionally, by effectively replicate the agile team members knowledge, as proved by the first evaluation results at section 5.3.3, the Metrix project confirmed that a hybrid system, integrating decision trees and an expert system, can prevent the knowledge loss and even be a possible tool to train new agile coaches.

However, since the study was conducted only inside JUMIA company, it is not safe to assume that a solution of this nature can prevent brain drain on a larger scale, as there is no substantial evidence to support that claim.

From a critical perspective, these results suggest that this area offers a clear path for further scientific investigation. The study demonstrates that it is possible to produce accurate, auto-generated performance recommendations in metrics analysis by combining decision trees, guided by historical data, with an expert system. This opens the door to further research that could build on the limitations identified in this study, potentially achieving even better results.

In practical application within companies, the solution showed its potential to address the common problem of brain drain by capturing human expertise in a knowledge base that can evolve over time. The high accuracy of the results also suggests that the system could be utilized for training new professionals, accelerating their learning curve and contributing to both individual development and overall company profitability.

From a personal perspective, the author believes that the proposed solution represents an innovative approach to the identified problem. It offers a pathway to automate team performance analysis while effectively mitigating brain drain and optimizing human resource allocation. By utilizing advanced metrics analysis techniques, this solution empowers teams to make data-driven decisions, thereby enhancing overall operational efficiency.

6.2 Future Work

Despite the above-mentioned promising outcomes, the solution also has its limitations. The lack of existing datasets and comparable solutions posed a significant challenge during development. The dataset used was assembled from a small set of human-generated reports within the company, which constrained the scope of the investigation.

While the system produced favourable results, it is important to note that the study was limited to a single company, reducing the diversity of workflows and behaviours analyzed. Additionally, while the system produced good results, human oversight is still recommended to ensure that the best possible recommendations are delivered to teams. Human reviewers can also provide context that the system, relying solely on data, may miss.

These identified limitations can be addressed in future work. Expanding the dataset would improve the reliability of the analysis. Incorporating a Natural Language Processing (NLP) module could add valuable insights, such as sentiment analysis and the intentions behind team interactions. Additionally, integrating the system with tools like Power BI or other interactive dashboards could enhance the user experience by providing more intuitive visualizations of performance metrics. The use of generative Artificial Intelligence (AI) could also be an interesting avenue for future exploration, potentially addressing the issue of missing context.

In conclusion, the positive results and practical applicability of the solution within the company, supported by extensive research demonstrating the benefits of metrics analysis across various sectors, show that the adoption of automated metrics analysis is not just a management strategy but a necessity for thriving in today’s rapidly evolving competitive landscape.

References

Aduwa, F., Lundquist, J., & Persson, F. (2023). *Implementing DDMRP : Supply Chain transformation Through Inventory Management*. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-199804

Ahmad, M. O., Lwakatare, L. E., Kuvaja, P., Oivo, M., & Markkula, J. (2017). An empirical study of portfolio management and Kanban in agile and lean software companies. *Journal of Software: Evolution and Process*, *29*(6), e1834. https://doi.org/10.1002/SMR.1834

Ahmad, M. O., Markkula, J., & Oivo, M. (2013a). Kanban in software development: A systematic literature review. *Proceedings - 39th Euromicro Conference Series on Software Engineering and Advanced Applications, SEAA 2013*, 9–16. https://doi.org/10.1109/SEAA.2013.28

Ahmad, M. O., Markkula, J., & Oivo, M. (2013b). Kanban in software development: A systematic literature review. *Proceedings - 39th Euromicro Conference Series on Software Engineering and Advanced Applications, SEAA 2013*, 9–16. https://doi.org/10.1109/SEAA.2013.28

Alaidaros, H., Omar, M., & Romli, R. B. (2018). *Towards an Improved Software Project Monitoring Task Model of Agile Kanban Method*. http://excelingtech.co.uk/

Anttila, S. (2016). *The hidden pitfalls of Kanban in software development*. https://aaltodoc.aalto.fi/handle/123456789/15160

Bezsmertnyi, O., Golian, N., Golian, V., & Afanasieva, I. (2021). Behavior Driven Development Approach in the Modern Quality Control Process. *2020 IEEE International Conference on Problems of Infocommunications Science and Technology, PIC S and T 2020 - Proceedings*, 217–220. https://doi.org/10.1109/PICST51311.2020.9467891

Bhadauria, V. S., Mahapatra, R. K., & Nerur, S. P. (2020). Performance Outcomes of Test-Driven Development: An Experimental Investigation. *Journal of the Association for Information Systems*, *21*(4), 2. https://doi.org/10.17705/1jais.00628

Binamungu, L. P., Embury, S. M., & Konstantinou, N. (2018). Maintaining behaviour driven development specifications: Challenges and opportunities. *25th IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2018 - Proceedings*, *2018-March*, 175–184. https://doi.org/10.1109/SANER.2018.8330207

Binamungu, L. P., & Maro, S. (2023). Behaviour driven development: A systematic mapping study ☆. *The Journal of Systems & Software*, *203*, 111749. https://doi.org/10.5281/zenod

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, *6*, 1124553. https://doi.org/10.3389/FRAI.2023.1124553/BIBTEX

Braun, C. M. (2021). *Breaking the Mold on Job Shops*. https://dspace.mit.edu/handle/1721.1/139180

BUDACU, E., & POCATILU, P. (2018). Real Time Agile Metrics for Measuring Team Performance. *Informatica Economica*, *22*(4/2018), 70–79. https://doi.org/10.12948/issn14531305/22.4.2018.06

Chekushina, E. V., Vorobev, A. E., & Chekushina, T. V. (2013). Use of expert systems in the mining. *Middle East Journal of Scientific Research*, *18*(1), 1–3. https://doi.org/10.5829/idosi.mejsr.2013.18.1.12345

Davis. (2015a). *Agile metrics in action: how to measure and improve team performance* .

Davis. (2015b). *Agile Metrics in Action: How to measure and improve team performance - Christopher Davis - Google Livros*. https://books.google.pt/books?hl=pt-PT&lr=&id=sjszEAAAQBAJ&oi=fnd&pg=PP13&dq=agile+metrics+analysis+tech+companies+benefits&ots=ezl1e0-AxK&sig=7y1jBk26C4dXVK9pSHkEnLVUJtk&redir_esc=y#v=onepage&q&f=false

Deloitte. (2021). *Global Process Mining Survey 2021*. https://www2.deloitte.com/kz/en/pages/risk/articles/global-process-mining-survey-2021.html

Enayati, M., Bozorg-Haddad, O., Pourgholam-Amiji, M., Zolghadr-Asli, B., & Tahmasebi Nasab, M. (2022). Decision Tree (DT): A Valuable Tool for Water Resources Engineering. *Studies in Computational Intelligence*, *1043*, 201–223. https://doi.org/10.1007/978-981-19-2519-1_10

Eremia, M., Tomsovic, K., & Cârtină, G. (2016). Expert Systems. *Advanced Solutions in Power Systems: HVDC, FACTS, and AI Techniques*, 731–754. https://doi.org/10.1002/9781119175391.CH15

Eriksson, P. (2023). *Effects on Software Quality and Collaboration with Behavior-Driven Development*. https://urn.kb.se/resolve?urn=urn:nbn:se:bth-24689

Fagarasan, C., Cristea, C., Mihele, C., Popa, O., Ciceo, D., & Pisla, A. (2023). Kanban in Software Development—The Role of Leadership and Metrics. *Lecture Notes in Production Engineering*, *Part F1162*, 369–384. https://doi.org/10.1007/978-3-031-15602-1_28/COVER

Farcic, Viktor., & Garcia, Alex. (2015). *Test-driven Java development invoke TDD principles for end-to-end application development with Java*. Packt Publishing.

Fernando, L., & Mattos, B. (2019). *Agile or Non-Agile, That Is the Question Designing a Decision Support System for an Agile Approach in Software Development Projects*.

Forsberg, K., Mooz, H., & Cotterman, H. (n.d.). *Visualizing Project Management Models and frameworks for mastering complex systems Third Edition*.

Franklin, A., Gantela, S., Shifarraw, S., Johnson, T. R., Robinson, D. J., King, B. R., Mehta, A. M., Maddow, C. L., Hoot, N. R., Nguyen, V., Rubio, A., Zhang, J., & Okafor, N. G. (2017). Dashboard visualizations: Supporting real-time throughput decision-making. *Journal of Biomedical Informatics*, *71*, 211–221. https://doi.org/10.1016/J.JBI.2017.05.024

Gotz, P. (2021). *The Professional Scrum Team*.

Greening, D. R. (2015). Agile enterprise metrics. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2015-March*, 5038–5044. https://doi.org/10.1109/HICSS.2015.597

Gundogar, E., Sari, M., & Kokcam, A. H. (2016). Dynamic bottleneck elimination in mattress manufacturing line using theory of constraints. *SpringerPlus*, *5*(1), 1–15. https://doi.org/10.1186/S40064-016-2947-1/FIGURES/6

Hachemi, I., & Bakhouche, M. (2024). The impact of the Kanban model on Toyota's agility. *Dirassat Journal Economic Issue*, *15*(2), 233–248. https://doi.org/10.34118/djei.v15i2.3926

Harrison, D., & Lively, K. (2019). Achieving DevOps. *Achieving DevOps*. https://doi.org/10.1007/978-1-4842-4388-6

Heidenberg, J. (2011). *Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches*.

Helsinki, A. W., & Wikström, A. (2019). *Benefits and challenges of Continuous Integration and Delivery-A Case Study*.

Hemalatha, C., Sankaranarayanasamy, K., & Durairaaj, N. (2021). Lean and agile manufacturing for work-in-process (WIP) control. *Materials Today: Proceedings*, *46*, 10334–10338. https://doi.org/10.1016/J.MATPR.2020.12.473

Hingole, R. S. (2015). Fundamentals of expert system. *Springer Series in Materials Science*, *206*, 31–39. https://doi.org/10.1007/978-3-662-44497-9_3

Hroncova, N., & Dakic, P. (2022). Research Study on the Use of CI/CD Among Slovak Students. *2022 12th International Conference on Advanced Computer Information Technologies, ACIT 2022*, 458–461. https://doi.org/10.1109/ACIT54803.2022.9913113

Issagaliyev, A. (2023). *Investigation of the challenges of agile project management at Infineon Technologies*. https://opus4.kobv.de/opus4-haw/frontdoor/index/index/docId/3875

Jancura, E. G., & Overbey, J. T. (n.d.). *EDP: Expert Systems in Training*. Retrieved September 29, 2024, from https://egrove.olemiss.edu/wcpa

Jarmusch, A., Cabarcas, F., Pophale, S., Kallai, A., Doerfert, J., Peyralans, L., Lee, S., Denny, J., & Chandrasekaran, S. (2024). *CI/CD Efforts for Validation, Verification and Benchmarking OpenMP Implementations*. 111–125. https://doi.org/10.1007/978-3-031-72567-8_8

Kadam, K. D. (2022). *Application of Lean in High-Mix Low-Volume Production Systems: A Case Study in the Architectural Lighting Industry*.

Kamath, D. (2023). *Improving Agile Development Prac-tices*.

Khaled Yacoub, M., Mostafa, A. A., & Farid, A. B. (n.d.). *A New Approach for Distributed Software Engineering Teams Based on Kanban Method for Reducing Dependency*. https://doi.org/10.17706/jsw.11.12.1231-1241

Khan. (2015). *Survey on development of expert system in the areas of Medical, Education, Automobile and Agriculture | IEEE Conference Publication | IEEE Xplore*. https://ieeexplore.ieee.org/document/7100354

Kirovska, N., & Koceski, S. (2015). *USAGE OF KANBAN METHODOLOGY AT SOFTWARE DEVELOPMENT TEAMS*.

Kothandapani, H. P. (2019). Drivers and Barriers of Adopting Interactive Dashboard Reporting in the Finance Sector: An Empirical Investigation. *Reviews of Contemporary Business Analytics*, *2*(1), 45–70. https://researchberg.com/index.php/rcba/article/view/170

Krishnan, Sangeeta. (2023). *Thriving in a Data World : A Guide for Leaders and Managers*. 152.

Kupiainen, E., Mäntylä, M. V., & Itkonen, J. (2015). Using metrics in Agile and Lean software development - A systematic literature review of industrial studies. In *Information and Software Technology* (Vol. 62, Issue 1, pp. 143–163). Elsevier B.V. https://doi.org/10.1016/j.infsof.2015.02.005

Lawrence, C., Tuunanen, T., & Myers, M. D. (2010). Extending design science research methodology for a multicultural world. *IFIP Advances in Information and Communication Technology*, *318*, 108–121. https://doi.org/10.1007/978-3-642-12113-5_7

Lee, Liu, & Jin. (2014). *Decision Trees: Theory and Algorithms | Request PDF*. https://www.researchgate.net/publication/266970946_Decision_Trees_Theory_and_Algorithms

Lee, T., Huang, Y. Y., & Lu, T. Y. (2014). A unification of behavior models. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 212–215. https://doi.org/10.1109/ICSESS.2014.6933547

Leskinen, M. (2015). *Towards a flow efficient ICT development process with Kanban: A Case Study*.

Lewis, N. A. (n.d.). *ARE LEAN MANUFACTURING EFFORTS REFLECTED IN CORPORATE FINANCES?*

Liang, Q. (2021). Continuous Delivery 2.0: Business-leading DevOps Essentials. *Continuous Delivery 2.0: Business-Leading DevOps Essentials*, 1–331. https://doi.org/10.1201/9781003221579/CONTINUOUS-DELIVERY-2-0-QIAO-LIANG/ACCESSIBILITY-INFORMATION

Lindstrom, N. B. (2023). It is No Blame Game! Challenges and Best Practices in Communicating Metrics in Software Development Organizations. *International Research Workshop on IT Project Management 2023*. https://aisel.aisnet.org/irwitpm2023/1

Liu, J. Y. C., Chen, V. J., Chan, C. L., & Lie, T. (2008). The impact of software process standardization on software flexibility and project management performance: Control theory perspective. *Information and Software Technology*, *50*(9–10), 889–896. https://doi.org/10.1016/j.infsof.2008.01.002

Loriot, N., & Lebossé, F. (2021). *L'automatisation au service de la défense et de la sécurité par conception dans un monde déconnecté*.

Majdzadeh, R. (2024). Big Data Revolution: Transforming Business Landscapes through Data-Driven Decision Making. *Social Sciences Spectrum*, *3*(1), 115–125. https://sss.org.pk/index.php/sss/article/view/31

Making Work Visible EXPOSING TIME THEFT TO OPTIMIZE WORK & FLOW DOMINICA DeGRANDIS FOREWORD BY TONIANNE DeMARIA IT REVOLUTION PRESS PORTLAND, OR. (2017). www.ITRevolution.com.

Malohlava, M., Hnetynka, P., & Bures, T. (2013). SOFA 2 component framework and its ecosystem. *Electronic Notes in Theoretical Computer Science*, *295*, 101–106. https://doi.org/10.1016/J.ENTCS.2013.04.009

Mauvius. (2022). *The-Official-Kanban-Guide_A4*.

Mikko Kurkela DevOps Capability Assessment in a Software Development Team. (n.d.).

Nogués, A., & Valladares, J. (2017). Business Intelligence Tools for Small Companies. *Business Intelligence Tools for Small Companies*. https://doi.org/10.1007/978-1-4842-2568-4

Owen, G., & Varhol, P. (n.d.). *Timeless Skills for Modern Testers: Communication, Collaboration and Creativity*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., …

Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, *372*. https://doi.org/10.1136/BMJ.N71

Pathak, S., Mishra, I., & Swetapadma, A. (2018). An Assessment of Decision Tree based Classification and Regression Algorithms. *Proceedings of the 3rd International Conference on Inventive Computation Technologies, ICICT 2018*, 92–95. https://doi.org/10.1109/ICICT43934.2018.9034296

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pekarcikova, M., Trebuna, P., Kliment, M., & Dic, M. (2021). Solution of Bottlenecks in the Logistics Flow by Applying the Kanban Module in the Tecnomatix Plant Simulation Software. *Sustainability 2021, Vol. 13, Page 7989*, *13*(14), 7989. https://doi.org/10.3390/SU13147989

Pereira, L., Sharp, H., De Souza, C., Oliveira, G., Marczak, S., & Bastos, R. (2018). Behavior-driven development benefits and challenges: Reports from an industrial study. *ACM International Conference Proceeding Series*, *Part F147763*. https://doi.org/10.1145/3234152.3234167

Petersen, K., & Wohlin, C. (2011). Measuring the flow in lean software development. *Software: Practice and Experience*, *41*(9), 975–996. https://doi.org/10.1002/SPE.975

Piela, J. (2017). *Key performance indicator analysis and dashboard visualization in a logistics company*. https://lutpub.lut.fi/handle/10024/147689

Poth, A., Werner, M., & Lei, X. (2018). How to Deliver Faster with CI/CD Integrated Testing Services? *Communications in Computer and Information Science*, *896*, 401–409. https://doi.org/10.1007/978-3-319-97925-0_33

Process Mining Market by Offering (Software (Process Discovery Tools, Conformance Checking Tools), Services), Mining Algorithm (Deep Learning, Sequence Analysis), Data Source (ERP Systems, CRM Systems), Vertical and Region - Global Forecast to 2028. (n.d.). Retrieved December 18, 2023, from https://www.researchandmarkets.com/report/process-mining?utm_source=GNE&utm_medium=PressRelease&utm_code=rq3bs3&utm_campaign=1906248+-+Global+Process+Mining+Markets%2c+2023-2028%3a+North+America+Takes+the+Lead+in+Market+Growth%2c+While+Europe+Dominates+with+the+Largest+Market+Share&utm_exec=joca220prd

Prutchi, E. S., de S. Campos Junior, H., & Murta, L. G. P. (2022). How the adoption of feature toggles correlates with branch merges and defects in open-source projects? *Software: Practice and Experience*, *52*(2), 506–536. https://doi.org/10.1002/SPE.3034

Raftoudi, & Christina. (2023). *Project Management Analytics*. https://repository.ihu.edu.gr//xmlui/handle/11544/30197

Ram, P., Rodriguez, P., & Oivo, M. (2018). Software process measurement and related challenges in agile software development: A multiple case study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11271 LNCS*, 272–287. https://doi.org/10.1007/978-3-030-03673-7_20

Ram, P., Rodriguez, P., Oivo, M., & Martinez-Fernandez, S. (2019). Success factors for effective process metrics operationalization in agile software development: A multiple case study. *Proceedings - 2019 IEEE/ACM International Conference on Software and System Processes, ICSSP 2019*, 14–23. https://doi.org/10.1109/ICSSP.2019.00013

Ram, P., Rodríguez, P., Oivo, M., Martínez-Fernández, S., Bagnato, A., Choras, M., Kozik, R., Aaramaa, S., & Ahola, M. (2020a). Actionable Software Metrics: An Industrial Perspective. *ACM International Conference Proceeding Series*, 240–249. https://doi.org/10.1145/3383219.3383244

Ram, P., Rodríguez, P., Oivo, M., Martínez-Fernández, S., Bagnato, A., Choras, M., Kozik, R., Aaramaa, S., & Ahola, M. (2020b). Actionable Software Metrics: An Industrial Perspective. *ACM International Conference Proceeding Series*, 240–249. https://doi.org/10.1145/3383219.3383244

Raquel, C., & Lopes, J. (2023). *Metrology Product Development for Industry and Science Software Project Management Best Practices From a Multiple Case-Study Analysis*.

Rutkowski, L., Pietruczuk, L., Duda, P., & Jaworski, M. (2013). Decision trees for mining data streams based on the mcdiarmid's bound. *IEEE Transactions on Knowledge and Data Engineering*, *25*(6), 1272–1279. https://doi.org/10.1109/TKDE.2012.66

Saleem, N., & Azad, A. N. (1994). EXPERT SYSTEMS APPROACH TO MINIMIZE THE EFFECTS OF BRAIN DRAIN. *International Journal of Commerce and Management*, *4*(3), 68–75. https://doi.org/10.1108/EB047295

Saleem, N., & Moses, B. (1994). Expert systems as computer assisted instruction systems for nursing education and training. *Computers in Nursing*, *12*(1), 35–45. https://europepmc.org/article/med/8149301

Saltz, J. S., Hotz, N., & Sutherland, A. (2022). *Achieving Lean Data Science Agility Via Data Driven Scrum*. https://hdl.handle.net/10125/80218

Sampene, A. K., Agyeman, F. O., Wiredu, J., Brenya, R., & Robert, B. (2022). *Artificial Intelligence as a Path Way to Africa's TransformationS*. www.jmest.org

Schneid. (2017). *Branching strategies for developing new features within the context of Continuous Delivery | Request PDF*.

https://www.researchgate.net/publication/332735192_Branching_strategies_for_developing_new_features_within_the_context_of_Continuous_Delivery

Senavirathne, N., & Torra, V. (2020). On the role of data anonymization in machine learning privacy. *Proceedings - 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2020*, 664–675. https://doi.org/10.1109/TRUSTCOM50675.2020.00093

Shaked, A., & Reich Phd, Y. (2019). Improving Coordination and Collaboration in Connected and Automated Vehicle Development Projects Using Model Based Process Design. *SAE Technical Papers*, *2019-January*(January). https://doi.org/10.4271/2019-01-0103

Shore, J. (2022). *The Art of Agile Development*.

Şimşit, Z. T., Günay, N. S., & Vayvay, Ö. (2014). Theory of Constraints: A Literature Review. *Procedia - Social and Behavioral Sciences*, *150*, 930–936. https://doi.org/10.1016/J.SBSPRO.2014.09.104

Skeie, T., & Skeie, T. (2015). *Does Limit on Work-In-Progress (WIP) in Software Development Matter?* https://www.duo.uio.no/handle/10852/42080

Sofia Barroso de Araújo, A., Doutora Anabela Carvalho Alves, P., & Doutor Fernando Carlos Cabrita Romero, P. (2022). *Universidade do Minho Escola de Engenharia Production Control System for Lean and Agile Processes*.

Staron, M., & Meding, W. (2015). *Monitoring Bottlenecks in Agile and Lean Software Development Projects-A Method and its Industrial Use*.

Storey, V. C., & Baskerville, R. L. (2022). Computational Science: A Field of Inquiry for Design Science Research. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2022-January*, 5768–5777. https://doi.org/10.24251/HICSS.2022.703

The theory of constraints. (2013). *The Routledge Companion to Cost Management*, 145–162. https://doi.org/10.4324/9780203101261-12

Trainer, E. H., & Redmiles, D. F. (2018). Bridging the gap between awareness and trust in globally distributed software teams. *Journal of Systems and Software*, *144*, 328–341. https://doi.org/10.1016/J.JSS.2018.06.028

Tsilionis, K., Wautelet, Y., & Heng, S. (2022). Building a Unified Ontology for Behavior Driven Development Scenarios. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13709 LNCS*, 518–524. https://doi.org/10.1007/978-3-031-21388-5_36

Understand the Theory behind Clean Architecture. (n.d.). Retrieved September 28, 2024, from https://blog.scalablebackend.com/understand-the-theory-behind-clean-architecture

Ünlü, C. (2008). *The effects of test driven development on software productivity and software quality*. https://open.metu.edu.tr/handle/11511/17784

Vacanti, D. S. (2016). *Actionable Agile Metrics for Predictability An Introduction*. http://leanpub.com/actionableagilemetrics

van der Aalst, W. M. P. (2022). Process Mining: A 360 Degree Overview. *Lecture Notes in Business Information Processing*, *448*, 3–34. https://doi.org/10.1007/978-3-031-08848-3_1/TABLES/3

Van Der Vegt, C. (2023). *Analysing and visualising data to improve the productivity level of an Agile organised company Bachelor thesis*.

Van Looy, A. (2024). From Emerging Technologies to Business Opportunities. *From Emerging Technologies to Business Opportunities*. https://doi.org/10.1007/978-3-031-59770-1

Vegt, C. R. van der. (2021). *Analysing and visualising data to improve the productivity level of an Agile organised company*.

Verbruggen, F., Sutherland, J., Martijn Van Der Werf, J., Brinkkemper, S., & Sutherland, A. (2019). *Process Efficiency-Adapting Flow to the Agile Improvement Effort*. https://hdl.handle.net/10125/60134

Viehhauser, D. (2023). *Chair of Industrial Logistics Simulation-based Bottleneck Identification in a Job-shop with Changing Product Mixture and Production Volumes*.

Visser, J. (2017). Building Software Teams: Ten Best Practices for Effective Software Development. *Communications of the ACM*, *33*(10). https://www.oreilly.com/library/view/building-software-teams/9781491951781/

vom Brocke, J., Gau, M., & Mädche, A. (2021). Journaling the Design Science Research Process. Transparency About the Making of Design Knowledge. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12807 LNCS*, 131–136. https://doi.org/10.1007/978-3-030-82405-1_15

vom Brocke, J., van der Aalst, W. M., Grisold, T., Kremser, W., Mendling, J., Pentland, B., Recker, J., Roeglinger, M., Rosemann, M., & Weber, B. (2021). Process Science: The Interdisciplinary Study of Continuous Change. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3916817

Wang, C. N., Vo, T. T. B. C., Hsu, H. P., Chung, Y. C., Nguyen, N. T., & Nhieu, N. L. (2024). Improving processing efficiency through workflow process reengineering, simulation and value stream mapping: a case study of business process reengineering. *Business Process Management Journal*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/BPMJ-11-2023-0869/FULL/XML

Weber, R. (2018). Design-science research. *Research Methods: Information, Systems, and Contexts: Second Edition*, 267–288. https://doi.org/10.1016/B978-0-08-102220-7.00011-X

Weflen, E., MacKenzie, C. A., & Rivero, I. V. (2022). An influence diagram approach to automating lead time estimation in Agile Kanban project management. *Expert Systems with Applications*, *187*, 115866. https://doi.org/10.1016/J.ESWA.2021.115866

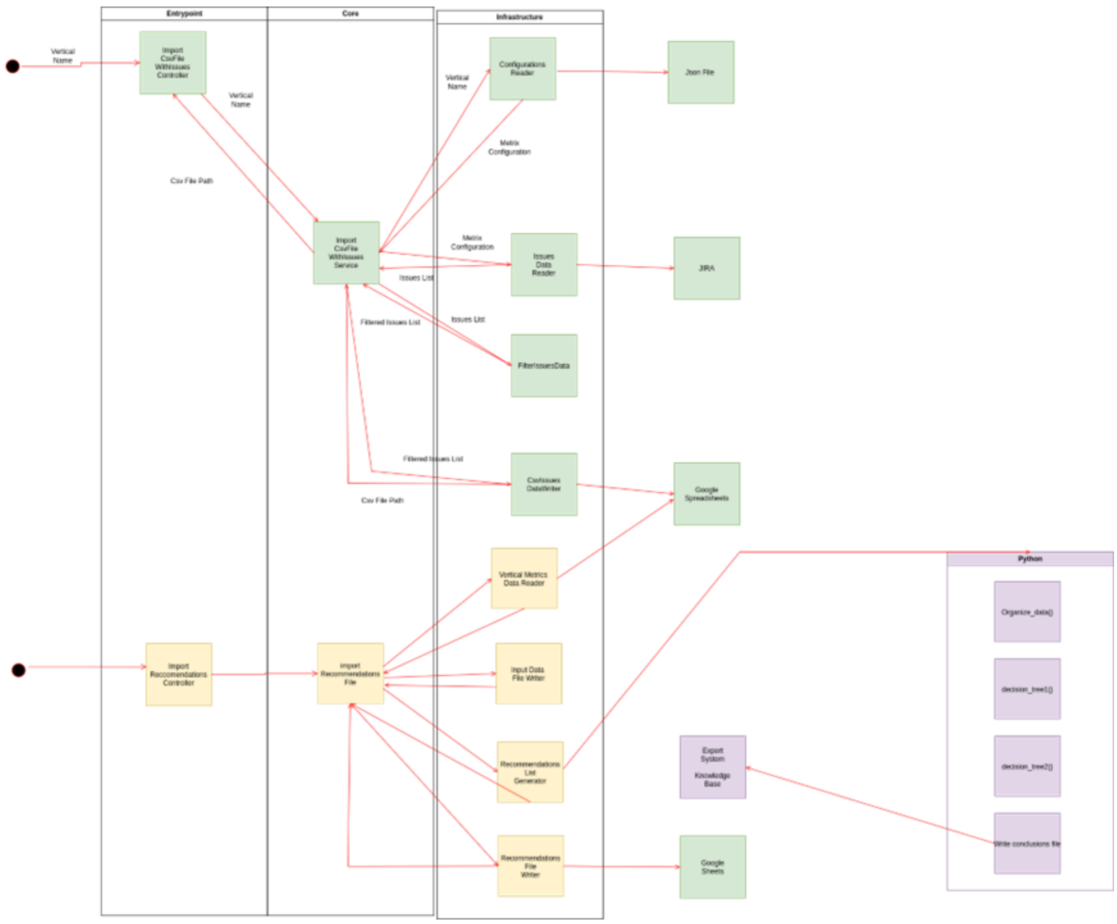
What is process mining? | IBM. (n.d.). Retrieved January 21, 2024, from https://www.ibm.com/topics/process-mining.

Zalevskaya, L. (2019). *Software development process improvements - Case QPR Software Plc*. http://www.theseus.fi/handle/10024/265364

Zeggaf, A. (2023). *Automated Data Collection and Dashboard Development : Enhancing Quality Monitoring and Analysis*.

Καθηγητής, Ε., & Λουκάς Τσιρώνης, : Κ. (2024). *Agile project management in big data projects*. http://dspace.lib.uom.gr/handle/2159/31140

Annex


Entrypoint
Core
Infrastructure
Vertical Name
Import CsvFile WithIssues Controller
Vertical Name
Configurations Reader
Json File
Vertical Name
Metrix Configuration
Csv File Path
Import CsvFile WithIssues Service
Metrix Configuration
Issues Data Reader
JIRA
Issues List
Issues List
Filtered Issues List
FilterIssuesData
Filtered Issues List
CsvIssues DataWriter
Csv File Path
Google Spreadsheets
Vertical Metrics Data Reader
Python
Organize_data()
Import Reccomendations Controller
import Recommendations File
Input Data File Writer
decision_tree1()
Export System
Knowledge Base
decision_tree2()
Recommendations List Generator
Write conclusions file
Recommendations File Writer
Google Sheets