People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

SAAD DAHLAB BLIDA 1 UNIVERSITY
Faculty of Sciences
Computer Science Department

End of Studies Project Thesis
To Obtain The Master's Degree in Computer Science

Option: *Software Engineering*

Business Intelligence And Machine Learning Platform For Enhanced Project Management at SONATRACH

Realized by:
Mr. MEGUENNI Aymen
Mr. ASSAMEUR Aymen Abdechafi

Academic Supervisor:
Dr. CHIKHI Nacim Fateh

Company Supervisors:
Mr. BOUHIDEL Sofiane
Mr. OULD KHAOUA Toufik

Defended On 02/07/2024 In Front Of The Jury Members:
President: Prof. OUKID Saliha
Examinator: Dr. MIDOUN Khadidja

Academic year: 2023 – 2024

Abstract

Since 1963, Sonatrach has been the leading gas and petroleum company in Algeria. As part of its digital transformation efforts, Sonatrach identified the need for a modernized project management system to enhance operational efficiency and data security. This thesis presents a comprehensive web platform tailored to Sonatrach's needs, integrating Business Intelligence (BI) and Machine Learning (ML) technologies. Our solution features dynamic BI dashboards with Tableau for improved decision-making and ML models for project classification. The platform streamlines information flow, improves data accessibility, and ensures robust data security, positioning Sonatrach ahead in the competitive petroleum industry. This work underscores the transformative potential of modern technologies in project management and offers a foundation for future enhancements.

Keywords: Business Intelligence, Machine Learning, Project Management, Sonatrach, Digital Transformation, Dashboards, Web Platform, Tableau

ملخص

منذ عام 1963، كانت سوناطراك الشركة الرائدة في مجال الغاز والبترول في الجزائر. كجزء من جهود التحول الرقمي الخاصة بها، حددت سوناطراك الحاجة إلى نظام إدارة مشاريع حديث لتعزيز الكفاءة التشغيلية وأمان البيانات. تقدم هذه الأطروحة منصة ويب شاملة مصممة خصيصا لاحتياجات سوناطراك، تدمج تقنيات الذكاء الاصطناعي والتعلم الآلي. تتميز حلولنا بلوحات قيادة ديناميكية باستخدام تابلو بغرض تحسين عملية اتخاذ القرارات ونماذج التعلم الآلي لتصنيف المشاريع. تسهل المنصة تدفق المعلومات، وتعزز إمكانية الوصول إلى البيانات، وتضمن أمان البيانات الحساسة، مما يضع سوناطراك في موقع متقدم في صناعة البترول و الغاز الاحترافية (مجال الطاقة عموما). يؤكد هذا العمل الإمكانات التحولية الجذرية للتقنيات الحديثة في إدارة المشاريع ويقدم أساسا صلبا لتحسينات مستقبلية.

الكلمات المفتاحية: ذكاء الأعمال، التعلم الآلي، إدارة المشاريع، سوناطراك، التحول الرقمي، لوحات القيادة، منصة الويب، تابلو

Résumé

Depuis 1963, Sonatrach est la principale compagnie gazière et pétrolière en Algérie. Dans le cadre de ses efforts de transformation numérique, Sonatrach a identifié le besoin d'un système de gestion de projets modernisé pour améliorer l'efficacité opérationnelle et la sécurité des données. Cette thèse présente une plateforme web complète adaptée aux besoins de Sonatrach, intégrant des technologies d'intelligence économique (BI) et d'apprentissage automatique (ML). Notre solution comprend des tableaux de bord BI dynamiques avec Tableau pour améliorer la prise de décision et des modèles ML pour la classification des projets. La plateforme rationalise le flux d'informations, améliore l'accessibilité des données et assure une sécurité des données robuste, positionnant Sonatrach en tête de l'industrie pétrolière compétitive. Ce travail souligne le potentiel transformateur des technologies modernes dans la gestion de projets et offre une base pour des améliorations futures.

Mots-clés: Intelligence économique, Apprentissage automatique, Gestion de projet, Sonatrach, Transformation numérique, Tableaux de bord, Plateforme web, Tableau

Author's Words

This work is dedicated to my parents, **Assameur Hamid**, **Bounemri Fazia**, **Bounemri Yacout** and **Bounemri Rabia**. Without their help, sacrifices, and constant support, I would not be where I am today. Their daily encouragement throughout my academic journey has shaped the person I am today.

It is also dedicated to my brother, **Assameur Mohamed**, and my sisters, **Assameur Kaouther** and **Assameur Nafissa**, who have inspired me and continue to do so.

I would also like to dedicate it to my uncle, **Assameur Said**, for his immense support throughout the internship period. His guidance has been incredible and immensely helpful during this time.

Finally, I would like to dedicate this work to all those who have helped us with this project, especially **Mr. Bouhidel Sofiane** and **Mr. Chikhi Fateh Nacim**. Their guidance has been an important cornerstone in reaching this milestone.

Aymen Abdechafi

Author's Words

"A" for Almighty, Thou who inscribed it in the Preserved Tablet, created Adam with Thy hand, and said to all, "Be", and so it was! Second to none indeed, He is worthy of my praise and thankfulness for granting me the power, wisdom, and patience to achieve this humble endeavor that is -hopefully- not our magnum opus just yet.

I feel most humbled looking back at all the sacrifices made by two people for me—two whom I idealized at first, then questioned, and finally understood. Only when I stood on equal adult ground did I truly understand, though I will never be their equal. Idealization is the state furthest from understanding. Mother, Father, thank you.

Isn't fate fascinating? Through the journey, we meet friends, adversaries, and people we love. Some are present, others are not. I feel immensely grateful for it all, which led to this very moment. All these people shaped the vessel that I am (and hopefully vice versa, in a good way). My brother, two sisters, close friends, and other friends and colleagues are clear images that come to mind. I thank them all for the positivity, inspiration, and encouragement they brought into my life and continue to do so. I am glad.

P vs NP? Ambition and pride? Resources? Or is it perhaps the blessings? Isn't it a pity to fall for the Übermensch temptation, a deal with the morning star, empty ideals, Is a bishop really worth more than a knight?! Regardless, at this moment, it all matters not, with no substance; all that is in sight is a completed work. I can finally put my guard down and rest a long, peaceful one. But make no mistake, "selflessness" is an understatement. None of this would have been possible without the compounding contributions of many people. My project partner, Aymen Abdechafi, is first among them—a contributor not only to getting the job done but in *style* as well, and that's what matters most. For it, I address all respect and salutes, wishing him all the well-deserved success for his future path, as well as to my other software engineering comrades and other talented acquaintances I have crossed paths with.

"In all chaos, there is a cosmos, in all disorder a secret order"; what an occasion to reflect and see the flashing past 24 years for what it is—a miracle, a single-celled organism has managed to evolve a little, and although late, starting adventure is still worth having. The future is right within our grasp; or not?

Aymen

Acknowledgements

We begin by expressing our deepest gratitude to Allah, the Most Merciful and Compassionate, for granting us the strength and wisdom to complete this work. We are immensely thankful to Mr. ASSAMEUR Said for his invaluable assistance in helping us quickly integrate into the company environment and culture during our initial days of internship. His guidance and insights into the internal structure were instrumental in our understanding of the organization.

We extend our sincere appreciation to Mr. BOUHIDEL Sofiane from Sonatrach, who generously dedicated his time and patience over the past 6 months to support us in delivering the highest quality work possible. Special thanks are due to Mr. OULD KHAOUA Toufik for proposing this challenging project and for believing in our ability to tackle it. We are grateful for his guidance, encouragement, and patience in addressing our numerous queries.

We also wish to acknowledge Mr. CHIKHI Nacim Fateh, our academic supervisor, whose expertise and mentorship have been invaluable throughout this project. His dedication to excellence and his exemplary skills in Software Engineering and as an Academic, have inspired and guided us not only during this project but also throughout our academic journey. We are particularly grateful for the freedom he gave us to be creative throughout the realization of this project.

Our heartfelt gratitude extends to the professors and staff of the Computer Science Department at the University of Blida for their unwavering support and guidance. We also wish to thank the software engineering mentors, senior comrades we crossed paths with inside university clubs like "ITC" and "CSC" in our early days in the computer science department, who guided us and helped shape our understanding and skills.

We are immensely grateful to Sonatrach's Production Division for their warm welcome and assistance during our internship.

We would like to express our appreciation to the judges who honored us by accepting to judge our work.

This work is dedicated to our families, the Assameur and Meguenni families, and to all the people who have believed in us, supported us, and shared in our excitement and challenges along the way.

Last but not least, we extend our thanks to our friends and all those who have shown us kindness and support throughout this endeavor. Your encouragement and belief in us have been a source of strength and motivation.

We dedicate this work to future generations, hoping it will serve as a stepping stone for their endeavors and aspirations.

Contents

List of Figures

List of Tables

List of Abbreviations

- **PM:** Project Management
- **ML:** Machine Learning
- **BI:** Business Intelligence
- **KPI:** Key Performance Indicators
- **OPEC:** Organization of the Petroleum Exporting Countries
- **EP:** Exploration And Production
- **HRM:** Hassi R'mel
- **HBK:** Haoud Berkaoui
- **HMD:** Hassi Messaoud
- **REB:** Rhourd El Baguel
- **GTL:** Gassi Touil
- **RNS:** Rhourd Ennouss
- **TFT:** Tin Fouye Tabankort
- **OHT:** Ohanet
- **STH:** Stah
- **INA:** In Amenas
- **ADR:** Adrar
- **DAT:** Direction Approvisionnement et Transport
- **DMG:** Direction Moyens Généraux
- **INF:** Direction Informatique
- **HSE:** Département Hygiène, Sécurité et Environnement
- **PMT:** Plan Moyen Terme / Mid Term Plan
- **RBAC:** Role Based Access Control
- **API:** Application Programming Interface
- **HTTP:** Hypertext Transfer Protocol
- **RDBMS:** Relational Database Management System
- **JWT:** Json Web Tokens
- **AI:** Artificial Intelligence
- **ORM:** Object Relational Mapping
- **TF-IDF:** Term Frequency-Inverse Document Frequency
- **MVC:** Model View Controller

General Introduction

Since 1963, **Sonatrach** has been the leading gas and petroleum company in Algeria, a pillar of its economy and energy sector. Over time, **Sonatrach** has developed to meet the growing demands of the global energy market. and similar to other long-standing companies, it faces the challenge of modernizing its technological infrastructure. For institutions such as **Sonatrach**, the call to implement digital transformation has become loud, which is essential for better operational efficiency and competitive advantage. Realizing its dire necessity to renovate its workflows and technological tools -specifically- within its production division, among others, it embraced digital transformation.

Why? The main purpose of this Thesis is to answer Sonatrach's need for a more intelligent and effective way of visualizing and managing information (in comparison to the current not-so-intelligent solution -for the lack of a better descriptive polite word-) regarding projects across its 11 regions of operation. This solution should consider the timely, central, and confidential aspects of information which is a challenging task, however, is essential to modernize the company. For this purpose, we have constructed a complete web platform adapted to the needs of **Sonatrach** in the aspect of project management.

Our offered platform is a centralized system for managing projects, fully equipped with modern Business Intelligence (BI) dashboards and powered with machine learning functionalities for the classification of projects. which is a big improvement over the current traditional, outdated methods that the company is using, which are heavily dependent on manual processes and other kinds of isolated communication channels. For this purpose, the company, **Sonatrach**, has entered into this territory to mainly streamline its information flow, improve data accessibility, and ensure strong data security as well, this step will put it ahead of the competition in the technologically dynamic scene of the petroleum industry.

The most important part of the BI functionality of the platform is the integration of **Tableau** technology since it holds powerful data visualization and a user-friendly interface that gives it the best possible experience in interaction with data, enriching the decision-making process. The project's incorporation of machine learning functionality further strengthens its ability to classify projects smartly and automatically (in order to predict projects families), as is necessary for handling the vast and complex data associated with **Sonatrach** and specifically that of the Production division.

The structure of our Thesis covers three critical components of our process: **"State of the Art"** The first one deals with the most current applications of BI technologies and machine learning in project management, with a specific focus on the petroleum industry. We present current trends, best practices, and potentials for transforming technologies in

operational contexts and strategic planning.

The second part of **"Study of the Existing Situation and Analysis of Needs"** shows a detailed review of the organizational structure of **Sonatrach** and the current processes applied by its production division. We collect important insights regarding the needs and challenges of this division through comprehensive analysis and engagement with current systems users. This is the fundamental work for creating a solution that is appropriate and effective.

The third part, **"Proposed Solution"** depicts our conceptual and practical approach toward the development of the new platform. We start with the -Theoretical- design aspect and then proceed toward the -Practical- implementation phases right from data acquisition up to the development of user interfaces. We elaborate on the architecture, functionalities, and processes for the integration of the system to prove that our solution addresses the identified needs and improves access to information, visualization, and confidentiality.

We finalize our work with a conclusion that aims to not only identify the technological gaps within the current systems of Sonatrach's production division (general problematic) but also how our solution offers a robust and far-sighted answer for the bridging of such gaps. We also state how we faced the challenges, which measures were adopted to overcome them, and how to improve and enhance the solution even further in the future, additionally, we include how such challenges met in our 6 months of internship in a professional environment pushed us to grow intellectually and sharpened our skills in both software engineering/architecture and potentially data engineering.

Objectives:

- Develop a centralized web platform tailored to Sonatrach's project management needs.
- Implement modern Business Intelligence (BI) dashboards for enhanced data visualization and decision-making.
- Integrate machine learning functionalities to classify projects efficiently and predict project families.
- Improve data accessibility, streamline information flow, and ensure data security across 11 operational regions.
- Provide a user-friendly interface with Tableau technology to enhance data interaction and decision support.

Part I

State of the Art

"Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road"

- Stewart Brand

Chapter 1

Project Management in the Petroleum Industry

1.1 Introduction

Project management in the petroleum sector needs skillful management to explore, extract, and refine effectively. Sketched below is an introduction that briefly explains the practices adopted by organizations such as Sonatrach. We will now pass through what some of these challenges are unique and how good project management practices become very pivotal. More so, this part highlights current trends and best practices that show how organizations adapt their operations to dynamic market conditions and regulations through innovative approaches.

1.2 Definitions

- **Project Management**: Project management involves "the application of knowledge, skills, tools, and techniques to project activities to meet project requirements" [1] It encompasses the effective coordination of resources and stakeholders to achieve project objectives within defined constraints.
- **Petroleum Industry**: The petroleum industry refers to the sector involved in the exploration, extraction, refining, and distribution of petroleum products, including crude oil and natural gas [2].
- **Investment Projects**: Investment projects are endeavors undertaken by organizations to allocate resources towards activities aimed at generating returns or benefits over a specified period [3].
- **Risk Management**: Risk management entails the systematic identification, assessment, and mitigation of potential risks that may impact project objectives and outcomes [4].
- **Resource Allocation**: Resource allocation involves the strategic assignment of available resources, including finances, manpower, and equipment, to various tasks or activities within a project [5].
- **Compliance**: Compliance refers to the adherence to laws, regulations, standards, and ethical principles relevant to the petroleum industry to ensure legal and ethical

conduct throughout project execution [6].

1.3 Introduction to PM in the Petroleum Industry

Project management in the petroleum industry applies established principles, methodologies, and tools for planning, mobilization, execution, and control of projects in oil and gas exploration production, and refining. This subpart gives an overview of some elements of project management in the petroleum sector.

1.3.1 Overview of PM in the Petroleum Industry

1. **Fundamental Principles of Project Management:**
   - Projects in the petroleum industry are typically complex, capital-intensive endeavors that require careful planning, coordination, and execution to achieve desired outcomes [3].
   - The project management process typically includes initiating, planning, executing, monitoring and controlling, and closing phases [1].
2. **Key Concepts in Project Management:**
   - **Project Lifecycle:** Projects in the petroleum industry typically follow a lifecycle that includes phases such as exploration, development, production, and decommissioning [7].
   - **Project Stakeholders:** Stakeholders in petroleum projects may include government agencies, investors, local communities, environmental groups, and regulatory bodies, among others [8].
   - **Project Constraints:** Constraints in petroleum projects may include budgetary limitations, regulatory requirements, environmental considerations, technological limitations, and geopolitical risks [3].

Stakeholder	Role
Government	Regulatory oversight, licensing, and taxation
Investors	Provide funding and expect returns on investment
Local Communities	Affected by project activities and environmental impacts
Environmental Groups	Advocate for environmental protection measures
Regulatory Bodies	Enforce compliance with environmental and safety regulations

Table 1.1: *Key Stakeholders in Petroleum Projects*

3. **Examples of Typical Projects:**
   - **Exploration Drilling:** Exploration drilling projects involve the search for new oil and gas reserves through the use of drilling rigs and geological surveys [6].
   - **Field Development:** Field development projects focus on the construction of infrastructure such as wells, pipelines, and processing facilities to extract and transport oil and gas from discovered reserves [9].

- **Refinery Construction:** Refinery construction projects involve the design and construction of facilities to process crude oil into refined products such as gasoline, diesel, and jet fuel [10].

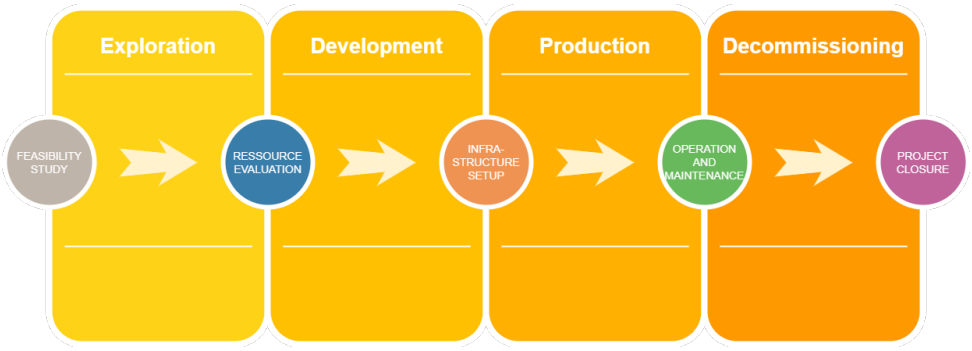

Figure 1.1: *Project Lifecycle in the Petroleum Industry*

1.3.2 Unique Challenges and Requirements in the Petroleum Industry

Project management in the petroleum industry incurs some unique challenges and requirements, this section will explore these unique aspects and implications for project management practices.

1. **Technical Challenges:**
   - **Complexity of Operations:** Petroleum projects often involve technically complex operations, such as offshore drilling in harsh environments or unconventional extraction techniques like hydraulic fracturing [11].
   - **Advanced Technology Adoption:** The industry continuously adopts cutting-edge technologies for exploration, drilling, and production, requiring project managers to stay abreast of technological advancements [12].
2. **Environmental Considerations:**
   - **Environmental Impact Assessment (EIA):** Oil and gas projects must undergo rigorous EIAs to mitigate environmental risks and ensure compliance with regulatory standards [13].
   - **Sustainability Initiatives:** Increasing emphasis on sustainability drives the need for environmentally responsible project management practices, including carbon footprint reduction and ecosystem preservation [14].
3. **Regulatory Compliance:**
   - **Stringent Regulations:** The petroleum industry is subject to extensive regulations governing safety, environmental protection, and resource extraction, necessitating meticulous compliance management [15].
   - **Geopolitical Considerations:** Projects often span multiple jurisdictions with varying regulatory frameworks, adding complexity to compliance efforts [16].
4. **Market-Related Challenges:**
   - **Price Volatility:** Fluctuations in oil and gas prices impact project economics, requiring agile project management strategies to adapt to market dynamics [17].

- **Geopolitical Risks:** Political instability, trade disputes, and sanctions pose significant risks to project continuity and profitability, necessitating risk mitigation measures [18].

How these challenges are managed is critical to assuring the success of projects and optimizing return on investment in the petroleum sector. Therefore, project managers who address technical complexities, environmental considerations, regulatory compliance, and market-related risks can increase project outcomes and return value.

1.3.3 Importance of Effective Project Management in Investment Projects within Competitive Companies

This subsection will deal with the critical importance of effective project management, especially for investment projects executed within companies like Sonatrach and the Production Division. Effective project management is one of the bases for the successful profitability of any production investment project [19].

Key Aspects	Description
Efficient Execution	Robust project management practices enable efficient execution of investment projects, ensuring that tasks are completed on time and within budget.
Goal Achievement	Project management facilitate the alignment of project activities with organizational goals, ensuring that project objectives are met effectively.
Optimal Resource Utilization	Effective resource management, a core aspect of project management, ensures that resources such as manpower, materials, and equipment are utilized optimally.
Risk Mitigation	Project management involves proactive identification, assessment, and mitigation of risks, thereby minimizing the impact of unforeseen events on project outcomes.

Table 1.2: *Key aspects driving project success and profitability*

Impact on Project Timelines, Budgets, and Resource Allocation

Effective project management directly affects the different aspects of implementing a project, such as time, budget, and resources. Project delays, cost overruns, and ineffective resource allocation in meeting a project's objectives can be managed by applying meticulous planning, scheduling, and control techniques [20].

Contribution to Risk Mitigation and Uncertainty Management

One of the essential functions of project management is identifying, assessing, and mitigating risks. The stakes are higher in an ambitious portfolio like the one of the Production Division due to the many capital investments and operational complexities. Project

managers can safeguard projects from potential threats and uncertainties through risk management strategies and contingency plans [21].

Figure 1.2: *Interplay of Project Management Factors*

1.4 Current Trends in Project Management

The future trends and best practices in project management drive the petroleum industry to evolve continuously in the quest for success with its projects. This section investigates the most recent advances in project management within the petroleum sector, addressing areas such as risk management, resource management, and compliance.

1.4.1 Advancements in Risk Management

Risk management plays a pivotal role in mitigating uncertainties and maximizing project success in the petroleum industry. Recent trends indicate a shift towards proactive risk management strategies that emphasize early identification and mitigation of potential risks [22]. Advanced risk assessment techniques, such as probabilistic risk analysis and scenario planning, are gaining traction to enhance project resilience against unforeseen events [23]. Organizations are leveraging sophisticated risk management software to streamline risk identification, assessment, and response planning processes, thereby improving project outcomes.

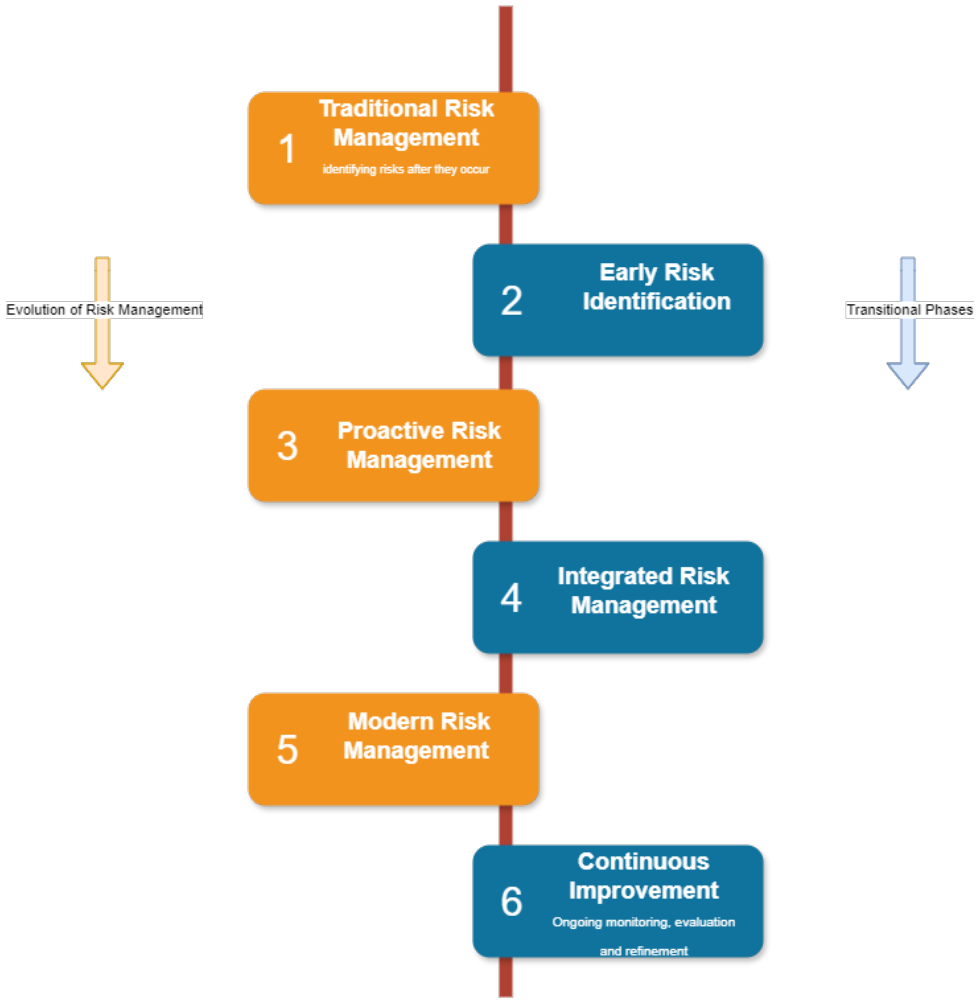

Figure 1.3: *Evolution of Risk Management in the Petroleum Industry*

1.4.2 Innovations in Resource Allocation

Efficient resource allocation is critical for optimizing project performance and achieving desired outcomes in the petroleum sector. Recent trends show a move towards dynamic resource allocation models that prioritize flexibility and adaptability [24]. Organizations are leveraging data-driven approaches and advanced algorithms to optimize resource utilization across multiple projects simultaneously. Integrated project management platforms equipped with resource management modules enable real-time visibility into resource availability, enabling project managers to make informed decisions and allocate resources effectively.

1.4.3 Compliance Management Strategies

The petroleum industry is highly regulated; organizational performance in compliance management is the core of their business operations. Development trends have been targeted toward integrated compliance management systems, which reduce processes for regulatory compliance and ensure conformance with industry standards [25]. Indispensable automated tools that track compliance will quickly ease processes in regulatory reporting, audit trails, and document management, thus minimizing the administrative burden on project teams. Organizations are investing in compliance awareness training and de-

velopment among project stakeholders to establish a culture of regulatory compliance.

Upcoming parts highlight the prevailing trends influencing project management in the petroleum industry. By adopting advanced techniques of risk management, resource management, and compliance strategies, organizations can ensure project effectiveness, mitigate threats, and achieve sustainable growth. we will discuss how leading organizations in the petroleum sector are adopting innovative project management techniques and tools.

1.4.4 Adaptation to Dynamic Market Conditions and Regulatory Requirements

The petroleum industry operates in a constantly evolving environment characterized by shifting market dynamics and stringent regulatory requirements. In this subpart, we explore how organizations in the petroleum sector are adapting to these challenges through innovative project management approaches.

Market-driven Project Management Strategies

Market-driven practices of project management are adopted by companies in the petroleum industry to respond to turbulent market conditions and utilize new opportunities that have been arising. Strategies include proactive market analysis, scenario planning, and agile project execution methodologies. When projects are handled strategically, an organization can closely monitor market trends, and customer demands to ensure that the project portfolio aligns with strategic objectives in light of rapidly changing market dynamics.

Agile Project Management Practices

Within the sector, wherein companies are increasingly trying to introduce more flexibility and responsiveness in executing projects, agile project management is becoming popular [18]. Agile principles use tools such as iterative development, continuous feedback, and adaptive planning so that project teams can change any or all of the project's scope, deliverables, and schedule at just about any time in reaction to these changing market conditions. Iterative methods foster collaboration, innovation, and satisfaction—all elements that facilitate the success of projects within dynamic environments.

Key Component	Description
Iterative Planning	Continuous refinement of project plans based on feedback and changing requirements
Continuous Feedback Loops	Regular feedback loops with stakeholders to gather insights, identify issues, and make adjustments
Adaptive Decision-making	Flexibility to adjust project scope, deliverables, and timelines based on evolving market conditions
Collaborative Work	Emphasis on teamwork, collaboration, and shared ownership of project goals and outcomes
Customer Value Delivery	Focus on delivering value to customers through incremental and iterative product development
Incremental Delivery	Gradual delivery of project deliverables in short iterations, allowing for early value realization

Table 1.3: *Agile Project Management Framework*

Compliance-driven Project Management Frameworks

Organizations in the petroleum industry are still very concerned with the issue of regulatory compliance. This is because the governing rules are stringent on the environmental, safety, and operational laws. Overall, this is ensured by putting up robust compliance management frameworks in their endeavors as they undertake project management. Such regulatory assessments, monitoring of the compliance state, and audit protocols must, therefore, form an integral part of compliance management frameworks to avoid compliance risks and to meet the industry standards.

1.5 Technological Landscape

1.5.1 Introduction

This chapter presents the evolution of the changing technological dimension of project management within the petroleum industry. We look in detail at these new advances in data analytics, remote monitoring, and automation technologies that have evolved to bring change in project management. in particular, we will consider the role that machine learning and business intelligence technologies play in optimizing project management processes. These tend to have transformational potential in the industry, from predicting costs and timelines to providing real-time insights into project performance.

This part, therefore, provides some case studies, showing examples of ML and BI integration within the PM system, which could offer a feel for improvement in project outcomes and decisions. Thus the goal of this chapter is to outline in a few words the technological landscape and what it implies for project management in this current context and that of the future of the petroleum industry.

1.5.2 Definitions

- **Technological Landscape:** The technological landscape refers to "the collective set of technologies, tools, and methodologies employed within a particular industry or domain" [26]. It encompasses the diverse range of hardware, software, and systems utilized to support organizational activities and processes.
- **Machine Learning (ML):** Machine learning is defined as "the science of getting computers to act without being explicitly programmed" [27]. It involves the development of algorithms and statistical models that enable computers to learn from data and improve their performance over time.
- **Business Intelligence (BI):** Business intelligence refers to "the technologies, applications, and practices for the collection, integration, analysis, and presentation of business information" [28]. It encompasses the tools and processes used to transform raw data into actionable insights for informed decision-making.
- **Integration:** Integration involves "the process of combining different components or systems into a single unified entity" [29].

1.5.3 Tools and Platforms for Project Management in the Petroleum Industry

In the petroleum industry, specialized tools are crucial for efficient project management, ensuring timely delivery and resource allocation.

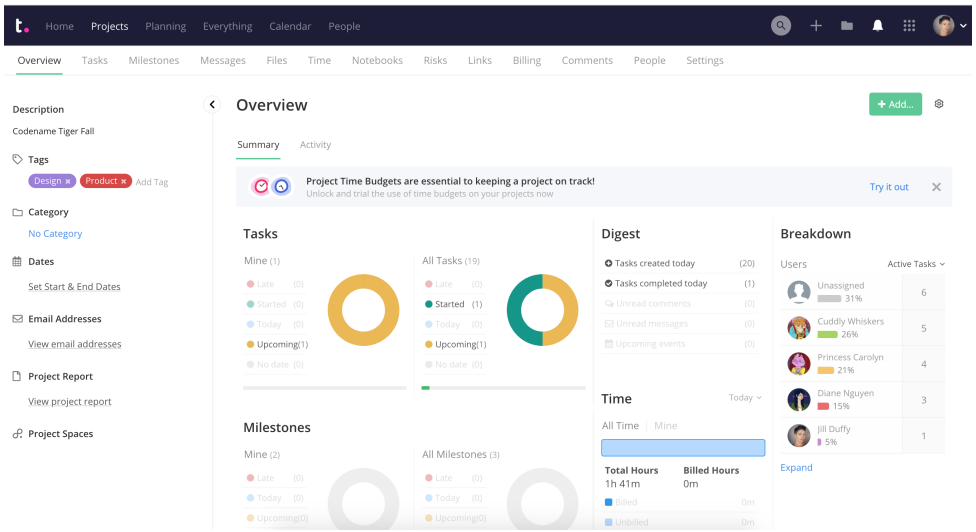

Figure 1.4: *Example Project Management Platform*[1]

The petroleum industry uses the Project Management Information System (PMIS)[2]. This system helps managers monitor progress, allocate resources, and communicate with stakeholders efficiently. Some commonly Used Tools for Project Management within the PMIS include:

[1] https://me.pcmag.com/en/old-project-management/14976/the-best-project-management-software-for-2023.
[2] PMIS is a centralized system that integrates scheduling, budgeting, and document management for efficient project execution and monitoring.

Tool	Description
Gantt Charts	Visual representation of project tasks and timelines
Earned Value Management (EVM)	Methodology for measuring project performance against planned objectives.
Risk Management Software	Tools for identifying, assessing, and mitigating project risks.
Document Management Systems	Platforms for organizing and sharing project-related documents.

Table 1.4: *Commonly Used Tools for Project Management*

1.5.4 Advancements in Data Analytics

This section addresses current developments in data analytic technologies, which critically support enhanced practices for managing projects in the petroleum industry. We shall see how such technologies are applied to make decisions, assess risk, and optimize resources in wide-based projects. We will also give some examples of data analytics applications relevant to project management in the petroleum sector.

For a project manager in petroleum, data analysis tools have become part and parcel of applying insight and prognosticative capabilities. Such tools would allow the project teams to analyze vast troves of data for any pattern, trend, or anomaly so that an intelligent and informed decision can be made.

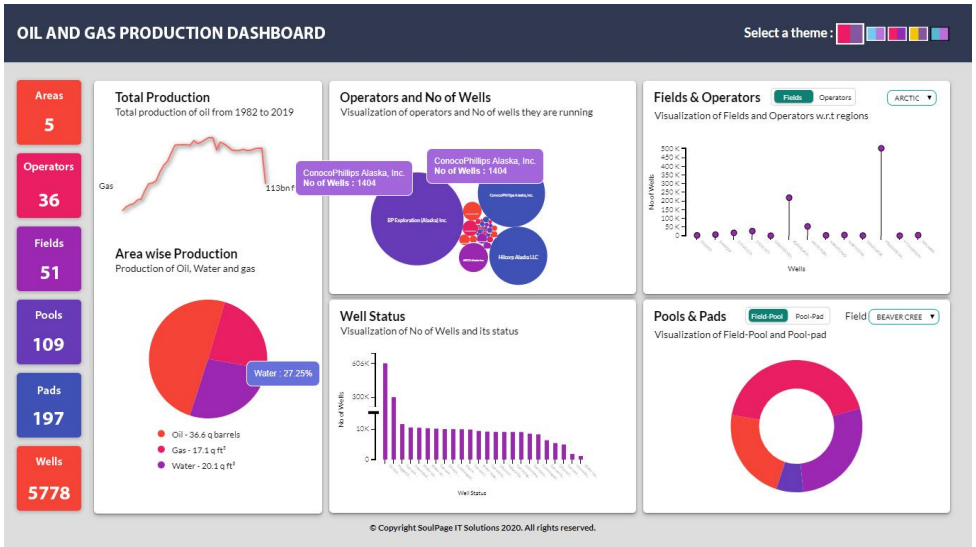

Figure 1.5: *A data analytics dashboard utilized in the petroleum industry for real-time monitoring*[3]

These mentioned smart decisions and effective risk mitigation abilities are done using historical data but also take into consideration the risk assessment matrix found below

[3] https://medium.com/@antonioneto_17307/what-to-consider-to-build-a-good-dashboard-d66c031bc7f1

Risk category	Likelihood	Impact	Mitigation Strategy
Supply Chain	High	Medium	Diversify supplier base to reduce dependency risks
Regulatory	Medium	High	Proactive compliance monitoring and reporting
Market Dynamics	High	High	Continuous market trend analysis

Table 1.5: *Risk Assessment Matrix*

These data analytics applications are deployed across domains related to project management in the petroleum sector: predictive models of equipment reliability that support optimized maintenance strategy, reservoir management with predictive analytics, and sentiment analysis related to stakeholder feedback and project satisfaction levels. Organizations adopting these applications can streamline their operations, and indeed enhance the performance of projects

1.5.5 Role of ML and Business Intelligence

In modern project management, particularly in industries like petroleum, integrating advanced technologies such as machine learning (ML) and business intelligence (BI) is crucial. ML enables computers to learn from data, enhancing decision-making by identifying patterns and making predictions [27]. BI involves methodologies for analyzing business data to aid decision-making [28]. In petroleum, ML and BI synergize significantly, offering project managers insights from vast datasets and real-time visibility into project performance. Some usual applications include predictive maintenance, anomaly detection, and optimizing production processes [30, 31].

- **Predictive Analytics:** ML algorithms analyze project data to forecast costs, timelines, and risks.
- **Real-time Monitoring:** ML enables continuous monitoring of project performance metrics, providing stakeholders with actionable insights.
- **Resource Optimization:** By leveraging historical data, ML facilitates optimal resource allocation throughout project lifecycles.
- **Risk Management:** ML algorithms prioritize risks based on historical data, enabling proactive risk mitigation strategies.

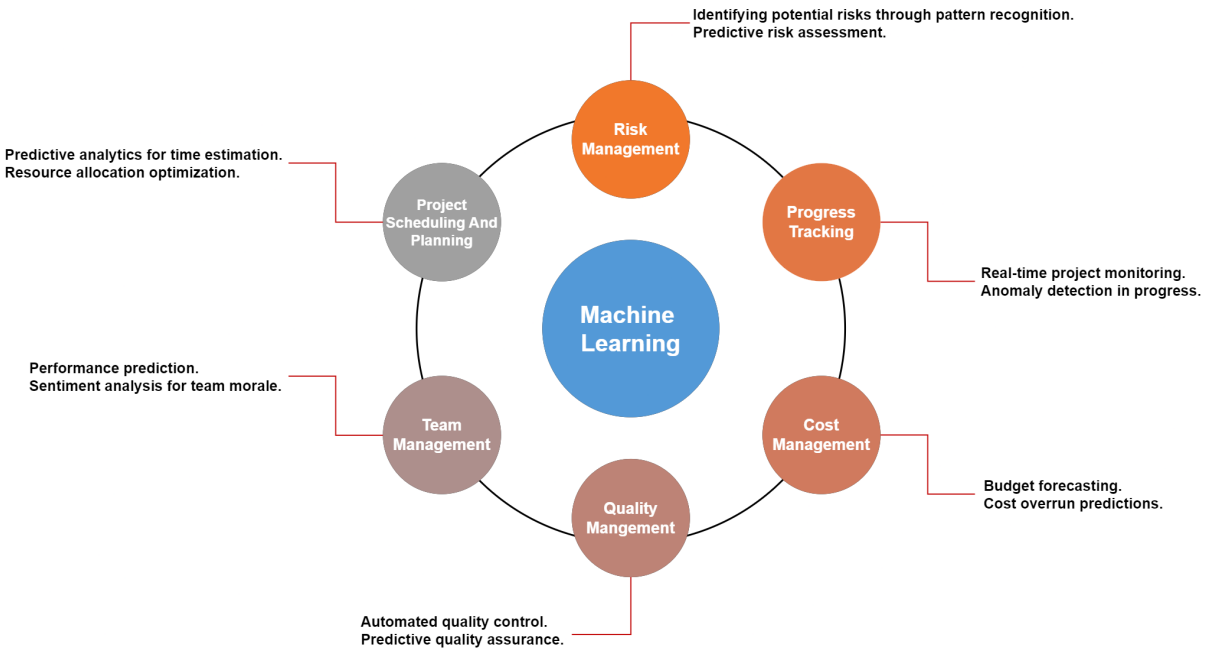

Figure 1.6: *Machine Learning Applications in Project Management*

Significance of BI Dashboards and Data Visualization

BI dashboards form the core center of project-related data visualization, together with an intuitive interface to help the stakeholders monitor different projects by visualizing progress, performance metrics, and key performance indicators. This provides real-time insight from different project execution perspectives to make proactively influenced decisions and resource allocation [32, 33].

In addition, these BI tools give stakeholders access to real-time analytics and visual information for a quick evaluation of project performance against set benchmarks and compliance standards. The capabilities of BI allow project managers to detect bottlenecks, foresee risks, and apply mitigation on time to assure success and compliance [34].

1.5.6 Integration of ML and BI case Studies in Project Management

1. **Real-time Performance Monitoring in Refinery Expansion Projects**:

   For example, integration of BI dashboards depicting real-time performance monitoring in expansion projects for the refinery. BI tools should collect and analyze from various sources, such as project schedules, allocations of resources, and financial metrics, if stakeholders are to be given an all-round view of the progress of a project and its KPI. Real-time visibility enables project managers to make proper decisions and, at the same time, change tactics immediately to ensure project success.

KPI Category	Description	Example Metrics
Schedule Adherence	Measure of project timeline adherence	Percentage of tasks completed on schedule
Resource Utilization	Evaluation of resource allocation and utilization	Labor hours spent vs. budgeted labor hours
Cost Performance	Assessment of project cost performance	Cost variance (CV), cost performance index (CPI)
Quality Assurance	Monitoring of project deliverable quality	Defect density, customer satisfaction scores

Table 1.6: *Key Performance Indicators (KPIs) Tracked via BI Dashboards*

2. **Case Study: Risk Prediction and Mitigation in Exploration Projects**:

   A good example where machine learning and business intelligence come together for risk prediction and proactive mitigation is the example of a multinational oil and gas company based out of Irving, Texas, United States: **ExxonMobil**. By applying advanced machine learning algorithms within its exploration projects, **ExxonMobil** analyzes immense datasets like geological surveys, seismic data, or historical project performance metrics. In its operations, **ExxonMobil** makes use of predictive modeling techniques like decision trees and neural networks for the identification of possible risks associated with, for instance, exploration activities that might involve increased reservoir uncertainty, geological complexity, or regulatory challenges.

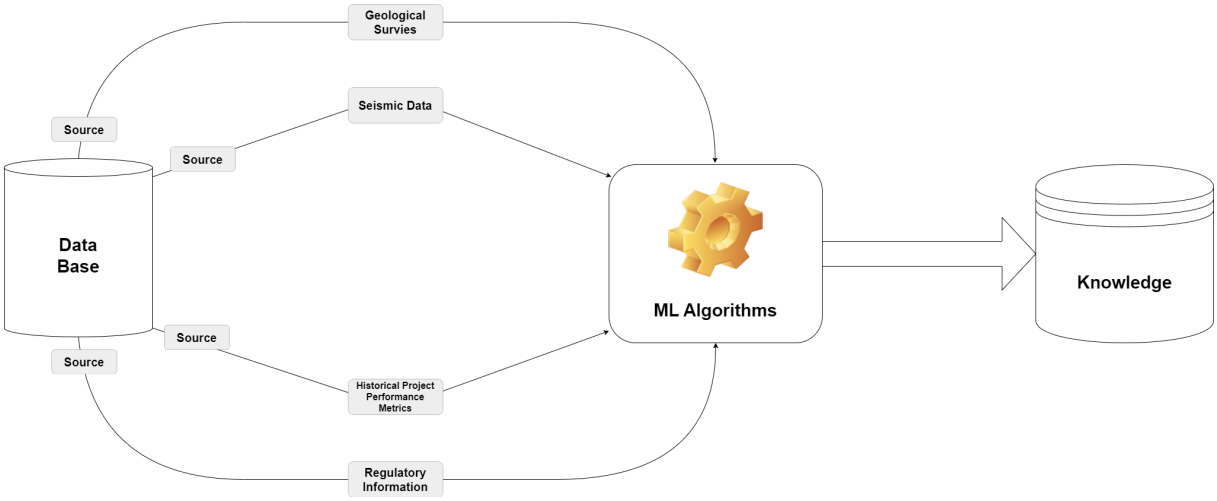

Figure 1.7: *Illustrating the components of ExxonMobil's risk prediction model*

1.6 Conclusion

Project management is the essence and soul of success in the petroleum industry, which is replete with challenges. Effective project management, particularly within the Production Division, is crucial for managing time, cost, and resources, ultimately leading to better organizational performance and achievement of target objectives. We have discussed unique aspects of project management tailored for oil and gas, including technical issues and regulatory compliance, and highlighted new trends and best practices such as improved risk management and resource utilization.

The technological landscape of project management in the petroleum sector is rapidly evolving, with machine learning (ML) and business intelligence (BI) technologies playing pivotal roles. ML algorithms enable predictive analytics, real-time monitoring, and resource optimization, while BI tools offer stakeholders intuitive dashboards for informed decision-making. Case studies demonstrate the positive impact of integrating ML and BI on project outcomes and decision processes.

However, the adoption of these technologies faces challenges such as organizational resistance and technical limitations. Addressing these challenges through tailor-made strategies and collaboration is essential. Digital transformation, driven by ML and BI tools, provides a significant innovative and competitive advantage in the petroleum sector. Implementing these technologies, alongside best practices, unlocks the full potential for successful project management, fostering creativity and growth in this dynamic industry.

Part II

Study of The Existing Situation And Analysis of Needs

"Every organization is perfectly designed to get the results it gets"

- Arthur W. Jones

Chapter 2

Study of The Existing Situation And Analysis of Needs

2.1 Study of The Existing Situation

2.1.1 Introduction

Sonatrach is integral to Algeria's energy sector, managing extensive investment projects in hydrocarbon exploration, production, transportation, and processing. An effective management of these projects is crucial to achieve strategic objectives, operational efficiency, and maintaining Sonatrach's leadership in the energy industry. This chapter focuses on the Production Division highlighting its mission, organization, and and its investment planning process.

By understanding the current operational landscape, this chapter lays the foundation for developing a solution to address existing challenges. The Following chapter will focus on requirements analysis, as it highlights challenges faced by the studies and planning department in visualizing and consolidating investment project data and underscores the need for an integrated data management and visualization solution to enhance Sonatrach's project management capabilities and support its growth and success.

2.1.2 Focus on the Production Division

Presentation of the Production Division

The Production Division is crucial for operating and advancing oil fields, gas and industrial facilities, and communication bases within SONATRACH. Under the jurisdiction of the Vice President of Exploration and Production (EP), this division has over 20,000 employees organized into:

1. **Management Departments:** Eleven management departments oversee strategic aspects of production operations, ensuring effective coordination and alignment with organizational goals.
2. **Regional Managements:** Eleven regional Managements are established to oversee production activities in designated geographical areas, facilitating localized management and operational oversight.

Mission of the Production Division

This structure allows efficient management and optimization of production operations across various facilities. The division supports SONATRACH's goals in the hydrocarbon sector through:

1. **Exploration and Evaluation of reserves:** The division conducts thorough exploration activities to identify and assess potential hydrocarbon reserves. Through geological and geophysical studies.
2. **Development of Oil and Gas Fields:** Once viable reserves have been identified, the Production Division leads the development of oil and gas fields.
3. **Operation and Production:** The division is responsible for the daily operation of oil and gas fields, overseeing activities such as drilling, well maintenance and production operations. It ensures the efficient recovery and processing of hydrocarbons to meet production targets, as well as an Effective management of production operations.
4. **Production Optimization:** Through innovative technologies and best practices, it seeks to maximize production rates, minimize downtime and optimize resource utilization, as well as continuously improve and optimize production processes.

By fulfilling these objectives, the Production Division contributes to SONATRACH's overall mission of sustainable energy supply, economic development, and environmental stewardship.

2.1.3 Production Division's Organizational Structure

Following its restructuring in 1984, SONATRACH retained the entirety of production, which has its headquarters at "Chemin de Réservoir - Hydra - Algiers", and deals exclusively with the production of liquid and gaseous hydrocarbons, in accordance with its designation.

At the headquarters, the following managements are present:

1. Legal Management
2. Health, Safety and Environment Management
3. Information Technology Management
4. Finance and Accounting Management
5. Personnel Management Department
6. Operations Management
7. Realization Management
8. Procurement and Transportation Management
9. General Resources Management

This organization chart illustrates the organization of Sonatrach's Production Division:

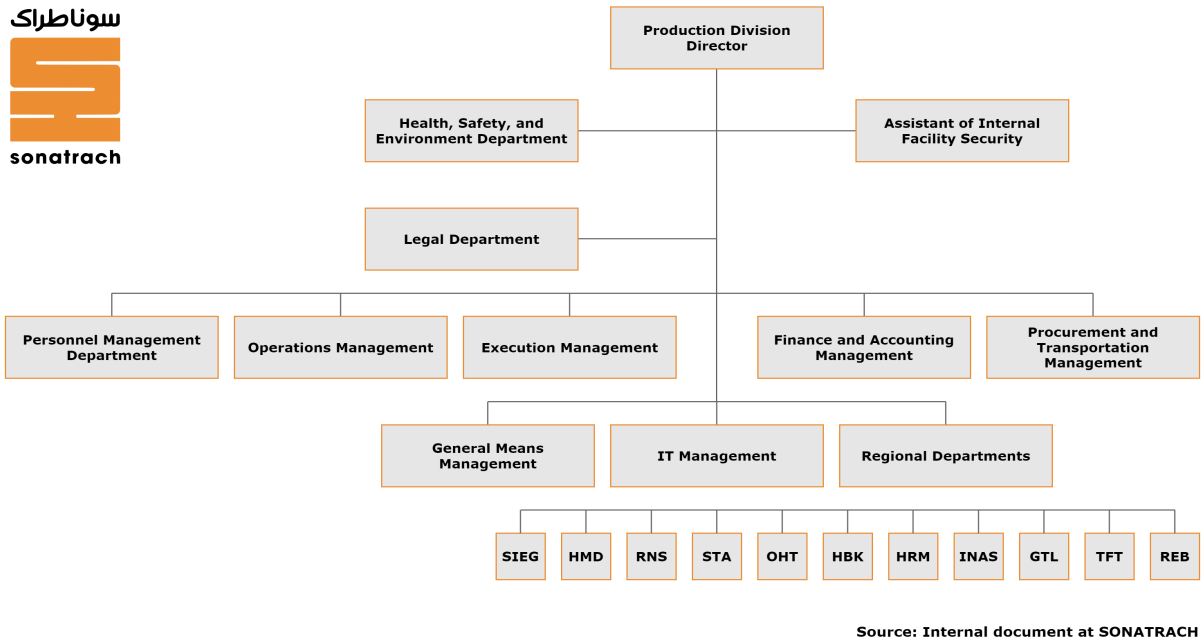

Figure 2.1: *SONATRACH's Production Division Organization Chart*

Each management within the Production Division plays a crucial role in supporting the division's mission and objectives, Sonatrach's regional management is instrumental in overseeing the division's activities in southern Algeria, particularly in the production sector. Recently, the regional management of "ADRAR" has been added in order to strengthen operations and expand Sonatrach's presence in the region.

The eleven regions are:

1. Hassi R'mel (HRM)
2. Haoud Berkaoui (HBK)
3. Hassi Messaoud (HMD)
4. Rhourde El Baguel (REB)
5. GassiTouil (GTL)
6. Rhourde Nouss (RNS)
7. Tin Fouyé Tabankort (TFT)
8. Ohanet (OHT)
9. Stah (STH)
10. In Amenas (INA)
11. Adrar (ADR)

These regions are crucial to Sonatrach's operations, contributing to efficiency, sustainability, and growth in the hydrocarbon sector. The Regional Managements ensure optimal performance and resource utilization across southern Algeria.

2.1.4 Operations Management Overview

The Operations Management of Sonatrach Production Division plays a key role in strategic planning and operational execution. It formulates development plans, oversees implementation, consolidates results, coordinates investment initiatives, and oversees regional operations, Operations Management is organised into four integral departments:

1. **Maintenance Department**

2. **Exploitation Department**
3. **Operations Assistance Department**
4. **Studies And Planning Department**

Since our primary focus is the Studies and Planning Department, we will be focusing our efforts primarily to this department.

Studies and Planning Department

The Department of Studies and Planning within Sonatrach stands as a foundational unit, dedicated to overseeing strategic planning with utmost precision. This department embodies a multifaceted approach, meticulously tailored to drive the realization of Sonatrach's strategic objectives. The Studies and Planning Department is responsible for the following tasks and more [35]:

1. Arbitration and establishment of the medium-term plan of the company (PMTE) in the presence of all regions and structures of the Production Division concerned by the projects.
2. Development, control, consolidation, and monitoring of the Monthly, annual, and multi-year plans of the investment component of the Production Division.
3. Preparation and consolidation of the Annual Execution Report of the Production Division, covering production, investment, financial exploitation, human resources, procurement and transportation management, and internal security.
4. Establishment of a monthly dashboard of management indicators of the Production Division.
5. Conducting technical and economic studies of investment projects in the Production Division.
6. Financial evaluation of projects, including cost analysis, financing assessment, turnover projections, consumption forecasts, and expenditure evaluations.
7. Conduct optimisation studies related to investment projects in the Production Division, aimed at enhancing operational efficiency and cost-effectiveness.

The Studies and Planning Department at Sonatrach is essential for strategic planning and investment management. It gathers data, creates financing plans, performs technical and economic studies, and oversees project execution to ensure efficient resource use and successful strategic initiatives [36].

The departments within Operations Management collaborate to ensure the smooth operation and continuous improvement of Sonatrach's activities. This organization chart illustrates the structure of the Operation Management.

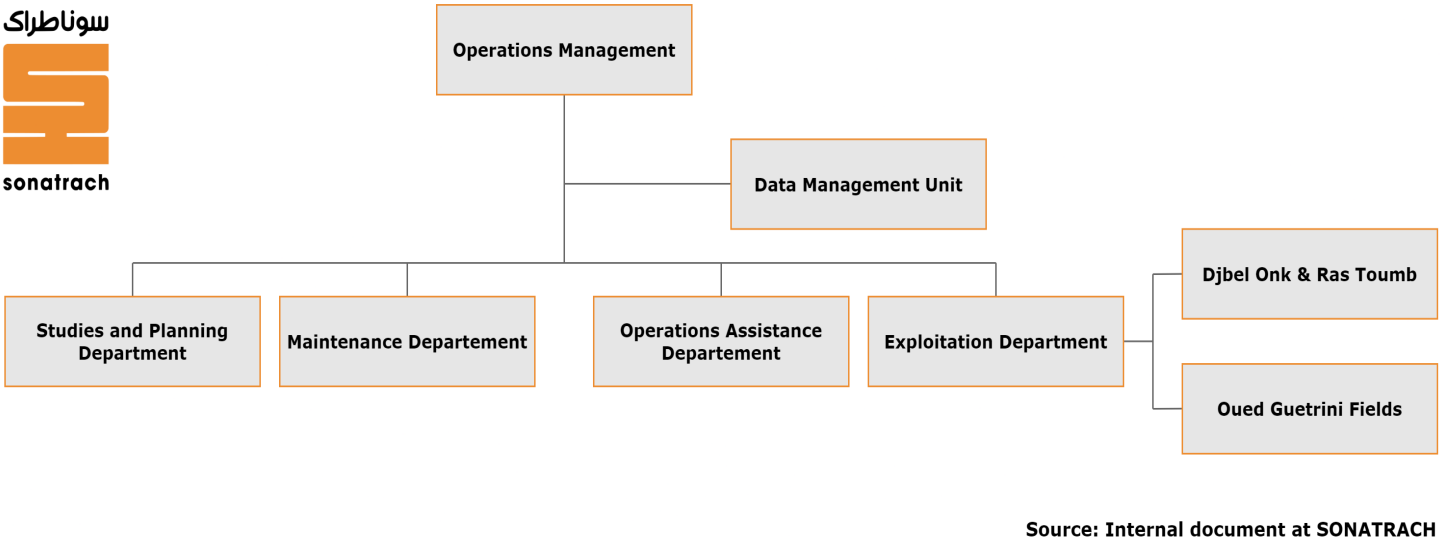

Figure 2.2: *SONATRACH's Operations Management Organization Chart*

2.1.5 Investments of the Production Division

Investment Projects Planning Process Details

1. **In June of year** $N-1$:
   The Studies and Planning Department initiates the budget campaign by sending templates related to the annual plan of year N and PMT N to $N+4$ to the relevant structures [37].
2. **During the period from June of year** $N-1$ **to the end of June of year** $N-1$:
   Structures under the Production Division develop their respective PMTs, fill in the templates already sent, and update project sheets [37].
3. **July of year** $N-1$:
   Arbitration meetings take place for at least fifteen days, one structure per day [37].
4. **During September of year** $N-1$:
   The planning department consolidates the PMTs of all structures. Once consolidation is complete, the department conducts an analysis of variances by project family and managing structure [37].
5. **End of September of year** $N-1$:
   Submission of the annual plan of year PMT of N year to $N+4$ year to the planning committee in the company in order to make informed decisions concerning investments projects [37].

The process involves all operational and functional structures under the Production Division [37]:

1. **Eleven (11) Regional managements** (HRM, HBK, HMD, REB, GTL, RNS, TFT, OHT, STH, INA, ADR)
2. **Headquarters structures** (DAT, DMG, INF, and HSE).
3. **Off-region sector** (Ras Toumb, Djebel Onk, and Oued Gueterini).

Investment Projects Families

The Studies and Planning Department in Sonatrach's Production Division groups projects into different families to improve operational efficiency, security, and sustainable development.

These project families address various aspects of the Company's activities and infrastructure [37]:

1. **Studies**: This includes all seismic studies, interpretation studies, as well as engineering studies for installation, processing, and hydrocarbon extraction projects. Each project, or the majority of projects, must start with a study, especially the large and expensive ones.
2. **Well Activity**: These are projects for drilling wells and well operations, such as 'work-over', 'snubbing', 'acidification', and 'fracturing'. including:
   - **Work-over**: There are two main types of well problems that require work-over intervention (using a drilling rig):
     (a) **Formation Problems**: Wells producing oil with excess gas, water, or sediments.
     (b) **Mechanical Problems**: Tubing anomalies (communication between tubing, either due to erosion or tube design defects, etc.).
   - **Snubbing**: Snubbing is a type of intervention on wells with anomalies.
   - **Acidification**: Acidification is a chemical washing method that involves the injection of chemicals or acids, thus cleaning the pores of the reservoir rock and facilitating fluid (crude) flow.
   - **Fracturing**: Fracturing involves enlarging the natural fractures of the reservoir rock or even creating fractures where they do not exist, provided the nature of the rock allows it.
3. **Specific Installations**: These are projects directly related to production, such as oil or gas processing projects, well connections, crude storage tanks, shipping pipes, well water injections, torch network renewals, etc. These projects are essential for the efficient and safe production of oil and gas within Sonatrach's operational areas.
4. **Maintenance of Installations**: The Installation Maintenance project family is dedicated to the upkeep of operational installations. This family encompasses a range of activities aimed at maintaining, repairing, and refurbishing infrastructure assets to ensure their reliability, safety, and longevity.By implementing these maintenance strategies, Sonatrach aims to uphold the operational integrity of its installations, ensuring continuous and safe production operations.
5. **General Installations**: These are projects indirectly related to production, such as road and track projects, environmental protection projects, production center laboratories, lighting network renewals, etc. These projects focus on the design, construction, and operation of general-purpose installations and facilities necessary for the smooth functioning of day-to-day operations
6. **Social Infrastructures**: These are projects aimed at enhancing the well-being and welfare of employees and local communities. They play a vital role in enhancing the

quality of life for employees and promoting a sense of community within Sonatrach's operational areas.

7. **Equipment**: This includes equipment directly related to production (technical equipment), which can be further classified into Gas Equipment and Petroleum Equipment.

In summary, organizing investments into project families enables Sonatrach to optimize resource allocation and project execution, and improve operational efficiency within the Production Division.

2.1.6 Conclusion

In conclusion, Sonatrach's Production Division is essential for the company's hydrocarbon exploration, production, and processing. Its comprehensive structure and regional managements in southern Algeria enable effective operations. Additionally, The Operations Management, with its departments, ensures efficient maintenance, production monitoring, operational support, and strategic planning.

Finally, the Studies and Planning Department, which is particularly important for investment planning, project evaluation, optimization studies and performance monitoring, contributes to operational excellence and strategic decisions. This organizational framework enables Sonatrach to optimize the use of its resources and promote sustainable growth in the energy sector.

2.2 Requirements Analysis

2.2.1 Introduction

The Studies And Planning Department within the Operation Management plays a crucial role in overseeing and managing investment projects across various regions and structures. These investment projects are vital for the growth, sustainability, and competitive edge of the organization. However, the department faces significant challenges in effectively visualizing and consolidating data related to these projects, hindering informed decision-making and strategic planning.

2.2.2 Importance of Investment Project Data Visibility

Investment projects are the lifeblood of the enterprise, driving innovation, expanding operations, and ensuring long-term profitability. These projects span multiple regions and structures, each with its unique set of requirements, timelines, and resource allocations. Effective monitoring and coordination of these projects are essential for efficient execution and optimal resource utilization.

Maintaining visibility into the progress, performance, and financial aspects of investment projects is crucial for the department. It enables timely decision-making, proactive risk management, and the ability to realign strategies as needed.

2.2.3 Limitations in Current Data Display Methods

Despite the importance of investment project data visibility, the current methods of displaying and consolidating this information are inadequate and fragmented. Reliance on

traditional tools like spreadsheets and static reports presents several limitations:

1. **Lack of Real-Time Updates:** Data in spreadsheets and reports is often outdated, failing to reflect the latest developments and changes in project status, timelines, or financial figures.
2. **Inconsistent Formatting:** With each region potentially using different templates and formats, consolidating data becomes a cumbersome and error-prone process.
3. **Difficulty in Comparing Data Across Regions:** The disparate formats and lack of standardization make it challenging to compare project performance and metrics across different regions, hindering effective benchmarking and resource allocation decisions.
4. **Limited Visualization Capabilities:** Spreadsheets and static reports offer limited visualization options, making it difficult to identify patterns, trends, and outliers within the data.

2.2.4 Challenges in Data Collection and Standardization

In addition to the limitations in data display, the Studies And Planning Department also encounters challenges in the process of collecting and standardizing investment project data from different sources. The current approach relies on a template distributed by the department, which serves as a guideline for filling in project-related information, However, this process is not without its shortcomings:

1. **Lack of Adherance to Templates:** Despite the distribution of a standardized template, some may deviate from the prescribed format or omit certain data fields, resulting in inconsistencies and incomplete information.
2. **Manual Data Entry:** The process of manually entering project data into the template is time-consuming and prone to human errors, further compounding data quality issues.
3. **Limited Validation Checks:** The templates often lack built-in validation checks, making it difficult to identify and rectify errors or inconsistencies in the data during the collection phase.
4. **Delayed Data Submission:** Due to the manual nature of the process, some regions may experience delays in submitting their project data, resulting in an incomplete or outdated dataset for analysis and decision-making.

These challenges in data collection and standardization exacerbate the existing limitations in data display, compounding the obstacles faced by the Studies And Planning Department in maintaining accurate and up-to-date visibility into investment project data.

2.2.5 Impact of Ineffective Data Display and Collection

The combined limitations in data display methods and data collection processes have far-reaching implications for the Studies And Planning Department and the enterprise as a whole:

1. **Hindered Decision-Making:** Without a clear, consolidated, and accurate view of investment project data, the Studies And Planning Department and regional managers struggle to make informed decisions regarding resource allocation, risk mitigation, and strategic adjustments.

2. **Potential for Delays and inefficiencies:** Inadequate data visibility and quality issues can lead to delays in project execution, as crucial information may not be readily available, easily accessible, or reliable. This can result in inefficiencies, missed deadlines, and increased costs.
3. **Sub-optimal Resource Allocation:** Lack of clear and accurate data comparisons across regions can lead to suboptimal resource allocation decisions, where resources may be misaligned with project priorities or regional needs.
4. **Reduced Competitiveness:** By failing to effectively leverage investment project data, the enterprise may miss opportunities for growth, innovation, and competitive advantage within the market.
5. **Increased Operational Risks:** inaccurate or incomplete project data can heighten operational risks, such as safety concerns, regulatory non-compliance, or financial losses due to poor decision-making.

2.2.6 Need for an Integrated Data Management and Visualization Solution

To address these multifaceted challenges and unlock the full potential of investment project data, the Studies And Planning Department requires an integrated data management and visualization solution. Such a solution should encompass the following key features:

1. **Centralized Data Repository**:A centralized data repository that consolidates investment project data, ensuring consistent formatting, data quality checks, and real-time updates.
2. **Automated Data Collection**:Automated data collection processes, such as web-based forms or integration with existing systems, to streamline data entry and minimize manual errors.
3. **Interactive Visualizations**: Interactive visualizations, such as dashboards, charts, and maps, that provide a clear and intuitive representation of project data, enabling easy analysis and comparison across all operational and functional structures.
4. **Advanced Analytics**:Integration of advanced analytics capabilities, such as predictive modeling and scenario analysis, to support data-driven decision-making and proactive risk management.
5. **User Access Controls**:Robust user access controls and permissions to ensure data security and privacy, while enabling collaboration and knowledge sharing among authorized stakeholders.

By implementing a comprehensive data management and visualization solution, the Studies And Planning Department can address the challenges related to data collection, standardization, and display. This, in turn, will enable accurate and timely decision-making, optimized resource allocation, and strategic alignment across all structures, contributing to the enterprise's overall growth, competitiveness, and long-term success.

2.2.7 Integrating ML and BI into Sonatrach's Project Management

Integrating machine learning (ML) and business intelligence (BI) technologies into Sonatrach's project management infrastructure presents several challenges, primarily due to the reliance on Microsoft Excel and the confidential nature of its databases.

- **Dependence on Microsoft Excel:**
  - **Challenge:** Heavy reliance on Excel leads to manual data entry, version control issues, and limited analytical capabilities. Transitioning to ML and BI tools may face resistance due to entrenched workflows.
  - **Consideration:** Address this by implementing change management and user training, highlighting the benefits of automated data analysis and real-time reporting.
- **Confidentiality of Databases:**
  - **Challenge:** Databases contain sensitive information, restricting access to external ML and BI providers.
  - **Consideration:** Implement data anonymization to share sanitized datasets while preserving privacy, and keep a close eye on maintaining security.
- **Organizational Resistance and Technical Limitations:**
  - **Challenge:** Resistance to change and technical limitations, such as legacy IT infrastructure, may hinder adoption.
  - **Consideration:** Engage key stakeholders, managers, and final product users to foster support and invest in training to enhance technical skills, developing legacy functionalities resembling the old usual system while slowly integrating into newer technologies is also an option.

2.2.8 Best Practices for Leveraging ML and BI within production division

- **Structured Data Representation:**
  - **Challenge:** Data spans multiple regions and project families.
  - **Best Practice:** Use a comprehensive data model to categorize projects, enabling efficient analysis.
- **Region-Specific Summary Tables:**
  - **Challenge:** Need for detailed project cost summaries by region.
  - **Best Practice:** Develop customized summary tables for each region, aggregating project costs.
- **Predictive Project Family Classification:**
  - **Challenge:** A specific functionality to predict the family of new projects is requested.
  - **Best Practice:** Implement ML algorithms for predictive classification based on historical data.

2.2.9 Conclusion

Best practices for leveraging ML and BI within Sonatrach's Production Division were discussed, emphasizing structured data representation, predictive project family classification, and integration with financial planning systems. Addressing these issues is critical to unlocking the full potential of the enterprise's investment initiatives. By implementing an integrated data management and visualization solution, the Studies And Planning Department can overcome current obstacles and adopt a more data-driven, collaborative approach to project management.

The proposed solution includes a centralized data repository, automated data collection, interactive visualizations, advanced analytics, and robust access controls, offering a comprehensive framework to tackle multifaceted challenges, embracing this solution will empower the Studies And Planning Department with accurate, real-time data, enabling informed decisions, optimized resource allocation, and proactive risk management. It will also foster collaboration and knowledge sharing among stakeholders, enhancing the enterprise's agility and responsiveness.

Sonatrach recognizes the pivotal role of investment projects in driving growth, operational efficiency, and sustainability. The Studies And Planning Department plays a crucial role in managing these initiatives. However, current methods pose significant challenges, hindering informed decision-making and strategic planning. By implementing the proposed integrated solution, Sonatrach can overcome these obstacles and unlock the potential of its investment data. This will facilitate informed decision-making, enhanced collaboration, and streamlined project execution, ensuring operational excellence and sustained competitiveness in the hydrocarbon industry.

The investment in this solution represents a strategic commitment to innovation and data-driven decision-making, strengthening Sonatrach's position as a leading energy provider while contributing to the socio-economic development of Algeria and the global energy landscape.

Part III

Proposed Solution

"Ideas are worth nothing unless executed. They are just a multiplier. Execution is worth millions"

- Steve Jobs

Chapter 3

Design

3.1 Purpose and Overall Behavior of the Application

The main objective of our web application is to automate the collection and summarization of investment project data stored in Excel files, providing a role-based experience for different users. This automation saves time, reduces manual effort, and minimizes errors.

The platform focuses on data visualization, project management, and integration with external systems like Tableau Cloud and the machine learning model. Role-Based Access Control (RBAC) ensures users access functionalities relevant to their roles, such as 'ADMIN', 'STAKEHOLDER', 'AGENT PLF OPR', 'REGIONAL MANAGER', and 'AGENT REGION X'. The 'ADMIN' role manages all platform aspects, including user accounts. The 'AGENT PLF OPR' role provides access to Tableau dashboards, project management features, and a prediction model form for classifying projects.

The application enhances efficiency by automating repetitive tasks, ensures consistency, increases accessibility, and supports scalability for large datasets and multiple projects. 'STAKEHOLDER' users get a dashboard for informed decision-making, while 'AGENT REGION X' users manage regional project data. Each role ensures users have the tools needed for their duties.

Overall, our application improves investment project data analysis workflow, ensuring reliability, efficiency, modularity, maintainability, and scalability, while enabling effective project management and integration with external systems.

3.2 System Requirements

This section outlines the functional and non-functional requirements of the application. Functional requirements specify the features and functionalities that the application must provide in order to meet the users' needs, while non-functional requirements define the quality attributes of the system.

3.2.1 Functional Requirements

The following functional requirements must be met by the application:

1. **Dashboard Creation:** The application must have a dashboard with different visualizations developed using the main investment projects data.
2. **Summary Table Generation:** The application must generate summary tables based on the extracted data, presenting key metrics and insights relevant to investment project analysis.
3. **File Upload:** The application must provide a user-friendly interface for the upload of Excel files. It is essential that users receive feedback on the upload process, including notifications for successful uploads or error messages for invalid files. Furthermore, the application must be capable of supporting large file uploads.
4. **Automated Data Collection:** The application should be capable of automatically extracting data from uploaded Excel files, including project details, financial information, and other relevant data points. Furthermore, data extraction should be able to support various file structures commonly used in project management.
5. **User Authentication:** It is imperative that users are able to securely log in to the application using their credentials. In order to protect user accounts from unauthorised access, it is necessary to implement strong authentication mechanisms, including password hashing. Furthermore, it is crucial that admin users are able to manage user accounts, including password resets.

3.2.2 Non-Functional Requirements

The application must meet the following non-functional requirements:

1. **Performance:** The application must be capable of processing data and generating summary tables within a reasonable time frame, even when working with large datasets.
2. **Scalability:** The application must be capable of accommodating an increasing number of users and a growing volume of data without any deterioration in performance.
3. **Usability:** The application must possess an intuitive and user-friendly interface, with clear navigation and responsive design. User interactions must be consistent across different sections of the application, following established design patterns and conventions.
4. **Maintainability:** The application's codebase should be structured in a logical manner and divided into modular sections, with clear adherence to coding standards and best practices. Comprehensive documentation should be provided, including code comments and API documentation.

3.3 Use Cases Overview And Descriptions

This section presents an overview of the primary use cases of the application through a global use case diagram. Each use case is described in detail, including the actors involved and the main flow of events.

3.3.1 Actors

1. **Primary Actors:**
   - **Admin:** The admin plays a crucial role in overseeing the platform's operations, including user management and data management tasks. As the primary administrator of the application, the admin is responsible for creating and managing user accounts, and overseeing the overall functionality of the platform.
   - **Agent PLF:** The agent PLF is a key user of the platform, primarily responsible for utilising its features to access and analyse investment project data. This actor interacts with the platform to upload Excel files containing project data, retrieve summary tables, access dashboards for data visualisation, and utilise the prediction model for classification.
2. **Secondary Actors:**
   - **Prediction Service:** The Prediction Service is a secondary actor responsible for deploying the developed prediction model. It plays a critical role in the platform by enabling the use of the predictive model for project classification. It also uses FastAPI to create endpoints, tightly integrating predictive analysis tasks with the application.

3.3.2 Global Use Case Diagram

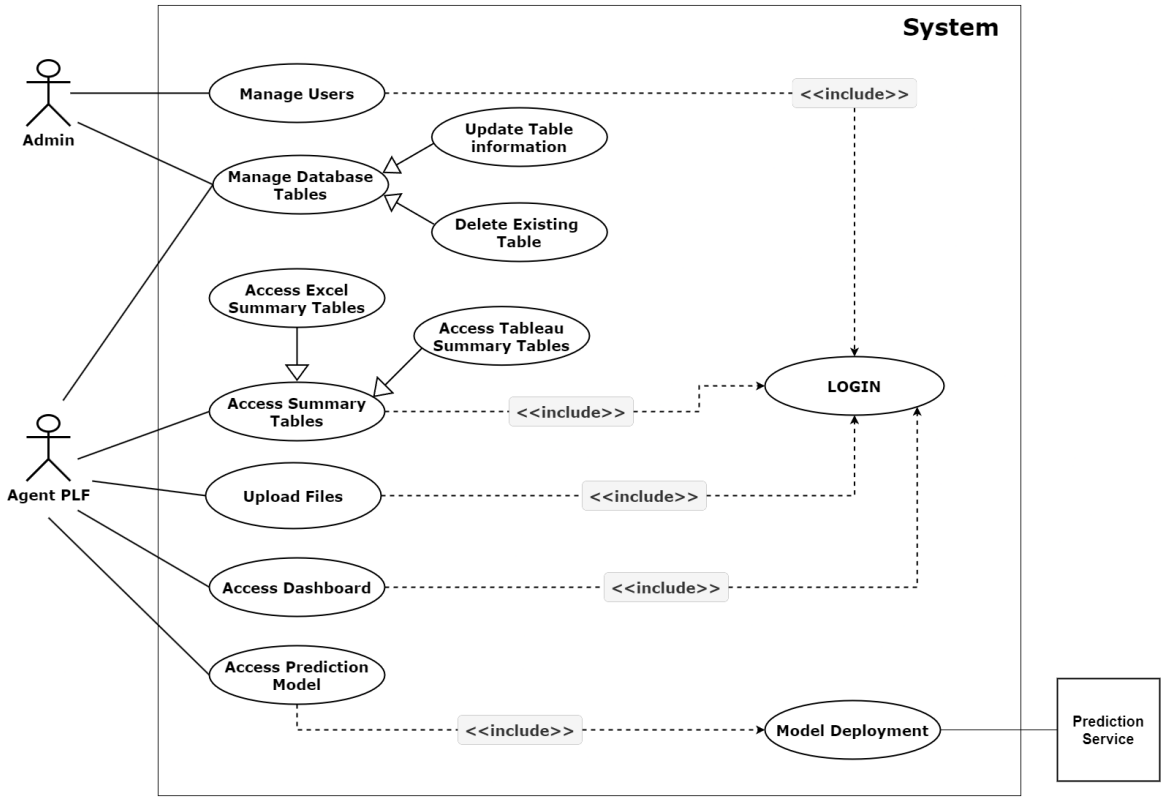

Figure 3.1: *Global Use Case Diagram*

3.4 Application Architecture and Components Descriptions

This section presents an overview of the architecture of the full application, including descriptions of its various components and their interactions. The application architecture is structured into distinct layers, which ensures modularity, maintainability, and scalability.

3.4.1 Overview

The application architecture is structured in a multi-tiered approach, consisting of the following layers:

1. **Front-end Layer:** This layer encompasses the components that are responsible for the presentation of the user interface and facilitating user interactions.
2. **Back-end Layer:** This layer handles business logic, data processing, and communication with external services.
3. **Database Layer:** This layer is responsible for the storage and management of the application's data.
4. **External Services:** In addition to the core components, the application interacts with external services for specific functionalities.

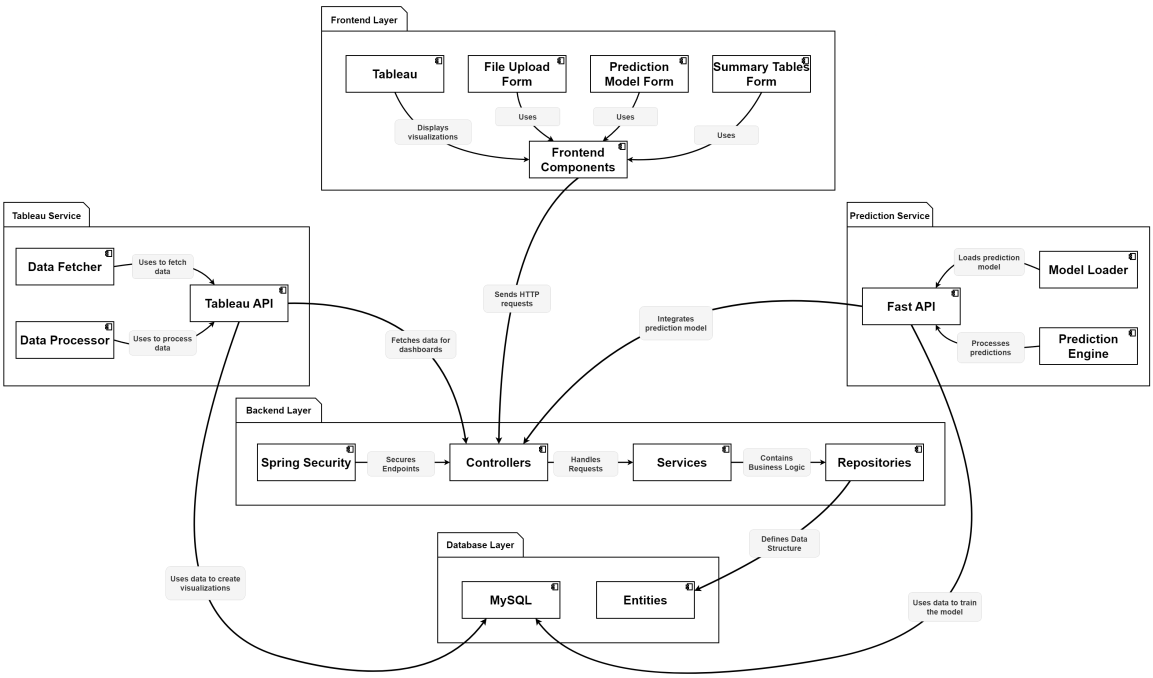

Figure 3.2: *Application's Architecture Overview*

3.4.2 Front-end Layer

This layer comprises components that render the visual elements with which users interact and handle the client-side logic.

1. **Tableau Dashboard:** The Dashboard component presents visualisations of investment project data. It employs Tableau, a sophisticated data visualisation tool, to generate interactive charts, graphs, and tables. Users can investigate and analyse data through a range of visualisations (*4.21*).

2. **File Upload Form:** The File Upload Form component enables users to upload Excel files containing data pertaining to investment projects. It provides a user-friendly interface for selecting and uploading files. Once a file is uploaded, the frontend communicates with the backend to process the uploaded data and initiate further actions (*4.22*).

3. **Prediction Model Form:** The Prediction Model Form component enables users to input data for predictions using a machine learning model. It presents a form where users can enter relevant information, such as the project label. Upon submission, the frontend communicates with the backend to process the input data and communicate with the model to make a classification and display the result to the user (*4.23*).

4. **Summary Tables Form:** The Summary Tables Form component displays generated summary tables based on the extracted data. It presents predefined summary tables according to predefined rules and standards. Users can view aggregated data, metrics, and statistics in tabular format, which facilitates data analysis and decision-making.

3.4.3 Backend Layer

The backend layer is composed of components that execute server-side operations, business logic and interact with the frontend layer and other services to fulfil user requests (*4.18*).

1. **Controllers:** These are responsible for handling HTTP requests from the frontend layer and routing them to the appropriate services. They implement request handling logic and facilitate communication between the frontend and backend.
2. **Services:** These implement business logic and perform data processing tasks. They encapsulate the core functionalities of the application and implement business logic and perform data processing tasks, orchestrating interactions between different components to fulfil user requests.
3. **Spring Security:** This provides authentication and authorization mechanisms to secure the application. It handles user authentication, access control, and session management, thereby ensuring that only authorised users can access protected resources.
4. **Repositories:** They manage data access and interactions with the database. They provide an abstraction layer for database operations, including CRUD (Create, Read, Update, Delete) operations and complex queries.

3.4.4 Database Layer

The database layer is responsible for storing and managing application data persistently. It comprises components that are responsible for defining the data structure and ensuring data integrity.

1. **MySQL Database:** The MySQL Database is employed to store data pertaining to investment projects, user profiles, and other application-related information. It utilises relational database management principles for the storage and retrieval of data.
2. **Entities:** Entities are defined as the data structure within the database. They represent entities such as users, projects, and summary tables.

3.4.5 External Services

In addition to the core components, the application interacts with external services for specific functionalities.

1. **Tableau Service:** The Tableau service serves to facilitate communication with Tableau for the purpose of data visualisation. The service retrieves data from the database and formats it in a suitable manner for visualisation in Tableau dashboards. The Tableau service may entail a number of tasks, including querying the database, transforming data into suitable formats for visualisation, and integrating with Tableau APIs for the creation and updates of dashboards.
2. **Prediction Service:** The prediction service interacts with an external prediction model to generate a classification and prediction on project family. It receives input data from users, processes it, and sends it to the prediction models for analysis. After receiving the prediction, the service formats and presents the results to the users through the frontend interface.

The external services in question play a pivotal role in the enhancement of the application's functionality, utilising ad hoc tools and resources for the purposes of data visualisation and predictive analysis.

3.4.6 Interaction Between Layers

The frontend, backend, and database layers work together to fulfil user requests, process data, and maintain application state.

1. **Frontend-Backend Interaction:**
   When a user interacts with the frontend interface, such as submitting a form or clicking a button, the frontend layer sends an HTTP request to the appropriate endpoint in the backend layer. Upon receiving the request, the backend layer routes it to the corresponding controller, which then invokes the appropriate service to handle the request.

   The service performs business logic operations such as data processing, and manipulation. It might interact with the database layer to retrieve required for processing. Once the necessary operations have been completed, the service returns a response to the controller, which in turn returns a response to the front-end layer containing the results of the operation. (***4.19***).

2. **Backend-Database Interaction:**
   The backend layer interacts with the database layer to retrieve, store and manipulate the application data. When a service needs to access data stored in the database, it interacts with the appropriate repository to perform database operations. The repository abstracts the database access logic, providing methods for CRUD operations

and querying data (*4.20*).

The service uses these repository methods to retrieve or modify data as required. It may perform complex queries or transactions involving multiple database entities to perform the requested operation. Once the database operation is complete, the service returns the result to the calling controller, which then formats it as a response to be sent back to the front-end.

3.5 Conclusion

In summary, the application should successfully automate the process of collecting data and creating summary tables from investment project data stored in Excel files. By using a structured architecture consisting of front-end, back-end, database and external service layers, the application increases efficiency and reduces manual effort in data processing.

The front-end layer provides a user-friendly interface that allows users to interact easily with the application. Key components such as the file upload form and the prediction model form facilitate data entry and interaction with the backend. The backend layer handles the core business logic, processing user requests and managing data through well-defined controllers, services and repositories. Integration with external services such as Tableau and predictive models further extends the capabilities of the application, providing advanced data visualisation and predictive analysis.

The database layer provides reliable and efficient data storage and retrieval, supporting application operations with a robust relational database management system. The interaction between the layers is designed to maintain data integrity, security, and performance, ensuring a smooth and efficient user experience.

This comprehensive approach to application architecture not only addresses the immediate need to automate data processing, but also provides a scalable and maintainable foundation for future enhancements and extensions. By automating repetitive tasks and ensuring consistency and accuracy, the application significantly improves the workflow of investment project data analysis, allowing users to focus on more strategic and value-added activities.

Chapter 4

Implementation

4.1 Introduction

Our project implementation phase is geared towards the development of a web platform that will be able to work for Sonatrach, along with the transformation of theoretical designs into a working system. Some significant technological choices include Tableau By Salesforce for Business Intelligence, Python for machine learning, and Spring Boot for backend development.

The subsequent sections detail the implementation process, starting with the implementation of the BI and machine learning integration part, backend development, frontend development, and lastly, a summary of the implementation phase.

4.1.1 Project Management Tools

Throughout our project, we have used essential tools for storing and keeping track of both the real-world implementation solution and the Thesis that represents a Documentation of it.

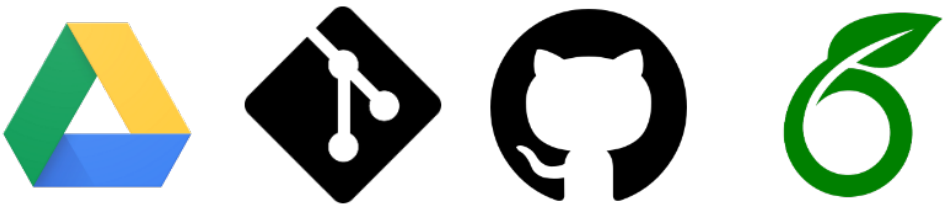

Figure 4.1: *Logos of the main Project Management Tools that were used*

1. **Google Drive:** Google Drive is a cloud-based file storage and synchronization service provided by Google. It allows users to store files in the cloud, access them from any device with an internet connection, and collaborate on documents in real-time [38].
2. **Git and GitHub:** Git is a distributed version control system used for tracking changes in source code during software development [39], GitHub is a web-based platform that provides hosting for Git repositories and offers collaboration features such as issue tracking, pull requests, and code review [40].

3. **Overleaf:** Overleaf is an online LaTeX editor that allows users to create, edit, and collaborate on LaTeX documents in real-time. It provides templates, version control, and project sharing features, making it a popular choice for writing academic papers, reports, and theses [41].

4.2 Business Intelligence Implementation

This is a project-focused implementation section that elaborates on BI solutions for Sonatrach. We will discuss the use of Tableau Prep, Tableau Desktop, Tableau Bridge, and Tableau Cloud, among others, explaining why we have chosen this specific stack. We go from cleaning data to dashboard developments, discussing key steps and challenges we met on the path to giving an inside look at our efforts with BI integration.

4.2.1 Overview of Technological Choices

Figure 4.2: *Logos of the BI Implementation tools*

1. **Tableau Stack:**
   - **Tableau Prep (Version 2024.1.0):** Tableau Prep is a data preparation tool that allows users to clean, shape, and combine data for analysis in Tableau Desktop or other data visualization tools [42].
   - **Tableau Desktop (Version 2024.1.0):** Tableau Desktop is a data visualization tool that allows users to create interactive visualizations and dashboards from their data [43].
   - **Tableau Bridge (Version 2024.1.0):** Tableau Bridge is a data connectivity tool that allows users to securely connect on-premises locally stored data to Tableau Cloud or Tableau Server [44]. This is necessary since it is not possible to connect a cloud service that has dashboards to a local data source (which is the local MySQL database on my Machine 1 (***4.2***)) due to firewalls and other security measures.
   - **Tableau Cloud (Version 2024.2.0):** Tableau Cloud (Previously known as Tableau Online) is a cloud-based platform for hosting and sharing Tableau dashboards and visualizations privately [45].
2. **MySQL Database (Version 8.0.36):** MySQL is an open-source relational database management system (RDBMS) that is widely used for storing and managing structured data [46]. The MySQL Workbench tool is used for designing, developing, and administering our MySQL database.

Justification for Technological Choices

Tableau Stack: We chose the Tableau Stack for several reasons:

- **Ease of Use and Powerful Visualization:** Tableau's user-friendly interface and advanced visualization options enable quick learning and effective data presentation, which are crucial for timely project implementation.
- **Access via Student Keys and Tableau Developer Program:** We were able to access Tableau Desktop and Tableau Prep using student keys provided by Tableau support, which last for a year. Additionally, we gained access to Tableau Cloud through the Tableau Developer Program, which provided us with a hosted site for our project needs.
- **Industry Standard:** Tableau is widely recognized and used in the industry, ensuring that our skills and solutions are relevant and up-to-date. Moreover, Tableau is currently the world's leading Business Intelligence (BI) tool, providing advanced analytics and visualization capabilities that are trusted by organizations globally.

MySQL Database: Our choice to use MySQL was based on:

- **Familiarity:** We have extensive experience with MySQL from our academic journey, making it the most comfortable and efficient choice for us.
- **Sufficiency for Project Needs:** MySQL adequately meets the requirements of our project without needing an unnecessary learning curve associated with more modern NoSQL databases. Its reliability and robustness are well-suited for our data management needs.

4.2.2 Implementation Process

Data Cleaning Operations

Data cleaning was one of the most essential steps to clean up raw data for further analysis and visualization. We retrieved an Excel file (.xlsx) from the Department of Operations and Planification: it constituted the Medium-term plan for the production division. In this table, there were some 1200 projects distributed over 11 regions. Each project has 56 attributes. These characteristics belonged to different types, like values in string form (for example, Region, Activité, Famille) or numeric values (like Coût Global Initial PMT 2022/2026, Coût Global Initial PMT 2023/2027) (*4.24*).

To clean and prepare the data, we used Tableau Prep Software, The following steps showcase the data cleaning operation performed:

1. **Connecting to the Excel File:** We connected the Excel file to Tableau Prep to begin the cleaning process.
2. **Removing Unnecessary Elements:** Unnecessary elements and rows that did not contribute to the analysis were removed to move the dataset to a higher standard.
3. **Handling Missing Values:** Missing values were identified and handled appropriately (either NULL cells case or 0's case). This included filling in missing data where possible.
4. **Adding a Project Key Column:** A new column for project keys was added. This column auto-increments to provide a unique identifier for each project.

5. **Defining Clear and Readable Attributes:** Attribute names were refined to be more clear and readable. This involved renaming columns to ensure consistency and clarity, a major example is the decoding of project activities where the letter 'A' was the equivalent of "Gaz" and 'B' for "Pétrole", another one was the coding of each Region which were coded using special letters, e.g. the Region "Hassi Messaoud" was coded as 'D', "In Amenas" as 'T'.
6. **Correcting Data Types:** Data types were corrected to ensure proper formatting and usability. String values were set as text, and numerical values were set as double.
7. **Exporting to MySQL Database:** After cleaning and preparing the data, it was exported as a table named "projects" into our "sonatrach" MySQL database (*4.25*).

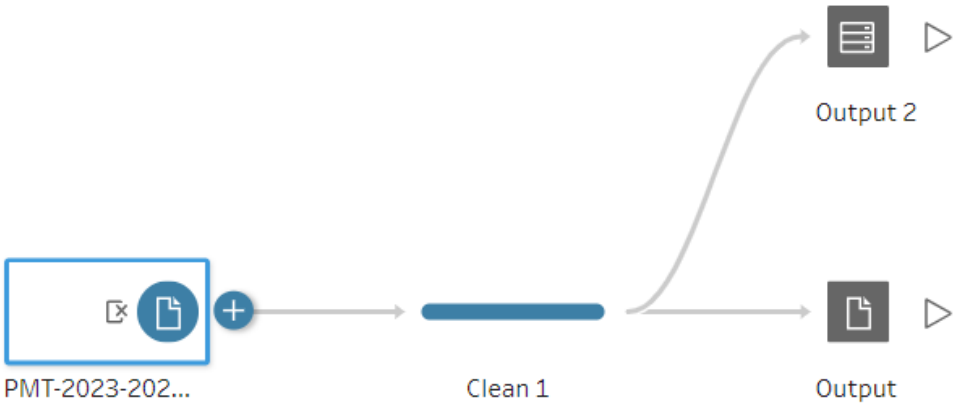

Figure 4.3: *Tableau Prep's Workflow*[1]

Creating Visualizations

The connection of the Tableau Desktop Software was done with the "projects" table in the MySQL database, which belonged to "sonatrach" database. It allowed the exercise of retrieving real-time data so that any alteration in the database could be realized in Tableau spontaneously, giving an updated and dynamic visualization environment.

The visualizations developed in Tableau Desktop were divided into two main categories: recapitulative tables and dashboards.

1. **Recapitulative Tables**
   - **Classification of Projects per Region Dimension:** Projects were classified by "Region," summarizing all numerical attributes with totals for each region.
   - **Classification of Projects per Project Activity and Family Dimensions:** Projects were classified by "Project Activity" (Gas or Oil) and "Family of Project" summarizing all numerical attributes with totals.
   - **Classification of Projects per Extensive Family Dimension:** Projects were classified by "Extensive Family"[2] summarizing all numerical attributes with totals.

[1] "Output" designs the new post-cleaning Excel file, "Output 2" is the same but in the form of a MySQL table.
[2] The "Extensive Family" dimension refers to detailed subcategories within the general project families. For example, the "activité puits" category is divided into sub-families such as workover/side track, short radius, work over d'abandon, work over mise en conformité, snubbing, stimulation, acidification, and fracturation.

- **Classification of Projects per Project Activity and Extensive Family Dimensions:** Projects were classified by "Project Activity" and "Extensive Family," summarizing all numerical attributes with totals.

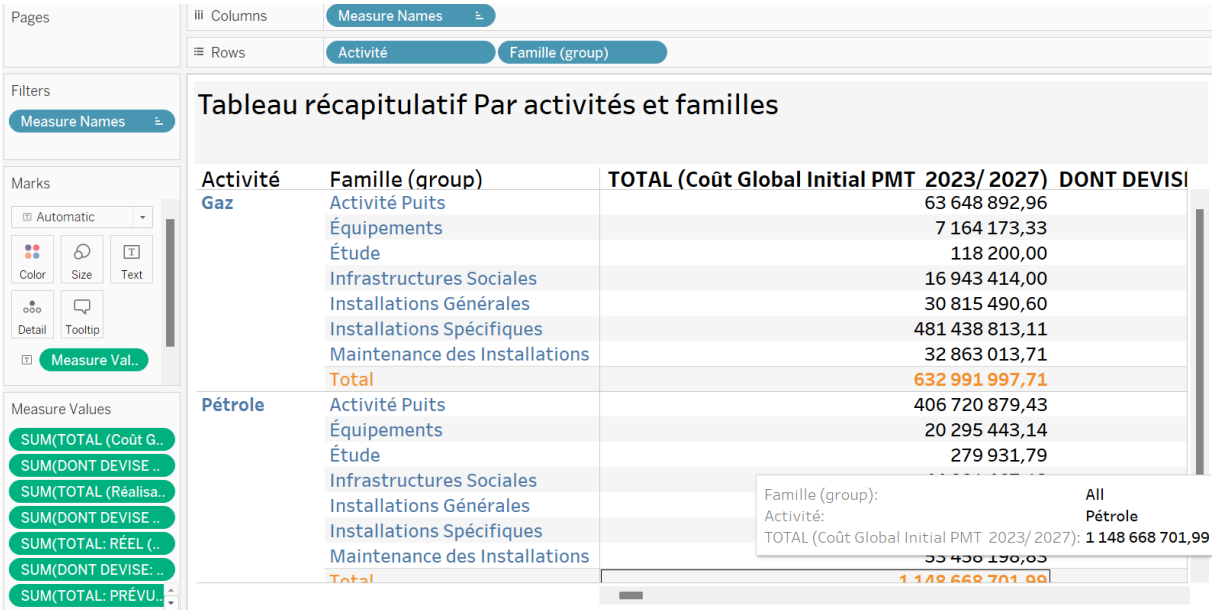

Figure 4.4: *The making of one of the Recapitulative Tables in Tableau Desktop*

2. **Dashboards**

   - **Map Display:** Regions and projects were displayed on a map, highlighting geographical distribution (*4.26*).
   - **Pie Charts:** Percentage distributions of projects by activity and family were visualized using pie charts.
   - **Horizontal Bar Charts:** Various metrics were represented using horizontal bar charts, offering a comparative view of project metrics.
   - **Line Charts:** Line charts were created to track the total monthly expenses on projects throughout the year in both local currency (DA) and foreign currency (USD).

Publishing to Tableau Cloud

After creating the dashboards in Tableau Desktop, the next step was to publish them to our Tableau Cloud website (*4.27*). Tableau Cloud offers multiple options for publishing the data that fuels these dashboards:

- **Uploading a Copy of the Database**: This involves uploading a copy of the database, specifically the tables feeding the dashboards, to Tableau Cloud servers or embedding the database within the workbook itself [47].
- **Using Tableau Bridge**: This software connects workbooks in a given Tableau Cloud website to a local data source on a machine that hosts the database server. Tableau Bridge provides either scheduled extract refresh on the data source or a live connection [48].

We opted for Tableau Bridge to maximize the confidentiality of the company's data and

prevent hosting sensitive information on external nodes. Additionally, we set the confidentiality of the workbooks to private, ensuring that only registered users on our Tableau Cloud website can access them. Even registered users have restricted access, limited to the dashboard visualizations without access to the original data source.

JWT Connected App Tool

Next step, we utilized the Connected App feature in Tableau Cloud. This feature enables the creation and management of trusted relationships between Tableau and external applications. Here's the process:

- **Accessing the Connected App Parameter**: We accessed the Connected App parameter in Tableau Cloud after consulting the official Tableau documentation and grasped comfortably how it works.
- **Generation of Confidential Fields**: For a connected app, Tableau Cloud generates three unique confidential fields: Secret ID, Secret Value, and Client ID (***4.28***).
- **Creating a JWT**: These fields are used to generate a JSON Web Token (JWT)[3] that our soon-to-be-developed application will use to authenticate with the Tableau Cloud API to retrieve its dashboards. We will be delving deeper into explaining this process in the next sections.

4.2.3 Challenges and Solutions

Data cleaning in Tableau Prep was labor-intensive due to the extensive refinement needed for quality assurance, dealing with inconsistencies, and missing values. Despite Tableau Prep's advanced features, the process was time-consuming.

Creating visualizations in Tableau Desktop was straightforward, thanks to its user-friendly interface and powerful capabilities. This part of the project went smoothly.

Publishing dashboards to Tableau Cloud was manageable, but configuring Tableau Bridge for data confidentiality was challenging. Ensuring data privacy and restricted access required careful attention and testing.

Implementing Connected App functionality in Tableau Cloud was complex. Setting up the connected app, generating JWTs, and using secret fields correctly required significant effort due to the complexity of the Tableau documentation.

In conclusion, while Tableau offers powerful BI tools, data preparation and advanced security configurations posed notable challenges. These were overcome through detailed attention and extensive use of Tableau's documentation and support.

4.3 Machine Learning Implementation

This section deals with project classification using machine learning. Our goal is to predict families of investment projects based on factors such as regions and project label. The diverse nature of projects in this sector requires a classification system.

[3] an open standard (RFC 7519) that defines a compact and self-contained way for securely transmitting information between parties as a JSON object. This information can be verified and trusted because it is digitally signed [49].

Previously, the studies and planning department filled an Excel file with project data, which led to classification differences with regions that caused delays. To address this, we are using the Random Forest algorithm, which is known for its effectiveness with complex data. This research aims to improve project classification and understand its driving factors.

4.3.1 Overview of Technological Choices

Figure 4.5: *Logos of the main Machine Learning Implementation tools*

- **Python (Version 3.11.6):** Python, a high-level, interpreted language, is acclaimed for its readability, simplicity, and robust library support. Known for web development, data analysis, AI, and ML, Python offers a variety of libraries for efficient data manipulation and model building [50]. Its straightforward syntax allows for rapid development and streamlined debugging, which is critical for complex projects like ours.
- **PyCharm Professional Edition (Version 2023.3.4):**PyCharm, an integrated development environment (IDE) tailored for Python, is known for its power. The Professional Edition offers advanced features such as integrated tools for databases, web frameworks, and scientific applications. Its error checking, and quick fixes increase code quality and productivity. With robust debugging and testing tools, PyCharm is essential for building reliable ML models and APIs [51].
- **FastAPI:** FastAPI, a modern web framework for API development with Python 3.6+, uses standard Python type hints for streamlined development. It prioritizes ease of use and developer friendliness. FastAPI excels at handling API requests and real-time data processing [52]. Used as a communication tool between our ML model and web application, FastAPI facilitates integration and seamless data exchange.

Libraries Used

During the development of the machine learning model we used several Python libraries, each playing an important role in various aspects of data manipulation, analysis, model implementation, and application deployment:

- **Pandas:** Utilized for data manipulation and analysis, providing flexible data structures and data analysis tools. Pandas is essential for handling and preprocessing the dataset, allowing for efficient data cleaning and transformation [53].
- **Scikit-learn:** Employed for implementing the Random Forest algorithm and other machine learning utilities. Scikit-learn offers a wide range of algorithms and tools for model training, evaluation, and validation, making it a key library for our project [54].

- **SQLAlchemy:** Used for database operations, offering a SQL toolkit and Object-Relational Mapping (ORM) capabilities for an effective integration with MySQL. SQLAlchemy simplifies the process of connecting to the database, executing queries, and managing data [55].
- **Uvicorn:** Used to deploy the FastAPI application. Uvicorn is a lightning-fast ASGI server, facilitating the efficient handling of asynchronous requests and the smooth deployment of the FastAPI application [56].

4.3.2 Data Description

Data Origin

The data used in this study is provided by the company's planning and studies department. It was initially stored in an Excel file called "PMT" (Plan Moyen Terme), which is used as a main source of information on investment projects throughout different regions.

The Excel file is sent to the regions and departments of the company as part of the data collection process. According to the template provided, each institution shall fill in project information. The file will be sent back for review after it is completed.

This method aims to collect project data from across regions and sectors in order to allow better analysis and allow decision-making.

Dataset Overview

A total of 1,289 investment projects, each containing at least 56 features, are included in the dataset, These features include different aspects of the projects, such as geographic information and financial details. It is important to note that with changes in project reporting requirements or data collection procedures, the number of features may change from year to year; Here is a detailed explanation of the features included in the dataset:

- **Activity Type:** Indicates whether the project relates to the company's gas or oil activities.
- **Region:** This represents the geographical area of the country where the project is located, providing a basis for regional analysis.
- **Management Structure:** Describes the organisation in which the project is being carried out, which, in some cases, may be the same as the name of the region or the specific perimeter.
- **Perimeter**[4]**:** Refers to specific areas or zones within a region where the project is happening, providing additional information about project location.
- **Perimeter's Code:** To facilitate perimeter identification, a standard coded representation of the perimeter name is provided.
- **Family:** This is used to classify a project into one of the various families based on its nature or purpose, facilitating project categorization.
- **Project Label:** Mostly the name for each project or a brief description, enabling easy reference and tracking.

[4] a total count of 191 Perimeters exists, distributed upon 11 regions.

- **Cost Features:** Include monthly costs for the first year and yearly costs for the following four years, providing insights into the financial aspects of each project over time.

4.3.3 Methodology

In this section, we outline the methodology employed to develop the machine learning model for predicting project families. This includes the algorithm chosen, feature engineering techniques, model training process, and evaluation metrics used.

For the Algorithm selection, **Random Forest Algorithm** was chosen for its suitability to handle high-dimensional data, capture complex relationships between features, and provide robust predictions. Its ensemble learning approach, which combines multiple decision trees, makes it well-suited for our task of project family prediction, and here's why:

- **Accuracy and Robustness:** Random Forest is known for its high accuracy, especially in handling high-dimensional data with complex relationships between features and target variables. Since accuracy is crucial for the classification task, Random Forest's ensemble of decision trees can provide reliable predictions.
- **Feature Importance:** Random Forest provides a measure of feature importance, which can be useful for understanding the relative importance of different features in predicting the target variable. This information can help in feature selection and interpretation of the model's behavior.
- **Robustness to Overfitting:** Random Forest tends to generalize well to unseen data and is less prone to overfitting compared to individual decision trees. By averaging predictions from multiple trees, Random Forest reduces the risk of overfitting, making it suitable for datasets with noise or outliers.

The steps involved in preparing the dataset for model training, including tasks such as handling missing values, encoding categorical features, processing text features, and splitting the dataset into training and testing sets are the following:

1. **Treating Categorical Features:** The dataset contains categorical features such as 'Région', which were encoded using one-hot-encoding. This approach converts categorical variables into a binary matrix, suitable for input into the Random Forest algorithm. The one-hot encoding was implemented using the "get_dummies" function in the pandas package.
2. **Text Feature Processing:** The 'Libellés' feature, containing project names, was processed using text vectorization to transform the textual data into numerical form, The TfidfVectorizer from the scikit-learn library was put to use in order to convert the 'Libellés' text into TF-IDF (Term Frequency-Inverse Document Frequency) vectors.
3. **Combining Features:** The one-hot encoded categorical features and the TF-IDF vectors of the 'Libellés' feature were concatenated to form a unified feature set. This combined feature set was then prepared for model training.
4. **Data Splitting:** The dataset was split into training and testing sets using a $\frac{6}{10} : \frac{4}{10}$ split to evaluate the performance of the Random Forest classifier. The "train_test_split" function from scikit-learn library was used for this purpose.
5. **Model Training:** As the Random Forest classifier was chosen for this project. The

classifier was trained on the prepared training set, utilizing the combined feature set of one-hot encoded categorical variables and TF-IDF vectors from the 'Libellés' feature.

4.3.4 Implementation Process

Connecting to MySQL Database for Training: Data was loaded from the MySQL database 'sonatrach' (***4.25***) using SQLAlchemy. After preprocessing, the cleaned data was used for model training (***Appendix***).

Model Development in Python: The machine learning model was developed using Python (***4.29***). The process included initial data preprocessing, applying the Random Forest algorithm, and validating the model using various metrics.

Deploying the Model Using FastAPI and Uvicorn: The trained "random forest model with high accuracy.pkl" was deployed using FastAPI, with Uvicorn serving the application (***4.30***). The setup included endpoints for predictions, handling input data, and returning structured results.

4.3.5 Model Performance

This section evaluates the machine learning model's performance in predicting project families.

To properly evaluate the model performance, we used the accuracy, along with a detailed classification report, and the confusion matrix.

Key metrics include precision, recall, and F1-score. Precision measures the accuracy of positive predictions, recall measures the ability to identify all positive instances, and the F1-score is the harmonic mean of both. The support column shows the number of true instances for each class in the test set.

The accuracy of the model is (**89.147%**), the classification report also indicates strong performance in classes like "Acidification," "Etude," and "Installations Spécifiques," with high precision and recall. Classes like "Work over d'abandon" had lower support, affecting performance.

Class	Precision	Recall	F1-Score	Support
Activité Puits	0.99	0.97	0.98	158
Infrastructures Sociales	0.70	0.70	0.70	20
Installations Générales	0.83	0.77	0.80	97
Installations Spécifiques	0.88	0.92	0.90	193
Maintenance des Installations	1.00	0.86	0.92	35
Équipements	0.53	1.00	0.69	9
Étude	1.00	0.50	0.67	4

Table 4.1: *Performance Metrics*

Confusion Matrix

The confusion matrix shows the actual versus predicted classifications for each class, with rows representing actual classes and columns representing predicted classes.

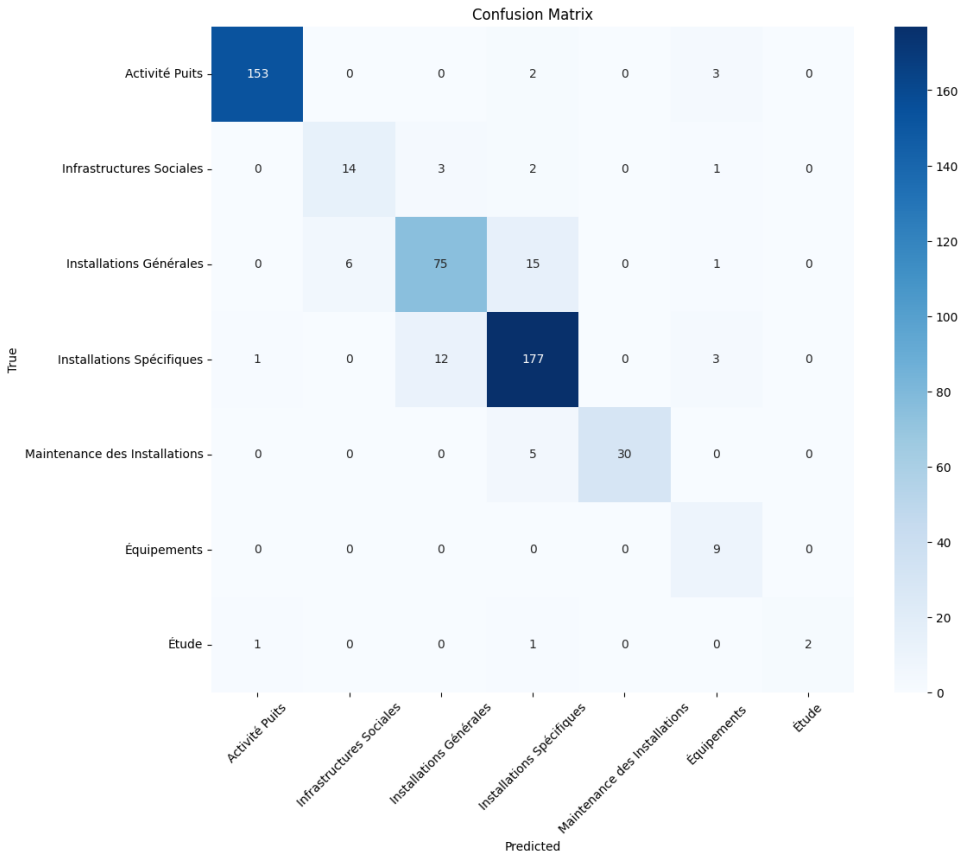

Figure 4.6: *Confusion Matrix*

The confusion matrix demonstrates the performance of the classifier across different categories. The classifier shows high accuracy for major classes like **Activité Puits** and

Installations Spécifiques, with most samples in these categories being correctly classified and minimal misclassifications. However, there are notable misclassifications in some categories, such as **Installations Générales** being frequently misclassified as **Installations Spécifiques**, indicating overlapping features that make these categories harder to distinguish. Additionally, rare classes like **Équipements** and **Étude** show perfect or near-perfect classification, but due to the small sample sizes, these results may not be entirely reliable. This suggests a need for more data or alternative methods to handle class imbalance.

4.3.6 Limitations

Despite the model's promising performance, several limitations were observed during its development and evaluation:

- **Imbalanced Classes:** Imbalanced class distributions can lead to biased predictions and lower performance in minority classes.
- **Feature Engineering Challenges:** Extracting meaningful features for project family prediction is difficult when the label feature lacks clear indicators.
- **Data Quality Issues:** Noise, outliers, or inconsistencies in the dataset can negatively impact model performance and generalization.
- **Limited Sample Size:** An insufficient dataset may not capture the full diversity of investment projects, limiting the model's generalization to unseen data.

4.4 Backend Development

Backend development is crucial in software engineering, managing data, ensuring security, and integrating with other services. In our project, the backend supports the BI and ML components, ensuring data flow and interaction between the frontend and analytical tools.

The backend is built with Java and the Spring Boot framework, chosen for its modularity and scalability in creating microservices. Tools like Hibernate for ORM and Spring Security for authentication enhance functionality and security.

This section delves into the technological choices for backend development, the libraries used, the implementation process, and the challenges encountered, offering insights into building a solid backend for a comprehensive BI and ML project.

4.4.1 Overview of Technological Choices

Figure 4.7: *Logos of the Main tools used for Backend development*

- **Java (Version 17.0.10):** Java is a versatile, high-level programming language known for its portability and performance. It runs on any machine supporting the Java Virtual Machine (JVM), making it suitable for large-scale applications [57]. We chose Java because we are comfortable with it, having used it extensively in the last 4 out of 5 years of our curriculum studies.
- **Spring Boot (Version 3.2.4):** Spring Boot is an open-source framework that simplifies the development of Spring applications. It reduces the need for manually creating configurations code associated with other Spring frameworks (ex. Hibernate, data JPA, Spring security..etc) and abstracts it away, and allows easy deployment with embedded servers like Tomcat [58].
- **Postman (Version 10.23):** Postman is a collaborative platform for API development, simplifying the creation, testing, and documentation of APIs with its user-friendly interface [59]. We used Postman to test API requests to the ML endpoint and the Tableau Cloud endpoint.
- **Spring Tool Suite 4 (Version 4.21.1):** Spring Tool Suite (STS) is an IDE for Spring applications, providing tools for implementation, debugging, and deployment [60]. It was used on Machine 1 (*4.2*).
- **IntelliJ IDEA Ultimate Edition (Version 2023.3.5):** IntelliJ IDEA is a feature-rich IDE for Java development, offering advanced tools for web and enterprise development, and support for frameworks like Spring [61]. It was used on Machine 2 (*4.2*).

Notable Dependencies

During the development of our backend, several dependencies[5] were used to enhance functionality and run several processes. Obvious dependencies that come to mind are:

- **Spring Boot Starter Dependencies:**
  - **spring-boot-starter-data-jpa:** Provides integration with Spring Data JPA for data persistence. It simplifies database interactions by using Java Persistence API (JPA) for ORM, allowing for efficient database operations [63].
  - **spring-boot-starter-security:** Adds security features to the application, including authentication and authorization mechanisms. This starter is crucial for securing the application against unauthorized access [64].
  - **spring-boot-starter-web:** Facilitates the creation of web applications, including RESTful services[6] using Spring MVC[7]. It includes everything needed to set up a web application quickly [67].
  - **spring-boot-starter-webflux:** Provides support for building reactive web applications with Spring WebFlux, which enables handling asynchronous requests efficiently [68].
- **poi-ooxml (Apache POI):** Apache POI is a Java library for reading and writing Microsoft Office documents, such as Excel files. This module specifically handles the

[5] In the context of a Spring Boot project, a dependency is an external library that the project needs to function correctly. Dependencies are managed using a build automation tool such as -in our case- Maven, which handles downloading, updating, and linking these libraries to the project [62].

[6] RESTful services adhere to the Representational State Transfer (REST) architecture, which uses standard HTTP methods and stateless communication to provide scalable and interoperable web services [65].

[7] Spring MVC (Model-View-Controller) is a framework within the Spring ecosystem that allows for building web applications by separating the application logic (Model), user interface (View), and control flow (Controller) [66].

Office Open XML (OOXML) formats used by newer versions of Excel [69].

- **mysql-connector-j:** A JDBC driver for MySQL databases, allowing Java applications to connect and interact with MySQL databases seamlessly [70].
- **json:** A library for parsing and generating JSON data. It simplifies the handling of JSON data structures, which are commonly used in web services and APIs for data exchange [71].

4.4.2 Project Structure

Initialization

The project was initialized using the Spring Initializr, a web-based tool that simplifies the setup of new Spring Boot projects by allowing users to select the necessary dependencies and configurations. By using this tool within the Spring Tool Suite (STS), we quickly generated a base project structure that included the main Spring Boot starter dependencies.

The project was organized into six main packages to adhere to clean code principles and ensure maintainability (***4.31***). These packages are:

- **models:** Contains entity classes that map to database tables. This package is the closest to the database layer.
- **repositories:** Contains repository interfaces that extend Spring Data JPA repositories. These interfaces handle data access operations.
- **services:** Contains service classes that implement the business logic. These classes interact with the repositories to perform operations.
- **controllers:** Contains controller classes that handle HTTP requests and responses. These classes interact with the services to process incoming requests and return appropriate responses.
- **configurations:** Contains configuration classes, including SecurityConfig, which define security parameters for the application.
- **components:** Contains additional components such as utility classes and custom beans that support the application.

In total, the backend consists of 26 classes, 2 repository interfaces, and 1 enumeration (UserRole). The classes are distributed across the 7 packages. Additionally, the default package "com.example.demo" contains -only- the main application class "PFEApplication.java" (***Appendix***).

At this stage, we used basic HTML pages as a view layer for testing purposes, focusing on functionality rather than aesthetics. The frontend development phase will be discussed in upcoming sections, it also goes without saying that a private GitHub repository is initiated and shared between project partners.

4.4.3 Database Connection and Design

The database connection was configured in the "application.properties" file using JDBC. This setup ensures that the Spring Boot application can communicate with the MySQL database (***Appendix***).

The database design includes two key tables: "projects" and "users", the "projects" table was previously created using Tableau Prep (***4.25***), it stores information about various projects.

The "users" table was created by Hibernate's JPA based on the "User" entity class. It stores user information such as usernames, passwords, and roles (***Appendix***). These entity classes represent the structure of the tables in the "sonatrach" database and are crucial for the application's interaction with it.

4.4.4 Microservices Communication

To make communication with the BI and ML microservices a reality, we designed specific controllers: "TableauController" for Tableau Cloud and "PredictionController" for the FastAPI ML model.

For the BI microservice, the "TableauController" class is crucial. It uses fields generated previously in the "Connected App" parameter in Tableau Cloud (***4.28***). These fields are stored in "application.properties" to avoid exposing them in the code base (***Appendix***).

and here's a snapshot of the TableauController class:

```
package controllers;
//Sevevreal necessary Imports //

@RestController
@RequestMapping("/TableauController")
public class TableauController {
     // the values needed for the connected application to generate the
    JWT to authenticate with Tableau cloud
    @Value("${tableau.online.username}")
    private String userId;

    @Value("${tableau.online.client-id}")
    private String is;

    @Value("${tableau.online.secret-id}")
    private String kid;

    @Value("${tableau.online.client-secret}")
    private String secret;

    @GetMapping
    public ResponseEntity<String> generateJWT() {
        // Validate input parameters :)
        validateInputParameters(userId, iss, kid, secret);

        String[] scopes = {"tableau:views:embed"};
        String token = createToken(userId, kid, secret, iss, scopes);

        // Return JWT as part of the response body
        return ResponseEntity.ok().body(token);
    }
//rest of the class//
```

The "TableauController" class uses these fields (after validating their correct formats)

to create a JWT valid for 200 seconds, following Tableau's guidelines[8] This JWT is then returned from the endpoint "/TableauController" and used by the frontend layer to authenticate with the Tableau Cloud API and retrieve dashboards. Extensive testing was conducted using Postman to ensure the validity of JWT tokens for authentication with Tableau Cloud. We also used browser debugging tools, especially the network logs, to verify the interactions.

For the ML microservice, the "PredictionController" class handles sending POST requests to the FastAPI application running on port 8000 (***4.30***) and receiving responses in return. This communication uses the JSON format. It is trivial to mention that the controller is linked to the view layer as well, where users enter data for the POST request and view the prediction results.

```
@PostMapping("/submit-project")
    public String submitProject(Project project, Model model) {
        // Send form data to FastAPI application
        String[] predictions = webClientBuilder.build()
                .post()
                .uri("http://127.0.0.1:8000/receive_project_data")
                .body(BodyInserters.fromValue(project))
                .retrieve()
                .bodyToMono(String[].class) // Expecting an array of
   predictions (String[])
                .block(); // Block to get the response synchronously
//rest of the PredictionController class
```

4.4.5 File Upload Functionality

Motivation

The decision to implement this functionality acknowledges the importance of human factors and potential resistance to change within the organization. So, The primary focus of this functionality is to automate the generation of summary tables from Excel files containing various data structures. This automation is aimed at streamlining the process of creating summaries from investment projects data, which is crucial for efficient management and informed decision-making.

The file format required for this functionality is **.xlsx**. A template is created and distributed across different regions and structures to ensure that investment projects data is consistently filled out. Users upload the file through a designated form, which they need to fill out correctly to ensure the data is processed and stored accurately. This form is accessible via the user interface of the application. Validation checks are in place to verify the file type and structure, ensuring that only correctly formatted files are processed. So This step includes checking for the correct format.

Once the file is uploaded, the system uses the input from the form to transform the coded values in the Excel file into their actual values. This involves parsing the file to extract the data and applying specific transformations to convert coded values into readable information. After these transformations, the data is stored in a designated table within the database, with each project being assigned a unique identifier. This ensures that

[8]The implementation of JWT authentication follows the official Tableau documentation, which provides guidelines on how to authenticate an external node (such as our Spring Boot app) with Tableau Cloud. For more information, visit https://help.tableau.com/current/api/rest_api/en-us/REST/rest_api_concepts_auth.htm.

the data is organized and easily retrievable for future use. The system also handles data inconsistencies and errors by implementing various checks and balances during the transformation process.

Predefined summary tables are generated from the stored data to provide users with insightful information for informed decision-making. These tables are structured to highlight key metrics and aggregated data points, making it easier for users to understand and analyze the information. The summary tables are displayed in a user-friendly interface, ensuring that the data is presented clearly and effectively.

Key classes (*4.32*) supporting this functionality include:

- **DatabaseTableService:** Retrieves tables from the database for table generation, ensuring data availability.
- **DataPreprocessor:** Decodes column values for data accuracy and understanding.
- **EmptyValuesFiller:** Handles missing data to ensure completeness and accuracy of generated tables.
- **InsertDataInDB:** Dynamically creates and inserts processed data into the database, ensuring correct data structuring and storage.

The integration of these functionalities provides a familiar and efficient method for users to generate and manage summary tables while facilitating a gradual transition to the Tableau-based system. This approach balances innovation with user comfort and organizational adaptability.

4.4.6 Security Mechanisms and Access Control

We implemented a secure login mechanism using Spring Security, focusing on password security and access control. User passwords are hashed using the BCrypt algorithm[9] before being stored in the database, ensuring that even if the database is compromised, the actual passwords remain protected.

```
    @Bean
    BCryptPasswordEncoder passwordEncoder() {
        return new BCryptPasswordEncoder();
    }
```

Access control is managed based on user roles, defined within the "SecurityConfig" class. Role-based access control (RBAC) allows us to assign permissions to roles and users to roles. This ensures that only authorized users can access certain endpoints. The "hasRole" helper method is used to check if a user has a specific role, facilitating access control within controllers.

```
// Helper method to check if the user has a specific role
private boolean hasRole(Collection<? extends GrantedAuthority>
    authorities, String role) {
     return authorities.stream().anyMatch(auth -> auth.getAuthority().
    equals(role));
}
```

[9] BCrypt is an adaptive hashing algorithm that uses factors that make password hashing computationally intensive, thereby providing solid security against password attacks [72].

4.4.7 Challenges and Solutions

Throughout the development of our system, we faced several challenges in which we opted for solutions in order to overcome them.

Integration Challenges: Integrating different components of the system posed challenges. Ensuring that the data flow between the front-end, back-end, Microservices, and database was smooth required careful planning and testing. We encountered several issues related to compatibility and data consistency, which we resolved by thoroughly testing each module independently before integrating them into the larger system.

Security: Securing user data and also ensuring proper access control were one big priority of ours, Implementing security measures, such as the use of BCrypt for password hashing and role-based access control, allowed avoiding potential vulnerabilities in a realistically confident manner. Despite these measures, security remains an ongoing concern that requires continuous monitoring and updates.

Bugs and Debugging: Like any other software project, we faced various bugs during the development of the Backend. Debugging these issues was time-consuming -to say the least- and required the use of various tools and resources. We relied heavily on community support from platforms like Stack Overflow and GitHub to find solutions or at least new insights into complex problems faced. This approach widely accelerated the debugging process.

One thing worth noting, is the fact that this project is version 1.0, indicating that it is the first fully functional iteration, While it is a significant milestone, we anticipate further refinements and updates based on user feedback and evolving requirements. Future versions will continue to improve upon the foundation laid by this initial release.

4.5 Frontend of The Final Product

Previously, it was mentioned that we have used basic HTML pages as a view layer to test the functionalities created in the Backend development phase, in this section we explore the efforts made to create a dedicated and intuitive view layer interface for our system (*4.33*).

4.5.1 Overview of Technological Choices

Figure 4.8: *Logos of the main Frontend Tools that were used*

- **Thymeleaf**: Thymeleaf is a modern server-side Java template engine compatible with Spring Boot (5). Its natural templating capability simplifies backend integration and data exchange [73].
- **CSS (Cascading Style Sheets)**: CSS describes the presentation of HTML or XML documents, essential for creating visually appealing and responsive web designs, and separating content from design for easier maintenance [74].
- **JavaScript**: JavaScript is a versatile programming language for interactive and dynamic web pages. It enhances user experience with real-time updates and interactions without reloading the page, crucial for modern web development [75].

4.5.2 Some Final Product Interfaces

Figure 4.9: *First Page: Login page*

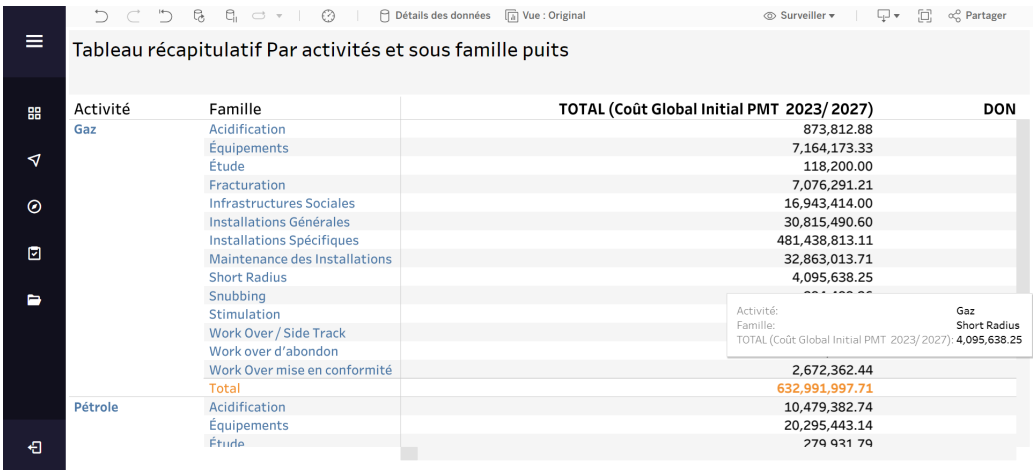

Figure 4.10: *A Recapitulative Table Display*

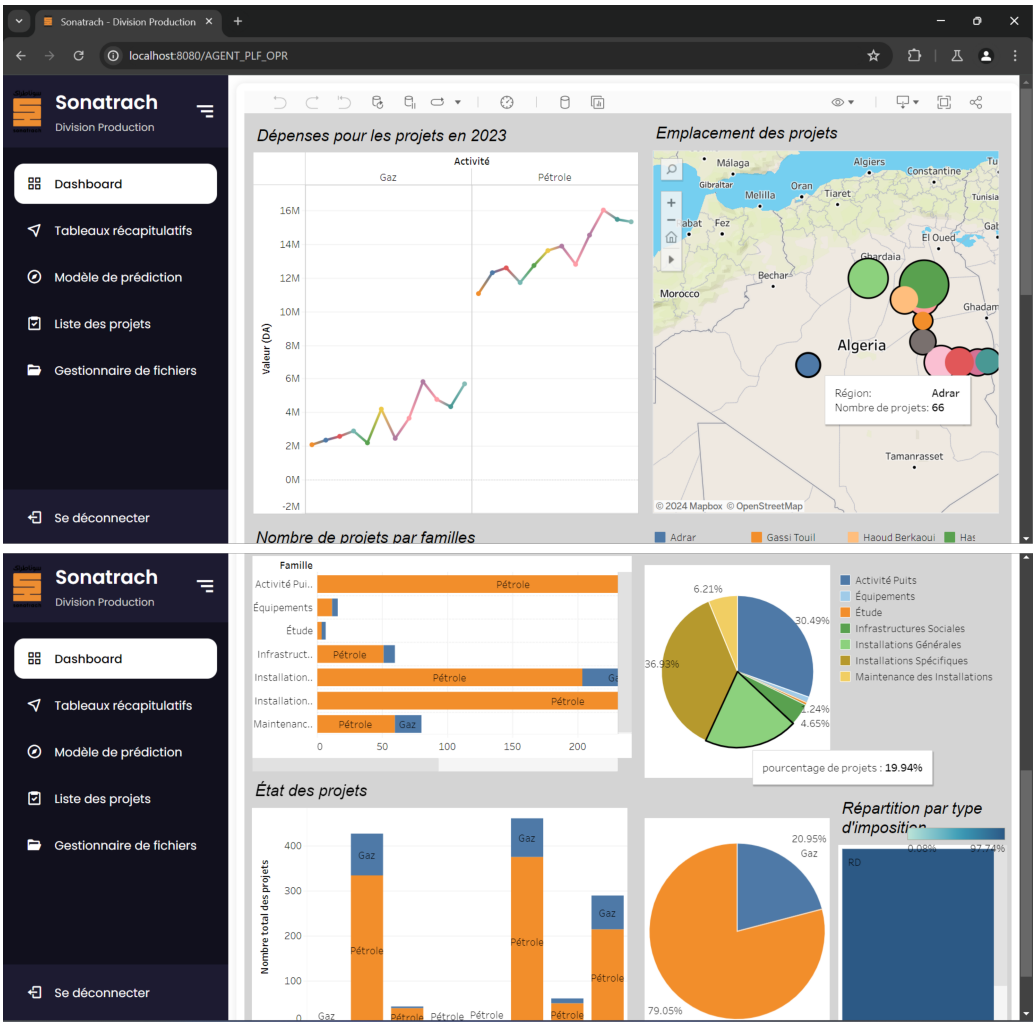

Figure 4.11: *Main Projects Dashboard*[10]

[10] All visualization elements (single or whole) can be downloaded in either pdf, png or pptx formats.

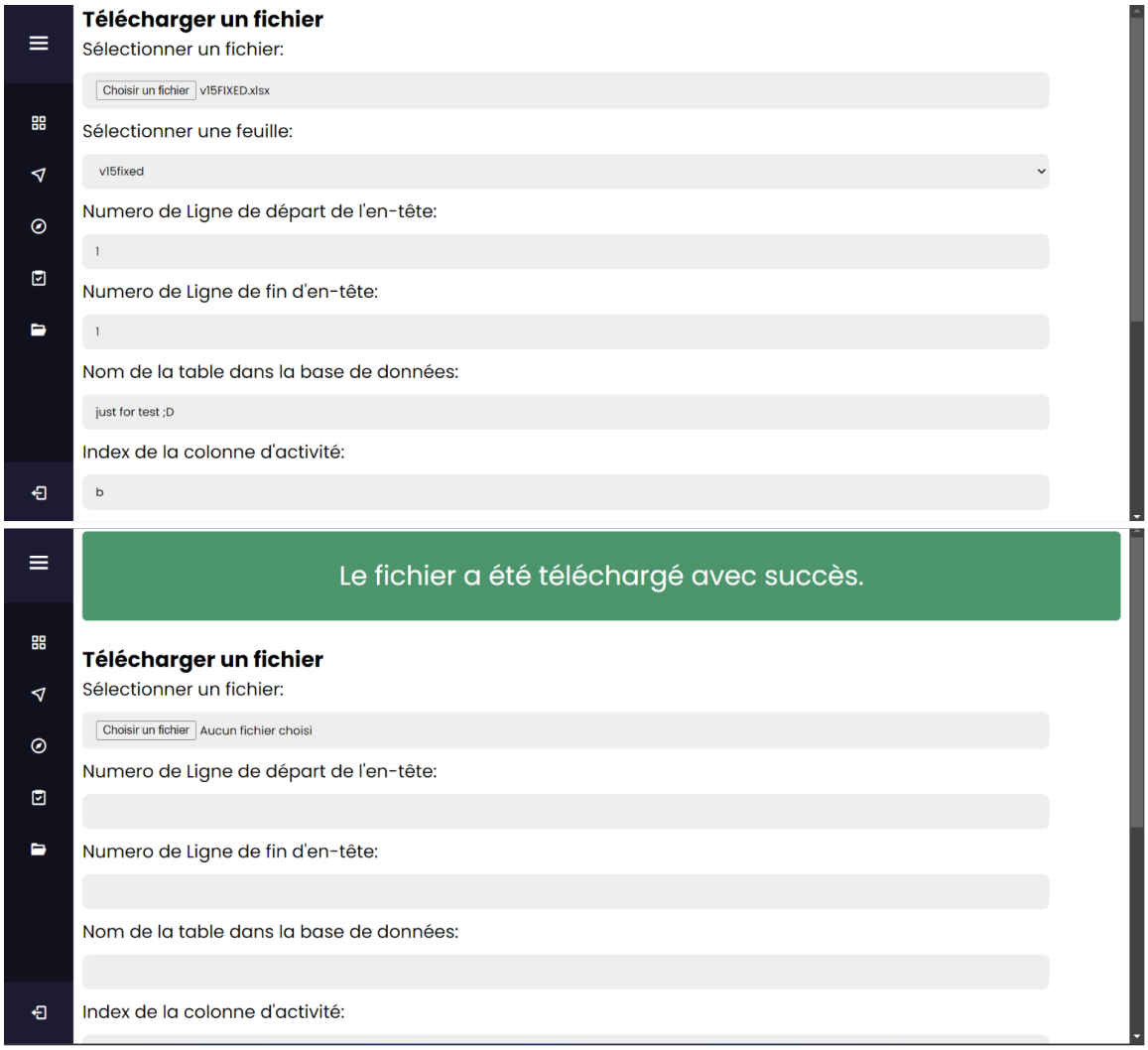

Figure 4.12: *Legacy Functionality: Excel File Upload Form*

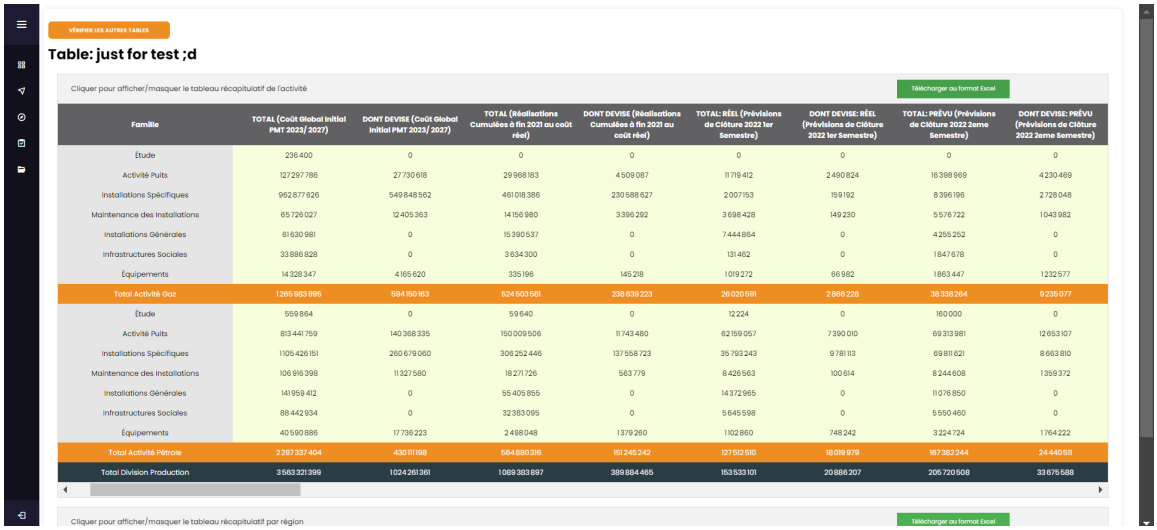

Figure 4.13: *Legacy Functionality: Recapitulative Tables of the Uploaded Excel File*[11]

[11] Resulting Recapitulative tables can be downloaded in xlsx format

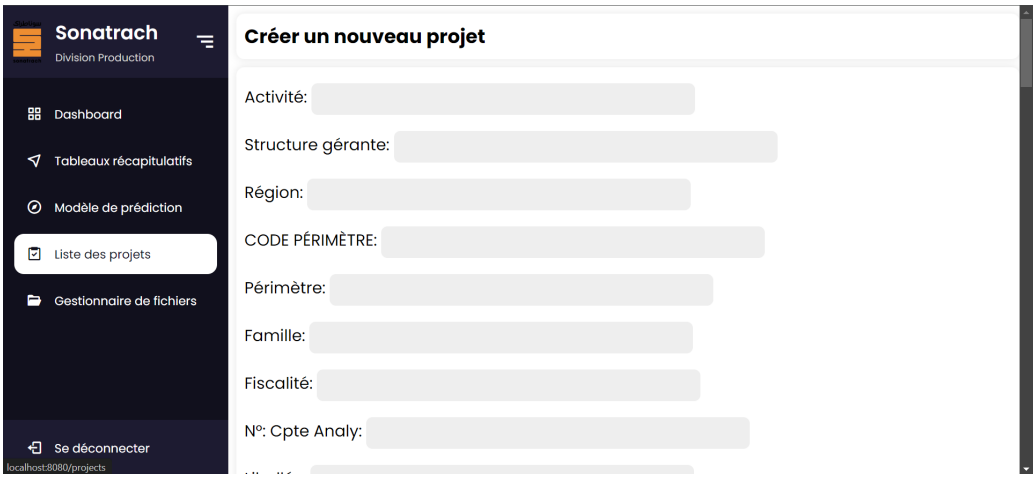

Figure 4.14: *Adding New Projects*

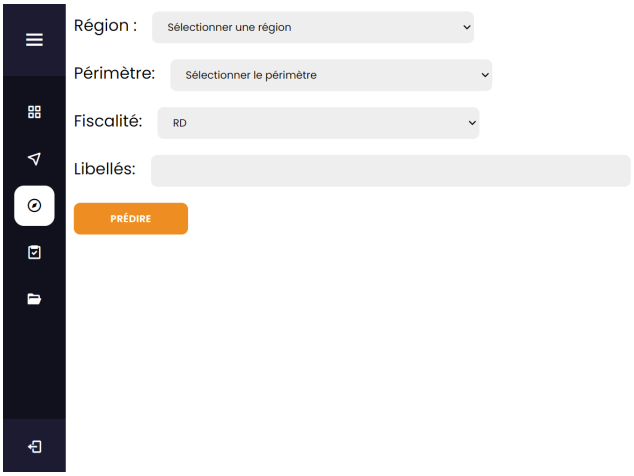

Figure 4.15: *Prediction Model Functionality*

Figure 4.16: *Administrator Dashboard: Users Accounts Control*

Figure 4.17: *Administrator Dashboard: Adding new Users*

4.6 Conclusion

Integrating BI and ML with backend and frontend components was complex but rewarding. Tableau enabled powerful visualizations, improving data-driven decisions. Python was key for developing robust machine-learning models, while Spring Boot ensured secure and efficient backend operations. Thymeleaf provided a responsive frontend.

These technologies were chosen for their flexibility, ease of use, and performance. Each component ensured seamless integration, handling complex operations, and delivering insightful visualizations. Deploying the ML model and Spring Boot application on different servers will enhance scalability and performance.

To further improve model performance, future research could explore:

- **Advanced Algorithms:** Investigating advanced machine learning algorithms like neural networks.
- **Class Imbalance:** Implementing oversampling or undersampling[12] to improve performance on minority classes.

For BI integration, future work could include:

- **Advanced Visualization Techniques:** Creating more detailed dashboards based on stakeholder feedback and other factors.
- **Migration to Tableau Server:** Considering Tableau Server for hosting visualizations to provide more control and local hosting options.

Additionally, to further enhance the user interface, more advanced JavaScript frameworks like React or Vue could be implemented. These frameworks would provide a more intuitive and dynamic interface, improving user experience and interaction with the application.

This implementation phase has laid a strong foundation for future improvements, creating a powerful and flexible system that meets current and future needs.

[12] Oversampling increases instances in the minority class, while undersampling reduces instances in the majority class.

General Conclusion

The subject matter of this Thesis has been focused on the modernization of **Sonatrach's** technology infrastructure in dealing with project information from its vast operations. In this regard, considering the need for creating a centralized, highly secured, and efficient system, we developed a comprehensive web platform through the infusion of state-of-the-art Business Intelligence and Machine Learning technologies.

We have already accomplished the integration of various technologies during the implementation phase of developing a scalable and robust system. We delivered dynamic visualizations to ease better decisions by using Tableau for BI. Python, with its rich libraries, enabled us to quickly design and deploy sophisticated ML models to better the system's effectiveness for categorizing and predicting projects. Spring Boot was run as the core of our back-end to ensure the security and easy integration of other components. We could design a responsive and intuitive user interface with Thymeleaf on the front end to ensure smooth user experiences.

Some of our highlight features include robustness and efficiency. The extensive data handling and maneuvering of complex queries are high-speed, thus allowing for timely insights into the data and an overall improvement in operational efficiency. The ML models integrated into the system assist in keeping the system up-to-date with the latest trends and data patterns, which allows it to be fine-tuned continuously with better project management strategies and techniques.

This is a learning curve for both of us team members in both academics and profession. We understood and learned a lot in integrating the different technological pieces into a unified whole. Specifically, these hurdles, from data cleaning to ensuring data security, provided us with adequate hands-on practice in solving problems and thinking critically. This project has also been very instrumental in our real-world training in software engineering and data analysis, while it lets us be efficient in project management skills and helps build our prowess in the tech field.

Finally, this project presents a forward-looking solution to the technological challenges of **Sonatrach**, It not only responds to current needs but also lays a strong foundation for future improvements. It would help **Sonatrach** retain its competitive edge in a changing energy marketplace, supporting continuous growth and success.

Bibliography

[1] Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Sixth Edition*. Project Management Institute, 2017.

[2] International Energy Agency. What is the petroleum industry? Retrieved from the International Energy Agency website.

[3] Harold Kerzner and Frank P. Saladis. *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. John Wiley & Sons, 12th edition, 2017.

[4] David I. Cleland and Lewis R. Ireland. *Project Management: Strategic Design and Implementation*. McGraw-Hill Education, 5th edition, 2007.

[5] John R. Schermerhorn, Paul Davidson, David Poole, Alan Simon, Peter Woods, and Stephanie L. Chau. *Management: Foundations and Applications*. John Wiley & Sons Australia, Ltd., 2nd asia-pacific edition, 2014.

[6] S. Chandrasekaran. *Petroleum Drilling and Production Regulation Handbook*. Springer, 2019.

[7] P. Eskerod and A. L. Jepsen. *Project Stakeholder Management*. Routledge, 2016.

[8] Timothy J. Kloppenborg and John A. Petrick. *Project Management Essential*. Sage Publications, 2016.

[9] M. Le Coz. *Oil and Gas Engineering Guide*. Editions Technip, 2017.

[10] S. Mokhatab, J. Y. Mak, and J. V. Valappil. *Handbook of Natural Gas Transmission and Processing*. Gulf Professional Publishing, 2019.

[11] ExxonMobil. Technical challenges in the petroleum industry, 2019.

[12] Schlumberger. Advanced technology adoption in oil and gas, 2020.

[13] BP. Environmental impact assessment guidelines for oil and gas projects, 2018.

[14] Shell. Sustainability initiatives in the petroleum industry, 2021.

[15] International Association of Oil & Gas Producers. Regulatory compliance standards in the oil and gas sector, 2020.

[16] Wood Mackenzie. Geopolitical considerations in petroleum projects, 2019.

[17] U.S. Energy Information Administration. Oil and gas price volatility: Trends and implications, 2021.

[18] Deloitte. Managing geopolitical risks in oil and gas investments, 2020.

[19] J. Smith, R. Johnson, and A. Williams. Advancements in risk management in the petroleum industry. *Journal of Petroleum Project Management*, 15(2):45–58, 2021.

[20] Peter Jones and Peter Taylor. *Project Management in Practice*. Routledge, 2018.

[21] David Hillson and Ruth Murray-Webster. *Understanding and Managing Risk Attitude*. Routledge, 2017.

[22] Julia Gebhardt, Katharina Hefter, Juliane Butters, Eva Kalteier, Matteo Coppola, Bernhard Gehra, Thomas Pfuhler, Jeanne Kwong Bickford, Abhinav Bansal, and Pierre Roussel. Mature risk management in uncertain times. *BCG Global ESG, Compliance, and Risk Report*, 2023. Accessed: 2024-04-06.

[23] Samantha Beavers. Three emerging trends in risk management. *Master of Management, Risk and Analytics*, 2022. Accessed: 2024-04-06.

[24] S. Jones and M. Brown. Innovations in resource allocation: A case study of the petroleum sector. *International Journal of Project Management*, 28(4):321–335, 2020.

[25] K. Peters and D. Miller. Compliance management strategies: Best practices for the petroleum industry. *Journal of Petroleum Compliance*, 10(1):12–27, 2022.

[26] S. Dixit and T. Daim. Mapping the technological landscape: Insights from technology roadmapping research. *Technological Forecasting and Social Change*, 146:127–140, 2019.

[27] Tom Mitchell. *Machine Learning*. McGraw Hill, 1997.

[28] Hsinchun Chen, Roger H.L. Chiang, and Veda C. Storey. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4):1165–1188, 2012.

[29] M. Janssen, Y. Charalabidis, and A. Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 32(4):63–71, 2015.

[30] A. Sircar et al. Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*, 6(4):379–391, 2021.

[31] Ianire Taboada, Abouzar Daneshpajouh, Nerea Toledo, and Tharaka de Vass. Artificial intelligence enabled project management: A systematic literature review. *Applied Sciences*, 13(8):5014, 2023.

[32] Yellowfin Team. What is data visualization and its importance in business intelligence? *Yellowfin*, 2023. Accessed: 2024-04-05.

[33] Tableau. What you need to know about bi dashboards. *Tableau*, 2023. Accessed: 2024-04-05.

[34] Hugh J. Watson and Barbara H. Wixom. The current state of business intelligence. *IEEE Computer Society*, 2007.

[35] SONATRACH Production Division. Document granted by Mr. Bouhidel Sofiane, 2024. Translated from French to English.

[36] Hadjer Bachasais and Imene Rebache. Conception et realisation d'une application web pour la gestion des contrats de la division production, 2023. Mémoire de Licence.

[37] Zakaria Khodri. Établissement des prévisions des coûts des projets d'investissements liés directement à la production, 2023. Mémoire de Master.

[38] Wikipedia contributors. Google Drive. https://en.wikipedia.org/wiki/Google_Drive. Accessed: May 21, 2024.

[39] Wikipedia contributors. Git. https://en.wikipedia.org/wiki/Git. Accessed: May 21, 2024.

[40] Wikipedia contributors. GitHub. https://en.wikipedia.org/wiki/GitHub. Accessed: May 21, 2024.

[41] Wikipedia contributors. Overleaf. https://en.wikipedia.org/wiki/Overleaf. Accessed: May 21, 2024.

[42] Tableau. Tableau prep. https://www.tableau.com/products/prep, 2023.

[43] Tableau. Tableau desktop. https://www.tableau.com/products/desktop, 2023.

[44] Tableau. Tableau bridge. https://www.tableau.com/products/bridge, 2023.

[45] Tableau. Tableau cloud. https://www.tableau.com/products/cloud-bi, 2023.

[46] MySQL. Mysql. https://www.mysql.com/, 2023.

[47] Tableau. *Publish a Data Source on the Web*. Tableau Software, 2024. Accessed: 2024-03-24.

[48] Tableau. *Keep Data Fresh Using Bridge*. Tableau Software, 2024. Accessed: 2024-03-27.

[49] JWT.io. *Introduction to JSON Web Tokens*. Auth0, 2024. Accessed: 2024-03-25.

[50] Python. https://www.python.org/. Accessed: 2024-04-02.

[51] Pycharm professional edition. https://www.jetbrains.com/pycharm/. Accessed: 2024-04-04.

[52] Fastapi. https://fastapi.tiangolo.com/. Accessed: 2024-04-04.

[53] The Pandas Development Team. Pandas: Python data analysis library, 2024. Accessed: 2024-04-05.

[54] The Scikit learn Development Team. Scikit-learn: Machine learning in python, 2024. Accessed: 2024-04-04.

[55] The SQLAlchemy Authors. Sqlalchemy documentation, 2024. Accessed: 2024-04-05.

[56] The Uvicorn Authors. Uvicorn: The lightning-fast asgi server, 2024. Accessed: 2024-04-05.

[57] Oracle Corporation. Java platform, standard edition documentation, 2024. Accessed: 2024-04-10.

[58] VMware, Inc. Spring boot reference documentation, 2024. Accessed: 2024-04-10.

[59] Postman, Inc. Postman learning center, 2024. Accessed: 2024-04-11.

[60] VMware, Inc. Spring tool suite 4 documentation, 2024. Accessed: 2024-04-10.

[61] JetBrains s.r.o. Intellij idea documentation, 2024. Accessed: 2024-04-10.

[62] Spring.io. Spring boot - dependency management, 2024. Accessed: 2024-04-17.

[63] VMware, Inc. Spring boot starter data jpa documentation, 2024. Accessed: 2024-04-17.

[64] VMware, Inc. Spring boot starter security documentation, 2024. Accessed: 2024-04-17.

[65] IBM Cloud Education. What is rest?, 2023. Accessed: 2024-04-19.

[66] Spring.io. Spring mvc framework, 2024. Accessed: 2024-04-19.

[67] VMware, Inc. Spring boot starter web documentation, 2024. Accessed: 2024-04-17.

[68] VMware, Inc. Spring boot starter webflux documentation, 2024. Accessed: 2024-04-17.

[69] Apache Software Foundation. Apache poi - the java api for microsoft documents, 2024. Accessed: 2024-04-18.

[70] Oracle Corporation. Mysql connector/j, 2024. Accessed: 2024-04-18.

[71] JSON.org. Json java library documentation, 2024. Accessed: 2024-04-18.

[72] Laurentiu Spilca. *Spring Security in Action*. Manning Publications, 2020.

[73] Thymeleaf. Thymeleaf documentation, 2024. Accessed: 2024-05-20.

[74] MDN Web Docs. Css: Cascading style sheets, 2024. Accessed: 2024-05-20.

[75] MDN Web Docs. Javascript, 2024. Accessed: 2024-05-20.

[76] New Relic. Imperative vs. declarative programming: How to choose, 2021. Accessed: 2024-05-02.

[77] Robert C. Martin. *Agile Software Development, Principles, Patterns, and Practices*. Prentice Hall, 2002.

[78] Spring Boot. spring-boot-starter-thymeleaf, 2024. Accessed: 2024-05-31.

[79] Thymeleaf. thymeleaf-extras-springsecurity6, 2024. Accessed: 2024-05-31.

Appendix

1. **Operations Management Overview**: There are three other departement within the Operations Management

   1.1 **Maintenance Department:** The Maintenance Department is responsible for the following tasks:

      1.1.1 Monitoring and consolidating the maintenance of rotating machinery and static equipment installed on production sites.

      1.1.2 Monitoring operational parameter readings and chemical analyses of oils to ensure better predictive maintenance.

      1.1.3 Implementation of the company's standardization strategy.

   1.2 **Exploitation Department:** The Exploitation Department is responsible for:

      1.2.1 Operational consolidation and daily monitoring of production, shipment, and pressure maintenance of upstream flows from the nine regional managements, the non-regional sector, and the sixteen groups.

      1.2.2 Daily, weekly, and ten-day reports and explanations of any deviations from the planned schedule.

      1.2.3 Analysis of variances by operating area in its own effort and in association.

      1.2.4 Annual company balance sheet by operating area, which includes the coherence of the four hydrocarbon chains (LPG, crude, gas, and condensate) from the well to the shipping pipe (internal or external sale).

   1.3 **Operations Assistance Department:** The Operation Assistance Department is responsible for:

      1.3.1 Monitoring maintenance of installations at well level, operations applied to wells, and contracts related to operations on wells for all regions.

2. **Machines specifications** During the length of our project, we have used two personal machines:

Machine	Specifications
Machine 1	CPU: Intel Core i7-8665U 1.90GHz-2.11 GHz GPU: Intel® UHD Graphics 620 RAM: 16GB OS: Windows 11 Pro 23H2
Machine 2	CPU: Intel Core i5-1135G7 2.40GHz-2.42 GHz GPU: NVIDIA T500 RAM: 16GB OS: Windows 11 Pro 23H2

Table 4.2: *Specifications of Machines*

3. **Programming Paradigms and Design Choices**

In designing our microservices architecture, we utilized both imperative and declarative paradigms[13] to optimize backend operations and interactions with external systems.

3.1 **Imperative Paradigm** We used the imperative programming paradigm to control the flow logic within the Spring Boot backend. This approach allows for explicit instructions and fine-grained control over backend processes such as user management, request processing, database operations, and security implementations.

3.2 **Declarative Paradigm** The declarative programming paradigm was employed for interactions with our BI and ML microservices. For the BI microservice, we integrated Tableau dashboards hosted on Tableau Cloud. For the ML microservice, we utilized a FastAPI application serving our machine learning model. The Spring Boot backend communicates with these services through API endpoints, focusing on what needs to be achieved rather than detailing how to achieve it.

[13] Imperative programming is a paradigm that uses statements to change a program's state, focusing on how to perform tasks with explicit instructions. Declarative programming, on the other hand, focuses on what the outcome should be, describing the logic without explicitly listing control flow steps [76].

4. Backend Architecture

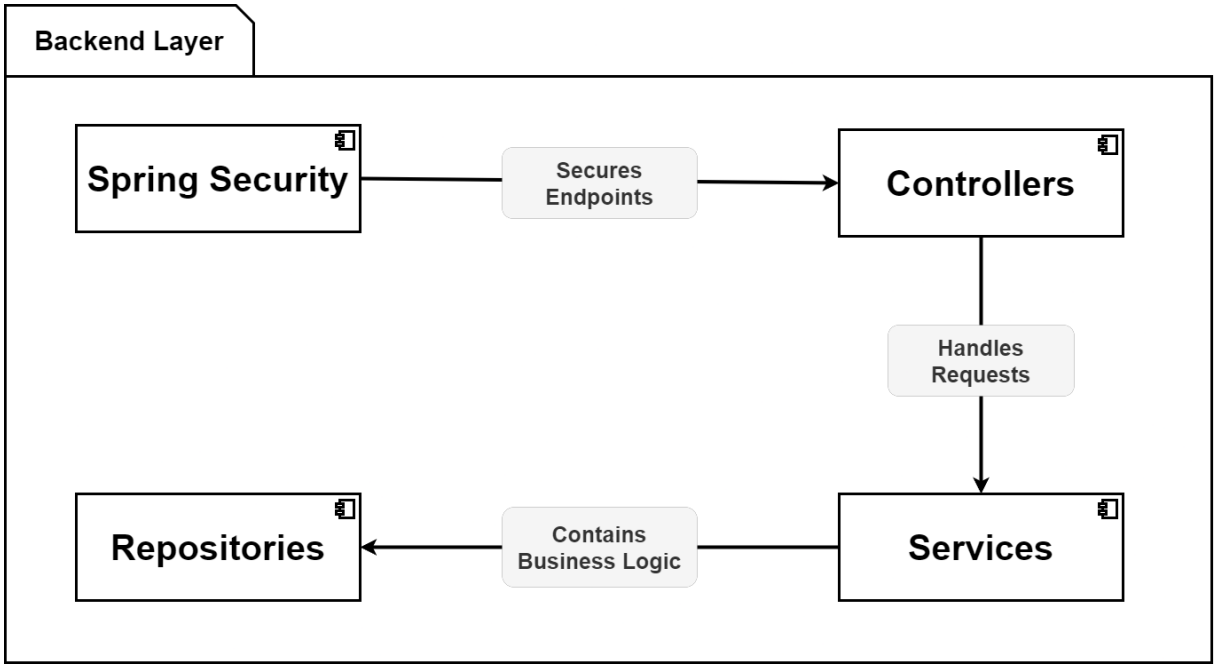

Figure 4.18: *Back-end Architecture*

5. Frontend-Backend Interactions

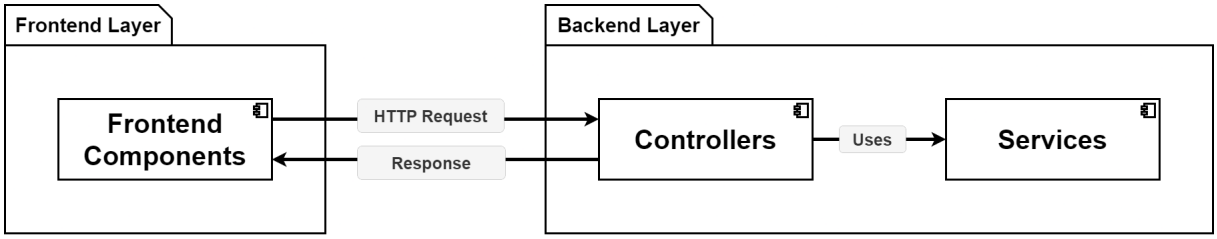

Figure 4.19: *Frontend-Backend Interaction*

6. Backend-Database Interactions

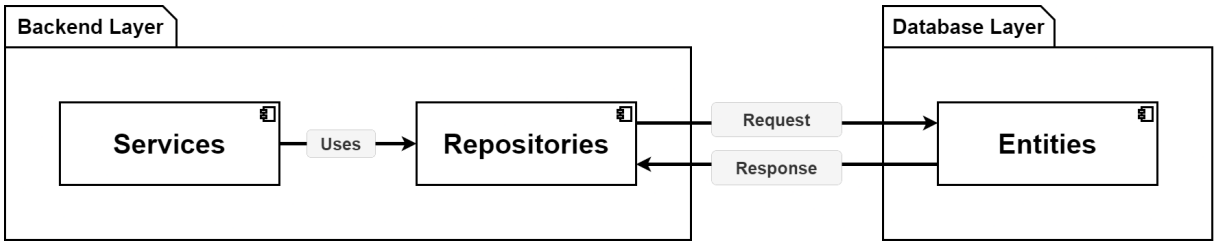

Figure 4.20: *Backend-Database Interaction*

7. Tableau Dashboard Retrieval Process

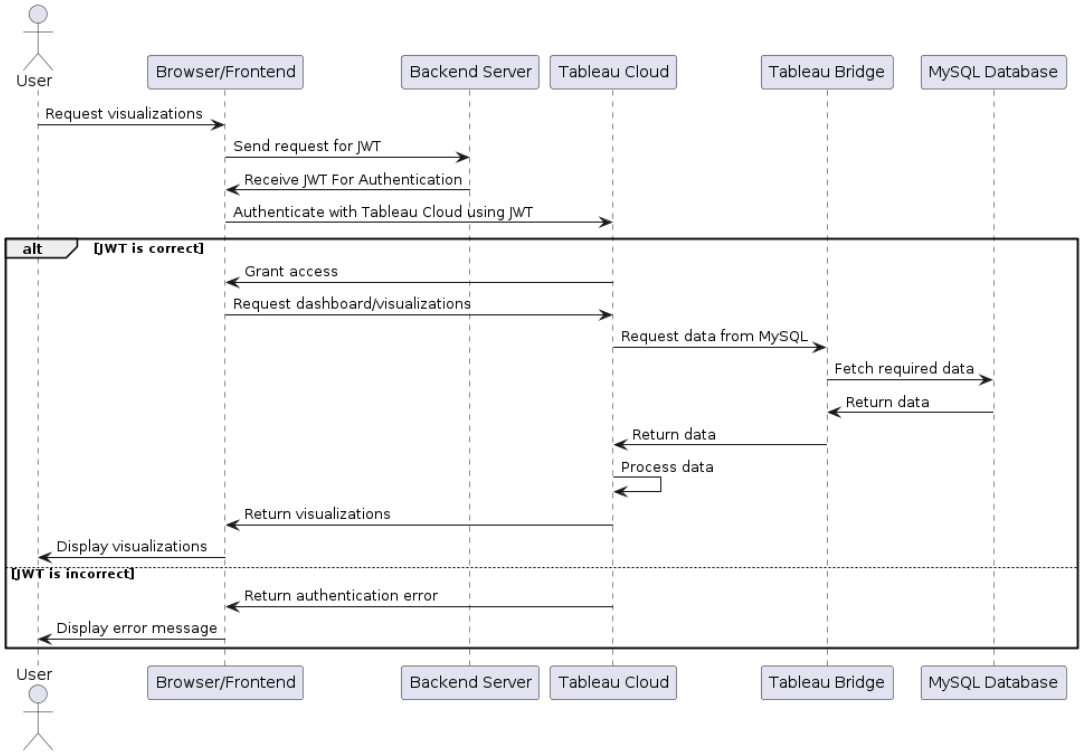

Figure 4.21: *Tableau Dashboard Retrieval Process*

8. File Upload Process

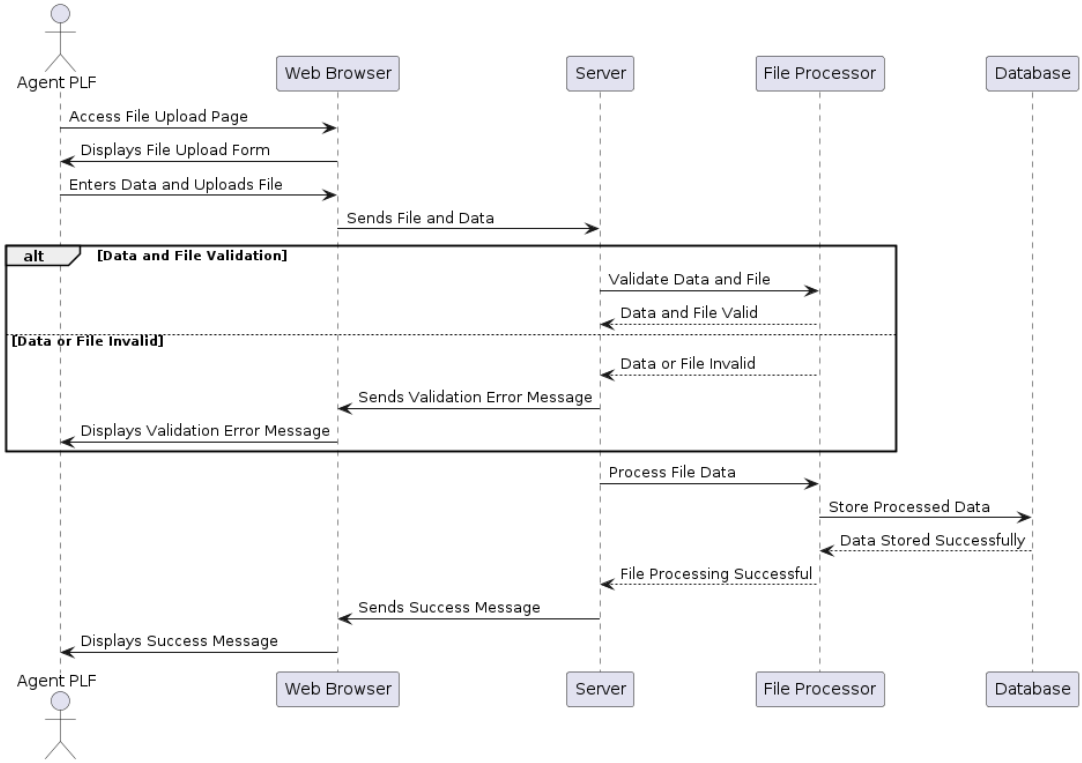

Figure 4.22: *File Upload Process*

9. Prediction Process

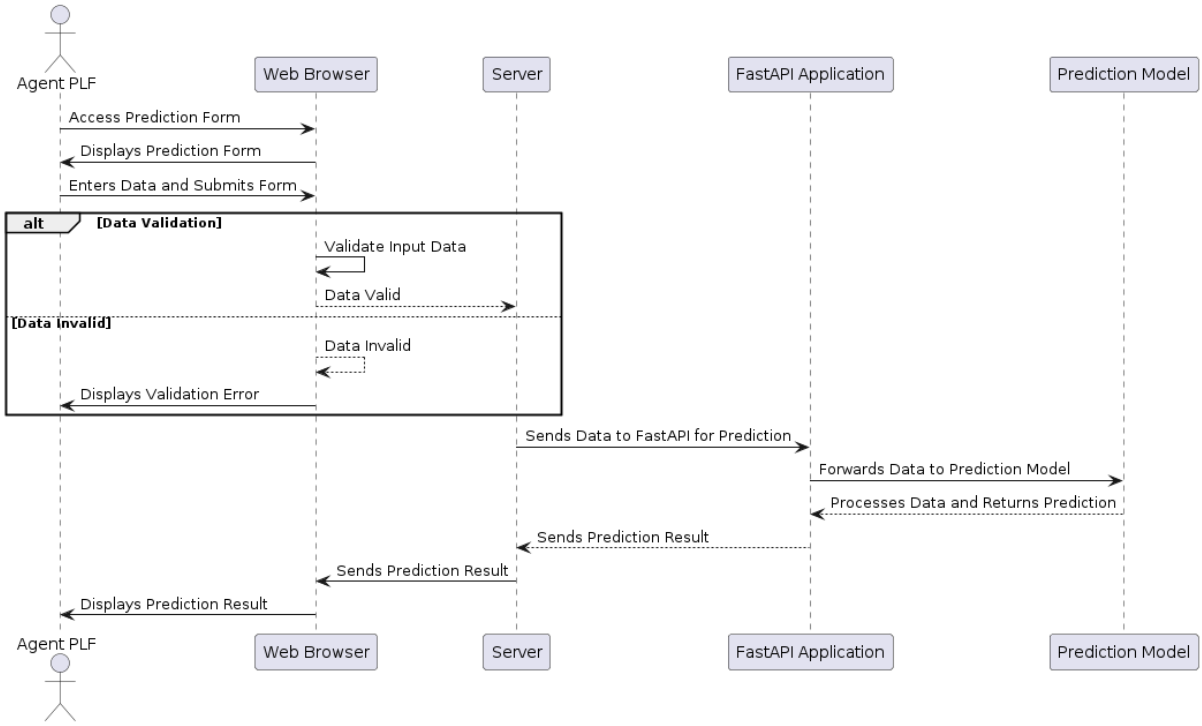

Figure 4.23: *Prediction Process*

10. The handed Excel file containing Projects data

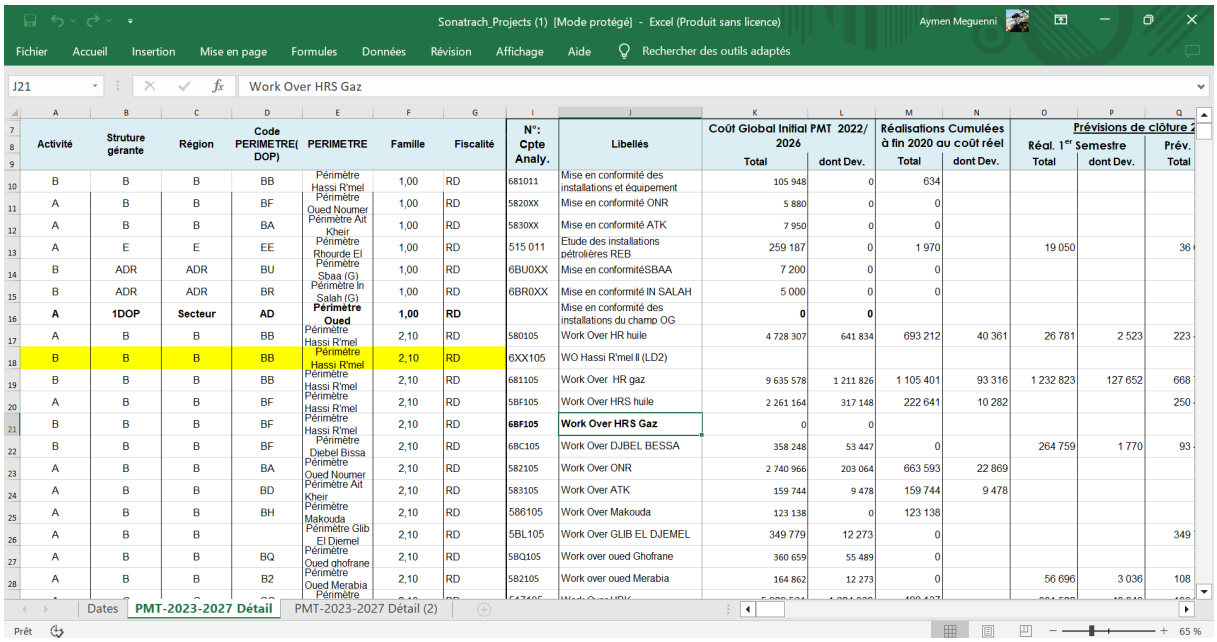

Figure 4.24: *Screenshot of the main Excel file provided of the medium-term plan projects*

11. "Projects" table in MySQL Workbench

project_id	Activité	Structure gérante	Région	CODE PÉRIMÈTRE	Périmètre	Famille	Fiscalité	Numero Compte Analytique	Libellés
1	Gaz	Hassi R'Mel	Hassi R'Mel	BB	Périmètre Hassi R'mel (P+G)	Étude	RD	681011	Mise en conformité des installations et équipe
2	Pétrole	Hassi R'Mel	Hassi R'Mel	BF	Périmètre Oued Noumer	Étude	RD	5820XX	Mise en conformité ONR
3	Pétrole	Hassi R'Mel	Hassi R'Mel	BA	Périmètre Ait Kheir	Étude	RD	5830XX	Mise en conformité ATK
4	Pétrole	Rhourd El-Baguel	Rhourd El-Baguel	EE	Périmètre Rhourde El Baguel	Étude	RD	515011	Etude des installations pétrolières REB
5	Gaz	Adrar	Adrar	BU	Périmètre Sbaa (G)	Étude	RD	6BU0XX	Mise en conformitéSBAA
6	Gaz	Adrar	Adrar	BR	Périmètre In Salah (G)	Étude	RD	6BR0XX	Mise en conformité IN SALAH
7	Pétrole	DOP (Secteur hors région)	Hors région	AD	Périmètre Oued Gueterini	Étude	RD	NULL	Mise en conformité des installations du champ
8	Pétrole	Hassi R'Mel	Hassi R'Mel	BB	Périmètre Hassi R'mel (P+G)	Work Over / Side Track	RD	580105	Work Over HR huile
9	Gaz	Hassi R'Mel	Hassi R'Mel	BB	Périmètre Hassi R'mel (P+G)	Work Over / Side Track	RD	6XX105	WO Hassi R'mel II (LD2)
10	Gaz	Hassi R'Mel	Hassi R'Mel	BB	Périmètre Hassi R'mel (P+G)	Work Over / Side Track	RD	681105	Work Over HR gaz

projects 1 ×

Apply

SQL Editor closed

Figure 4.25: *Snapshot of the resulting "projects" table in MySQL Workbench*

12. Map of projects distribution

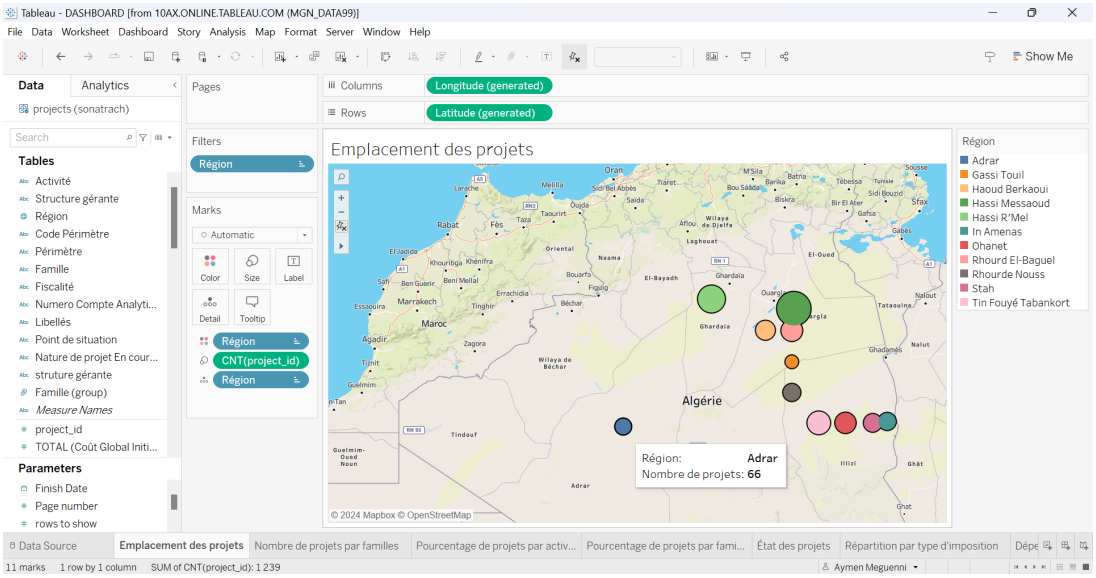

Figure 4.26: *Developement of one of the Visualization elements: Map of projects distribution*

13. Tableau Cloud interface

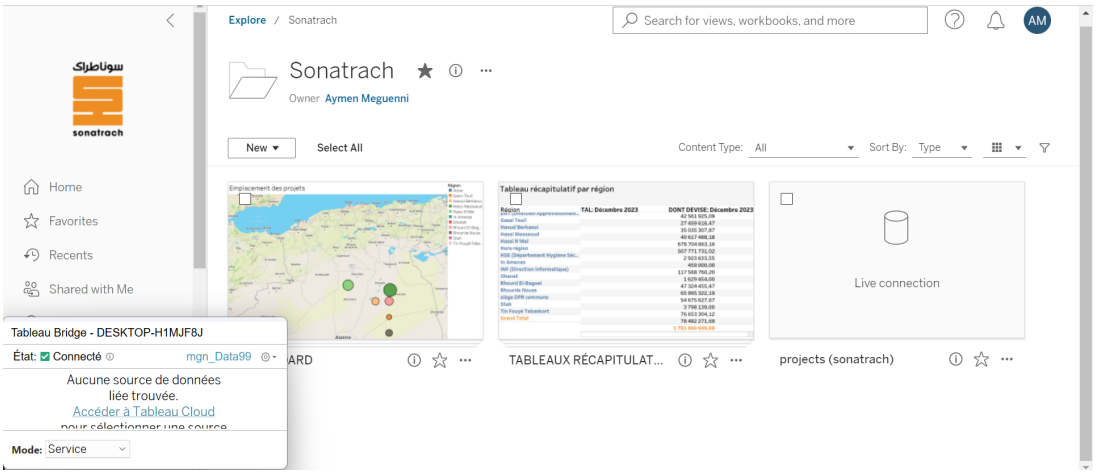

Figure 4.27: *Workbooks uploaded to Tableau Cloud using a live connection to local data source*

14. Generated Fields in Tableau Cloud Settings

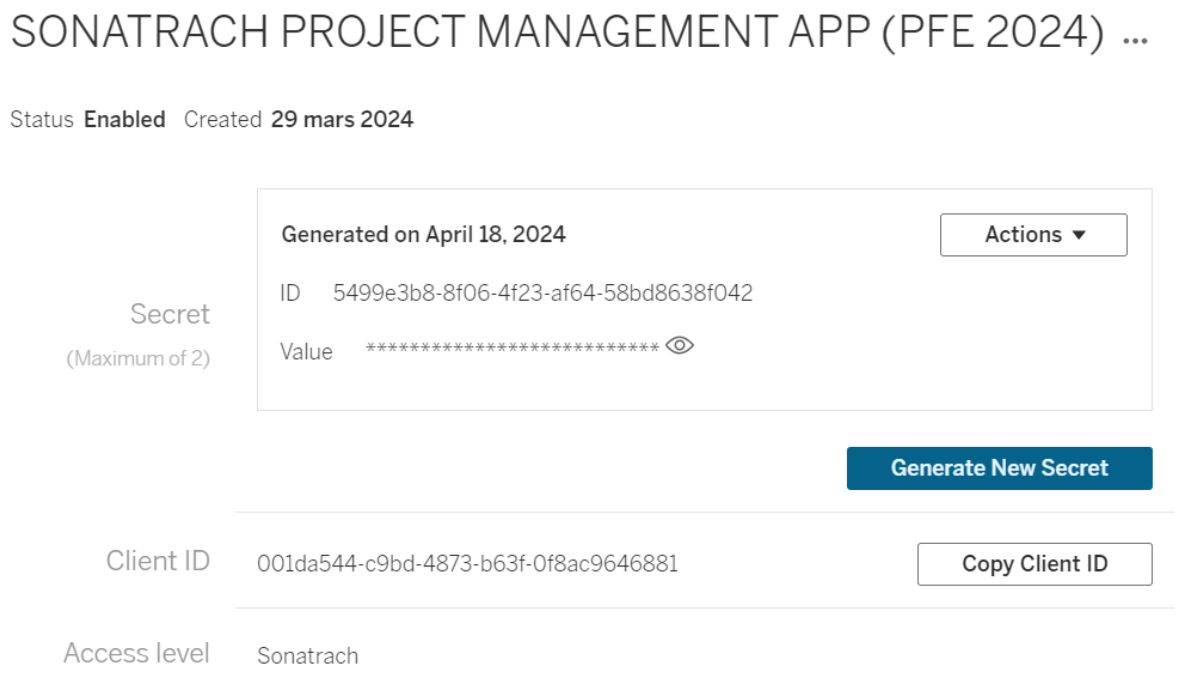

Figure 4.28: *Connected App settings page in Tableau Cloud*

15. Connection with database in Python code

```
# Info
user = 'root'
password = \\Mysql server password\\
host = 'localhost'
port = '3306'  # Default MySQL
database = 'sonatrach'
# Create connection with the database
engine = create_engine(f'mysql+mysqlconnector://{user}:{password}@{
    host}:{port}/{database}')
```

16. ML implementation project in Pycharm

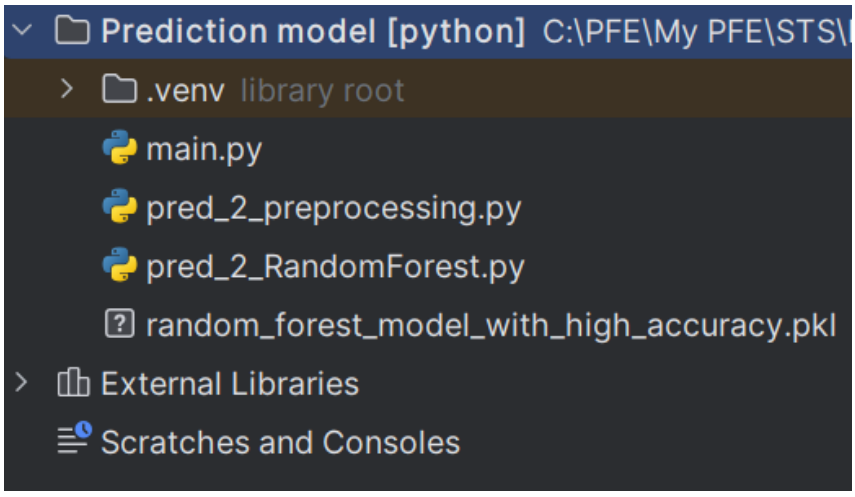

Figure 4.29: *Structure of the ML implementation project with resulting Model*

17. starting the FastApi Application

Figure 4.30: *Uvicorn server is running and listening for requests on all available IP addresses at port 8000*

18. Used Engineering Model

We adopted a hybrid model combining Waterfall and Agile methodologies to maximize efficiency and ensure timely delivery[14].

Initially, we used the Waterfall model for 15 weeks, focusing on design, development, testing, and deployment preparation. This ensured strict adherence to deadlines and a clear project roadmap.

In the development phase, we transitioned to Agile, allowing flexibility and responsiveness to feedback. Iterative cycles and weekly feedback sessions with supervisors provided valuable insights for continuous refinement.

The hybrid model balanced the need for initial structure with the flexibility required during development, helping us meet deadlines and align the final product with stakeholder expectations.

19. Data Base Connection Settings

```
spring.datasource.url=jdbc:mysql:/localhost:3306/sonatrach
spring.datasource.username=root
spring.datasource.password= // MySQL server password here :) //
spring.jpa.hibernate.ddl-auto=update //to automatically update the//
          //database schema based on models package entity classes//
```

[14] A software engineering model is a structured approach used to plan, develop, and manage software projects. It includes methodologies and practices that guide the software development process, ensuring that the final product meets the required standards and specifications. Common models include Waterfall, Agile, and Hybrid, each offering distinct advantages based on project needs.

20. Backend Code Structure

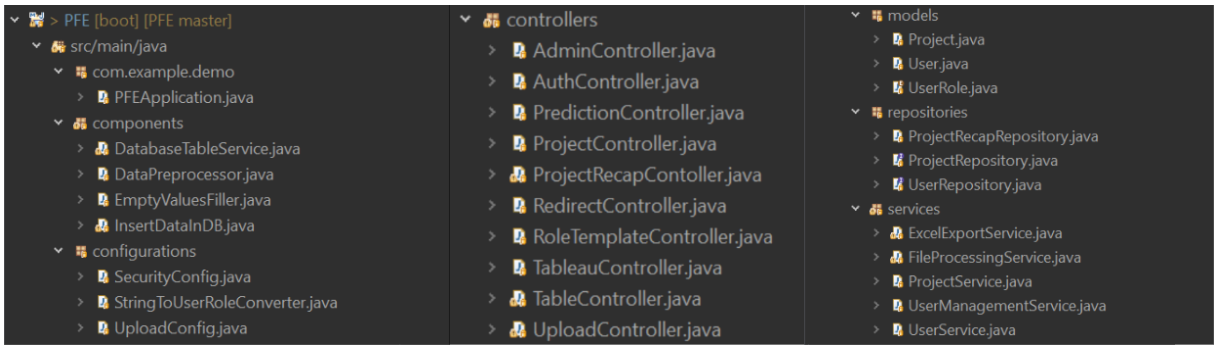

Figure 4.31: *Complete Backend Structure in Spring Tool Suite*

21. Code of the Main Class

```
package com.example.demo;

import org.springframework.boot.SpringApplication;
import org.springframework.boot.autoconfigure.SpringBootApplication;
import org.springframework.boot.autoconfigure.domain.EntityScan;
import org.springframework.context.annotation.Bean;
import org.springframework.data.jpa.repository.config.
    EnableJpaRepositories;
import org.springframework.web.client.RestTemplate;

@EntityScan(basePackages = {"models"})
@EnableJpaRepositories(basePackages = {"repositories"})

@SpringBootApplication(scanBasePackages = {"com.example.demo", "
    repositories","controllers","models","services","configurations",
    "components"})
public class PFEApplication {

  public static void main(String[] args) {
    SpringApplication.run(PFEApplication.class, args);
  }

     @Bean      //used for the Tableau Controller//
    RestTemplate restTemplate() {
          return new RestTemplate();
      }
}
```

22. Followed Development Principles

The project structure follows the Single Responsibility Principle (SRP), which is a fundamental aspect of the SOLID[15] principles in object-oriented programming. SRP states that a class should have only one reason to change, meaning it should have only one job or responsibility. By structuring the project into separate, independent layers (models, repositories, services, controllers, configurations, components), each package and class has a distinct responsibility, which improves code readability, maintainability, and testability. This separation of concerns also facilitates easier

[15]SOLID is an acronym for five design principles intended to make software designs more understandable, flexible, and maintainable. The principles are Single Responsibility, Open/Closed, Liskov Substitution, Interface Segregation, and Dependency Inversion [77].

debugging and modification since changes in one layer do not directly affect the others.

23. User Class Code

```java
@Entity
@Table(name = "users")
public class User implements UserDetails {

    @Id
    @GeneratedValue(strategy = GenerationType.IDENTITY)
    @Column(name = "USER_ID")
    private Long id;

    @Column(name = "username", nullable = false, unique = true)
    private String username;

    @Column(name = "password", nullable = false)
    private String password;

    @Enumerated(EnumType.STRING)
    @Column(name = "role", nullable = false)
    private UserRole role;

    // Constructors, getters, and setters
}
```

24. Tableau Cloud Connection Settings

```
tableau.online.username=sonatrachprojects@gmail.com
tableau.online.client-id=001da544-c9bd-4873-b63f-0f8ac9646881
tableau.online.secret-id= //secret id value//
tableau.online.client-secret= //client secret value//
```

25. Key Classes for the Legacy functionality in STS

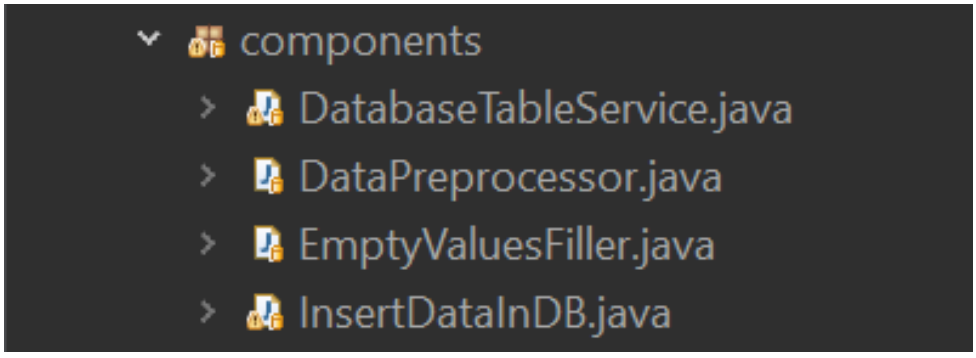

Figure 4.32: *Key Classes for the Legacy functionality*

26. Frontend Code Structure

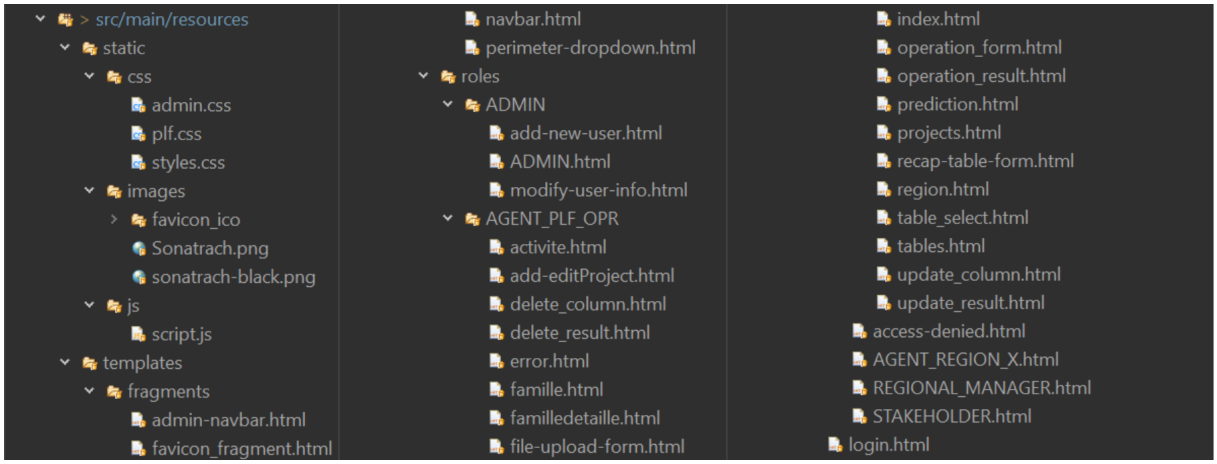

Figure 4.33: *Complete Frontend Side Code Files Structure*

27. Notable Dependencies for The Frontend

- **spring-boot-starter-thymeleaf**: This is a starter for building MVC web applications using Thymeleaf views with Spring Boot. It simplifies the integration of Thymeleaf as the view layer in a Spring Boot application, providing defaults for rapid development and enabling the efficient generation of HTML content from server-side Java code [78].
- **thymeleaf-extras-springsecurity6**: This library provides integration between Thymeleaf and Spring Security, facilitating the creation of secure web applications. It allows the use of Spring Security features directly in Thymeleaf templates, such as conditionally displaying content based on user roles and permissions, enhancing the security and user experience of the application [79].