MASTERS IN MANAGEMENT (MIM)

Masters Final Work

DISSERTATION

MACHINE LEARNING TECHNOLOGIES IN THE GREENTECH INDUSTRY: A CASE STUDY

CARLO RENZETTI

MARCH - 2024

MASTERS IN MANAGEMENT (MIM)

Masters Final Work

DISSERTATION

MACHINE LEARNING TECHNOLOGIES IN THE GREENTECH INDUSTRY: A CASE STUDY

CARLO RENZETTI

SUPERVISOR: PROF. ARTUR MANUEL BARROS DA CUNHA

JURY:
PRESIDENT: PROF. JOSÉ MANUEL CRISTÓVÃO VERÍSSIMO
RAPPORTEUR: PROF. JESUALDO CERQUEIRA FERNANDES
SUPERVISOR: PROF. ARTUR MANUEL BARROS DA CUNHA

MARCH - 2024

ACKNOWLEDGMENTS

I want to use this section to thank all the people who were involved during the completion of this dissertation and more generally in my master's studies journey:

I will start with my supervisor, Arthur Cunha, who supported me and followed me in writing this document, even during tough times and despite his many commitments.

I would also like to thank my team at Company X, it was a pleasure to work with you and to learn new things every day thanks to your expertise. In particular, a special thank you to the Machine Learning Engineer who sacrificed his time for the interview. The same applies to Company Y's CTO. I wish I could mention you, but that would ruin the whole purpose of the anonymity

The most classical acknowledgments: my girlfriend Martina and my family who sent me their support from far away and wasted time listening to my complaints.

It doesn't matter if I met you in Lisbon or Berlin or Italy or elsewhere, I want to say thanks to everyone I can consider and call "friend". You shaped my master's journey and gifted me with enjoyable memories.

Lastly, I want to share an appreciation for the cities that hosted me during these last two years: Lisbon and Berlin. Two cities opposite to each other but each made me fall in love with it in different ways. They strongly influenced me and my future.

I will now face my next steps with energy and hope, my two traits that I wish I would never lose.

This master's dissertation, titled " Machine Learning Technologies in the Greentech Industry: A Case Study" authored by Carlo Renzetti, is subject to an embargo due to the inclusion of confidential information from X GmbH & Y GmbH. The embargo period for this dissertation extends for 10 years from the submission date, starting on October 15, 2023.

During the embargo period, access to the full content of this dissertation will be restricted and not available for public access or distribution. The purpose of this embargo is to protect confidential and proprietary information belonging to X GmbH & Y GmbH, as disclosed in the document.

Authorized individuals, including the members of the master's jury, may have access to the restricted content for academic purposes.

After the embargo period ends, this dissertation will become accessible to the public through the ISEG library.

ABSTRACT

As sustainability gains popularity and becomes more and more relevant in today's economy, so does Artificial Intelligence and more specifically Machine Learning. New climate challenges arise, and the environment's health declines, stimulating new solutions and new approaches to be put into effect. This is what resides at the core of the GreenTech industry, which takes into consideration different technologies and tries to leverage them to challenge sustainability issues. Artificial intelligence and Machine Learning could play an important role in achieving this goal.

The following work aims to highlight the importance that a Machine Learning approach can have on the GreenTech industry. This dissertation elaborates on the point of contact between these two trends, using a real-life business case as an example. The case revolves around an ML consulting agency and its client, a BlueTech company that produces data products regarding the seas' health status. The company object of scrutiny operates in the context of the Sustainable Development Goals (SDGs) as it aligns with goals 13,14 and 17. The analysis of the case shines a light on the possibility of improving an early-stage BlueTech company with specialized Machine Learning consulting and implementation. To conclude his work the author reflects on the results of the case and the conflicting effects of the application of AI and ML. The outcome of the case study is positive as ML solutions improve the efficiency and scalability for the client. At the same time, the application of technologies such as AI and ML requires a lot of energy consumption, and it is therefore polluting. The duality of those technologies makes them beneficial and harmful to the environment at the same time.

Keywords: Artificial Intelligence, GreenTech, Machine Learning, Sustainability

Abbreviations and Acronyms

- Artificial Intelligence (AI)
- Carbon Dioxide (CO2)
- Chief Technology Officer (CTO)
- Cloud Computing (CC)
- Greenhouse Gas (GHG)
- Machine Learning (ML)
- Machine Learning Engineer (MLE)
- Machine Learning Operations (MLOps)
- Master Final Work (MFW)
- Sustainable Development Goals (SDGs)
- Underwater Hyperspectral Imaging (UHI)

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

1.1. Theoretical Background

Much has changed since the day Alan Turing thought of an intelligent machine, the so-called "Universal Turing Machine" (Copeland & Proudfoot, 2007). It was 1936 and Turing's abstract machine consisted of a scanner and a limitless memory-tape that moves back and forward past the scanner; the scanner reads the symbols on the tape andwrites further symbols (Copeland & Proudfoot, 2007). That was the first time that the concept of Intelligent Machine (or Artificial Intelligence) was conceived, and the foundations of modern technology were grounded. Successively, Machine Learning wasborn as a stream in Artificial Intelligence, which is gaining popularity in recent times in the field of computing and data analysis that will make applications behave intelligently (Jhaveri et al., 2022). ML is essentially about one approach, making machines that can learn to perform tasks (Rebala et al., 2019) and one characteristic thatmade this technology so popular in recent years is the fact that ML techniques are being deployed to solve real-world problems in daily lives providing insights into structures and patterns within large datasets, creating models by learning from existing datasets to predict or forecast outcomes or behavior (Rebala et al., 2019). The GreenTech sector is involved in solving real-world problems as well, and this dissertation will show through an empirical business case, the point of contact between those two industries.

1.2. Relevance of the study

Sustainability is becoming a predominant topic in most industries worldwide and the most forward-thinking companies are now considering it and leveraging it as a competitive advantage (see Chapter 2.4). As new challenges arise, new solutions and new approaches arise as well, generating new markets and successful companies. This work aims to analyze a specific business case hoping to shine light on the progress thatis being made towards a sustainable future. This study tries to show how it is possible to leverage a newborn technology to solve a recently highlighted problem and also to elaborate on the importance of the consulting industry and its approach, which can really make a difference to small/medium enterprises. The protagonist of the case is a company that works in the BlueTech niche market and whose values align with the well-known Sustainable Development Goals (SDGs). In particular, the SDGs involved in the case are Life Below Water (n.14), Climate Action (n.13), and Partnerships for the Goals (n.17). Considering what was said above, this dissertation is most relevant to workers in the Greentech or Machine Learning sectors that want to take inspiration from the case, but also to whoever is interested in knowing more about thosetwo industries. Lastly, the document can be beneficial to any person looking for practical cases of SDGs application, especially regarding the goals mentioned above.

1.3. Goals of the Dissertation

The idea at the core of this Master's Final Work is to analyze a current real-life trend using empirical evidence, theoretical literature, and personal experience.Considering the different types of Master Final Work available, the most fit option to reach the goals is the Dissertation which allows the author to write about real-world business cases following an established structure.

This work aims to fulfill different goals:

1. Acknowledge the existence of the Greentech industry as a current prevailing trend and the presence of applied ML in it. This objective involves proving the success of both industries individually but also of the point of contact between the two.

2. Showcase a real-life Machine Learning application in a GreenTech company. The business case analysis aims to strengthen the other goals of the dissertation and it represents the core of the document. Conclusions and generalizations revolve around the business case.

3. Prove the possibility of applying ML as a solution to Greentech problems, thanks to its flexibility and adaptability. It is not possible to ensure the absolute success of Machine Learning solutions in the GreenTech industry through a single case study, but it is feasible to prove that the implementation of ML can have positive outcomes and improve GreenTech companies, as it has happened at least once before.

4. Drawing more general conclusions and generating inputs for a reflection on the topic. The last paragraph aims for this target, developing a thought process that goes beyond the borders of the single business case. The limits of the business case will have to be taken into consideration, to address a wider discussion point.

1.4. *Dissertation's structure*

This document is divided into five main chapters, each of which is further divided into logical subsections. Following the Introduction, the document also includes a Literature Review, the Methodology explanation, the Case Analysis, and lastly the Conclusions chapter. The Introduction is divided into a few subsections that contain a first contextualization of the dissertation topic, the goals of the dissertation, and this same paragraph about the structure of the dissertation. The Literature Review responds to the necessity of diving deeper into the contextualization of the overarching subjects involved: Machine Learning, the Green Economy, and the GreenTech. Each one of those covers at leastone subsection. The Methodology is a short chapter that elaborates on the two different methods used to recover the necessary information to write this Master Final Work, the Document Analysis and the Expert Elicitation methods. Diving deeper into the core of the dissertation, the Case Analysis presents the main players in the business case which arethe consulting company X and the GreenTech start-up Y. Continuing with the analysis, It was highlighted what were the main issues for the client (in this case Y) and what approach the consulting company (X) decided to take to face those. Lastly, the Conclusion includes topics such as the output overview, the implementation plan, and the parallel between the expectations and the actual results. It was also inserted a paragraph stating the author's opinion and personal thoughts coming from a careful review of the study, along with some retrospective improvements on how the processes could have been better.

CHAPTER 2 - LITERATURE REVIEW

2.1. The structure

The literature review constitutes an essential element of this dissertation, aiming to explore existing scholarly works and research relevant to the focal point of the study. By diving into the accumulated knowledge on the topic, this review seeks to contextualize the research within the wider academic discourse and identify areas where new contributions can be made.

The literature can be categorized into distinct thematic strands, each representing a pivotal aspect of my MFW. One prevalent theme revolves around Artificial Intelligence and Machine Learning where studies have explored the differences betweenthe two. After a brief introduction to the climate crisis and green topics, where academic literature is certainly not lacking, the review transitions to sustainability as it moves towards its relationship with ML: the point of connection is the Greentech industry. To conclude the Literature Review, it is fundamental to explain what Greentech is and why this upcoming sector is getting more and more popular. The aim of the Literature Review chapter is to set the basis for the analysis, explaining the main factors at playand how they have been researched before, facilitating a more detailed exploration later on. As the work intends to study a real and concrete case of ML implementation in the Greentech industry, it is important to be sure that the roots of the topic are well set, and that the assumptions made are proven by previous work.

2.2. Artificial Intelligence

To build this study it is useful to start from the foundations, that is defining Artificial Intelligence and its goal in today's world. As Stanford professor John McCarty said in 1955, Artificial Intelligence can be defined as the science and engineering of making intelligent machines, especially intelligent computer programs (McCarthy, 2007). AI was also defined as a science and a set of computational technologies that are inspired by the ways people use their nervous systems and bodies to sense, learn, reason, andtake action (Monett et al., 2020). The world of AI is very broad, and it can be applied to many different sectors and situations: this versatility is one of the characteristics that makes this technology one of the most innovative and popular solutions on the market. The past decade has witnessed the fast-paced development of artificial intelligence (AI) in solving longstanding problems, and AI has played an indispensable role in profoundly transforming business, transportation, finance, and healthcare, to name but a few (Zhang et al., 2022). Each application of AI in a specific sector heavily influences other sub-sectors, such as the most known FinTech, HealthTech, and GreenTech. In addition to its versatility, some of the major advantages of AI are the automation of processes, the efficiency of data analysis, and problem-solving (Chhaya et al., 2020). Nowadays, AI is pervasively used in a broad array of applications across a variety of areas to benefit the whole of society, such as recommendation systems, fraud detection, autonomous driving, social media analysis, and business analytics (Zhang et al., 2022). Success from AI adoption can take the form of enhanced work outcomes, better technological tools, improved employee behavior, and more efficient results (Ramachandran et al., 2022).

2.3. Machine Learning

As we find ourselves in the post-industry 4.0, most processes are now digitalized, and the importance of data is evident to everyone. Due to its high operational and strategic potential, notably in generating business value, “big data” has recently become the focus of academic and corporate investigation (Fosso Wamba et al., 2015). This is where Machine Learning, a specific branch of AI, shines and offers companies added value. To better define it, Machine learning (ML) is a field of computer science that studies algorithms and techniques for automating solutions to complex problems that are hard to program using conventional programming methods (Rebala et al., 2019). In shorter terms, Machine learning is the study of computational methods to automate the process of knowledge acquisition from examples (Bose & Mahapatra, 2001). The terms AI and ML are often used interchangeably, so it is not always easy to distinguish and define what kind of implementation is being done, but what is clear is that ML is a subfield of AI (Rebala et al., 2019). Artificial intelligence (AI) is a much broader field of study than machine learning (ML). AI is all about making machines intelligent using multiple approaches, whereas ML is essentially about one approach – making machines that can learn to perform tasks (Rebala et al., 2019). Machine learning techniques play a preponderant role in dealing with massive amounts of data and are employed in almost every possible domain. Building a high-quality machine learning model to be deployed in production is a challenging task, for both, the subject matter experts and the machine learning practitioners (Quemy, 2020). Some of the major advantages of ML involve prediction making, continuous improvements, and easy identification of trends and patterns (Chhaya et al., 2020).

In recent years AI, and especially ML, have gained a lot of popularity, becoming one of the most upcoming technologies. This is because technology has broken the human efficiency bottleneck, decreased repetitious labor, and increased work efficiency (Ramachandran et al., 2022). Artificial Intelligence also improves staff retention and supports the acquisition of new customers, improving employee experience, increasing productivity, and assisting HR professionals in becoming more informed advisors (Ramachandran et al., 2022). Given the numerous benefits outlined, it is safe to say that the integration of AI and ML into business operations to interact with workers, get feedback, and utilize that information to develop a meaningful employee engagement plan will pay off in the long run (Ramachandran et al., 2022).

For a broader adoption and scalability of machine learning systems, the construction and configuration of machine learning workflow need to gain in automation (Quemy, 2020). That being said, ML algorithms solve the problems in an indirect way, by first generating a model based on processing the dataset and then predicting the label of a new input data point by executing that model (Rebala et al., 2019). This way ML algorithms tend to be more accurate than human-created rules since ML algorithms will consider all data points in adataset without any human bias due to prior knowledge (Rebala et al., 2019), ML is nowbeing applied to many industries including GreenTech, as in the business case object of the study.

2.4. Sustainability and Green Economy

To reach the goal of the dissertation a predominant topic to mention is Sustainability and the Green sector. The concept of sustainability is strongly linked to the equally famous sustainable development, such that the terms are often used as synonyms, even in the academic and scientific fields (Ruggerio, 2021).

As of today, there is an ongoing debate about the contradictions that come from the sustainable development definition; because of the natural limitations of our planet, resources, and therefore the economy, some scholars challenge the reality of a sustainable development solution, and others offer different approaches such as the degrowth and "buen vivir" (Ruggerio, 2021). For the sake of the study, the author will side with the school of thought that believes in sustainable development and the possible rise and success of Green and circular economies, acknowledging at the same time the struggles and difficulties in finding a correct definition of the concept of sustainability. The World Commission on Environment and Development in 1987 tried to define sustainable development as development that meets the needs of the present without compromising the ability of future generations to meet their own needs (Commission on Environment, n.d.), and to have a more complete understanding of the issues involved, it could be useful to also keep in mind the complexity of the factors involved (from political to social) and the intergenerational and intragenerational equity of the sustainability concept (Ruggerio, 2021). Despite the contrast of the academic scene on definitions, everyone aligns in stating that in the last decade, the sensibility andattention of both the general public and business world on Green topics has drastically increased. A proof of this behavior can be found in the increase of Green products sold (and therefore produced), where as a green product we mean those products that use fewer resources, have lower impacts and risks to the environment, and prevent waste generation already at the conception state (Jan et al., 2019). Green products have seen an increasing trend of 73% only from 2009 to 2010 and increasing since (Jan et al., 2019). Multiple possible factors contribute to the just mentioned rise, such as the visible effects of the well-known climate change, the ongoing political discussion, and the attention-seeking actions of green activists (Jan et al., 2019). Alongside the growth ofthe subject in media and everyday life, a new business sector has emerged, and the success of green products has proved them to be valuable not only from a societalpoint of view but also from a business perspective. Renewable energies, responsible waste disposal, water quality improvements, and other environmentally friendly initiatives are now viewed as lucrative business opportunities. Investors of all types findthese ventures attractive, and they no longer burden governments alone (Dorschel, n.d.)This mechanism called green capitalism boosts faster growth as competition pushes companies to invest in R&D, look for new innovative solutions, and focus on always more specific topics or solutions (Dorschel, n.d.).

2.5. The GreenTech sector

As environmental protection becomes a global consensus, many industries are trying to replace pollution-heavy technologies with more environmentally friendly, green technologies (Zheng et al., 2022). Considering these developments, the natural outcome is the birth of a new sub-sector, the so-called GreenTech. With this term are indicated all types of environmentally friendly technologies, which can be used in products, services, and processes that deliver value using fewer resources and producing less pollution than current standards (Kim et al., 2023). A mix of advanced technology to green solutions can be also referred to as CleanTech. Some of the technologies that can be part of the

GreenTech movement are the previously mentioned Artificial Intelligence and Machine Learning. In the context of GreenTech, it makes sense to also introduce one of the main players involved, the climate-tech startups. With this term, we refer to environmentally sustainable startups that either use environmentally friendly technologies to mitigate the harmful effects of their business model, indirect benefits, or even create sustainable benefits, e.g., by generating renewable energies, direct benefits (Luisa et al., n.d.). The definitions of the terms mentioned so far, such as CleanTech, GreenTech, and ClimateTech, are not well established and they tend to overlap therefore for the sake of the research it will be simplified, and only refer to the interested industry and companies as GreenTech.

2.6. Going forward

This concise literature review encompasses the key subjects of this dissertation, focusing on technology, sustainability, and their intersection.

The primary technology under consideration is one of the most innovative on themarket, Artificial Intelligence, which can automate processes and read through huge amounts of data. Machine Learning, which turned out to be a sub-category of AI, is even more indicated to be the protagonist of this study, as it can train models to work with fresh sets of data and develop educated conclusions.

On the environmental side, it has been proven that the wide topic of sustainability and sustainable actions has a crucial role in today's business world. Nowadays it cannot be seen as a burden to carry on the company's shoulders or a disturbing topic to avoid, but instead, it represents a new field of opportunities. Green topics have risen in popularity as soon as companies understood that they can be profitable and benefit both themselves and the environment. GreenTech (or CleanTech) companies can leverage a set of different technologies such as AI or ML or others, to work towards sustainable goals. These types of technologies can promote firm growthand advance their production technology by reducing emissions and consuming less energy (Kim et al., 2023). GreenTech solutions are adopted by businesses even when they turn out to be not profitable as they still carry advantages such as enhancing competitiveness and creating new jobs and new perspectives (Kim et al., 2023).

CHAPTER 3 - METHODOLOGY

The goal of this chapter is to elaborate on the methods and techniques used in this work and to explain how the author was able to justify and come to conclusions. Two different methods were used to recover relevant sources and to structure theanalysis: the document analysis qualitative research method and the expert elicitation method. As the study has a strong focus on a real-life case, it was decided to use approaches that allowed to leverage the author's insights on the business case. The author's contribution and assessment derive from qualitative sourcing and analysis, meanwhile, every quantitative aspect of the dissertation derives from the work of experts and professionals involved in the case.

3.1. The Document Analysis Method

Given the nature of the Study and the availability of significant documents, qualitative methods turn out to be the best fit as a research methodology. Document analysis is a systematic procedure for reviewing or evaluating documents, both printed and electronic (computer-based and Internet-transmitted) material (Bowen, 2009). Some of the major advantages of this method include efficiency, availability, and costeffectiveness (Bowen, 2009). On the other hand, the method implies biased selectivity, as documents are likely to be aligned with corporate policies and procedures and with the agenda of the organization's principals (Bowen, 2009). This specific method was chosen because of the already existing material about the case object of the dissertation. The consulting company X produced different types of documents while working on the case, as is common practice in the sector. Concretely, a slideshow in the form of a PowerPoint presentation was created as the final document for the Audit phase. The Audit phase culminated in a presentation made by the consulting team to the client'sdata team and access to this document was given to employees of the consulting company. The document was largely used, in addition to personal interviews and other minor files, to first grasp a general idea of the case and to dive deeper into the technical details later. The deck consists of around 30 slides that cover all the aspects of the Audit, from the first approach to the suggested solutions. In the document, it's possible to find the main issues related to the company as well as the answers from the consultants in detail.

A second relevant document that was used as a source, is the PowerPoint presentation born by the combined efforts of Company Y's CTO and one of Company X's ML Engineer. In occasion of an event focused on GreenTech and AI scale-ups, the two experts produced a slideshow which was then presented at the event. This second slide deck covers more generic information about the two companies but also about their collaboration, partially overlapping the Audit document.

3.2. The Expert Elicitation Method

Document analysis is often used in combination with other qualitative research methods as a means of triangulation, and such sources can include interviews, participant or non-participant observation, and physical artifacts (Bowen, 2009). The second qualitative method the author used is the expert elicitation method, which mostly consists of one-on-one interviews with experts on a specific topic. This approach is often used when there is a lack of data or, as in this case, an expert's knowledge can provide valuable insights (O'Hagan, 2019). Expert elicitation can be described as a way in which expert opinion and judgment enter statistical inference and decision-making or also as the process of expressing expert knowledge on a specific topic (O'Hagan, 2019).

It was possible to recover first-hand information as the author was actively working beside professionals who operated on the GreenTech application case that was analyzed in this dissertation, so considering the availability of such high-quality sources and the structure of the document, this method proved to be a perfect fit. To properly apply expert elicitation and avoid cognitive bias, it is necessary to use defined protocols and structured questioning (O'Hagan, 2019). The technique chosen to use for this specific case is based on a semi-structured interview; the factors involved in the study, such as the complexity of the program development and implementation, suggested this approach as it allows to focus the area of research, but it also allows the interviewee to elaborate on what he believes is most important (Adeoye-Olatunde & Olenik, 2021). A semi-structured interview is a qualitative type of interrogation that permits interviews to be focused while still giving the investigator the autonomy to explore pertinent ideas that may come up in the course of the interview, by developing more on whatever he considers valuable (Adeoye-Olatunde & Olenik, 2021). The interviewees chosen in the study are two profiles highly involved in the case: a MLE from the consulting team and the CTO of Company Y, on the client's side. The MLE interview was carried out in person, while the CTO's was done via phone call. Both interviews were first recorded and then transcribed. The scripts are attached to this document as Attachment 1 and Attachment 2. The interviews are purely qualitative, therefore there was no need for any data analysis software. Having statements from both sides of the deal on the technical level and macro level helps the author to overcome possible bias in the answers.

CHAPTER 4 - CASE ANALYSIS

4.1. Reasoning behind a practical case analysis

Before starting with the actual analysis that represents the core of this work, it is important to quickly review the goals of the Dissertation. The topic of this dissertation is the analysis of the case study and in particular the results when it intersects the GreenTech sector and advanced technologies such as Machine Learning. The Literature Review above showed the increase of popularity in the Green Economy opportunities, the attention towards ML, and the rise of environmentally sustainable solutions. Considering all the information available, one of the goals of the Dissertation is to show a practical real-life case of the positive synergy between sustainability and technology, where ML solutions are applied to a GreenTech problem (see Chapter 1.3). The goal is not to generally prove the advantage given by AIsolutions to GreenTech companies, as one single practical and first-hand experience case does not have enough statistical significance, but rather to contribute to the display of successful points of encounter between those two worlds. One occurrence does not allow the generalization of conclusions nor assumptions, nonetheless proves the possibility of AI advantages to GreenTech companies as it has happened at least once before. In academic work, case studies as a research method are widely used because they may offer insights that might not be achieved with other approaches. (Rowley, 2002) For example, practical and real-life cases consider also unpredictable variables, human-related or not, which can't be imagined in fully theorical works. In addition, the reasons behind the choice of a single case analysis reside in personal interest and data availability; as the author was in direct contact with the people involved in the case,

there was the chance to interview them and access detailed data, leading to the rise of a personal interest in the matter. A real business case, such the one analysed in the study, can be a proof of the effect that ML solutions can have on sustainability involved companies.

4.2. The players involved

On a macro level, there are two main organizations involved in the case and those two entities interacted with each other through a business-based relationship. For privacy reasons, both companies have requested to remain anonymous and will therefore referred to as “Company X” and “Company Y”, or simply “X” and “Y”. The names andspecifics of the companies are not relevant to the goals of this work. One of the two players is a GreenTech startup, Company Y which represents the client in the settings ofthe case. On the other side Company X, a tech consultancy firm specialized in ML solutions, covers the role of the service provider. Through a standard remuneration-service exchange those two companies joined forces, to first analyze and then improve the processes and products of the startup, Y. As Company Y is working in the GreenTech space and X is specialized in supporting clients in those same industries, the match and collaboration between the two was optimal. On both sides, the professionals involved played a big role in the success of the case. X proposed a team of three experienced consultants: a Senior Machine Learning Engineer as Team leader accompanied by another Senior Machine Learning Engineer and a Machine Learning Engineer as Team members. On the receiving side, Y assigned the project to its Data team composed of a Machine Learning expert, a Data Scientist, and two back-end Software Developers. The CTO and Co-Founder of the BlueTech company was also involved in the process.

4.2.1. The BlueTech Company

Y is a young innovative company, born in Germany with over 20 employees as of today. At the core of the company, there is the Ocean and its wellness, as the main driver for the founders was to preserve the quality of the world's waters. The founders acknowledged the fact that healthy oceans are key to billions of people in terms of food, jobs, and climate regulations and from there they built the idea of a start-up thatpositions itself in the BlueTech, a sub-category of the wider and more famous ClimateTech. BlueTech is best described as a cross and multidisciplinary ecosystem, which seeks to solve problems related to the ocean, marine, and maritime sectors (Cooper, 2019), and specifically, Y has decided to focus on one element of the oceans that strongly affects the whole water ecosystem: the seafloor and its meadows. The seafloor meadows' quality and health have a huge impact on the overall ocean ecosystem as the presence and type of seagrass can determine huge differences in the life underwater (van Nugteren et al., 2009). As Company Y itself affirms, the seafloor mirrors how healthy our oceans are and needs to be protected and not exploited. The company operates in the broader market of the Blue Economy, a sector that can be generally defined as an ocean economy that aims at the improvement of human well- being and social equity, while significantly reducing environmental risks and ecological scarcities (Lee et al., 2020). A more well-rounded definition is made by the World Bank, which described Blue Economy as the sustainable use of ocean resources for economic growth, improved livelihoods, and jobs while preserving the health of the ocean ecosystem (De Fontaubert C, 2017). These definitions already elaborate on the social and sustainable value of this specific economy: the companies that work in this sector, including Y, align with certain values that reflect the Sustainable Development Goals, especially Life Below Water and Partnerships for the Goals (see Figure 1).

4.2.2. The ML Consulting Company

The journey of X started in 2018 when the founders of the company recognized the need for a company that did not just create data science models but that realizes lasting Machine Learning solutions with business impact. Since then the company grew significantly opening hubs in the Netherlands and in Germany and focusing on specific industries. One of the strongest attributes of X is its experience in the GreenTech Industry, supporting clients and implementing models fit for the occasion. Keeping this in mind, the negotiations with Y ran quickly and smoothly. Like many other consultancies, X operates by offering different services to the client, starting with an Audit and proceeding with the development and implementation of the solution chosen, if agreed upon by the client. This allows X to acquire as much knowledge as possible about the customer and act accordingly. It was also possible to have a precise overview of the case and have firsthand sources of information thanks to this work method. In this specific collaboration, X approached Y with a first Audit and suggested how to proceed with the development and implementation.

4.3. Company Y's challenges

To understand the goal of the dissertation, it is important to first determine the company's core business and its main characteristics. The importance of the ocean's seafloor has been already mentioned above and to assess its well-being Y utilizes its products based on hyperspectral imaging and data analysis. As the CTO confirmed in his interview, Company Y recovers the data themselves in-house, as there are no precise datasets regarding the seafloor meadows already available online. To be more precise, Y recovers data through an emerging technology, underwater hyperspectral imaging or in short UHI (see Figure 2), which is an extension of hyperspectral imaging technology in air conditions, and is undergoing rapid development for applications in shallow and deep-sea environments (Liu et al., 2020). There are also other methods to retrieve data on the ocean seafloor and those include satellite imagery, hydrographic imagery leveraged through sonar devices, and manual imagery by in-situ divers.

Once the data has been recovered, it is possible to have a more significant view of the seafloor thanks to orthoimage, elevation, coverage, biomass, carbon stock, and seagrass meadows health. Through data analysis, each one of those parameters can be structured and it becomes possible to compare different coastlines and seafloor zones from different parts of the world. One problem that quickly arose involved the scalability of Y's data collection as it is expensive and time consuming to retrieve precise and valuable information through UHI. As the company's scope grew, so did its necessity to expand to different coastlines of the oceans, and to do so an improvementof the model and infrastructure is needed.

Some of Y's original issues were clarified by the MLE Consultant in his interview: the problems at Y did not reside in its model algorithm which was already good, but rather in the infrastructure and its maintainability. Each component of production was rather isolated, making collaboration hard.

4.4. Company X's approach

X is a well-established IT consultancy that works with best practices and standard procedures. Those include specific structures for its Audit projects and precise followups. The company's Audit are divided into Deep-dive modules, some applied by default in every business case and others chosen and customized with each different client. The Machine Learning Engineer involved in the case confirmed in his interview that the Audits at Company X usually last four weeks, involving at least two ML Engineers. During this period of time, the engineers interview people covering different positions to establish the role of Machine Learning technology in the business. The Audit of Y included the default audit modules Use-cases analysis, Infrastructure, and Roadmap in addition to the modules Algorithms and MLOperations analysis.

The structure of the Audit was chosen keeping in mind the goal and the scope of it, as well as the ambition of the BlueTech startup. Company Y works towards scaling up its commercial activities in the upcoming years by strengthening its data product and approaching new markets, so the goal of this audit was to create a future blueprint for the data & AI technology that allows Y to scale its seagrass use case. This blueprint should be durable in the long run, yet flexible enough, considering that Y's role within the market and its products are developing. The scope of the project was defined as a short-term and ready-to-start action, including the already mentioned blueprint for a scalable data and ML infrastructure, an assessment of the current algorithm in place,and ideas for future improvements.

For each module analyzed, Company X's modus operandi is to state the current situation, the challenge to solve, and the proposed solution. This way it is easier for the client to understand the direction to take and for the consulting company to build a step-by-step Roadmap of actions to take. In this case, as in every other Audit, X's team inspected the current tech situation, highlighted things to keep because they were going well, and suggested what could be improved and what is considered mandatory to improve to reach the scaling-up goal.

4.4.1. Infrastructure

Brett Frischmann defines infrastructural resources as resources that are required as inputs to support downstream activities. This also applies to data (Davies et al., 2019). More specific to the case, a machine learning pipeline uses training data to generate a model, which can then be used to perform inference with serving data (Haque et al., 2022). The process implies several challenges to address as data is typically large, it may arrive continuously, and in the latter case, it may also arrive in (incomplete) chunks (Haque et al., 2022). Moreover, data can contain errors hard to detect that need to be caught early, before they propagate downstream and taint models (Haque et al., 2022).

The first part of this deep-dive module consisted of an assessment of the current infrastructure for the seagrass use case and its possible improvements. When X started the audit, the data was scattered in different buckets or data containers, and the procedure was dependent on manual triggering. This can lead to islands of automation and to risks in case of increasing data load. In its assessment, X approved the first step of gathering data from different data sources but suggested changing the intermediate and temporary data bucket storage. In cases of this nature, the most common process bottleneck is the manual triggering part that turns out to be a mandatory point of improvement to scale up the data load. X's team was able to build a new blueprint to implement to solve the main issues identified starting from this analysis. The machine-learning engineers involved also suggested specific software solutions, proved by years of industry knowledge.

4.4.2. ML Ops

It is necessary to introduce the concept of ML Ops and its definition before elaborating on the Audit results. Machine Learning Operations, or in short ML Ops, comes from the meeting of DevOps and Machine Learning and it is a method that is highly used by engineers worldwide. Development Operations means a set of practices with the main purpose of minimizing the needed time for a software to release, reducingthe gap between software development and operations (Symeonidis et al., 2022). Connecting the two concepts at its base, MLOps stands for the deployment of ML models in production (Symeonidis et al., 2022).

The goal was to enhance Company Y's MLOps structure as its capability was not scalable, because of difficulties in training increasing data load, and lacking of automated monitoring. At its core, the Machine Learning engine provides valuable predictions as the code was well-structured according to best-practice software principles, but other aspects need improvements. Company X underlined the challenge of local model training approaching memory and computational limits as the data load increases and strongly warned Y about the manual handling of data extraction and the necessity of setting up an automated monitoring system. Once again, a blueprint tosolve the problem was created, and supporting software was suggested.

4.4.3. Algorithms

Machine Learning relies on different algorithms to solve data problems, but there is no single one-size-fits-all type of algorithm that is best to solve a problem. The kind of algorithm employed depends on the kind of problem you wish to solve, the number of variables, the kind of model that would suit it best, and so on (Mahesh, 2018).

Keeping this in mind, Company X's purpose was to test the solidity of Y's code and to determine if there was a better approach possible. This is a common procedure for the consulting agency X. The algorithm that has been written to assess and detect the seagrass and ocean meadows was the object of scrutiny. A well written code should be able to use the data given to determine specific information about the ocean seafloor, which will then be used to assemble Y's products.

It was determined from the Audit that the model is performing well, but there was manual action involved which can be limiting when scaling use cases.

The solution suggested to solve the main issues was to increase generalization by data augmentation and active learning, but also to switch to a new model approach. According to the Machine Learning Engineer who worked on the case, a model that is able to leverage the cloud to work on the data implies that the same people can work on the same thing at the same time. The model would become more scalable.

4.4.4. Roadmap

The scope of the Audit includes a roadmap, which can be used to make the following steps concrete and actionable. The goal of consulting X's roadmap is to guide a client through the months following the Audit, aiming to solve the issues that came up and improve the company. The roadmap is also useful to determine the exit point for X: the cooperation between two companies is not meant to be long lasting as the client's data team is supposed to take over the restructuring and continue with the day-to-day tasks. The Audits made by X, include a high-level roadmap with activities for the upcoming six months and a three months short-term action plan with detailed steps. The actionplan is divided by focus area, by yearly quarter, and by responsible team. In the case analyzed, the areas involved were the ones studied in the inspection so Infrastructure, ML Ops, Algorithm, and the responsible team could be either X's team, Y's own data team, or a joint group. The roadmap covered only until Company X's exit point fromthe project.

CHAPTER 5 – CONCLUSIONS

This final chapter aims to summarize the results of the Audit as well as dive deeper into the opinions and suggestions of the experts. The author elaborates on the comparison between the expected results and the actual output, along with possible improvements to these case and similar ones. As mentioned before, a single case cannot be statistically considered valid to prove and generalize the benefits of AI on Greentech-related issues, but it can demonstrate the existence of potential advantages, given that it has occurred at least once before. Taking into account what has just been written, the author draws conclusions and uses the analyzed case to derive usefulassumptions applicable to other similar situations.

5.1. Case results

The Audit phase was divided into three different major areas (see Chapter 4) and each one of them has led to specific results and solutions to their own problems. These results were presented by X's team to Company Y, including analysis, suggestions, and detailed workflows. Following the Audit, X worked on the development of the solutions suggested, focusing on infrastructure and a cloud component implementation. In addition to this, it is common practice for the consulting team to plan periodic follow-upmeetings and detailed sessions to train the client's data team to operate the new tech stack.

The results of the Audit and Development are very much connected to the already present tech situation and the general health of the company. When interviewed, the MLEngineer of Company X provided an overview of the initial situation at Company Y. Without diving into the technical specifics he mentioned the machine learning case, so the way they used their machine learning model and how they trained it was already really good but the infrastructure could be optimized as it was divided in isolated modules. Keeping this in mind, the machine learning model had little room for improvement while the general scalability and cloud component addiction could be improved the most. Another improvement that can be useful for AI companies is a proper model registry, a centralized hub that allows to compare all the models in the same sheet. The absence of this tool can cause the issue of distinguishing between the actual models and the experimental ones. The team at X supported the creation of this tool in the development phase.

The case highlighted that when talking about Infrastructure, a challenge can be to build a future-proof, scalable infrastructure that utilizes data with an orchestrated and consistent approach. In this situation, the obstacles are the manual triggering of processes and the scattering of data, which can be solved by extending the data lake to handle increasing data loads and setting up specific ML-related software. Useful software that are suggested by company X for data management includes MongoDB, AWS Lambda, and Glacier. The collective of those three software can be easily integrated into an already existing company's tech stack, facilitates automatic scaling, and allows low-cost and long-term storage of images. Apache Airflow is suggested as an orchestration tool, a software solution or framework that helps coordinate and manage the various components and processes involved in a machine learning pipeline.

It is designed to streamline and automate the deployment, execution, and monitoring of machine learning tasks. Lastly, AWS Batch and Kedro can support in working with large quantities of data at the same time.

The challenge of MLOps was to build a scalable framework that could operate with multiple use cases, beyond the specific Seagrass use case. To do so, the new MLOps BluePrint strongly relied on the software suggested for the infrastructure improvement.In addition to the ones already mentioned, EvidentlyAI can be used to monitor the data drift and ML performance of the models deployed. It can be very helpful to create reports to assess the health of the Machine Learning Operations.

Regarding the Algorithms, the goal was to find a way to leverage data to make generalizable models and assess their quality quickly. Data augmentation and Active learning can be used to carry out the task. Data augmentation applies relevant images transformations to enrich dataset with synthetic data, while Active learning focuses on selecting the most informative data points for labeling and making the most out of labeling efforts. These two methods enable a company to use the same model for multiple campaigns or use cases. X's solution to improve scalability, generalization and long-term efficiency is to implement a cloud solution that could solve the problem of isolated island of responsibility. Moving in this direction, the Algorithm augmentation overlaps the Infrastructure improvements; mentioning the MLE in his interview, wrapping a model into an AWS Batch function means that it now runs on the cloud. This approach adds different advantages to a computer vision trained model, such as a customized Resnet18 model. Cloud computing can potentially provide a more cost- effective alternative to acquiring and maintaining large-scale systems operations in- house (Müller et al., 2015). The overall cost savings enabled by cloud computing have even been referred to as the principal advantage of cloud computing (Müller et al., 2015). Another benefit of cloud-based applications is that they avoid multiple versions of the same document existing side by side in an organization (Müller et al., 2015). making it possible to work on a task at the same time.

5.2. Results analysis

Looking at the results, the Y business case can be considered a success, as it ended with positive outcomes. The overall health of the company and its products improved and the roots for future expansion and development were set.

In his interview, the MLE, mentioned two parameters used by X to determine that can be used to determine the success of the collaboration which were how much data can be processed in how little time or in parallel and how fast does a data product goes from development to production. He also stated that both of those parameters highly improved thanks to Company X's support. Another criteria that was considered, was theoverall happiness of the client towards the consultancy's services. This factor is subjective and therefore hard to measure, but he confidently assured about it. Additionally in favor of the satisfaction of the client with the Audit, Company Ydecided to hire the consultants to work on the development phase as well. The occurrence of this event can serve as evidence supporting the validity of the client's happiness.

The management of the project was one of the factors that contributed to the favorable results of the consultancy. The firm knowledge and modus operandi were key in running the audit and development phases smoothly. In the consultant's words, X uses fixed and well-tested procedures in its approach with a new client, so that the audit consists of a four-week process in which two engineers interview multiple people in the company at different levels to establish the use cases of the machine learning case as well as the impact of this on to the business. The results of the audit will then lead to the development of the solutions if the satisfaction of the client is guaranteed. Having a set procedure and constant work methods gives X an incredible advantage as the consultants' team can work within a structured framework. This can also lead to specificknow-how that can make the difference, such as the means and quality ofcommunication. An example of positive communication can be the communication withY's data team which was very smooth throughout the whole project, thanks to theteam's relationship with Y's CTO and the four members of Y's data team that were involved. The interaction between the two teams consisted of open communications channels, two deep-dive meetings per week, and a daily startup meeting hosted by Y's Scrum Master.

In cases of this nature, the data team plays an important role, as it is essential to have the right set of skills in supporting the consulting team's efforts. In the GreenTech company indicated in this work, the team involved in the collaboration is well set up asit is composed of a Machine Learning Expert, a Data Scientist, and two back-end Software Developers. According to the expert interviewed, the back-end developers function as data engineers, creating the pipelines that let your data flow in your company where you need it to be, cleaning and transforming the data in the rightformat, while the machine learning engineer works on the model and has the knowledgeto know what can or cannot be done with Machine Learning implementations. Lastly, the data Scientist role can work on the final part of the process, operating on the dataand turning it into useful outcome. As said by X's ML engineer, to have all three of this expertise in a data team is optimal and often quite uncommon, therefore Company Y employs a high-class team. A fruitful data team was relevant in achieving the results stated.

When asked about the importance of data teams in GreenTech companies, the CTO of Company Y replied it is crucial to have a functioning data team, and it is not apeculiarity of GreenTech companies, but it applies to most businesses that leverage AI. The importance of having a data team is related to the fact that to solve specific problems, it is necessary to train and customize generic models, to make them applicable to each distinct challenge. The two interviewees agreed that it is possible to easily recover online generic ML Models already tested, but it is mandatory to have experts that can adapt the model, such as the identification of seagrass meadows on the seafloors, to apply them to specific issues.

5.3. Possible Improvements

In every project, there is room for improvements, and it is useful to acknowledge them so that can be applied in future business cases of a similar nature. To analyze the case after its completion and asking the opinion of the experts involved is a possible way to determine and highlight changes for the better. According to company X, there is not much that could have been done differently as the output was valuable, and the initial status of Y was already quite good. Although the MLE did not find possible upgrades on the technical side, he talked about obstacles in the project management approach. X's team adjusted to Company Y's way of working which consisted of an Agile method and periodical meetings. The consultant took part in the daily meetings of Y which were coordinated by a Scrum Master and often turned out to be too specific on marine biology or other product related topics. In his opinion, those meetings were mostly useless and unnecessary to engineering work but useful to show collaborations on as many fronts as possible. From the engineer's reply, it is possible to assess that improvements in ML cases in the GreenTech industry can be related not only to technical issues but also to communication and collaboration obstacles, as both of those aspects of the project cover an equally important part. In the specifics of X, the enhancement to make for future cases is to carefully evaluate the importance of the client's processes and to previously determine if they are relevant to the project or not. Adjusting to the client's way of working can have beneficial effects, but it can alsocause hurdles of such nature.

5.4. Additional Information

The Y – X business case had some peculiarities that according to the first interviewee are uncommon in similar cases. During the Audit phase of the collaboration, X'sconsultants determined more information about its client's data team and machinelearning setup. Both the data team composition and ML algorithm at Y were already well structured (see Chapters 5.1 and 5.2) before the beginning of the project. This advantage from Y allowed the consultants to focus on a different set of issues, supporting current growth but also future development. In his words, the unusually already well-defined skillset at Y made it so that X did not have to focus on advising on what kind of expertise to have in-house or outsourced and on how to improve the machine learning algorithm. In the consultant's experience, this peculiarity distinguished the collaboration from other X cases in the GreenTech industry.

The significance of having a high-quality data team for Company Y was underlined by its CTO. According to him, in AI companies and more specifically in focused sectors such as BlueTech, it is impossible to operate with data without having strong in-house expertise.

It is common in theses and academic research to have limitations that can change and influence the final product. This can specifically happen with practical cases as the information available is often not published online and must be retrieved by contacting decision-makers, persons of interest, and workers involved. The Y case falls in this category as the only available information online are limited to the companies' websites. As an employee of X, the author had access to useful documents such as the Audit final slideshow that contains all the information presented by the consultants tothe Y team. He was also present at the presentation made by Y in collaboration with X regarding their collaboration. During their speech at the event, a ML Engineer from X and the CTO from Y talked about the case. The slideshow of the presentation was lately made available to the employees at X. Lastly, it was possible to arrange the interviews with the persons of interest mostly thanks to the personal relationship between them and the author.

The scarcity of reliable information published on the encounter between Greentech and ML and the technicality of the development phase were the limitations that actually impacted the work the most. Also, the anonymity requested by the companies generated an additional challenge in the case study. To preserve the privacy of those involved, it was necessary to omit or generalize technical details and sensitive information that emerged from the interviews and the documents. For the same reasons, some parts of the interviews were omitted (see Attachment 1, 2) and only the dialogue cited or source of reference was inserted.

The role at Company X, the interest in the project, and his network allowed the author to gather useful information, and simplify and summarize technical matters. At the same time, the peculiarities and uncommon advantages of Y were not significant enough to obstacle an overview of the GreenTech market and the understanding of the case review and analysis.

5.5. Machine Learning in GreenTech

This work was introduced by literature references and presented a practical case of a successful collaboration between two young and innovative companies. The purpose of the document is not only to showcase this collaboration but also to draw conclusions from it, extending the interest to different aspects of the topic. To do so, it is useful to first review how the goals of this dissertation were met. It was possible to highlight the importance of the GreenTech industry and ML as current trends thanks to the Literature review and the practical case. At the same time, the successful output of the case proved the possibility of improving a GreenTech company operating in the Blue Economy with customized ML solutions and consultancy.

The case of the X – Y collaboration revolved around Artificial Intelligence consulting and overall results can be considered favorable. In the interviewee's opinion the case can be considered a success because, besides the happiness of the client, there were improvements in parallel data processing and in the speed of production of data products. From a sustainability perspective, it was not determined the direct impact of the improvements because of the multitude of factors involved, but it can be assumed that a faster production of data products (at parity of quality), as well as a more efficient functioning of the company, can only be beneficial. A cloud solution was implemented which can be overall beneficial for the business as it enables innovative new services and business models that decrease time to market, create operational efficiencies, and engage customers and citizens in new ways (Müller et al., 2015). On the other hand, it could also be source of a higher energy consumption (Yenugula et al., 2024).

Even though this case supports the theory of a beneficial correlation between ML and Greentech products, it does not prove its absolute validity, as there are many factors to be considered. In the analysis of the business case, the positive outcomes of a Machine Learning solutions approach were the protagonist, as this is what was derived from the sources reviewed. But in optic of a more general conclusion, it is important to highlight the problems and issues related to the usage of this technology.

Artificial Intelligence implementations can be used to support sustainable goals, but they also contribute to damaging the environment. AI (both in the sense of training models and of uses) can consume vast amounts of energy and generate greenhouse gas (GHG) emissions (García-Martín et al., 2019), making the outcome of its application debatable in most cases. Although AI is a promising and powerful technology for sustainability, its application creates a huge carbon footprint, which represents a direct form of rebound effects (Nishant et al., 2020). AI is being applied to GreenTech companies for sustainable purposes and at the same time it somehow contributes to the negative aspects it is trying to solve. During the creation and implementation of AI solutions, the model training is the phase that requires the most computational power and has a bigger impact on the environment. A single AI model training can emit carbon dioxide equivalent to the lifetime of five cars (Hao, 2019). Systems based on AI need a lot of processing power thereby increasing energy consumption. A high amount of data raises the cost of servers and the high amount of electricity required to keep data centers cool (Dwivedi et al., 2022). Also, ML can rely on Cloud Computing (CC) which contributes highly to greenhouse gas emissions, and it is estimated that data centers will consume 8% of global electricity by 2030 (Yenugula et al., 2024). An organization will consume more energy as a result of the implementation of AI, so the impact of AI on CO2 emissions must be a key factor in decision-making (Dwivedi et al., 2022).

Another limitation regarding Machine Learning in GreenTech is correlated with the hard predictability of the variables involved. Human-related variables are complicated to incorporate into ML models because they are sometimes unpredictable and still evolving. Thus, it is difficult to adopt a deterministic approach in which outcomes can be fully determined because we cannot precisely estimate potential climate changes (Nishant et al., 2020).

That being said, the overall result of AI implementation is usually considered fruitful as these technologies have the potential to minimize waste production, use less energy, and improve recycling and industrial symbiosis prospects, among other sustainability advantages (Kar et al., 2022). Determining the potential of digital technology to propel successful sustainability activities among companies in the future is much needed area (Kar et al., 2022).

Another point of contact worth mentioning between GreenTech and AI is the profitability of the technology. When interviewed, the CTO of Y mentioned that one of the biggest challenges for the GreenTech startups relies upon investments. In his opinion, in the GreenTech industry, it is not always clear from the beginning where money can be earned, and this could make investors more reluctant to invest. There are specific categories of investors that are focused on sustainability-driven companies, but they also value the economic aspects more than the sustainability ones. According to the CTO, this mechanism is especially strong in the marine or BlueTech industry. On the other hand, it has been mentioned in the Literature Review that AI and ML can be useful to the overall support and increase companies' performances (see Chapter 2.2, 2.3). The logical consequence of these two different statements suggests that the integration of AI solutions into GreenTech companies that have uncertain economic returns may attract greater attention from investors. Considering all the characteristicsof these technologies mentioned so far, there is the possibility that artificial intelligence could make GreenTech companies profitable and therefore more appealing to investments.

It is difficult to determine with absolute certainty that Machine Learning has a positive impact on the GreenTech industry and sustainability in general, as the elements to be considered are many and sometimes in opposition with each other.Each case should be evaluated on its own, but it can be said that there are manyinstances where value is created with the addition of AI. Besides the single case study object of the study, there are a multitude of cases available to investigate where this happened. For instance, Company X's consultant highlighted in his interview scenarios of companies using computer vision speech recognition or satellite imaging, where ML could be effective. The positive collaborations are so many that new industries and markets are born out of it, indicating a development path for the years to come. As the health of the environment declines, technological progress advances making it logical toleverage this last one to address worldwide problems. Despite its issues and challenges, Machine Learning could be the key to upholding a more sustainable and greener future.

REFERENCES

Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. *JACCP Journal of the American College of Clinical Pharmacy*, *4*(10), 1358–1367. https://doi.org/10.1002/jac5.1441

Agarwal, N. D., & Kumar, V. V. R. (2020). Three decades of green advertising – a review of literature and bibliometric analysis. *Benchmarking*, *28*(6), 1934–1958. https://doi.org/10.1108/BIJ-07-2020-0380

Bose, I., & Mahapatra, R. K. (2001). Business data mining - a machine learning perspective. *Information & Management*.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Chhaya, K., Khanzode, A., & Sarode, R. D. (2020). Advantages and disadvantages of Artificial Intelligence and Machine Learning: a literature review. *International Journal of Library & Information Science (IJLIS)* , 9–10. http://www.iaeme.com/IJLIS/index.asp30http://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1JournalImpactFactor

Commission on Environment, W. (n.d.). *Report of the World Commission on Environment and Development: Our Common Future Towards Sustainable Development 2. Part II. Common Challenges Population and Human Resources 4*.

Cooper, J. (2019). *Growing Capability and Capacity in the Bluetech Startup Ecosystem*.

Copeland, B. J., & Proudfoot, D. (2007). *Artificial Intelligence: History, Foundations, and Philosophical Issues*. http://www.AlanTuring.net/tunny

Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. (2019). *The State of Open Data: Histories and Horizons*. African Minds and International Development Research Centre.

De Fontaubert C. (2017). The Potential of the Blue Economy: Increasing Long-Term Benefits of the Sustainable Use of Marine Resources for Small Islad Developing States and Coastal Least Developed Countries. *The World Bank*. https://www.researchgate.net/publication/317952452

Dorschel, R. (n.d.). *Global Warming as Investment Climate: Green Tech and the Translation of Needs (Working Paper)*.

Dwivedi, Y. K., Hughes, L., Kar, A. K., Baabdullah, A. M., Grover, P., Abbas, R., Andreini, D., Abumoghli, I., Barlette, Y., Bunker, D., Chandra Kruse, L., Constantiou, I., Davison, R. M., De, R., Dubey, R., Fenby-Taylor, H., Gupta, B., He, W., Kodama, M., … Wade, M. (2022). Climate change and COP26: Aredigital technologies and information management part of the problem or the solution? An editorial reflection and call to action. *International Journal of Information Management*, *63*. https://doi.org/10.1016/j.ijinfomgt.2021.102456

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and DistributedComputing*, *134*, 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007

Hao, K. (2019). Training a single AI model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review*.

Haque, S., Eberhart, Z., Bansal, A., & McMillan, C. (2022). Semantic Similarity Metrics for Evaluating Source Code Summarization. *IEEE International Conference on Program Comprehension*, *2022-March*, 36–47. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Jan, I. U., Ji, S., & Yeo, C. (2019). Values and green product purchase behavior: The moderating effects of the role of government and media exposure. *Sustainability (Switzerland)*, *11*(23). https://doi.org/10.3390/su11236642

Jhaveri, R. H., Revathi, A., Ramana, K., Raut, R., & Dhanaraj, R. K. (2022). A Review on Machine Learning Strategies for Real-World Engineering Applications. In *Mobile Information Systems* (Vol. 2022). Hindawi Limited. https://doi.org/10.1155/2022/1833507

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. In *Journal of Cleaner Production* (Vol. 376). Elsevier Ltd. https://doi.org/10.1016/j.jclepro.2022.134120

Kim, K., Nonnis, A., Özaygen, A., & Kogler, D. F. (2023). Green-tech firm creation in Germany: the role of regional knowledge. *International Entrepreneurship and Management Journal*, *19*(1), 97–120. https://doi.org/10.1007/s11365-022-00808-3

Lee, K. H., Noh, J., & Khim, J. S. (2020). The Blue Economy and the United Nations' sustainable development goals: Challenges and opportunities. In *Environment International* (Vol. 137). Elsevier Ltd. https://doi.org/10.1016/j.envint.2020.105528

Liu, B., Liu, Z., Men, S., Li, Y., Ding, Z., He, J., & Zhao, Z. (2020). Underwater hyperspectral imaging technology and its applications for detecting and mapping the seafloor: A review. In *Sensors (Switzerland)* (Vol. 20, Issue 17, pp. 1–21). MDPI AG. https://doi.org/10.3390/s20174962

Luisa, L., Elisa, H., & Killian, M. (n.d.). *Success Factors of Climate-Tech Startups A Study of VC-backed Climate-Tech Ventures in the Nordics*.

Mahesh, B. (2018). Machine Learning Algorithms-A Review. *International Journal of Science and Research*. https://doi.org/10.21275/ART20203995

Mccarthy, J. (2007). What is Artificial Intelligence? *Stanford University - Computer Science Department*. http://www-formal.stanford.edu/jmc/

Monett, D., Lewis, C. W. P., Thórisson, K. R., Bach, J., Baldassarre, G., Granato, G., Berkeley, I. S. N., Chollet, F., Crosby, M., Shevlin, H., Fox, J., Laird, J. E., Legg, S., Lindes, P., Mikolov, T., Rapaport, W. J., Rojas, R., Rosa, M., Stone, P., … Winfield, A. (2020). Special Issue "On Defining Artificial Intelligence"—Commentaries and Author's Response. *Journal of Artificial General Intelligence*, *11*(2), 1–100. https://doi.org/10.2478/jagi-2020-0003

Müller, S. D., Holm, S. R., & Søndergaard, J. (2015). Benefits of cloud computing: Literature review in a maturity model perspective. *Communications of the Association for Information Systems*, *37*(1), 851–878. https://doi.org/10.17705/1cais.03742

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, *53*. https://doi.org/10.1016/j.ijinfomgt.2020.102104

O'Hagan, A. (2019). Expert Knowledge Elicitation: Subjective but Scientific. *American Statistician*, *73*(sup1), 69–81. https://doi.org/10.1080/00031305.2018.1518265

Quemy, A. (2020). Two-stage optimization for machine learning workflow. *Information Systems*, *92*. https://doi.org/10.1016/j.is.2019.101483

Ramachandran, K. K., Apsara Saleth Mary, A., Hawladar, S., Asokk, D., Bhaskar, B., & Pitroda, J. R. (2022). Machine learning and role of artificial intelligence in optimizing work performance and employee behavior. *Materials Today: Proceedings*, *51*, 2327–2331. https://doi.org/10.1016/j.matpr.2021.11.544

Rebala, G., Ravi, A., & Churiwala, S. (2019). *An Introduction to Machine Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-15729-6

Rowley, J. (2002). Using Case Studies in Research. *Management Research News - Emerald Insight*.

Ruggerio, C. A. (2021). Sustainability and sustainable development: A review of principles and definitions. In *Science of the Total Environment* (Vol. 786). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2021.147481

Symeonidis, G., Nerantzis, E., Kazakis, A., & Papakostas, G. A. (2022). MLOps - Definitions, Tools and Challenges. *2022 IEEE 12th Annual Computing and Communication Workshop and Conference, CCWC 2022*, 453–460. https://doi.org/10.1109/CCWC54503.2022.9720902

van Nugteren, P., Moodley, L., Brummer, G. J., Heip, C. H. R., Herman, P. M. J., & Middelburg, J. J. (2009). Seafloor ecosystem functioning: The importance of organic matter priming. *Marine Biology*, *156*(11), 2277–2287. https://doi.org/10.1007/s00227-009-1255-5

Yenugula, M., Sahoo, S. K., & Goswami, S. S. (2024). Cloud computing for sustainable development: An analysis of environmental, economic and social benefits. *Journal of Future Sustainability*, *4*(1), 59–66. https://doi.org/10.5267/j.jfs.2024.1.005

Zhang, X., Chan, F. T. S., Yan, C., & Bose, I. (2022). Towards risk-aware artificial intelligence and machine learning systems: An overview. *Decision Support Systems*, *159*. https://doi.org/10.1016/j.dss.2022.113800

Zheng, S., Jiang, C., Fu, X., Ge, Y. E., & Shu, J. (2022). Subsidies for green technology adoption under uncertain demand and incomplete information. *Omega (United Kingdom)*, *112*. https://doi.org/10.1016/j.omega.2022.102675

ATTACHMENTS

ATTACHMENT 1 – Interview X

MLE - Interview

Transcribed on: 01/03/2024, 15:25:06

Total recording length: 00:40:21

In this transcript, the Machine Learning Engineer is referred to as "MLE" and the Author as "Author"

Author: { 0:02 }

OK. Hello. Hi, MLE. Thank you for being here. I will now ask you a few questions about the Company Y case. Yeah, well first of all, if you can present yourself just your name and your role at the company.

MLE: { 0:21 }

Yeah, of course. I'm MLE. I work as a machine learning engineer at Engines which is a machine learning consultancy company which, since every machine learning consultancy company does something else, may be good to state that we do the architecture design but also data governance design and then implementation of yeah, cloud infrastructure that enables AI cases as well as the AI use cases themselves.

Author: { 0:54 }

OK. Yeah.

{ 0:54 }

So you've already kind of mentioned a bit, but what is X, modus-operandi for an audit, how does it work? What are the steps? All right.

MLE: { 1:07 }

Yeah. When we get a new client, we really focus on bringing an AI model to production because we usually think most or too many AI cases are just being you know, thought of, designed and then never implemented. So usually when we get a new client, we'll start with an audit which is a four-week process in which two engineers interview multiple people in the company at different levels to establish basically the use case of the machine learning case as well as the impact of this on to the business.

{ 1:50 }

And then based on that audit we suggest what would be the best way to help that company or to scale up or get there or be more successful or maybe also just reduce workload by implementing AI into their workflow

Author: { 2:10 }

OK. Thank you. And I'm talking about, I want to talk now about more specifically about the Y business case, Y is a client of X and you guys just did an audit on them. What was the tech situation at Y when you started the audit? And what were the main problems that you found? Found out, yeah.

MLE: { 2:41 }

We actually we did an audit and a development with them. So first the audit then worked with them for a couple of months to actually implement of the things that we found and we looked at three things at Y in particular which were the algorithm,

{ 2:57 }

so the actual use case of the machine, of AI, the infrastructure that would enable you to use this machine learning model and the ML OPS procedures, so machine learning operations which is just like DevOps but then for machine learning infrastructure. So those are certain procedures what you can do to make the use case of machine learning models smooth, let's say in industry. Sorry, what was actually the question? you asked for a particular finding or what we looked at only?

Author: { 3:34 }

Yeah, what you looked at and what turned out to be the main problem. Just if you could summarize.

MLE: { 3:40 }

So those: what we look at and then main findings. Their machine learning case, so the way they used their machine learning model and how they trained it was already really good.

{ 3:50 }

So the way they used it, the kind of model that they used the the data quality was good as well And their overall let's say accuracy, although that's not necessarily the metric I think of, 1-score would be more particular in this case, but just the the accuracyof the material model was good as well. So we didn't focus much on that in the actual development. We instead looked more at the infrastructure and found out that things weren't very maintainable. They were very modulized, so isolated. Yeah, islands of responsibility with one person being responsible for a certain part, which makes the collaboration between people in the company quite hard. So we changed that centralizing things and made it bring things to the cloud so that the same people can work on the same thing at the same time and also just makes it so much more scalable.

{ 4:49 }

So if you want to scale up operations, then you can fulfill still working on a larger amount of data that you collect, Yeah. And then one or two more things in machine and corporations related to stuff like make the…, create a proper model registry. So when you train a model, make sure that you can track well what your experiments were, as well as then keep all those models in a centralized hub where all the people that work onmodels can find the models and compare the models on more like 1 sheet which is a… which is a good idea.

Author: { 5:34 }

OK. Yeah. Yeah. Thank you. And out of out of those three areas that youmentioned earlier, so the infrastructure, the MLops and then the algorithm by by itself, which one do you think was the one that had the biggest issues, that needed to to work on it the most and maybe was also the harshest to to work with?

{ 5:58 }

I mean, you kinda already answered to that.

MLE: { 6:01 }

Yeah, Yeah, I do think that's infrastructure just because infrastructure is a hard case often especially if you have an industry with a lot of requirements or a lot of different processing steps. So in the case of Y, they have to make images of the sea ground and then create a data product that shows you six different things of seagrass. So tells you whether there's seagrass or not. But then if there's seagrass, it also tells you its density and its health and the altitude in which it grows and stuff like that. So if you want to have so many different results and working with images in the 1st place, then you havea lot of challenges to follow.

{ 6:52 }

Because you need to like have at least a couple of preprocessing steps as well as some later processing set to create every single layer. Which means that you need to be able to shift your data around quite a lot, assuming that not one person does it all in one pipeline, which is never the case. So the infrastructure really requires you to be able to shift large amounts of data, which images typically are and not break, while having several people work on it and be able to collaborate on it at the same time as well. So yeah, in terms of infrastructure at Plan Blue, it was very modularized where sort of every expert created, like was building and working on their own part of the data pipeline or product pipeline, whatever you want to call it. And within their own system they would be like doing those steps all on their own computer and then uploadsomething to a sort of shared storage.

{ 7:59 }

Which makes things messy, because that means every next person that requires this output, or this data product, this processed or half processed product will have to wait for the other person to finish. And even when they're finished, then they don't really know where that first person pushed their product to and when it will be done. And suchstuff like that. So if you bring that all to the cloud, then you can automize things by placing automized triggers and stuff like that. That collaboration is much more efficient.

Author: { 8:39 }

OK, So what about the model . You said that the machine learning use case was already quite good at Y. Can you tell me a bit more about the model they were using?

MLE: { 8:56 }

They used the Resnet 18 model, which is very common in machine learning in computer vision. If you're working in industry and you do something with computer vision, you don't want to reinvent the model usually because there's a lot of really good models out there that are free, accessible and just you can just download them. So if you were to like build a layer by layer and again you really spend your resources not very well because you would be competing with global players like Google and Microsoft who have built these kinds of models. So Resnet is one of the ones that is well known because it won the competition some win, some win probably in 2018, I'm not sure, might have also just 18 layers, it could be either one. But just like that, there's for example the dense-net and U-net, although that's more like a type of model to be honest and the Alex-net.

{ 9:58 }

So those you know could have, could have all worked, but in this case it's the Resnet due to the residual connections. But that's just minimistic stuff we don't care about. But what I'm saying is that model already was good in computer vision, so it could recognize probably like 100 different objects and at least would be like everyday objects. Like this is a dog, this is a mobile, this is a lollipop or something. And then this very capable model was basically trained to become Dumber, to be able to specialize in one case, in one case only. And that was differentiated between is the seagrass or is this sediment or let's say not seagrass. So to do that you take that very capable model and keep feeding it labeled input of the same kind, which is seagrass or not seagrass.

{ 10:52 }

And then eventually it will, you know become really good at that task because while it was very generalized and good in vision things, first it would have been bad if you just use it on seagrass. So because that wasn't one of the thing but you can transfer learn that model is what it's called to get this new kind of model and what you then you call the first model that you have your foundation model, then you transfer learn it with label data and what you gain is a fine tune model and that's what they did and they worked well.

Author: { 11:37 }

OK.

{ 11:39 }

So it worked well and that's what they use and what they did and just to just to get back and clarify what X did is to add the cloud component to this already working model, Is that correct? to make it also more scalable and as we said to improve the infrastructure?

MLE: { 12:02 }

Yes, implementation-wise all we did is wrap it into a AWS batch function, which is literally just means instead of running it on your computer, you run it in the cloud. That means you can run it on not just. I don't know. On your computer, it may be able to look at like 10 different images at the same time, let's say. But in the cloud you can make that is infinitely high, right? You can just give it more processing power and then you can say like, oh, I have 1000 images or like more like a couple of million images, and I want you to look at 1000 at the same time. And that's possible. So that's what we did. We did give them some recommendations for what they could do with their model later,because at the moment they want to scale up operations, but the quality of their model isgood enough.

{ 13:00 }

Eventually they will have other temperatures, like maybe they want a model that can recognize seagrass not just in the Mediterranean Sea but also in the Philippines and Fiji and everywhere else. And you can imagine that with water quality and seagrass type, that would bring new challenges. So then you want your model to become more generalizable, and we get them some tips how that could be done. We also gave them some tips how they can reduce the amount of time they spend on labeling images. That Isaid about transfer learning. The model you need labeled images and there are certain things you can do to make it like, well, you need less labeled data, or rather, the datathat you label is more efficient. Yeah.

{ 13:53 }

And those were all suggestions to keep in mind when they come to those kinds of challenges later in their industry, no?

Author: { 14:05 }

OK. What about the team, the data team at Y? Do you think they had all the roles they needed and all the skills, the skill set that they needed? Or there were some particular skill sets, some positions that they were lacking at the time.

MLE: { 14:26 }

Yeah, interesting question.

{ 14:31 }

They were fairly well set up, I think because they had a machine learning expert, let's say, plus a data scientist and then two software developers that acted as well back- end engineers, but therefore also a little bit like data engineers. So yeah, depending on what your company wants to do, I'd say it's always good to have a data engineer becausethat person knows how to create the pipelines that let your data flow in your company where you need it to be. And then a well, machine learning expert or machine learning engineer, I mean all these terms, they are not protected. So you come into one company,it means one thing, you come into a different company, it means the other thing. But then you have somebody who like one person that knows where how to do the pipelines for the the data funneling and then one person that knows what to do with it in terms of creating the models and actually use it thinking about what is something that can bedone with machine learning and what is something that cannot be done.

{ 15:55 }

So, for that you want somebody who knows more about machine learning. And then you need somebody who is good with infrastructure because otherwise you might be able to build pipelines and those sort of push your code, but like can select your code, let's say. So, data engineering can transform the data into the right format, let's say that's what data engineers or data analysts often do. They clean the code, they select theproper code and they mix and Max basically. But that is a little bit different from creating the infrastructure. So, the computer could be your computer could be theinfrastructure part. But also, in this case a AWS batch function is part of the infrastructure and a storage component like a database is part of the infrastructure. And if you have a data engineer or data analyst, they can do good things in cleaning data.

{ 16:53 }

The machine learning experts know what to do with the data but neither of them can function if you don't have any street to transport the data on and that's what the person for the infrastructure creates. So yeah, for machine learning case in the industry, you kind of want to have expertise in those three and the very least and they had that. Although it's very uncommon that, for example for them there was 2 software developers that did back-end things in before. So, they were capable of doing all these ML OPS things, which means they built. We can create the infrastructure as necessary for for machine learning. But they came from more traditional background like everyone does because there's very little people that create infrastructure but only do it for machine learning application. That's just too new for that to be a thing already, not long enough for people to have that in their CVs.

{ 17:51 }

So I think in general in the industry, if people say they do infrastructure stuff but they're specialized in machine learning that's very rare thing, people search for it. A lot more common to train your software developers towards becoming machine learning experts as well.

Author: { 18:15 }

OK. And more about the project by itself and how it ran. So how would you say was the communication with the with the client? With Y. Was it smooth? Did you have any troubles? How often maybe did you meet?

MLE: { 18:36 }

Generally, very smooth because we had an already good connection to the CTO of the company through one of X main stakeholders who was also an investor at Y. So there was already a pre-established relationship there and it's honestly CTO was just a very nice guy, called Guy.

{ 18:59 }

On top of that we always try to well we have to collaborate with our clients. So we had quite intense communication but also interactions with the 2 developers that I talked about which are like back-end developers let's say and the two data site, the two people, the data team which are now set as machine learning experts in data scientists basically. So those four people were our main correspondents and we had weekly meetings with them like twice a week, deep-dive really working together on something plus of course open communication channels that either could ask questions together all the time and that was what it was really needed and useful.

{ 19:50 }

And then there was also the week, like the daily startup settings, startup meetings, which is just an agile way of working, but the whole team work meets in the morning and talks about what they did yesterday, what they will do today and and then have a scrum master to to do all this. It's very typical. And yeah, for me, I thought that was mostly useless to our engineering work. So I participated, but I also often working on onour work already because I was just like, yeah, it was just unnecessary for my work to learn what the bio marine scientist would write into their paper today or not. But yeah, it was just part of our work that we like to you know really show collaboration as many fronts as possible and like be there for them to ask potentially questions about us.

{ 20:57 }

And therefore it was it was quite natural for us to adopt their Agile working as well and join those meetings too.

Author: { 21:07 }

OK. Yeah. Thank you. So you said earlier that, yeah, the first phase was the audit, but then there was also kind of the development phase which then turned out to bemore of, as we said, improvement of the already well done model with the cloud component. Was there anything else that you did after the audit, that you guys implemented the plan blue?

MLE: { 21:39 }

Yeah. Most of what we did was more related to infrastructure and analogs.

{ 21:45 }

So what we did was we created a unified infrastructure with the tool called Terraform and this is a infrastructure as code tool, which means you create one script which then once you run it deploys all the infrastructure. So you can everything you do in that script you can do manually and that's usually how it is. And that way it would be you go to WS, so Amazon Web Services and you click on EC2 instance and then it will create you a compute node and then you click on the database API from them and then that will create you a database. And then you click a lot of other menus in the console tocreate the entire visitor bit by bit and always try to connect them as you do it.

{ 22:43 }

The infrastructure as code is a little different, where you write one code and say like, I need a database, I need AEC 2 instance, I also need an API gateway on top of it and the VPC around it and all those lovely components that most people are completely unaware of being necessary for everything we do in applications and Internet and so on. What not, yeah. And then once you've written that entire script, you, you run it and it will deploy it all, which means that the next time a new person or they want to create a new environment of infrastructure in a cloud, they have already script that they can just copy and do again and can immediately do that. And that ties into what I earlier said about companies. When they scale up, you want to make things standardized so that it's easier to on and off-board people and to in general scale up production. Because yeah, growing your team often just means you have more chaos. And more chaos means you want to make things easier and thereby kind of reduce the chaos again.

Author: { 23:54 }

OK, so basically the development in the infrastructure that you did was to create this Terraform script which is the hardest part to do at first and then once it's done it's easier and it's smoother to run than other options.

MLE: { 24:11 }

Yeah, indeed that's the biggest thing. But then there were also a lot of components that we just added as like you said, the wrapping the machine learning case, thealgorithm into a AWS Batch function. AWS Batch itself is also part of the infrastructure. So that way you see how these things are really connected like that is something that didn't have and we created for them by writing into that script that I said about talked about and that therefore supercharges there and I'll use case with the algorithm let's say. So it's really, yeah, interwoven. Although we talk about these as three different concepts, they're very much overlapping.

Author: { 24:55 }

OK. Yeah.

{ 24:56 }

Now question which is a bit more generic in the whole sector, about the whole sector; so how much added value do you think machine learning solutions can bring to acompany like Y for a company like that in the green tech industry? I'm saying machinelearning can be an overall game changer or more something like partial improvements, details, definition or maybe just same level of performance and just simply moreflexibility and more adaptability or scalability.

MLE: { 25:36 }

Yeah, that is very general. In general, I would say it always can be a game changer for certain. It really depends on your industry of, I mean of your company. The industry Green tech sure, generally works as well, but the greatest,

{ 26:00 }

so how big the impact of machine learning really is in the end, it’s often how many tasks do you have in your company that are easy to automate or that can be automated at all. And then yeah, how many of those tasks have to and how do you have and how many, how important are there for your final thing. So with Plan Blue, it was quite massive. Like if they had, they could create their product by having 10 specialists working full time on labeling images into seagrass and no seagrass, right? And they would have come up with the Finder product, which is the same.

{ 26:45 }

So by training a machine learning algorithm to do that instead, to label things to be seagrass and no seagrass, you know, I I mean, I haven't really done the math, but so The thing is, instead of having 10 people work full time for a couple years, you just need to have one person work for a couple weeks to label enough data to train the model. And then you don't even need that person for this task anymore. But that person can then, you know, write white papers or do things that machine learning algorithms just don't do as well or shouldn't do, let's say. Which is like doing, bringing for since that person's obviously an expert on seagrass, you know. That person can then focus on doing things that really need expertise, writing papers, looking at things, advancing the general approach of Y. So that changes the entire, like employ 10 people for infinitely long, as long as you want to have polar or have one person for a little while. And then you have a model that can do the job of not just 10 but 100 people easily because you can just scale it up once the model is good. You can. Yeah.

{ 27:59 }

So it's just, it really depends in GreenTech, what kind of process you have in your pipeline that can be given to a machine. Well, that will come down to, you know, is it a computer vision application? Then something like this where you can label things are very very common. If it's more about speech recognition or so, then there are models for this. If you have for example a different case where you let's say would have more than enough data, you have like way too much data in fact. And you have only a few experts. But you know, you can't process that, all that data with the machine learning algorithm because it really requires some human touch, which is, you know, quite often the case basically in everything so far.

{ 28:57 }

And then minus the few things that machine learning has done so far, that still can be really help with machine learning. Because although you don't want the task to be done machine learning, if you have too much data or too much things that you could be working on, then machine learning might be a really good case to funnel and shift through this amount of data that you could be working on and only present the most valuable cases towards the human worker that actually works on them. So there's a lot of those are just two ways, right? How machine learning can really supercharge your your business and it is, yes. That's also why our job is a job at the moment, because thinking about where can machine learning actually bring value is a is a task in itself. But yeah, in most cases there is something that can be optimized.

Author: { 29:51 }

OK.

{ 29:52 }

So it's it's not too much about, yeah, the GreenTech industry by itself, it's more about the task that each company does, right? So how much you can automate. And so the connection to green tech might be the computer vision. I will say there is, there are alot of green tech companies that work with computer vision or satellite imaging. And so those types of companies can benefit the most maybe from machine learning implementation.

MLE: { 30:24 }

That's probably too general. In GreenTech there's sub-sectors called Earth Observation and that one is very yeah, it deals it. Hence machine machine learning. Computer vision tasks deal really well or lend itself really well to this sub task Earth observation. Because Earth Observation always has imagery in there. You want to recognize things in that imagery and you usually have a lot of images so that one can really have computer vision.

{ 31:00 }

In other cases, you maybe want to have an electrical grid, and electrical grids themselves work by the way to… They have the energy, they try to bring it to where it's needed and it's a lot of play. It's really a game to bring your energy where it's needed so that you don't have excess energy that you have to either throw away or sell on the cheaper or try to go and sort. So their machine learning is really good because it's such acomplex system, that nobody can. Humans get really good at it of course, because they get this gap feeling and they have a very good processing unit of themselves let's say. But machine learning can take these very complex tasks as well, really good, and then predict where will this energy be needed. So that's a different thing in computer in GreenTech which has needs prediction algorithms but not computer vision algorithms.

{ 31:53 }

So you cannot say and I certainly wouldn't know which kind of algorithm can help which industry the best or which part of green tech would have the greatest benefit from machine learning. I really think basically everyone working in almost every industry can can get benefit from machine learning and green tech is just certain things have already been figured out to work really well. Computer vision and and earth observation. Let's say predictive algorithms in energy trade business or energy market market systems. What else are green tech companies? Probably things to observe emissions from cars and other industries. That's again computer vision. Yeah. Or predictive algorithms. Yeah, I don't know. Tell me an industry and I can philosophize how potentially a machine learning algorithm works.

Author: { 32:57 }

Right, yeah.

MLE: { 33:01 }

Sorry, it doesn't really lend to a nice general statement

Author: { 33:04 }

I mean it is like this, yeah. Then this is about Y, going back to Y, but it's also quite more about machine learning implementation in general. Say, what parameter did you consider in the Plan Blue case, and therefore what are the parameters in similar cases that can be considered to assess if the implementation was helpful or not?

MLE: { 33:41 }

You mean the AI application I suppose? Particularly.

Author: { 33:47 }

Yes, I mean after the audit and the development and the whole cooperation, how did you guys from X assess that it was a success or not? Like what parameters did you consider?

MLE: { 33:58 }

Of the… now by the project… do we assess whether the machine learning case is a success or do we assess whether our collaboration with Plan Blue was a success?

Author: { 34:10 }

I will say I was talking about the project, but I guess also the collaboration is very much linked to it.

MLE: { 34:16 }

Yeah, I mean, OK, collaboration as well. So OK, collaboration object. Well, from our perspective of course it is how happy are they, like how happy is the CTO with what we build and how happy are they engineers and how much do they beg us to stay and continue working on basically. But that's very, very individual of course.

{ 34:39 }

I guess we evaluated by seeing the before and after picture and the parameters there would be how much data can be processed in how little time or in parallel from Y, which is now much more than it was before. It was limited to how much could Yara's machine single computer run, and now it can be as much as they want to spend money on providing the heavy computing machines in the cloud.

MLE: { 35:16 }

So that's really a change. And then how fast does a data product goes from development into production? Would be another?

Author: { 35:30 }

OK.

MLE: { 35:31 }

So when they create a new model, how long does it take for them to train it to find out which one is the best model and then use that model in production?

{ 35:39 }

Which is, yeah, supercharged by using ML flow for a model, registry, model tracking, and then a version control system to link development, the development environment to the production environment, and then some kind of software to trigger or orchestrate these working processes, which in this case was Air Flow. So that all makes it faster for them to just iterate on models and get them into production. I think those two are probably, how much can they process and how fast can they create new models that can process these things. It's kind of the most important.

Author: { 36:25 }

And both of those parameters that you mentioned were improved?

MLE: { 36:30 }

Yeah, Oh yeah. Massively, because that was our whole focus, to help them scale up production. So that was our goal, to make things, yeah, capable, capable or scalable. So work on more easy-to-maintain, robust infrastructure that more people can work on or can deal with the higher workload.

{ 36:57 }

And yeah, centralized things so a bigger team could collaborate.

Author: { 37:03 }

OK. So I have, yeah, we'll say one last question. So yeah, well I had another question but you kind of answered that already more or less. So how important it is to have a well-built data team, data infrastructure and machine learning implementation process or machine learning structure in a GreenTech company. But you already kind ofmentioned this in the other answers, so it is quite crucial as of now. So my last questionis going to be if there is anything that you would have done differently about the project, like possible improvements in a future project similar to this with company X that is similar to plan Blue.

MLE: { 38:00 }

Yeah, interesting question. What depends on the on the company of course, in terms of Y, they had a already well set up data team. So I guess in a different case we would maybe focus more on advising them on what kind of expertise they want to have either in-house or out of outside the company from experts. But the, yeah, creation of the data team and the skills that they would require to maintain a machine learning algorithm in production, that is, yeah, that is what's unusually well defined already in Y.

And I'm sure in a different project that would have a stronger focus, I also think it would be well.

{ 38:58 }

And then we have to go through the other things again, check the machine algorithm presumably because I think Y was a little bit of a special case there. They won't, there will be easier improvements to the machine learning algorithm itself aswell. So then we would have a stronger focus on that, on that as well. Yeah, I think that is what I would expect from a different case that those things that were now unusually good already will probably become stronger focus.

Author: { 39:30 }

OK. So in in the case of Y is actually not much you would have done differently because Y by itself was already quite good and you guys did all you could.

MLE: { 39:41 }

Yeah, honestly I wouldn't change much in the case of Y. Maybe it's also too early to say, yeah, because if they now come back to us and like have you and say like what you build didn't work out then, of course, I know.

{ 39:57 }

But now we just created what we thought was best and we're really happy with it. So it's hard to come up with things now that could have been done better because otherwise we would have probably done that and I would have already come up with them.

Author: { 40:11 }

Right. OK. Yeah. Well, that was it. Thank you so much for your time of course.

MLE: { 40:20 }

Of course.

--- End of transcript ---

ATTACHMENT 2 – Interview Y

CTO - Interview

Transcribed on: 13/03/2024, 18:09:21

Total recording length: 00:14:28

In this transcript, the Chief Technology Officer is referred to as “CTO” and the Author as “Author”

Author:

Hi, CTO. How are you?

CTO:

Hi

Author:

Okay, yeah, so I will try to be quick and I'll explain to you what... what ishappening. So basically I am writing, as I already mentioned to you, I'm writing my master thesis from the University of Lisbon and I'm here on exchange in Berlin, so I'm finishing here.

……………………………………………………………….

Author:

How do you retrieve the data that you're going to to implement?

CTO:

Sorry, how do we retrieve the data?

Author:

Yeah that then you're gonna use for for generate your your product. So the data that you launch into the whole machine learning and artificial intelligence…

CTO:

We gathered them ourselves, we collect them ourselves.

Author:

OK, yes

CTO:

Those are our own data.

Author:

Yes. And what are the techniques? How do you…through divers? Is that the main source of data?

CTO:

For now, yes, it goes a long way to explain that, and it's because it's not divers, it's a payload system that we carry on, on a on a system that can be diver operated now, but can also be attached to underwater vehicle.

Author:

OK.

OK then. Well, OK. I will ask you a question be more, yeah, as we said more..more general. As a CTO of a green tech company, how much important you think it is to have well-built data team and data infrastructure in the sector? Do you think it does make a difference?

CTO:

To have a data team specifically in green tech?

Author:

Yes, yes. A well-built yes. How much does it count? How much does it matter?

CTO:

Well, we need to process a large amount of data, so you need a data team and it's a very specific problem. It's not, especially in green tech, I think the problems are very specific about another LLM or whatever, all this good AI hike that's going on everywhere.

But the data that you cannot find, that there are no data sets on the Internet that you can you can use to pre train your models or something. So you need a data team focused on your data, and it's not something you can get off the shelf somewhere else and you have to develop it yourself. So you need a data team. Without a data team you cannot do anything.

Author:

OK. Yeah. It's mandatory. You cannot… You have to customize. Definitely. Yeah. Thank you. Then… So when you set the whole… Yeah. Here, sorry, here I'm trying to cut out all the questions that are more related to the cooperation, so we can leave those out.

OK, Yeah.

What do you think were the major difficulties in in starting a green tech company a and setting up… also on the IT side? What do you think were the major difficulties? Maybe you had obstacles from regulations or, I don't know, whatever.

CTO:

No it's mostly getting that the investment climate for green tech is not that easy. There's a lot of investors that say they are impact investors but they're not. They still want this… The these are still aimed at generating money and then green tech just not always immediately… it's not always clear where money can be earned and that makes that makes it difficult as the green tech startups sometimes.

Author:

OK.

So the difficulty relies on the first phase also, to get the the whole company starting because as you said the investors…

CTO:

Also later, right? It's like getting funding into something. Well, every startupdepends on funding. Without funding you can operate and that, and that is not always easy.

Author:

OK. Yeah. Yeah. OK..

CTO:

But that that really depends. That really depends on the on the startup. We are in marine, not just really but in the marine off-stage we operate, that is difficult.

Author:

OK. Because of the specificy. Yeah, makes sense.

OK. So, yeah, I think it should be… Thank you for your…

……………………………………..

Author:

Thank you so much… thank you so much, bye bye.

--- End of transcript ---

ATTACHMENT 3 – Images

Figure 1 – The SDGs in the Blue Economy (Lee et al., 2020)

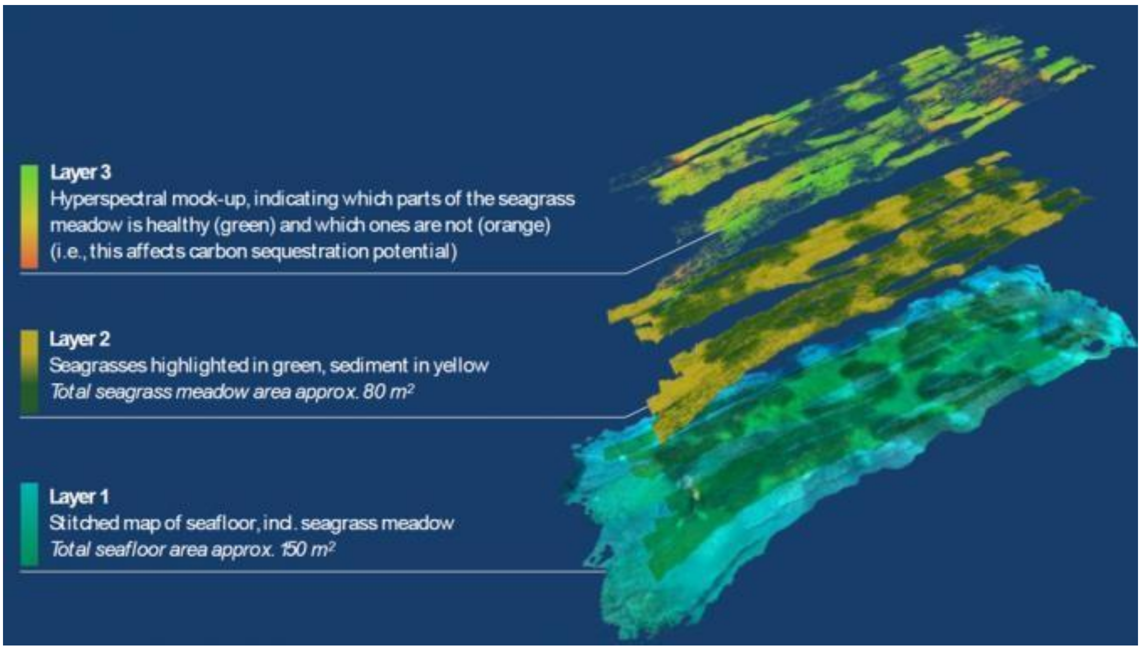

Figure 2 - Example of seagrass meadows UHI