17-01-2021
Final Version

GRADUATION THESIS

AgroVision
AGRICULTURAL SOFTWARE

What are the best AI solutions for AgroVision to implement in their data warehouse for better analysis and prediction of the data?

By: Amots Oko, 436647@student.saxion.nl
Agrovision Supervisor: Ignat Fisser
Saxion Supervisors: Ate Zuithoff, Jos van de Pol

Executive Summary

The main problem of this research revolved around Agrovision not using Machine Learning (ML) to analyse the data which is received from their clients (the farmers). The absence of ML analytics at Agrovision meant that the farmers could not work efficiently and effectively in an industry coming under increasing constraints. The objective was to investigate the most appropriate ML solution according to Agrovision's requirements, test it, evaluate it and finally implement it, in order to enhance Agrovision's value proposition and add business value to the farmers.

The ICT Research Methods was the principal methodology utilised during the research. The different activities encapsulated in this methodology provided comprehensive measures to gain relevant knowledge. Additionally, this methodology offered a framework for the research to be conducted in because of its triangulations and validations abilities.

The research resulted in finding the most suitable ML solution for Agrovision. This was followed by an *Implementation Plan* that would have supported integrating the new ML solution into Agrovision's software architecture and a *Change Plan* which would have abated any risks associated with this integration. Unfortunately, the project's implementation phase did not occur, but the research and accompanying plans have ensured the ease of future realisation.

The project's outcome gives a clear positive answer to the main question – a solution was found which has the maximum compatibility with Agrovision's requirements, an implementation plan was created to ensure alignment of all aspects and a change plan was depicted to negate the risks involved. This has laid a solid foundation for its future continuation and objective.

Table of Content

1. Introduction

Artificial Intelligence (AI) and ML stem back from the 1950s (Marr, 2016) but has only been widely used in the last decade. Many organisations have employed or are exploring the possibility to use AI & ML. Analysis such as Big Data, Multi-Variable or Prediction are to mention just a few. These days many organisations collect large amounts of data from many different data points as Internet of Things (IoT) becomes more extensively utilised. However, more traditional Business Intelligence (BI) analysis cannot handle these amounts of data or the large quantities of variables. ML analytics uses algorithms to search for patterns in the data, giving insight into business questions, thus creating Business Value.

Agrovision has been a software provider to the Agricultural sector for over thirty years. As part of their value proposition they also offer an analytics service, but this did not include ML. The lack of ML analytics can potentially hinder their clients' progress as they might fall behind other farmers who do benefit from the insight of ML analytics. This situation has the potential to force Agrovision's clients to search for alternative solutions that include ML. Agrovision has recognised the need to incorporate ML analytics as part of their value proposition and has agreed to create this project to bridge the gap.

The main research question of this project is – "What are the best AI solutions for AgroVision to implement in their DWH for better analysis and prediction of the data?" This main question's objectives were to find and implement an ML analytics solution that can be assimilated into the current Agrovision enterprise architecture and provide its customers with additional business value.

The methodology used in this project was the ICT Research Methods. These methods provided a framework and guidelines to conduct the research as well as validate the findings. Each method has a set of activities adapted explicitly to conducting research in IT. All the five methods were employed during this research, demonstrating its versatility and the validity of the results.

This thesis initially states the beginning situation of the project and the problem that gave rise to this project. It continues with the research question and the deliverables. The organisational context and theoretical framework follow that. Chapter four specifies what research methods were used for each sub-question and the project's phasing. Chapter five contains the results of the project per sub-question, which indicate the actual process the project has undertaken. The last chapters comprise of the conclusions, recommendations and discussions.

1.1. Problem Analysis

As the world population continues to grow, the need to produce more food grows with it, but the areas for this food production are becoming smaller. There is a clear need to extract more yield from every cultivated land section. One way of doing so is through ML analytics which gives a better analysis of the land and its return than the more traditional analysis used before. Whether it is used for decreasing diseases, forecasting the weather, forecasting the coming crop or the number of offspring to be born, ML analytics has the ability to transform a farm to becoming ultra-effective and efficient. For these reasons, Agrovision would like to include ML analytics as part of their value proposition.

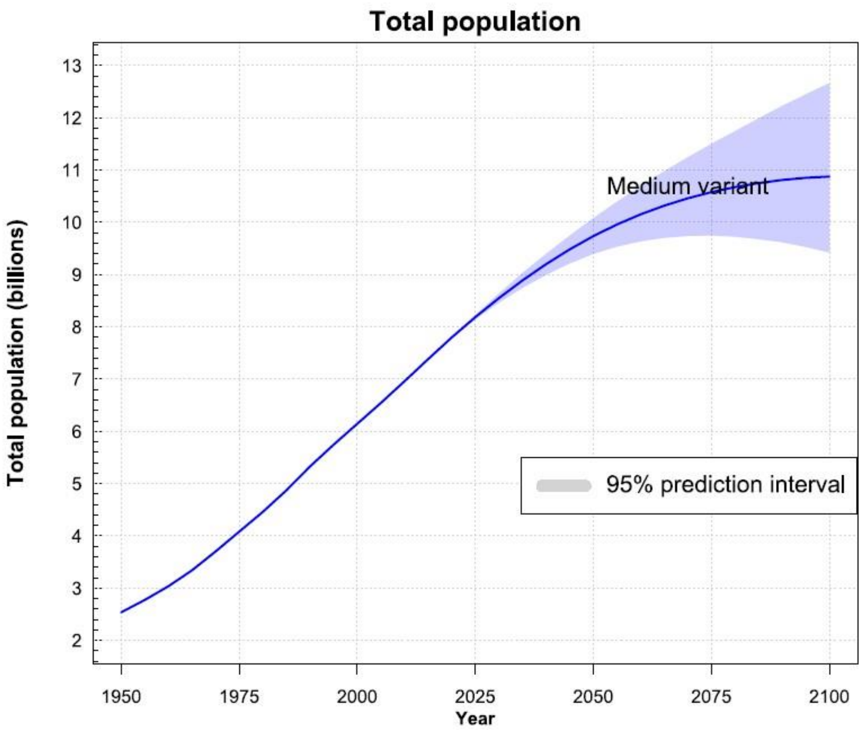

Figure 1 – world population prediction according to the United Nations (United Nations world population prospects 2019, n.d.)

<u>Starting Situation:</u>
At the start of the project, Agrovision was collecting data from many different enterprises through the different applications they provide. The pig husbandry applications data was collected in a database from which a data-warehouse (DHW) was created for analytics purposes.

<u>Analysis Disadvantages:</u>
Analysis of the data was being done at Agrovision, but there was no use of ML analytics. Furthermore, ML analytics has not been researched or applied. The analysis at Agrovision is done using Microsoft Power BI, but there were some disadvantages. On top of that, there is also an analytics section in PigVision (a software supplied to the pig-husbandry industry by

Agrovision), but that has its disadvantages (discussed in 5.2.). One of the most significant disadvantages being that no predictions could be made. Only historical results are a substantial limitation as there is no demonstrated implication on the forthcoming future. As there was a demand for this from the Agrovision customers, the most apparent solution was an ML application.

Analysis Requirements:

From Agrovision's client perspective, there is a great need for ML analytics. The farmers would like to know, within reasonable certainty, what will occur in the near future at their farm. Issues such as sows coming in heat, disease management and feed management are just some of the predictions ML analytics can provide, but at the start of the project were not being done. From Agrovision's perspective, there is a service which is demanded by its clients and is being offered by other enterprises around the world and more specifically by its competitors.

Problem Description:

The situation described above creates a problem for both Agrovision and the farmers. The farmers would like to continue working with Agrovision, but are falling behind compared to other farmers who do have the ML analytics service. Agrovision is facing the prospect of losing clients who would like to have ML analytics but are not receiving it. This problem has been affecting both sides since Agrovision's competitors, such as Afimilk (Afimilk, n.d.) or Allflex (Allflex, n.d.), started offering ML analysis as a service.

Recent Agrovision Innovations:

Agrovision has been aware of this predicament for some time now and recently have started taking steps towards alleviating it. In the past twelve months approximately, Agrovision started implementing new databases for the pig-husbandry. These databases are called ADEX (Agrovision Data Exchange). A plan is slowly being implemented for ADEX pig (one of the ADEXs being implemented) to collect data from all the pig-husbandry solutions, which Agrovision offers. On top of ADEX, a DHW was created for analytics purposes. Because of these steps and the growing ML analytics demands, the time was right to begin researching and eventually implementing, an ML analytics solution at Agrovision.

Expected Impact of ML Solution:

The new ML analysis would answer questions such as: should a particular type of feed prove more effective than others in the health of the animals or their growth rate? Another example is the effect of weather on the animals - what is the optimum temperature, humidity or rainfall? For example, a known fact is that colder climate yields more milk in dairy Friesian cows (Oko, 2020).

AI and ML analytics are becoming more widely used than ever before. Agrovision has recognised this business need and needed to make sure they are a step ahead from most of their competition by applying it as part of their value proposition. This will give the farmers the ability to work more effectively and efficiently to make sure they can compete in the ever-evolving field of agriculture.

1.1.1. Problem Statement

Agrovision would like to know which ML solutions can be implemented on their DWH to better analyse and predict the data, creating a better level of service by Agrovision and, subsequently, more added value to their clients.

1.2. Main & Sub-Questions

The main and sub-questions in a project define what the research is about and what is being found out. The main question is formed at the beginning of a research project because it stipulates its principal goal. It is the project's core and serves as its primary driver. The sub-questions are also defined at the beginning of the project. These sub-questions define and describe the project's objectives, deliverables and work methods. In this project, the sub-questions' order delineates the project's progression, as one sub-questions is the direct result or is the obvious continuation of the preceding sub-question. The sum of all the sub-questions is the answer to the main question:

Main question:
What are the best AI solutions for AgroVision to implement in their DWH for better analysis and prediction of the data?

Sub-questions:
1) Which AI solutions exist in the market? Research into the available solutions in the market based on initial requirements from Agrovision. This provided a narrowed down list of potential solutions, out of an excessively substantial number of existing ones.
2) What are the limitations of the existing Agrovision data analysis solutions? This question was merged with question three because the existing solution's limitations are precisely the new solutions' requirements. This helped define the existing problems and obtain the potential solution's requirements. It acted as the guideline for choosing the most appropriate solution.
3) What are the requirements of Agrovision and their clients from such solutions? As stated before, this sub-question was merged with sub-questions two.
4) Based on solution characteristics and stakeholders' requirements, which AI solution is the most suitable for Agrovision? This sub-question was the direct result of the sub-questions before it. This is the decision on the best ML solution for Agrovision with which Agrovision continued working with.

5) What is the best architectural software design for the new AI solution prototype? In order to implement the new ML solution, an architectural design had to be made. According to this design, the solution's implementation was done (among other criteria). The result of this sub-question is the implementation plan.
6) What is the best change strategy for the different business processes which will be affected by the new solution? There is an inevitable change and resistance to this change when implementing new software. A change strategy was formed to mitigate this resistance and inform and educate all personnel concerning this software.
7) How to implement the chosen AI solutions in the Agrovision DWH according to the change strategy and architectural design? This is the last step of the project, as it is the result of all the research done previously. It is the implementation of the chosen ML solution according to the implementation plan and in consultation with the change plan.

1.3. Deliverables - Agrovision

Sub-question one – a list of the different researched solutions, their characteristics, advantages and disadvantages.
Sub-question two – a record of the disadvantages of the current analysis system.
Sub-question three – a list of requirements from the new ML solution.
Sub-question four – the chosen ML solution based on its characteristics and the Agrovision requirements.
Sub-question five – an implementation plan for the chosen solution. This plan will include the new software architectural design, IT governance, IT alignment, and other new solution aspects.
Sub-question six – a change report that, together with the implementation plan, will serve as the guideline for installing the new ML solution.
Sub-question seven – actual implementation of the new ML solution.

1.4. Deliverables – Saxion

- Plan of Approach – a document that defined the project's different variables, such as subject, goals, deliverables, scope and risks, and the work to have been carried out.
- Concept graduation report – an interim deliverable of the project report. The purpose was to receive feedback of report work carried out until that point. Improvements were made to the report according to this feedback.
- Final graduation report – the final report handed in at the end of the project. This report will be checked and will be graded. This grade will later be consulted for the final graduation grade.
- Reflection – a personal and professional reflection of the graduate about the project. The professional reflection was part of the final graduation report and the personal reflection was a separate document. It was done at the end of the project.

- Presentation – a presentation to the different stakeholders as well as other interested parties of the project. Was done at the end of the project and performed as a summary of the work carried out and the project’s conclusions.

1.5. Summary

This chapter contains the project’s initial definitions – why does this project exist (problem analysis) and the plan to solve this problem (research questions) is part of it. The other part consists of all the agreed deliverables to the project's stakeholders. These were all defined before the project commenced in the Plan of Approach (Oko, 2020). The next chapter describes the *organisation* within which the project will take place.

2. Organisational Context

The purpose of defining the organisational context is to understand what the reasons are behind Agrovision's structure, strategy, and other different aspects that concern this project. Moreover, once defining all these aspects, the changes that need to occur because of this project's objectives are clear and obvious.

Agrovision Value Proposition:

Agrovision was founded in 1986 as a software company, developing software for the agricultural sector. Over the years, Agrovision has bought more enterprises which were integrated into the main company. Nowadays, Agrovision makes software for dairy cattle, pig-husbandry, poultry and crops in thirty locations worldwide. Furthermore, Agrovision offers finance software for the agricultural sector with which farmers and accounting firms can work together optimally (Over ons - Agrovision, n.d.). Finally, Agrovision offers software for what the company calls AgriBusiness - the businesses which buy the products of the agricultural sector (for example Albert Heijn). This software is developed specifically for the AgriBusiness. It provides full traceability of the purchased products.

Agrovision has 175 employees in three locations in Europe:

- Deventer, Netherlands (also the head office)
- Oudenaarde, Belgium
- Tørring, Denmark

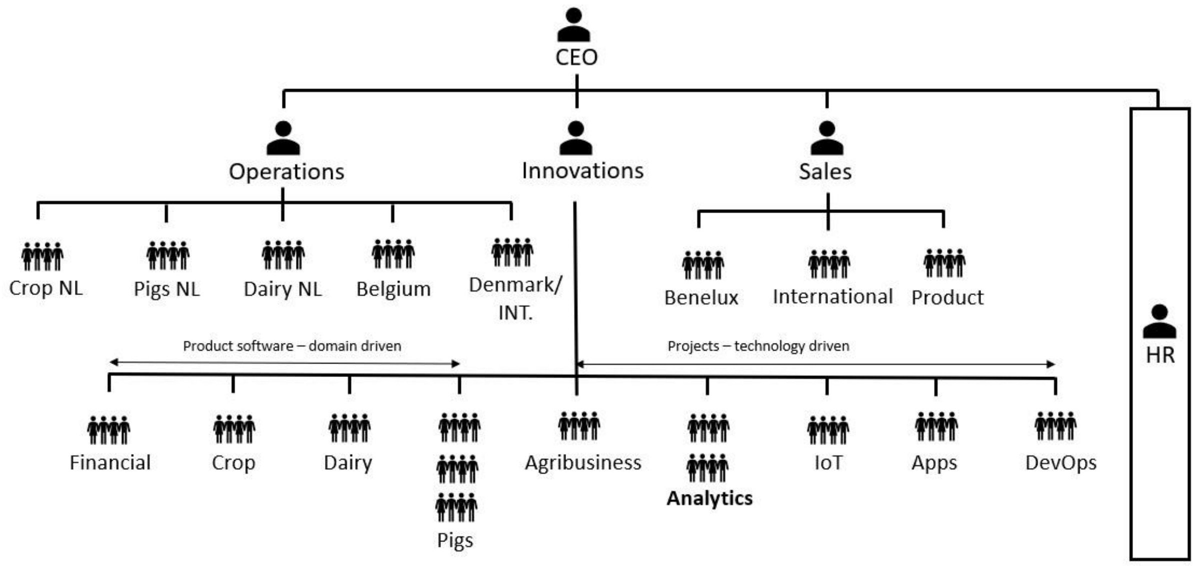

Figure 2 – organisational chart Agrovision

Agrovision’s Structure:
As shown in the chart above, Agrovision has a relatively flat structure. According to Jacob Morgan from Forbes Magazine (Morgan, 2015), this is the most common organisational structure today. There is a strong emphasis on collaboration and communication in Agrovision, which serves the need to develop and maintain new software. Management at Agrovision is there to support the employees and power of authority is pushed down instead of pushing down orders or communication messages. This structure, for all the above reasons, enables Agrovision to be an innovative company which can maintain its core value proposition, but at the same time continue to improve on in and even integrate new enterprises with relative ease (something which has happened in the past and, no doubt, will happen in the future).

The Innovations Department:
Because Agrovision is a multinational company, the primary language used is English. The company contains different departments such as management, sales, human resources, operations and innovations. The innovations department oversees product (software) development and maintenance, both internally and externally. The Innovations department is divided into Product and Project teams. The product teams are in charge of the different product which Agrovision offer – Pigs, Dairy, Crops and Finance. Except for the Pigs product which has three teams, each other product has one team. The project teams, which are technology-driven, span different products, such as Analytics (two teams), Mobile App (one team), IoT (one team), Agribusiness (one team) and internal systems (one team). Scrum is the principal framework with which the teams work. The team in which this project was being carried out is one of the analytics teams, in the innovations department, as the assignment is about analysing the data.

Problems Faced and Solutions:
As mentioned before, Agrovision has bought and integrated different applications, but that has created some problems, most notably the data which is received for these applications has no uniformity. For example, in the pig-husbandry sector, Agrovision has four main application (and many other smaller applications) – Pigmanager, FARM, Ceres and PigVision. When the project commenced, the data collected from these applications was in a database created in 2008. For a multitude of reasons, this database no longer served the current and future requirements of Agrovision. A new database needed to be created and all the data needed to be unified. Roughly a year ago, Agrovision created ADEX – Agrovision Data Exchange, a database that collects and unifies the data from the different application. Furthermore, a data warehouse (DWH) was being realised for analysis and reporting purposes. This implementation is still ongoing, but it provided Agrovision and their clients a sustainable solution since then.

The Need for AI:

Moreover, Agrovision has been following the developments of AI in general and ML in more detail because it considered ML as the next step in data analysis. At that point in time, the data analysis was being done, but not using AI. The analysis was lacking in many aspects, which results in insufficient service to the clients. When the project began, there have not been any studies of AI within Agrovision or have there been any implementation of it in the DWH or any other aspect of Agrovision.

As ML has become more widely used, it can serve as a real advantage for the enterprises who use it as a predictive or analytical tool. Agrovision was no exception to this. There was a growing demand from clients, as well as internally, to provide this analysis. The clients will benefit from a much more in-depth analysis of their data as well as predictive abilities, and Agrovision will benefit from a much better service level provided (Oko, 2020).

3. Theoretical Framework

This chapter explores the concepts behind the project. It defines and explains these concepts to make them clear and understandable. The concepts explored here are at the heart of the project and serve as part of its primary focus.

3.1. Artificial Intelligence

Artificial Intelligence (AI) has become one of the most sought-after IT resources in the past few years. According to Thomas W. Dinsmore from DataRobot, the total AI worldwide spending in 2021 is expected to reach $58 billion (Dinsmore, 2018). Until recently, an analysis was done to display past figures and was based on a relatively small number of variables. With advances in computing power, IoT and the growing need to incorporate many more parameters, a more advanced solution was needed. Nowadays, AI is found in many applications - from autonomous cars to marketing, healthcare, manufacturing and many other fields. AI is making a significant change in our lives, even if we do not always realise it. Google uses AI to suggest appropriate advertising to their clients, supermarkets to give a more market suited product assortment and social media to recognise trends.

3.2. Machine Learning

Machine Learning (ML) is a subset of AI. It uses algorithms to learn from data autonomously. ML has only been widely developed and used in the last two decades, though its origins go back as far as the 1950s according to Forbes magazine (Marr, 2016) when Alan Turing created the “Turing Test” which was designed for a computer to fool a human it is also a human. Since the 1950’s many advancements were made with ML, such as the “nearest neighbour” algorithm was written in 1967, IBM’s Deep Blue beats chess world champion in 1997, Microsoft introduced Kinect in 2010 which can track human movement and allows humans to interact with computers and Facebook develops DeepFace in 2014 which can recognise humans on photos the same level as a human can. The definition of ML is - computer algorithms are used to learn from data and information autonomously. In machine learning, computers do not have to be explicitly programmed but can change and improve their algorithms by themselves.

Currently, there are many ML solutions which are available. Some are more specific for one platform or another and some intend to appeal to a broader market share. Some are for smaller databases and some are for ‘big data’. ML has become a very integral part of our daily routine and is probably here to develop further and possibly take an even more substantial foothold in our lives.

3.3. Data Mining

Data mining is a technique used in conjunction with machine learning. Data mining is a way to develop intelligence (i.e. actionable information or knowledge) from data that an organisation collects, organises and stores. It is a process that uses statistical, mathematical and ML techniques to extract and identify useful information and subsequent knowledge (or patterns) from large sets of data. Data Mining extracts data and applies algorithms to search for patterns within the data. These patterns can be in the form of business rules, affinities, correlations, trends or prediction models.

Data Mining is tightly affiliated with many other disciplines, such as statistics, AI, ML, management science, information systems, visualisation and databases. Using all these disciplines, data mining extracts useful information and knowledge from datasets (Sharda et al., 2018).

Figure 3 – data mining environment (Sharda et al., 2018)

3.4. Cultural Typology

During the 1980s scientists started paying attention to the concept of ‘culture’ in organisations. Culture was not given proper consideration because it was deemed self-evident, and factors that already exist in organisations. Organisational culture can be defined as values, believes and hidden assumptions that organisational members have in common. Quinn & Cameron developed the four typologies to understand the cultures that shape and drive many aspects of an organisation. There are four cultures which Quinn & Cameron Recognised (Quinn et al., n.d.):

a) Clan culture – loyalty and traditions hold the organisation. Leaders are mentors or even father figures. Emphasis on long term benefits of human resources. Giving great value to the customers' needs and care for the people. Teamwork, participation and consensus are considered as core ideals.
b) Hierarchy culture – a very formalised and structured working environment. Work is defined by procedures and is efficient, coordinated and organised. Formal rules and policies hold the organisation together. Long term focus on stability and results. Reliable value proposition delivery, smooth planning and low cost is the definition of success.
c) Market culture – result-oriented organisation. The people are competitive and goal seekers. The leaders are tough, demanding and encourage competition. Emphasis on winning holds the organisation together. Long term focus on measurable targets and goals. Success is defined as market share and market penetration.
d) Adhocracy culture – dynamic, entrepreneurial and flexible organisation. Leaders and employees are risk-takers. Experimentation and innovation hold the organisation together. The long-term emphasis on growth and new resources. There is encouragement and freedom to initiate. Delivering the value proposition is considered a success.

This typology is a tool used to analyse an organisation’s cultures, giving insight into their orientation and strategy. One of its uses is in change management to understanding the organisation and therefore assist in defining the necessity of the change.

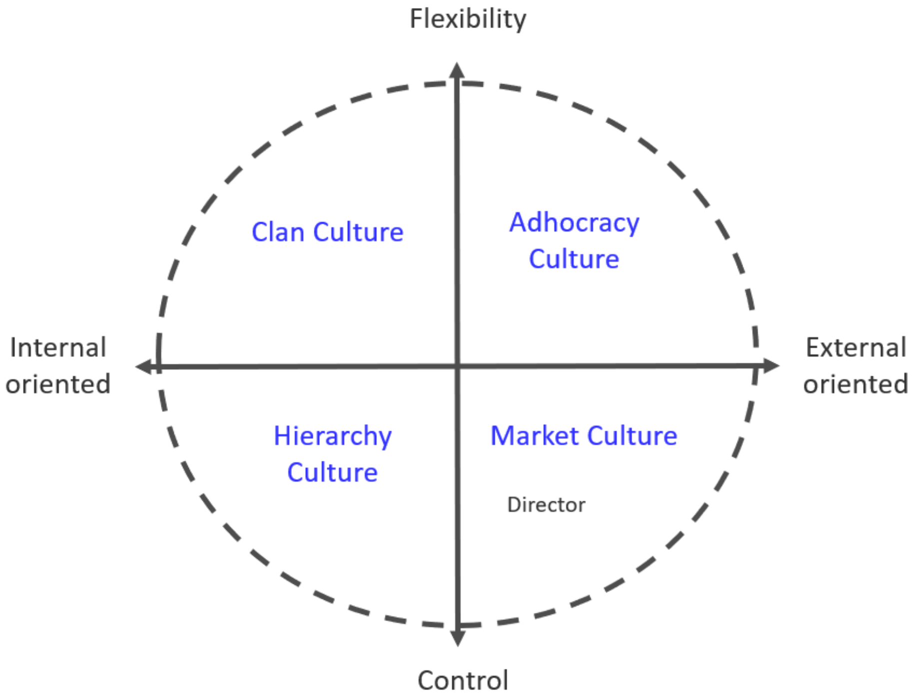

Figure 4 – cultural typology by Quinn & Cameron

3.5. Customer Value Strategy

Michael Treacy and Fred Wiersema formulated three 'strategies of value' delivered to customers. These three strategies can help define which strategy an organisation employs and, should the current strategy prove ineffective, which other strategies might prove more successful. The three strategies are (Treacy et al., 1993):

a) Operational Excellence – the organisation's objective following this strategy is to be the leanest, efficient and effective organisation in the market. These organisations continually strive to reduce overheads and optimise business processes. Moreover, the endeavour to send their product in the highest convenience and at the lowest price.
b) Product Leadership – these organisations' objective is to produce and deliver the best products in the market. These organisations must be highly creative, bring their products to the market as quickly as possible, and continually pursue new and better solutions to their products.
c) Customer Intimacy – these organisations' goal is to deliver the best and most precise product according to the customer's increasingly subtle definition. The cost and effort spent in the present will pay-off in the long run with their customers' loyalty.

When plotting these three strategies on Quinn & Cameron's typology, it looks like the figure below.

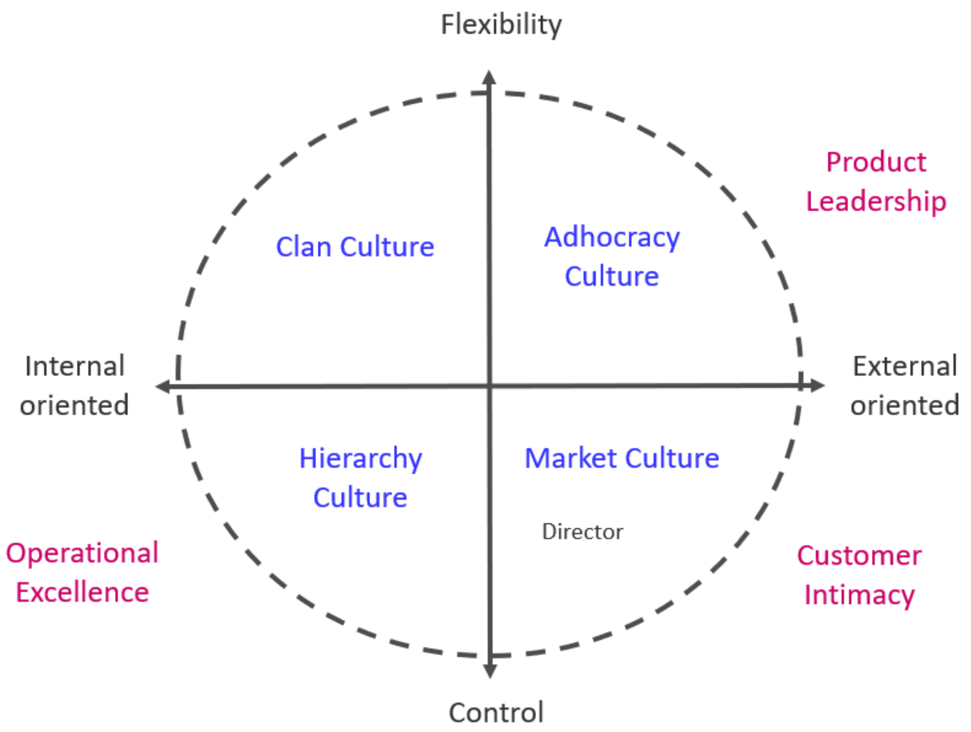

Figure 5 – cultural typology with Treacy & Wiersema Customer Value Strategy

3.6. Summary

As mentioned before, these three concepts of AI, ML and Data Mining formed this project's basis. The initial research revolved around different AI or ML solutions available, it continued into the architectural design of the chosen solution, then culminating with the practical use of Data Mining using the solution. For those reasons, it is crucially important to have a clear understanding of these concepts and the associations between them.

Moreover, the change plan had to be defined with different models to clarify the current and future desired situation better. These two models helped define what needs to change and the methods for the change. The next chapter will address the different *research methods* used during the project and the *project phases* and *research types*.

4. Research Methods per Sub-Question

Why ICT Research Methods:

As the name implies, the ICT Research Methods were made for ICT project where research is involved, which precisely applies to this project. The toolkit offers a set of possible research methods, a framework and guidelines to select the appropriate (combination of) methods. Furthermore, The ICT Research Methods has robust validation capabilities when applying multiple methods. Here is a brief description of each method (Van Turnhout et al., 2013).

Methods Description:

Library – this method is meant to familiarise the researcher with what work has been carried out and the findings. This method's most important values are "review of the literature" and "building on the work of others". This method typically results in an overview of existing work.

Field – this method is designed to get to know the environment the research is conducted in. Organisation, personnel and software are some of the environment variables that this method explores. This method's result is an outline or a detailed picture of the project's environment.

Workshop – this method is about researching possibilities. The emphasis in this method not to rely on past work, but to explore and create new possibilities within the project. Common results of this method are prototyping or developing an existing solution.

Lab – this method is essentially about testing and validating. Whether it is testing an idea, a theory or a piece of software, it is usually to check the tested subject's validity. The results of this method are conclusions and validations through the conducted tests.

Showroom – benchmarking is the main aim of this method - how does the chosen solution compare to others or how do peers view this solution. Giving or getting feedback to another principal aspect of this method after the benchmarking has been done. This results in the justification of the solution, how it may differ from others, or integrate better.

Below is a description of each sub-question and its features, a definition of the research methods which were used (all below described methods and activities are referenced from the following: Bonestroo et al., 2018):

4.1. Which AI solutions exist in the market?

Objective:

This sub-question's objective was *qualitative* research into which ML solutions are available in the market, based on preliminary requirements from Agrovision. This objective resulted in a *report of different ML solutions* available in the market and their characteristics.

Library Method How & Why:

The first method used for this sub-question is the *Library* research method. The work revolved around finding what ML solutions are available in the market which fall within the initial requirements of Agrovision. At first, researching those solutions was done on the internet, which is *Literature Study*. This activity involved finding keywords using a search engine, knowing how to select the essential information and applying the right filters to the search. The keywords were – “machine learning solution providers” on the Google search engine. The applied filters were the provider's size and familiarity with it. The more familiar and more prominent providers were considered first. An additional filter was to eliminate direct competitors of the Agrovision platform. This activity was employed because it is the most suitable for researching large amounts of existing information.

At the same time, *Expert Interviews* were conducted, which helped narrow down the search and gave it more focus. This activity is about finding the right experts, keeping an open mind and using interviewing techniques for asking the right question. Once contact was made with each vendor, more *Expert Interviews* took place with each prospective solution to find out more specific details. If any of the solutions did not adhere to Agrovision’s requirements, it was not considered further. This activity took place because it could narrow down the list of available solutions according to the expert views

Lab Method How & Why:

The second method used is *Lab* and more specifically, *Non-Functional Test* activity. Where available, the solution was downloaded and tested to determine whether it fulfils requirements related to usability, reliability, performance and supportability. In other words, did the usability of this solution conform to Agrovision’s initial requirements? This activity was employed because it was perfect for pitching the new solutions against the requirements.

Validity & Reliability of Findings:

Using these two methods, it was possible to make sure the chosen solutions match up to Agrovision’s needs and therefore, were considered candidates for future use. The *Expert Interview* validates the *Literature Study* activity by ensuring the findings are applicable according to the experts in this field. Furthermore, using the *Non-Functional Test* activity validated the findings even further by using the solution and reaffirming the findings once again.

4.2. What are the limitations of the existing Agrovision data analysis solutions? & What are the requirements of Agrovision and their clients from such solutions?

Objective:
These two sub-questions were merged as it was discovered during the research, they were only one question, meaning the limitations of the existing system are precisely the requirements from the new system (a further explanation can be found in section 6.2). The objective of these sub-questions was qualitative research of what the limitations of the current analysis solutions at Agrovision are and derived from that, the requirements of the new system. It will result in a *report of the shortcomings of the current analysis solution and the new solution's requirements,* which will help determine which is the most suitable ML solution for Agrovision.

Library Method How & Why:
Two methods were used for these sub-questions. The first was *Library,* where *Expert Interviews* were conducted. Experts assist in answering particular questions which are relevant to the topic. In Agrovision there are experts who either create and maintain the analysis system or experts who know what the requirements from such a system are, but not being met. Those experts were interviewed and have provided their insight into the problem. This activity was used because it gave first-hand insight into the different problems from different perspectives, giving rise to the different requirements.

Field Method How & Why:
The second method was *Field,* within which *Interviews* took place. This method is remarkably similar to the previous one, but differs in that the people interviewed are stakeholders who relate to the subject in one way or another, but are not considered experts. For the project, personnel involved in the pig-husbandry within Agrovision were interviewed to better understand where the shortcomings and therefore requirements are. His activity was used because it gave more insight into the problems and requirement from even more perspectives than the previous activity.

Validity & Reliability of Findings:
Using these two methods ensured personnel gave their (expert) opinions from many different perspectives, which resulted in numerous different views and a comprehensive record of disadvantages/requirements. Moreover, because of these different opinions come from different views, the methods used validated each-other, which ensured a high-quality product.

4.3. Based on solution characteristics and stakeholders' requirements, which AI solution is the most suitable for Agrovision?

Objective:
In this sub-question, the objective was *the chosen AI solution* implemented at Agrovision. What is more, this was the direct result of the previous three sub-questions.

Workshop Method How & Why:
Two methods were used for determining this sub-question. The first method was *Workshop*, where tow activities took place:
Initially, it was *Gap Analysis*. Collecting all the different disadvantages of the current analytics system and comparing it to the new ML system's requirements created a clear and obvious plan of where the analytics system was then and where it should be. Following that, was the activity *Multi-Criteria Decision Making*. An implementation of a new ML solution concerned many areas of Agrovision, and as such, many criteria were involved. The ability to process the different criteria was crucial for the right decision making, which was also why a proper relay of the information was crucial (the *Pitch*). This method was used because it allowed the exploration of the different solutions, which provided further insight into how suitable they are for Agrovision.

Showroom Method How & Why:
The second method was *Showroom,* where *Pitch* was utilised. A *Pitch*, which contained all the information collected during the previous sub-questions including recommendations of the best-perceived solution/s, was given to a selected team chosen to make that decision. The *Pitch* activity was used because it enabled a team of experts to make the decision rather than a single person.

Validity & Reliability of Findings:
The three activities, spanning two methods used here, made sure the right decision for Agrovision was taken. The gap between the current and desired situation was defined and a decision was made based on all presented criteria and expert knowledge of the team. Furthermore, the information was relayed in a straightforward manner, which made the justification of the choices much more straightforward.

4.4. What is the best architectural software design for the new AI solution prototype?

Objective:

This sub-question's objective was to design a new IT architecture that included the new ML solution in the Agrovision systems. The product was an architectural software design or an *Implementation Plan*. This implementation plan ensured IT governance was maintained and business & IT alignment continued. This design was the main guideline for later implementation.

Library Method How & Why:

Just like all sub-questions before it, this one employed two research methods. The first was *Library* which consisted of two activities:

Literature Study was conducted to research software architecture in general and research Agrovision documentation to get to know its architecture specifically. In parallel, *Expert Interviews* were done to ask all the necessary questions that arose. Furthermore, there was a need to understand where the new ML solution fitted in the architecture and how it connected to the system's different components. There was a "to and fro" between the two activities as some insights which were made brought rise to other questions which needed answering, and vice-versa. These two activities were used because there was a need to gain knowledge of this subject and because the *Expert Interviews* can validate the *Literature Study*.

Workshop Method How & Why:

The third activity was *IT Architecture Sketching* which is part of the *Workshop* method. As its name implies, it revolved around sketching ideas on how the new ML solution integrated into the current architecture. At the beginning of the design process, this activity took place to have a preliminary idea of the possible locations and what changes needed to occur.

Validity & Reliability of Findings:

A proper methodology is essential during software architectural changed. A small mistake can cause a big problem in the future, so every effort must be taken to avoid it. For that reason, the described activities had to occur, which ensured all procedures were followed and all errors avoided.

4.5. What is the best change strategy for the different business processes which will be affected by the new solution?

Objective:
For this sub-question, the objective was a *change plan* documented in the *change report*. When implementing new software within an enterprise, changes must occur to accommodate it. These changes can be in many forms, for example, work processes, guidelines, personnel restructuring, training and the like. Furthermore, the changes can occur internally (within the enterprise), externally (the external stakeholders associated with the enterprise) or in both. The change report documented all the changes that took place, the risks involved with each change, and all the actions taken to offset them.

Library Method How & Why:
Part of the research into the change plan was done by *Literature Study* – what are the change strategies which are documented in written literature. This activity is within the *Library* research method. The *Literature Study* activity was used because there was a need to gain extra knowledge about the subject.

Field Method How & Why:
On top of the above activities, some *Observations* were done to determine which changes occurred and what risk was involved with each change. Once the changes were identified, a *problem analysis* of each change was done to thoroughly understand its impact. Both of these activities are within the *Field* research method. These observations were done because they helped validate the different change processes.

Validity & Reliability of Findings:
The two different methods were deployed here with a clear plan in mind. *Library* gave an overview of the subject, *Field* gave an overview of the environment and validated the processes. Moreover, both internal and external resources were consulted for the solution to be as comprehensive as possible. All method together validated the findings as each one further reaffirmed the preceding method.

4.6. How to implement the chosen AI solutions in the Agrovision DWH according to the change strategy and architectural design?

Unfortunately, the implementation did not occur because of lack of time (the reasons are explained in the Results section). The methods were still kept here to indicate that a plan was made for the implementation and these methods would have been used according to that plan.

Objective:

The objective of this sub-question would have been to *implement* the chosen ML solution in the Agrovision environment. The main implementation guideline was the Implementation Plan and accompanying it was the change plan.

Lab Method How & Why:

As all the research was done prior to this, it was not a pure research activity, nevertheless many different methods needed to be employed. A big part of the *Lab* method is testing, which had to have taken place during implementation – *Security Tests* would have been done as well as *System Tests*, *Usability Tests* and *Component Tests*. On top of that *Hardware Validation* would have been done - even though the solution would have been in the cloud, the payment is calculated (amongst other factors) by how many CPU's (hardware) would have been used and for how many hours per day. These different tests would have been used because that was the best technique to get a high-quality product.

Workshop Method How & Why:

Another activity which took place is *Requirement Prioritisation*, part of the *Workshop* method. As there were many requirements from the new solution, a prioritisation was made of which ones would have been implemented first according to the users and expert contribution. This activity was used, as mentioned before, because there was a need to decide which element would have been implemented first.

Library Method How & Why:

Furthermore, *Expert Interviews* (*Library* method) took place to ensure the right architecture would have been implemented according to the plan. If there had been any necessary changes, the same experts would have assisted in making sure those changes still adhere to the enterprise architecture. This method would have been used because of its validating factors and the use of the experts' knowledge.

Validity & Reliability of Findings:
Even though all the research was done before this stage, there were still many activities which would have taken place. The number of activities would have been even more extensive than the previous sub-questions, and that was to make sure the implementation was done correctly and adhered to the previously written documents (implementation and change plans). Moreover, three methods would have been involved in this sub-question, which was for the same reason as the activities – to ensure the implementation was done as efficiently as possible. Lastly, all activities would have been repeated more than once, which would have created a Deming Cycle, which enabled to validate every step and continuously improve the implementation.

4.7. Triangulation and Quality Control

Because different methods are used in the different phases, each method validates the other as they are different by design. When using at least three research methods, it creates a validation process. Each method has its own research technique or approaches the research from a different angle. These different techniques verify each-other through this different approach and through different results. For example, creating a *Prototype* that is part of Showroom can validate a Lab test, which can validate a *Library* research. Even though not all phases used three methods, the different combinations of the complete project methods are the validation factors. This type of validation ensures the quality of the research, and therefore the quality of the results, are kept at the highest level.

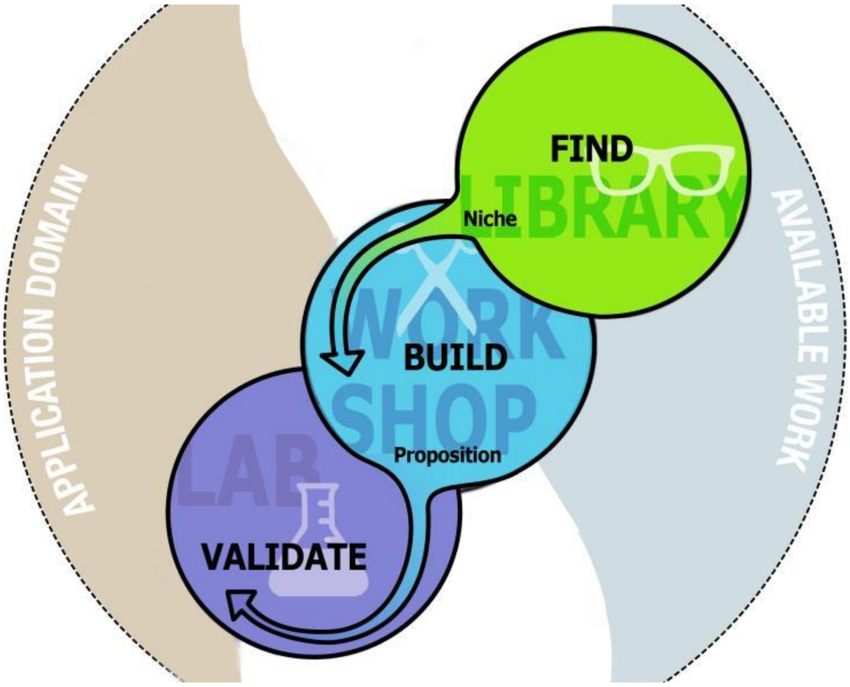

Figure 6 – ICT triangulation validation using three methods

4.8. Project Phasing

The process of the project and its methodology were fundamental because they defined how the process was conducted and which frameworks were employed to execute these processes. Below is a detailed explanation of the phasing, methodology and deliverables according to the phases (sub-questions per phase) and according to the stakeholders (Agrovision and Saxion).

Almost all projects nowadays are divided into phases – project's lifecycle. It is common practice to phase projects because it allows reflecting on what was done up to that point and deciding if to stick to the original plan in the next phase or alter it. In other words, phasing gives time for reflecting and future decision making (Grit, 2011).

This project's lifecycle was divided into five phases. Each phase had its own objectives and deliverables (all research type references from Grit et al., 2015):

1) Initiation Phase – the objective was to determine the plan of the project. The deliverable was the Plan of Approach. This stage mostly involved *qualitative* research as there was a large amount of data gathered and then narrowed down to its relevance to the project. Furthermore, a small amount of quantitative research was done, mainly in project management (risks & quality). This phase did not include any sub-questions as it was concluded prior to the research phase.
2) Research Phase – the objective was to conduct all the research regarding the current analytics solution and the newly proposed solution. The deliverables were:
    a. Reports of the current analytics system's limitations and the new ML solution's requirements.
    b. A list of possible ML solutions according to requirements from Agrovision and characteristics of the solution.

   This phase employed *qualitative* research as there was a large amount of information which needed to be narrowed down by in-depth interviews and group discussion. Sub-questions one, two and three are included in this phase.
3) Determination Phase – the objective and deliverable were to determine the best ML solution for Agrovision according to the previously conducted research. Once again, *qualitative* research was the primary research type as, much like the phases before, it involved writings and descriptions, but not so many numbers or figures. This phase contains sub-question four.
4) Designing Phase – the objective was to design the new ML solution's integration into Agrovision. There were three deliverables at this stage.
    a. an Implementation Plan – a new software architecture design which includes the new solution.
    b. A change plan – a report of the expected changes that the new solution created and how to mitigate these changes.

This stage employed both research types. Most of the research was *qualitative* in nature, as there were large amounts of information, but only the relevant information had to be extracted. Furthermore, *quantitative* research was done as cost, throughput and other figures had to be explored. This phase comprises of sub-questions five and six.

5) Implementation Phase – the objective would have been to implement the new solution (prototype) in Agrovision. The deliverable would have been an implemented new ML solution in Agrovision according to the two design documents (implementation and change) in the previous phase. Because this was based on the previous phase, the same two research types apply, for the same reasons. Sub-question seven is the main focus of this phase.

4.9. Summary

In this chapter, it became clear how the research was conducted per sub-question, the research methods, and how these methods assist in defining the research and validating it. Furthermore, the project's different phases were defined and what type of research was conducted in each phase. The next chapter details the *research results* and the work carried out to fulfil the project's objectives.

5. Research Results & Analysis

Below are the answers to the sub-questions. These answers are the research results and the work carried out during the assignment. The results are arranged by sub-questions. The sub-questions correspond with the research timeline, which was conducted, meaning sub-question one was researched first and sub-question seven last.

5.1. Results - Sub-Question One

Many ML solutions are competing for their share in the market. The array of different solutions is staggering as many enterprises create their solution either for internal use or for retailing.

<u>Initial Agrovision Requirements:</u>
The research had to be adapted to suit the initially researched requirements of Agrovision (see sub-question three). This means the solution had to be able to analyse large amounts of data and connect to the different software platforms Agrovision operates in.

<u>Broad Search was Narrowed Down:</u>
A comprehensive search was done on the internet to have a preliminary impression of available solutions and a quick scan of their different characteristics. Once the initial Agrovision requirements were applied, that narrowed down the list. Simultaneously, as the internet search, contact was made with a Saxion teacher who specialises in data mining to get a clearer understanding of the ML environment and get some suggestions of prospective solutions (Wesselink M., Personal Communications, 16 September 2020).

Once all information was processed, a list of nine different solutions was decided upon:

1) IBM Watson Studio
2) SAS
3) Qlik Sense
4) Oracle Machine Learning
5) HP Vertica Advanced Analytics
6) KNIME
7) Python/R
8) Python/R within Power BI
9) Microsoft Azure Machine Learning Studio

<u>Selection Explanation:</u>
It can be seen that all the names on the list are very recognisable. That is not a coincidence. As Agrovision need a robust solution which can handle its requirements, these solutions can offer such a level of reliability and support should it be needed. An exception to this is

KNIME. It is a smaller company based in Switzerland and although their solution is very comprehensive, it is still not as recognisable as the rest. The inclusion of KNIME in the list was to demonstrate that there are other solutions in the market which are not so well known but still offer a suitable solution. Power BI is on the list, even though it is not a 'pure' ML tool, because the current analytics at Agrovision is done in Power BI so the familiarity with the software can be a significant factor in the decision process.

Solutions not Selected:
Other recognisable names were not included in this list, such as Google or Amazon. Agrovision's primary software platform is Microsoft Azure. As both Google and Amazon are direct Azure competitors, there was no sense of researching or suggesting these solutions.

Working with Each Solution:
There were initial attempts to work with every solution using tutorials found on the internet, except the ones known to the graduate (Power BI, Qlik Sense). These attempts were made to get to know the solutions better and determine if they still conformed to the Agrovision requirements.

Contact with Solution's Vendors:
Simultaneously, contact was made with all solutions suppliers to get further information. With SAS and HP Vertica's exception, all vendors replied, and the necessary information was acquired. The information gathered revolved around each solution's advantages and disadvantages, from the Agrovision perspective and its cost. Because SAS and HP Vertica did not reply, those two solutions were not considered further after several attempts at contacting them.

Pitch of the Different Solutions:
All seven solutions were presented to a group within Agrovision who were selected to decide which solution to choose (Decision Team). There were six people in this group (including the graduate) with different capacities. Two were software developers with a keen interest in ML and, because of that, have requested to be a part of this project. Two more people are part of the AMT (Architecture Management Team). This team oversees all software architecture within Agrovision; hence every new software being added must be approved by them. The last person in the team is the assignment supervisor who obviously has to be involved. The table in Appendix A registers the information gathered from each solution and presented to this selected team. The table consists of advantages, disadvantages and cost of each researched solution. The advantages and disadvantages are from the Agrovision perspective and, as such, are very subjective and biased towards Agrovision's purposes. The costs (were deleted from the table for privacy reasons) do not consist of exact figures. As most of the solutions charge per usage of the solution's

resources, it was challenging to estimate. The figures which were given were the lower and upper cost estimates.

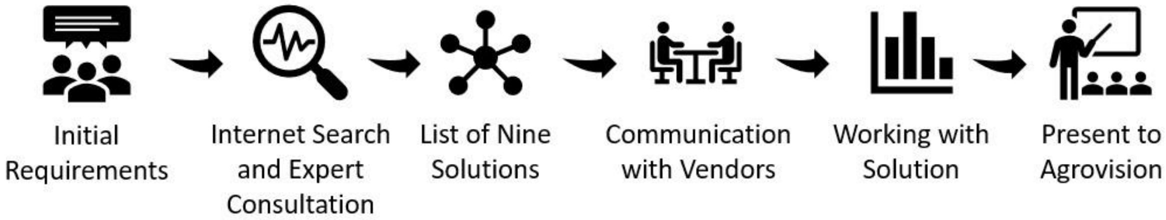

Figure 7 – research process of the ML solution

5.1.1. Summary

As mentioned before, there are many solutions available in the market. Because of that, the search for the perfect one for Agrovision was not an easy one. There were many requirements to consider and many characteristics to contemplate. The solution had to be robust enough to handle the amount of data and be reliable both in terms of software and in business terms, which means the solution and the company offering it will not dissolve in the near future.

There were deliberately disregarded solutions, such as Amazon or Google, because they are direct competition to Microsoft Azure. Others were researched, but not presented to the team because the company was not well known or too small (with the exception of KNIME as mentioned above). Some solutions were overlooked for no other reason than lack of knowledge. It is reasonably certain there are some excellent solutions which were not researched for that reason. Of course, there were solutions which did not respond to communications so were not considered further for that reason.

However, the seven solutions which were recommended all qualified as viable possibilities for Agrovision. They all fulfilled the initial requirements, which meant they could handle the amount of data, contain the right connectivity, and stand the test of time. These solutions represented the different options which were available and in a later time, one was chosen as the best solution which Agrovision can work with.

5.2. Results - Sub-Question Two & Three

When these sub-questions were initially formulated, it was thought that they would be two completely different questions which will provide two completely different answers. In most circumstances that would have been correct, as the limitations and disadvantages are typically different (though sometimes overlap) to the requirements. However, in the context of Agrovision and this project, these two sub-questions are one and the same. This stems from the fact that all the shortcomings of the existing analytics system are the new ML analytics requirements. If there were the slightest difference between the two, then the questions would have been separated, but the fact that they are identical means they had to be merged into one.

<u>The Need for ML:</u>
As mentioned before, Agrovision uses Microsoft Power BI as their analysing tool. Power BI has many abilities and is developing as one of the world's leading Business Intelligence tools (Marr, 2020). However, it still has some limitations; namely, it is not recognised as an ML tool. Furthermore, when creating Power BI reports, it can get extremely complicated if there is a need to use data from many different sources, such as IoT devices, weather data and the like. Moreover, there is one application which Agrovision offers the pig husbandry sector, which is called PigVision. Within PigVision, there are some analytical capabilities, but very few farmers use them for reasons which will be discussed later.

<u>Choice Reasons for Expert Interviews:</u>
In order to analyse the limitations of the current analytics system and therefore, requirements of the new system, several interviews took place with key stakeholders. These stakeholders are either considered experts or are related to the analytic system by some means. They are from four different sectors of Agrovision – sales, operations, product management and analytics:

- Analytics – the people who work with the current analytics system, create and maintain the reports and are also in contact with clients if there is a need for new reports or modify an existing one. They get specific analytics information about the farmers' demands and where Agrovision's analytics do not fulfil their expectations.
- Sales – this sector has constant contact with the customers. They are aware of the distinctive customers' demands from the solutions Agrovision provide. Feedback about analytics and reporting is part of this information.
- Product Management – This assignment's scope is pigs, so naturally, the pigs' product managers will be consulted. They oversee all the pig's products that Agrovision offers, therefore are very aware of any shortcomings to the pigs' data analytics. Product Managers of pigs have excellent technical knowledge of pig husbandry as well as the knowledge of Agrovision's value proposition. This

combination gives them a complete picture of where the shortcomings of Agrovision's solutions, including analytics.

- Customer Service - have constant contact with the clients, which gives them a highly effective view of the analytics system's weaknesses. Much like sales, they get direct feedback from the clients about analytics, amongst other topics.

These four sectors were chosen for interviews because of their exclusive analytics system viewpoint. Because of these different viewpoints, a clear understanding of the limitations of Agrovision's data analysis solution and the requirements of the new system can be obtained:

<u>Analytics Interview:</u>
When talking to Arnold Wisselink, one of the analytics personnel, it became apparent that farmers would like to have reports that will help them make future strategic and managerial decisions. This requirement is currently not met as it requires looking at future implications in the current data. This disadvantage to the current system is also a requirement from the new solution. ML should have predictive analytics as part of its capabilities. Furthermore, it was discussed that the current reports are based on relatively few parameters as it is too complicated to incorporate too many of them in the current analytic system, even though there are many more available parameters. At the same time, it is a disadvantage of the current system; it is a requirement from the new solution - the new ML solution has to integrate many parameters into its analytics (Wisselink A., personal communications, 04-09-2020).

<u>Sales Interview:</u>
The conversation with sales revolved around the fact that the current analytics are only done on past results, meaning there is no predictive analytics. Because there is a growing demand for predictive analytics from the clients, it is a significant disadvantage which has to be addressed. Also, the number of parameters used for the current reports is relatively small. At present, there are many more data sources (weather, feed intake, drink intake, weight and the like) which should be integrated into the analytics but are still not being used. Both these limitations were mentioned before by Arnold Wisselink and are, of course, also requirements from the new solution (Browers J., personal communications, 28-09-2020).

<u>Product Management Interview:</u>
When talking to the Agrovision's product managers of pigs, the most apparent issue discussed was the absence of predictive analytics. The farmers would like to use analytics as a tool, meaning they would like to make decisions based on data. The discussion went further about the need to incorporate IoT devices into the analysis, which is currently not being done. It was mentioned that there are many IoT devices which generate data that can

be used for analysis, such as coughing monitors, back-fat monitors and many others. The coughing monitor can sense the number of pigs coughing, which can help detect disease. As the name implies, the back-fat monitor senses the amount of back fat on live pigs. This can help determine the total amount of fat according to market demands. Much like the two previous conversations, this one revolved around the exact same issues, which are both limitations and, at the same time, requirements (Søgaard C., personal communications, 30-09-2020).

Customer Service Interview:
The conversation with Customer Service was about PigVision, which is one of the leading applications Agrovision offers to the pig husbandry industry. There is an analytic section in PigVision that can analyse data from different parameters, such as weight, water intake, food intake, temperature, etc. This tool was made to give a solution for leveraging data towards future managerial and strategic decisions. However, there are a few problems with this analytics section:
When PigVision analytics started, the different sensors' price to measure the data was exceedingly high, which meant it was not financially viable for farmers to buy them. By the time the sensors' price went down and their use became more commonplace PigVision analytics was old and outdated. Both reasons lead to PigVision analytics to be used by only one farm since its inception. Once again, the same limitations were discussed and the same need for a new ML solution was stated (Said Fredsted P., personal communications, 01-10-2020).

5.2.1. Summary

Determining the disadvantages of the current analytics system, and therefore the new system's requirements, was based on interviews with different people. These people represent different departments and therefore represent the full picture of how Agrovision as an enterprise views these disadvantages and requirements. Consequently, the fact that all people mentioned the same two issues meant that the conclusion is blatantly apparent:

1. There is a need to use predictive analytics so the farmers can use existing data to assist them in making future managerial and strategic decisions.
2. There is a need to integrate the growing number of parameters into the analytics system.

These two limitations of the current analytics system meant that there is a need to look for a different analytics system that can create these tools and integrate many parameters. In other words, the current system's limitations are precisely the requirements of the new ML analytics solution.

ML solutions can be used for predictive analytics, which can give answers to questions such as: "When to buy more feed?" or "When will a sow (a female pig) be ready to be inseminated?” These questions, and many other predictive measures, can be used as tools for the farmers to plan their operations in a much more accurate manner, saving time and money.

Furthermore, ML solutions were made for incorporating many data sources into their analytics capabilities. That will enable integrating all the data collected from different devices (IoT) and using the data for analysis. This will enable to answer many questions which are left unanswered using the current system, for example: “What is the optimal weight of a sow to deliver its litter?” or “What is the optimal temperature for pigs to grow?”

5.3. Results - Sub-Question Four

This sub-question is the direct continuance of sub-question one as it commences from the exact point sub-question one had finished – the presentation of the seven solutions to the team. Furthermore, it is the result of sub-questions Two and Three because all the requirements and characteristics of the new solution, which were researched in those sub-questions, have impacted the conclusion which is the most suitable solution for Agrovision.

The Team's Decision:

As mentioned in sub-question one, a team was assembled to decide which ML solution is most suitable for Agrovision (the Decision Team). A meeting with this team was held, and a pitch was presented with all the solutions' findings (Appendix 1). Each solution was presented with its pros & cons and a short explanation. After the team weighed all the points presented, a discussion ensued about the essential points that Agrovision should consider. At the end of the meeting, it was decided upon two most suitable solutions – Microsoft Azure Machine Learning Studio and IBM Watson Studio. The team agreed to convene in two weeks after these two solutions were researched further (New AI Solution (meeting), Personal Communication, 18th September 2020).

Contact (or lack of) with IBM and Microsoft:

During the next few days, communication was conducted for more in-depth research of both solutions. At a very early stage, it became evident that Microsoft has no interest in providing any answers except if payment was involved, whereas IBM was more than willing to give as much information as possible. This stems from how the two enterprises are set up to communicate with their clients:

Microsoft rarely converses directly with their clients, instead doing so by certifying local companies through a series of test and exams to communicate on their behalf. On the other hand, IBM has either their own branch in a country or has a main representative who carries out the communication. When concerning this project, the main difference was that IBM was ready and willing to converse, whereas Microsoft was not unless payment would have been made.

Following this more in-depth research of the two solutions, another meeting was scheduled to determine which of the two will be chosen. It became apparent during that meeting that the team needed even more information. There were concerns about automation of the learning models and about its abilities to connect to the different platforms Agrovision use. For this reason, the team could not reach a final decision. It was agreed that an additional meeting should be held with both Microsoft and IBM to get the answers for these questions (Final Decision AI Solution (meeting), Personal Communication, 6th October 2020).

Team Meeting with IBM:
A meeting with IBM took place, including a member of Agrovision AMT (Architecture Management Team), who is also a member of the decision team. During this meeting, all the questions were asked, and the answers were more than satisfactory. That meeting concluded that IBM's Watson Studio is more than a fitting solution for Agrovision. Microsoft, on the other hand, as mentioned before, would not converse directly, but only through a certified company and with payment being made.

5.3.1. Summary

The last meeting of the Decision Team took place on the 26 of October. Even before this meeting, it was pretty obvious to all involved that the preferred solution is IBM's Watson Studio. However, it was necessary to meet in order to make sure everyone is in complete agreement. Of course, during the meeting, it was agreed that Agrovision will employ Watson Studio as its preferred ML solution (Final….Final Solution (meeting), Personal Communication, 26 October 2020).

The only concern raised during the meeting was the costs involved, as it is not clear yet what the cost will be. However, as with all other solutions considered, it was established before that the actual cost will be revealed only after working with the solution. Moreover, the cost can be determined during an initial pilot project with IBM which they have assured will be conducted at no charge (though no charge will be made the cost can still be determined).

5.4. Results - Sub-Question Five

Every new solution that needs to be incorporated into an existing system must be connected in one way or another. A few steps need to be taken to connect that solution. Here it is specified how the new IBM Watson Studio would have been connected to Agrovision's existing architecture. Unfortunately, as was written before, the new solution's actual implementation did not take place, but the new architecture was made and an implementation plan was formulated (because most of the plan was not implemented, for grammar simplicity, it is written in the present or past tense).

The body within Agrovision responsible for assimilating new software (amongst other things) is the AMT (Architectural Management Team). Thus, it was only natural that the team that will assist in designing the software architecture for the new solution would consist of personnel in the AMT. A meeting was scheduled with three individuals from the AMT (two of which were in the Decision Team) to understand the enterprise architecture of Agrovision and have a brainstorming session on the initial steps to take in the design process.

5.4.1. Agrovision Enterprise Architecture

Agrovision is busy migrating its infrastructure from a "physical" or on-premise one to a cloud-based architecture. This migration holds several advantages, such as in-built security, automatic backups and extra functionality (Azure services and applications). However, as can be seen from the Enterprise Architecture (EA) in figure 8, most of Agrovision's infrastructure is still in the two physical servers in Deventer and Enschede (Initial Design Steps (meeting), Personal Communication, 27th October 2020).

<u>The Physical Servers (technology layer):</u>

There are two "physical" servers which Agrovision utilise. Enschede is the primary server, and it functions as the host server for Agrovision's applications. This server has its own backup and a switch for connectivity. The server in Deventer is at the Agrovision offices and is used as a backup for the server in Deventer. Once again, it has its own switch for connectivity. Both the Deventer and Enschede location switches are connected through a 10GB speed connection. This architecture has served Agrovision for a few years but is being phased out, as explained in the next paragraph.

<u>Microsoft Cloud Server (technology layer):</u>

Microsoft cloud server or Azure is a hosting facility that includes more than 200 enterprises (Microsoft Azure, n.d.). Azure aims at giving enterprises a complete solution for all its requirements, from management tools to analysis tools, to hosting facilities, internet services and so forth. Agrovision's strategy is to migrate all its IT infrastructure to Azure. This is an ongoing process which has started approximately one and a half years ago and will continue for the foreseeable future. Currently, Agrovision uses many of Azure's applications, including the software architecture, described in section 5.4.2. However, only one application that its clients can use is MyAgrovision, the new Agrovision platform. Agrovision's plan is to have only MyAgrovision as a replacement for the rest of its applications, thus it will operate only with Azure.

IBM Cloud (technology layer):
The IBM cloud is added here even though, in reality, it is not connected to Agrovision's architecture. This is to depict how it will be connected in the future once Agrovision will assimilate Watson Studio. All the different services which will connect or will be associated with Watson Studio are in Azure (Staging DB, DWH, SASS, reporting services and so forth).

Application Layer:
Agrovision's application layer comprises of the different applications Agrovision offers its clients. In practice, there are more than a hundred applications which Agrovision offer, but most are very small and are in the process of being assimilated into the larger applications (the ones listed in figure 8). The application layer offers the listed services to Agrovision clients and to Agrovision itself when maintenance, updates or support are needed. The registration service enables the relevant Agrovision clients to register their activities with the government.

Business Layer:
The business layer consists of the different actors associated with Agrovision's application, including Agrovision as mentioned before. These actors use all the services from the application layer for their particular purposes. The exception to this is the government actor who is only associated with the registration service as previously explained.

Summary:
The figure below is the high-level enterprise architecture of Agrovision. There are no figures with more details because Agrovision wants to keep it private. Agrovision have a lot of data from its clients, which can be used by Agrovision, but still belong to its clients. Should this data be revealed, it will have dire consequences. Nevertheless, this figure provides an insight into Agrovision's enterprise architecture, its future strategies (IBM cloud), how the clients benefit from it and will have further benefits once the migration to the Microsoft cloud is completed.

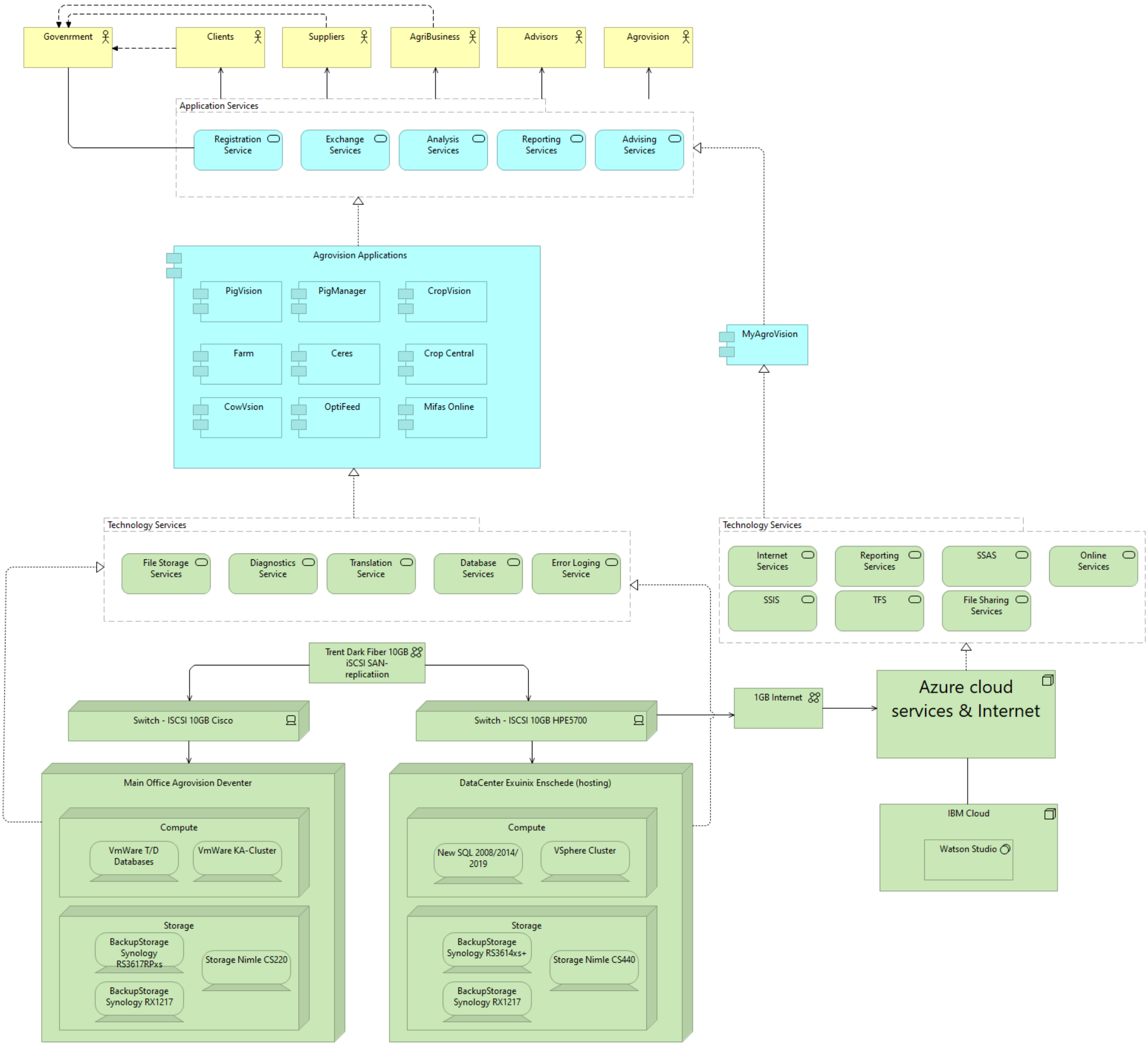

Figure 8 – Agrovision Enterprise Architecture including the IBM cloud

5.4.2. Software Architecture Starting Situation

Before the description of how the new Watson Studio should be integrated into the existing Architecture, there is a need to understand the architecture before this integration:

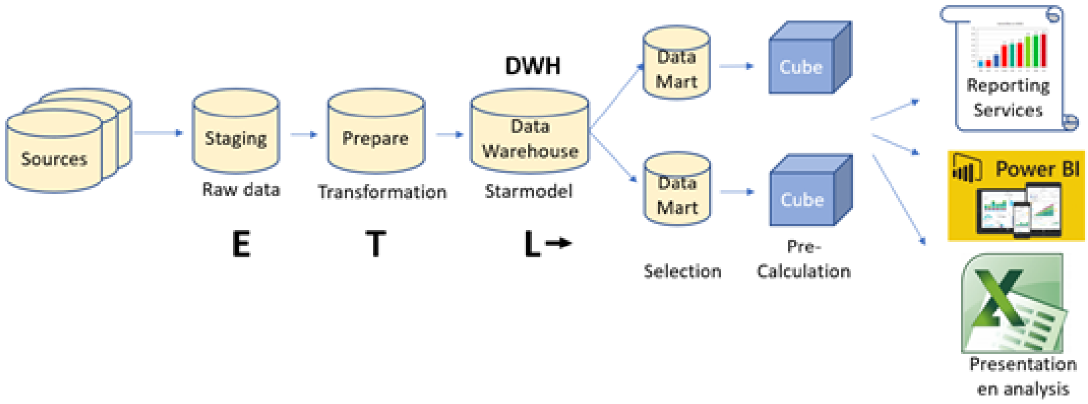

Figure 9 – the software architecture of ADEX DB and DWH

As can be seen from the figure above, the “sources” (on the left) is the ADEX DB which holds all the data collected from the farmers. The “Staging” DB extracts the relevant tables from ADEX without additional filters, relations or calculations by an incremental update. This is done to reduce the load on the main ADEX DB. The “Prepare” DB transforms the data it collects from the Staging DB into ‘star schemas’. These star schemas are where the ‘business rules’ are defined and applied for better and easier reporting. This is done in the Prepare DB (T) to free the DWH from these calculations and, therefore reduce the load on it. Following that, the data, including the star schemas, are loaded in the DWH through an incremental load. In the DWH there are many “Data Marts” from which slicing and dicing (cubes) of data can be done for the different reporting applications (on the far right). This is the ETL process's general schema between ADEX and the DTW. Once the architecture is clear, it is easier to understand how the new solution was integrated (Beijer J., Personal Communication, 6th November 2020).

5.4.3. Initial Software Architecture Brainstorming

The below sketch was made to depict how Watson Studio (mistakenly named DrWatson in the sketch) will connect to the existing Agrovision architecture. As shown in the sketch, the ML model was created in the IBM cloud (black circle). Following that, the model was exported to the Agrovision Staging DB as either an R or Python script (the orange rectangle). The model runs in the Staging DB, as it is strategically placed to take the data from the specific Data Mart, in the DWH (black square, names “pigs DWH”), and sends the results back to the DWH for reporting in Power BI (Initial Design Steps (meeting), Personal Communication, 27th October 2020).

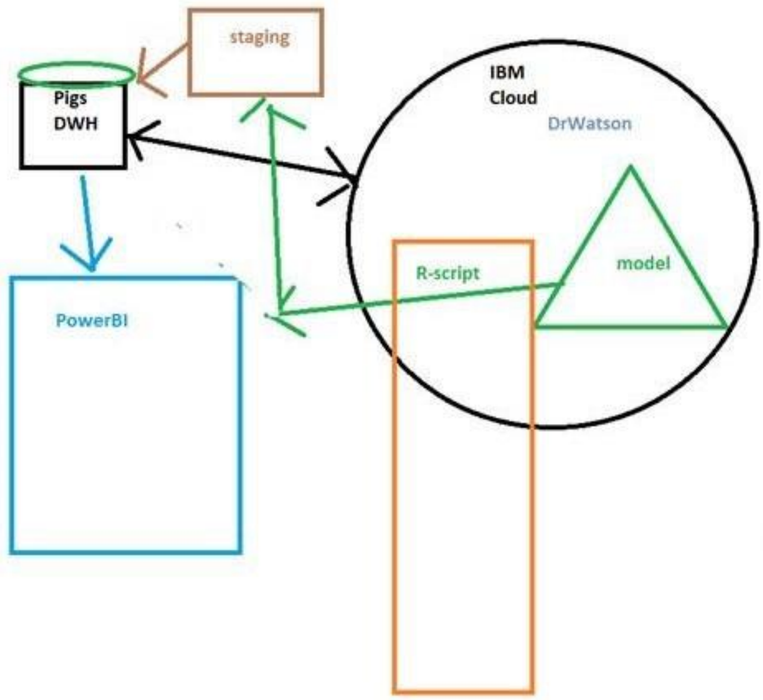

Figure 10 – initial sketch of architectural software design as made during the meeting

5.4.4. New Software Architecture Design

This new architecture design was made based on the previous brainstorming sketch (with assistance from Bakker A., Personal Communications, 04 December 2020). It was agreed that Watson Studio would be in the IBM cloud as Agrovision favour such architecture. As can be seen, Watson Studio receives the data from one or more cubes explicitly created for it. The model will be developed in Watson Studio in the IBM cloud. After the model is created and results are satisfactory, the model will be exported to the Staging DB and run there. Even though the model will run in the Agrovision environment, if any modifications are required in the model, they will be done in Watson Studio. After the modifications are completed, the model will be exported again to the Staging DB in the Agrovision environment. The data for the model in the Staging DB will come from the same cube as Watson Studio is exploiting. The results will be transferred to the DWH and eventually into one of the Presentation & Analysis platforms from the staging DB. The connectivity of Watson Studio to the Agrovision environment is done by tools built-in to Watson Studio. How the systems work is specified in the next section where the processes will be described.

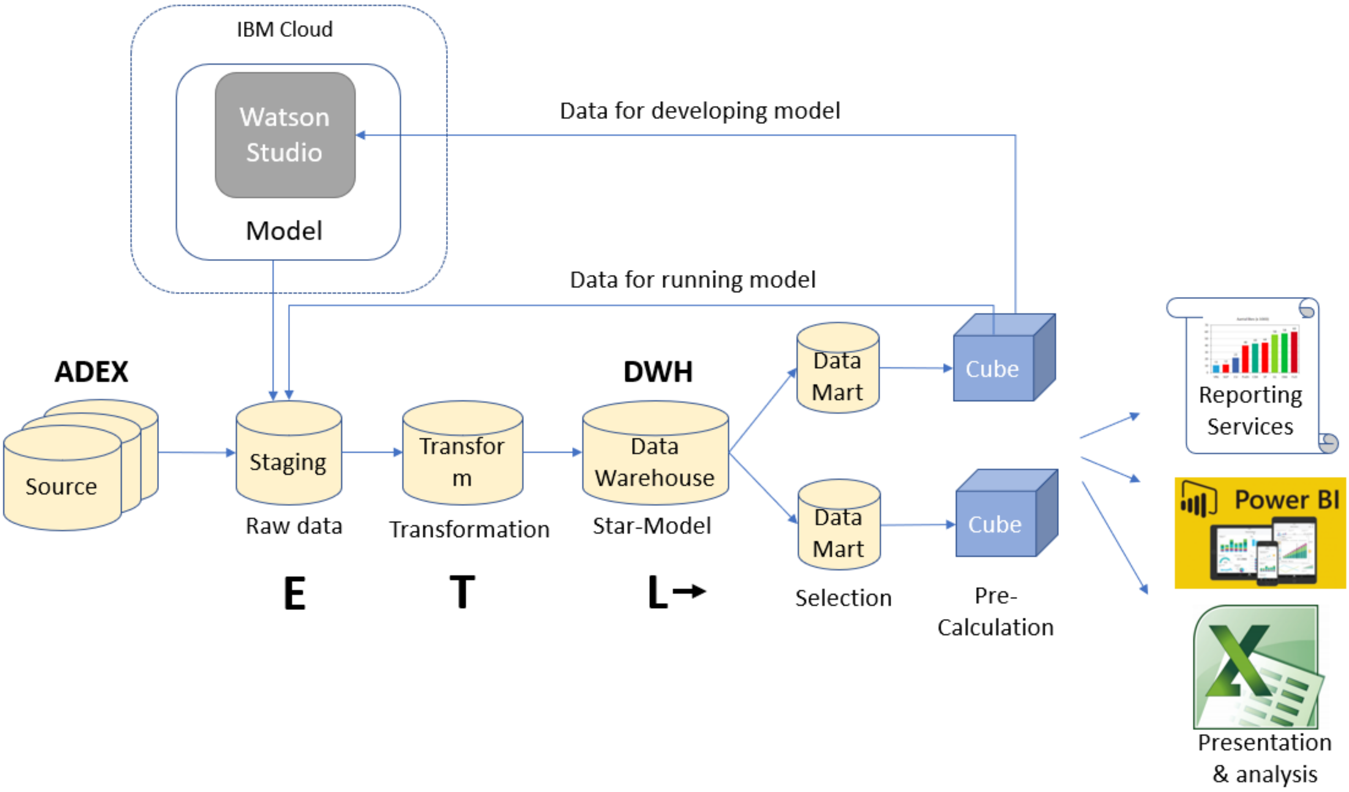

Figure 11 – software architecture design including Watson Studio

5.4.5. Current Report Creation

Because the primary processes within Agrovision that concern the new ML analytics platform is the repot creation and reporting to the clients, there is a need to describe how the current processes are so that the changes can be defined.

<u>Process Description:</u>
As shown in figure 12, the process starts with an analysis request from the client. Once the request is received, there is a need to investigate and analyse the query. If more information is required from the client, there is a need to contact and get answers for the relevant questions. Following that, a use-case is created from the query and KPI's are defined, which the analysis will revolve around.

There are instances where a version of the report already exists, in which case, the report is just modified to suit the client's requirements. This enables to omit many activities during the process, making it much shorter.

If the report does not exist, a data-mart needs to be created in the DWH, where all the relevant data is stored. From this data, an analysis cube is created in the SSAS service to be able to slice & dice the data according to the requirements. Power BI imports the cube and creates the report from it. As with every new product, the report must be tested for the results' reliability. Once the testing is done, the report can be uploaded to the relevant Agrovision platform, and the client can be notified. From that point, the client can benefit from the report's analysis.

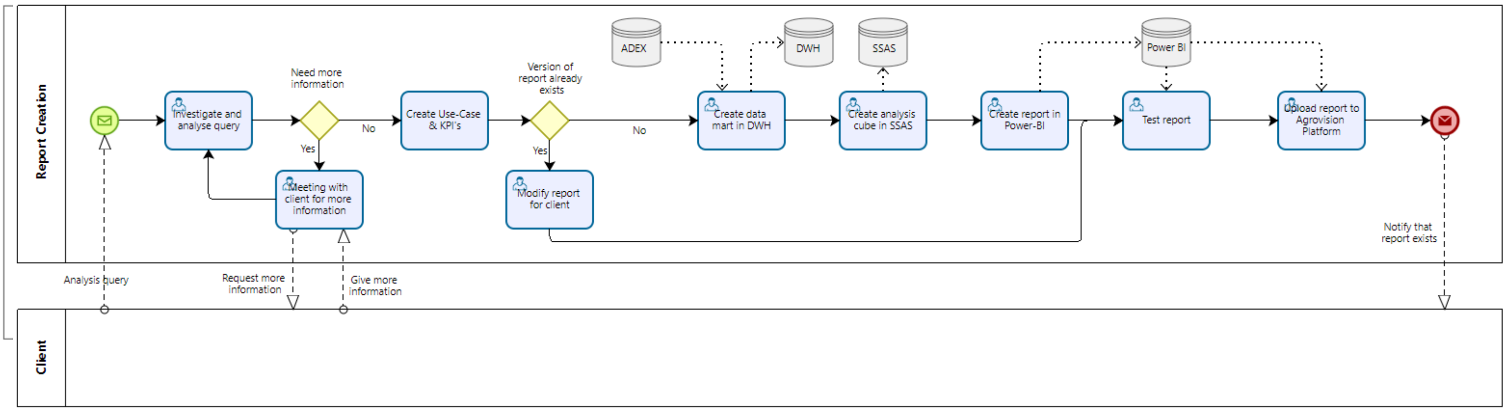

Figure 12 – BPMN of report creation

5.4.6. Implementation Plan

In order to implement the new ML solution, a plan has to be formulated. The plan is to ensure that the architecture, as well as the business alignment, are kept and time (and money) can be saved. This plan aims to describe every step that needs to be taken in as many details as possible, obtaining its information from both the architecture design and the process illustration.

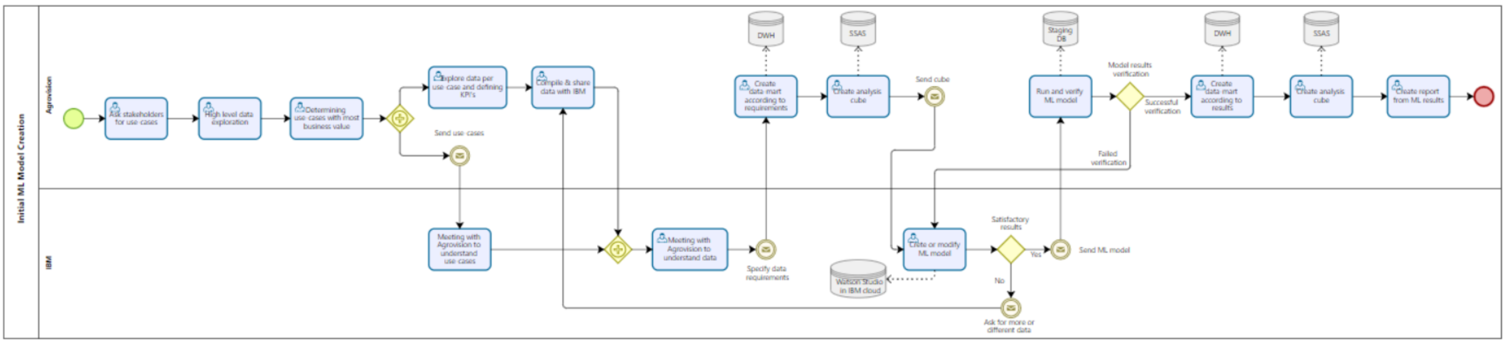

Figure 13 – BPMN of implementing ML analytics (as described below)

5.4.6.1. Use Cases

The first step of implementation is to understand the business needs. For that purpose, use-cases are created. These use-cases serve as the foundation and the drivers for the analysis because that is where business needs are being realised through data analysis. If the analysis is successful, there are further insights into these use-cases which create the business value. To reveal which use-cases have the most business value for Agrovision, experts need to be consulted, and once there is a satisfactory list, it can be submitted.

5.4.6.2. Cataloguing and Preparing the Data

To answer the use-cases, the data must be present for it to be analysed. The data has to be explored to determine whether there is enough of it to use for the analysis and determine whether it is the correct data. This process actually started in the previous step when one of the criteria to choose the right use-case was to consider which data is available.

<u>Overall Perspective:</u>
Initially, the data is looked at from an overall perspective. This is to determine if there is any data at all, answering the use-case. This step was done by the stakeholders in the previous step, but is also done by IBM who will need to create the model later. Once it is established from an overall view that the data does exist, there is a need to understand it.

<u>Data Understanding:</u>
The tables in ADEX and the DWH are looked at and there needs to be a clear comprehension of each column. This step takes place with stakeholders from both Agrovision and IBM. The stakeholders from IBM are the ones who will later create the model and the ones from Agrovision are experts on the data. These experts have the knowledge of what every column denotes, what does the data in the column signifies and what is its granularity (is it per month, per day, per animal, per farm and the like)

<u>KPI's:</u>
Furthermore, there is the need to consider which KPI's (Key Performance Indicators) exist in the data. These KPI's are the most relevant business indicators in the use-case data. They are also one of the most critical items to consider through the entire process. That is because they are initially used to determine if the correct data is available, later they are used to create the cube. Following that, they are used to verifying the results and lastly are used to display the results. An example of a KPI for the "forecasting the number of piglets to be born" use case, is the number of born piglets. Without this KPI, this use-case cannot be analysed, verified or displayed.

<u>Creating Data-Marts:</u>
The DWH is comprised of star-schemas. Each star schema has one fact table in its core and is connected to several dimension tables. The fact table contains data which can be counted. The number of piglets born to which sow and which boar (male pig) is an example of a fact table. The dimension tables contain data which describes the fact table. For example, different parameters of each piglet being born (female/male, in which farm and so forth). Each star schema is called a data-mart and it is independent of all other data-marts in the DWH. From these data-marts, cubes are created, which is described in the next section.

5.4.6.3. Cube Creation

A cube is created for every data-mart in the DWH as was described before. In Agrovision the cubes are created using SSAS – **S**QL **S**erver **A**nalytics **S**ervice. SSAS is an OLAP (**O**n**l**ine **A**nalytical **P**rocessing) service by Microsoft which provides data mining capabilities to an enterprise by creating cubes for analysis. Agrovision chose SSAS to create the cubes because of its flexibility and aggregation abilities.

Cube Description:
A cube is a three-dimensional model which contains data. The data is arranged in the cube to enable to do different 'slicing and dicing' of the data. The cubes in Agrovision are optimised for aggregating data, making complicated calculations as instantly as possible and authenticating different users' authorisation (Rooks W., Personal Communications, 14 December 2020).

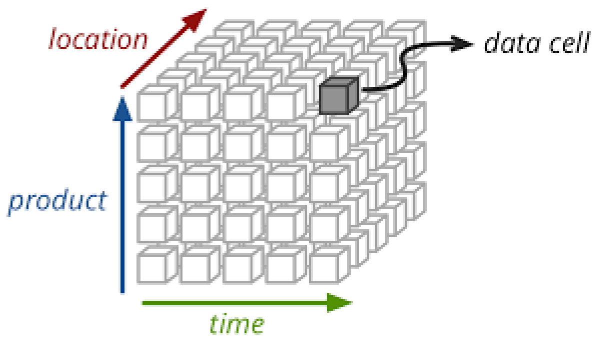

Figure 14 – illustration of an analytics cube

Cube Creation:
Unfortunately, the cubes for the three use-cases were not created, making it impossible to describe them. Although, these cubes would have used all the principles, both general and specific to Agrovision, described above. If there is a need to revisit the previous step to create the cube correctly, it can be done, creating an iterative process. The next step is to create the ML model using Watson Studio.

5.4.6.4. Creating and Modifying Model

Once all the data is in place and is arranged in the cube form for easy data mining, the ML model can be created. IBM's Watson Studio was previously chosen as the preferred tool for creating the model. Furthermore, it was previously agreed that the solution would be in the IBM cloud.

Watson Studio Description:
Watson Studio is an IBM-provided solution for building, running and managing AI models (IBM Watson Studio, n.d.). It uses a graphical interface with 'nodes' that can be dragged and

dropped in the appropriate location and be connected to create the model. A node has specific functionality such as importing data or applying a particular algorithm. In Watson Studio, the model can be created using Python or R languages, should that be the user's preference.

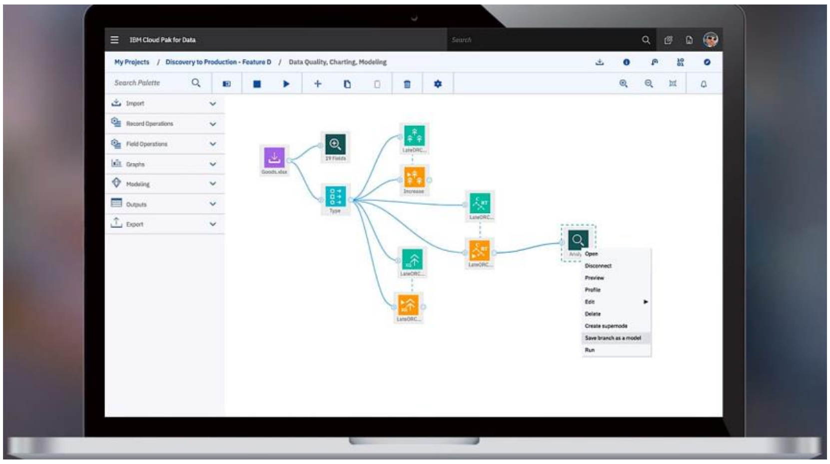

Figure 15 – screenshot of IBM Watson Studio with several nodes, which create the model

<u>Model Creation:</u>
Of course, the model was not created and as such will not be possible to describe in detail, but the steps which need to be taken for the creation are described:
The first step is to import the data from the created cubes. Once the data is in Watson Studio, it is possible to explore and transform it. Activities such as deleting rows with missing values or adding missing values by estimations can be done with dedicated nodes. Following that, a feature called AutoAI in Watson Studio can be utilised. AutoAI analyses the data and creates possible model pipelines and other data transformations, parameter settings and different algorithms that can be applied. By doing so, AutoAI enables the user to simplify the model building process. It is unnecessary to use AutoAI, but the potential gains cannot be ignored easily. Once AutoAI has finished, the results can be assessed and the best possible model containing the most suitable algorithm/s can be chosen. Further assessments need to occur to have an initial validation of the model. Should changes need to be done, nodes can be replaced or their function modified. This process is usually iterative as there is a need to check every part of the model to improve the results. Even though the process steps are simple to carry out, in practice, it can take months and even years to perfect the model, depending on the provided data and complexity of the use-cases.

<u>Iterative Process:</u>
Usually, it is challenging to predict how data will create the best model in advance. Typically, there is a need to go back to the previous steps to change the DWH data, which in turn, will change the cube. This process enables to evaluate different data configurations and gauge

which are the best for use by the model. Once the model gives satisfactory results, verifying it can begin.

The Biological Factor:
As mentioned before, creating the model can be a lengthy process. However, an additional factor needs to be considered, which can make the process even longer. This is the biological factor – the fact that the analysed subject is of biological nature, such as crops or animals, create further complications. There is an elevated level of unpredictability in producing biological goods. There are countless factors which affect the product, such as weather, humidity, light, soil, nutrition and the like. Many of these factors cannot be controlled, and even if they are, it does not guarantee perfect results. This is because of the genetic influence of the product. A specific crop, much like a specific animal, will react differently to precisely the same conditions. Some will produce more and some will produce less, which increases biological goods' unpredictability. This has to be taken into account when creating the model which complicates and prolongs the process.

5.4.6.5. Verifying Results

Much like the previous process the verification stage can be a lengthy one. Depending on the type of results, different methods can be applied.

Past Results:
If the results concern past figures, such as in use-case two (where is there a need to replace a sow?), there is a need to look at when sows were replaced in the past and for what reasons. If the results match these past occurrences, then the model is correct. If the results do not match, there is a need to change the model or create a new one altogether.

Future Results:
If the results are a prediction, such as in use-case three (predict coming litter according to the weight of a sow or back-fat), there is a need to look at future litters as they happen, compare the weight and back-fat figures, and verify the if the physical results match the model results. Of course, if the model is not correct, it needs to be modified, or a new one must be created.

Iterative and Prolonged Process:
The need to test both the figures and the physical occurrences, in the past and future, coupled with the biological factor, makes the verifying process very prolonged. There has to be enough verification data because only a few occurrences or figures are not sufficient for testing. This is especially compounded once testing future occurrences. Time has to pass and enough occurrences to occur to have the necessary figures. Furthermore, every time the model results are not correct, there is a need to go to previous steps to adjust the parameters and apply them. These necessities to verify results and the iterative nature of

this step can make this process very prolonged. If the model's results are satisfactory, there is a need to export the model to the Agrovision environment.

5.4.6.6. Exporting the Model

Once the results are satisfactory, the model is exported to the Staging DB's in the Agrovision environment. The export is done using either Python or R after which the Python/R plugin within SQL Server Management Studio is used to run the model. This will save money and resources as every use of the IBM cloud is costly and there is no need to use extra resources to send the data and receive the results from the IBM cloud.

<u>Model and Results in Staging DB:</u>
The decision to export the model into the Staging DB and run it there, was taken during the "Initial Design Steps" meeting (see section 5.4.2.). This will alleviate the extra load on ADEX. Furthermore, this DB is in a strategic location to access the cubes' data needed by the model and send the results to the DWH. These results are sent to the DWH and eventually the cubes through the same channels as all the other data transferred from Staging to DWH. This will maintain the same data transfer rules, which conforms to Agrovision's guidelines.

Once all this work is complete, the results can be displayed to the different stakeholders, which is the next step.

5.4.6.7. Displaying Results

In Agrovision the primary tool used for displaying reports and results to stakeholders is Power-BI, though some stakeholders like to arrange their results in their preferred tool. A cube is created for every report to ease analysis slicing and dicing. Power-BI has connectivity to various sources and the SSAS service is one of them. The analysis cube can be created in Power-BI, but Agrovision decided to use SSAS for reasons stated in section 5.4.4.3. Agrovision chose Power-BI because it is the dedicated Microsoft solution for Business intelligence and reporting (Agrovision uses Microsoft products by default).

<u>Cloud Base Reporting:</u>
Furthermore, Agrovision has started implementing a cloud reporting service to its clients, which can be accessed through a web browser, unlike the previous solutions which were through a dedicated program. These programs are part of Agrovision's value proposition and are currently the principal source of reporting, but the long-term plan is to move all reporting to the same cloud reporting service.

<u>Customised Reporting:</u>
The reports of the ML analytics are, much like other Agrovision reports, personalised for each client. If the client is a farm, this farm can see reports specific to it or related to it. Benchmarking reports is one example of such a report. If the client is a consultant, the

reports are customised for their needs. For example, access reports only specific to the consulted farms.

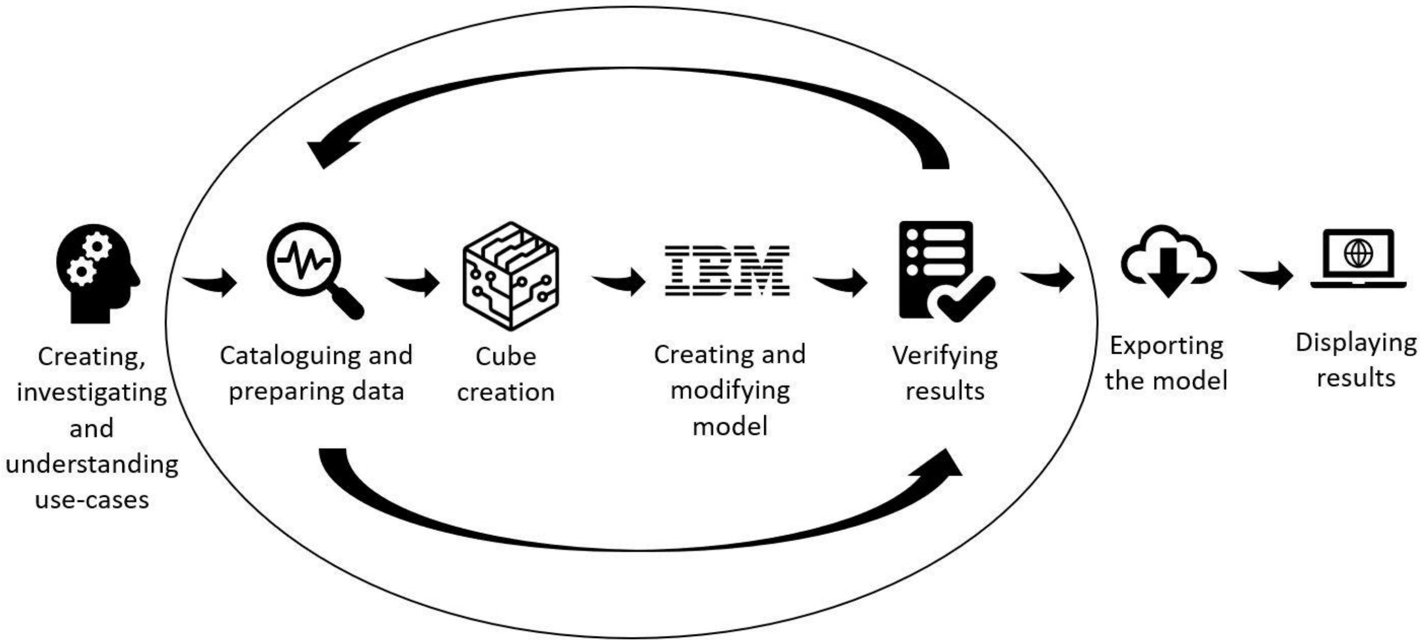

Figure 16 – Implementation process

5.4.7. Future ML Model Creation Process

There are some similarities to the existing report creation process, but many steps are different and need to be described. Moreover, when Agrovision personnel are proficient enough with creating the models without assistance from IBM, the IBM function in the process will not exist anymore. The differences between the processes are described below and shown in the BPMN diagrams.

<u>Difference Between Processes:</u>
Once the implementation plan is finished and the initial models have been created, the process of creating further models will change. Below is the description of this process. The main difference between the two processes is that IBM is not involved in this one. That is because Agrovision analytics personnel will be proficient enough to create the models without assistance from IBM.

<u>First Steps:</u>
The first steps in creating the model are the same as creating the standard reports. The client has a query which can be answered by ML analytics. There is a need to investigate the query, and if more information from the client is needed, contact has to be made. This is followed by creating a use-case and KPI's for the query. The use-case is the basis for the

continuance of the process, and the KPI's are the factors which will determine the analysis criteria. Moreover, if a similar model already exists and just needs to be modified, many steps can be omitted, saving time and money. Much like the creation of a new model, this modification takes place in Watson Studio in the IBM cloud.

Creation of the Model:

Should the model and its associated reports do not exist, there is a need to create them. Although before that, a data-mart has to be created in the DWH containing all the relevant facts and dimensions for the model. A cube can be created in the SSAS service from this data-mart, which will enable the ML model to slide & dice the data as necessary. Although Agrovision's analysts create the ML model, it has to be created in Watson Studio in the IBM cloud. Once it is created, it is exported to the Staging DB and is run from there.

Testing the Model:

If the model's results are not satisfactory, there is a need to modify the model in Watson Studio. Furthermore, if there is a need to modify the data, the data-mart must be altered to reflect these needs. Should the results prove satisfactory, they need to be verified by comparing them to factual results obtained from the farm. If the verification is successful, the (sub)process of compiling the report can start. However, if the verification fails the model, and possibly the data, have to be further modified. This (sub)process is iterative in nature, as seen in the BPMN model.

Creation of the Report:

Once all the model's tests are finished successfully, the results need to be accessible by the client. As Agrovision uses Power BI to display analysis results by default, it is used here as well. A new data-mart is created in the DWH, which grabs the ML model results from the Staging DB. An analysis cube is created in the SSAS service and is sent to Power BI, where the report is made. Once the report is ready, it is sent to the Agrovision front-end platform. The client is notified that the report is there and can benefit from the additional business value.

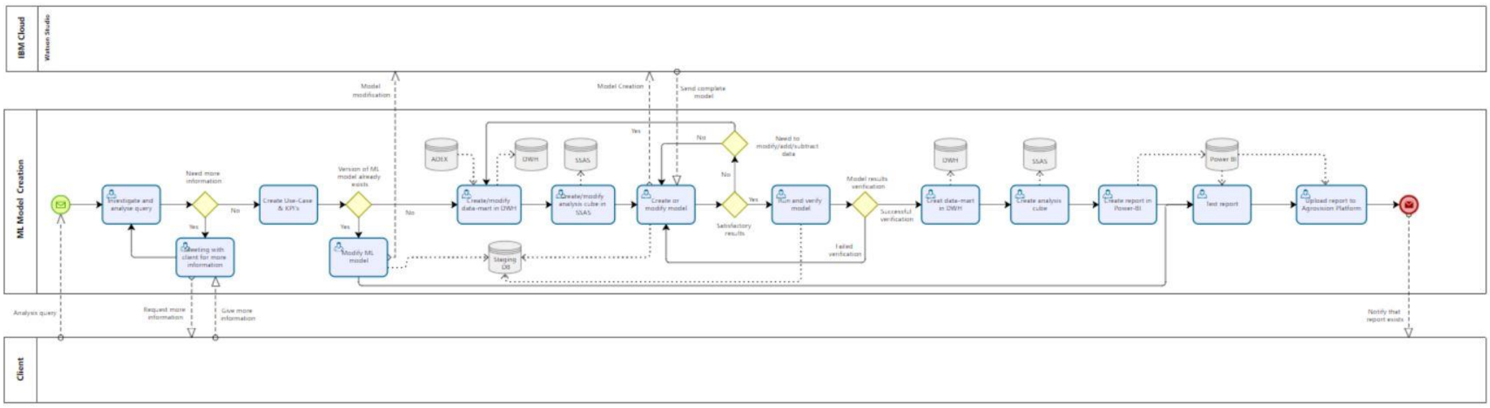

Figure 17 – BPMN of future ML model creation process

5.4.8. Summary

This implementation plan holds vital importance in this project. All the research which was carried out previously is implemented by using this plan. It has detailed instructions and guidelines that define the implementation's objectives and keep all stakeholders aligned with these objectives, the process, and the final results. This project plan, coupled with the change management plan described in the next section, form the main guiding principle for realising the new ML solution.

5.5. Results - Sub-Question Six

Change management is aimed at strategic changes in an enterprise to reduce or eliminate risks and personnel resistance to this change. Although the changes which have taken place because of the new ML solution are relatively small, it is still a change in strategy as this project is the first steps Agrovision took in the direction of AI.
The change plan's main two questions are - "What has to change at Agrovision?" and "How will this change be achieved?" These questions will be answered in the change analysis and subsequent steps below (all of section 5.5 is referenced from - Voortman, 2017. Unless otherwise stated).

5.5.1. General Analysis

The first topic to consider is the principal approach which drives the change. Although, there are some core competencies which make Agrovision unique, such as the knowledge of agriculture and its integration with IT which is the *Inside-out approach*, the main strategies of Agrovision put the customer at its forefront and is heavily influenced by its market and competition which makes for an *Outside-In approach*. In other words, while there are some elements of an inside-out approach, the outside-in approach is the most dominant at Agrovision.

<u>No Mission or Vision:</u>
While Agrovision has no Mission or Vision statements, its strategy is apparent to its employees and customers. As mentioned in chapter two, there are many changes Agrovision has implemented and is still continuing to implement. The integration of ML into Agrovision is part of these change and is part of the primary strategy to improve its value proposition and to become the "Google of Agriculture" (Veldhuis B., Personal Communications, 13 November 2020).

<u>Agrovision Structure and the Effect of Covid-19:</u>
In order to achieve this, Agrovision has become a relatively flat organisation that enables it to evolve and innovate at a rapid rate, but at the same time keep its customer base. Part of its ability to innovate revolves around result-driven teams who can adjust themselves to ever-changing goals. Furthermore, in part, Agrovision has become a virtual organisation because of multiple offices in multiple countries (some scrum teams span across three countries). Even though this has the potential to hinder development, Agrovision, with the use of technology, is still managing to remain as innovative as ever. The same technology has assisted during Covid-19 when working at the office was not possible. Even at times like these, Agrovision has continued its work as planned and even increased its sales in the last quarter of 2020 (Rasmussen J., Personal Communications, 13 November 2020).

Change Strategy Design:

For creating the change strategy, a model was used as a *framework* for a *systematic change management approach*. This model was used as a common language between different stakeholders as well as an agenda for mutual discussions. As discussed previously, even though Agrovision has no Mission and Vision statements, its strategy is clear and all employees and customers and all are aligned and attuned to it.

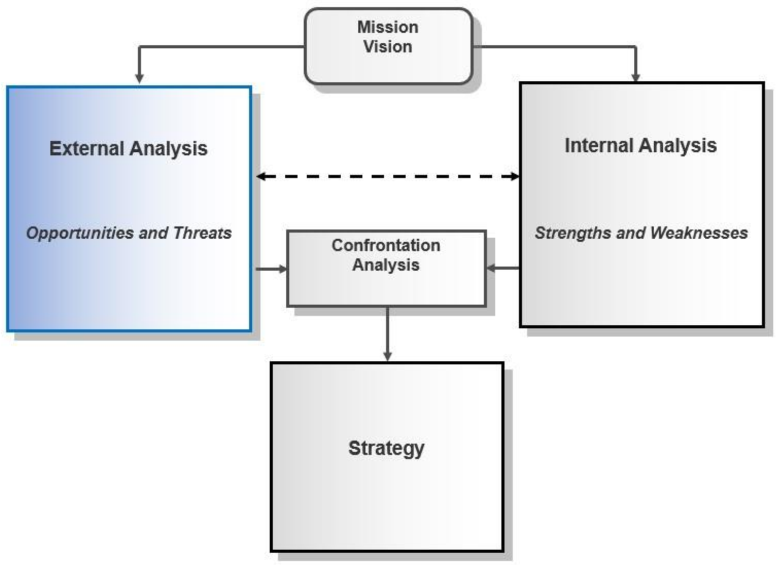

Figure 18 – the model for formulating the change strategy – starting with External Analysis

5.5.2. External Analysis

The first step of this model is external analysis. However, because this project consists of a relatively small change of strategy, there is no need to conduct a complete external analysis. Moreover, as mentioned before, the strategy of Agrovision is clear and obvious, so there is no need to create any change.

The only topic worth mentioning here is the lack of ML analytics which is a disadvantage of Agrovision compared to its competition, as mentioned in the problem analysis. Nevertheless, as this disadvantage is well documented, there is no need for additional analysis or further mentioning its effect on the strategy.

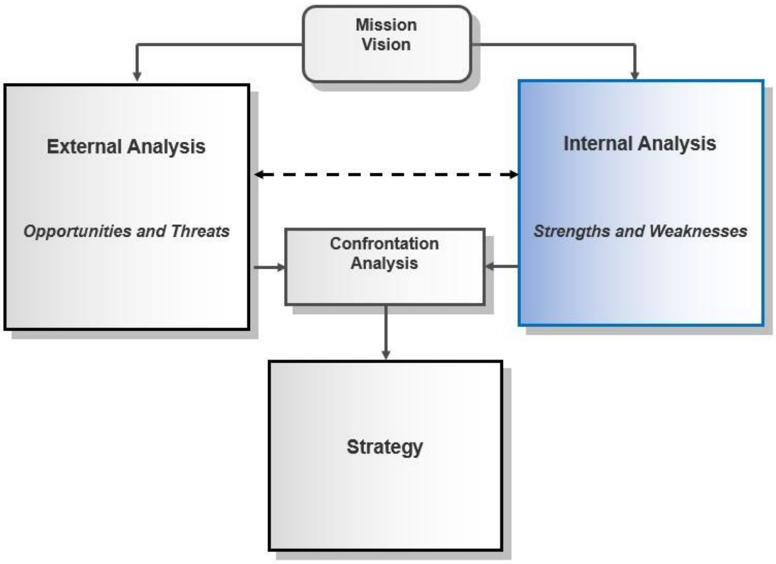

Figure 19 – the model for formulating the change strategy – continuing with Internal Analysis

5.5.3. Internal Analysis

The analysis which concentrates on the internal organisation is an integral part of the change analysis. Understanding the forces which shape the organisation, its culture and structure, amongst other topics, is essential for the change plan. Although some of these topics were discussed in previous chapters, it is essential to bring them all together for the analysis to be successful.

<u>Organisational Structure:</u>
As shown in Figure 2, the Agrovision organisational structure is relatively flat. This enables direct communication between employees at different levels, making it easy to innovate and adapt when needed. Although there is still an evident hierarchy, the managers are there to support the employees. This structure has enabled Agrovision to innovate and be flexible on the one hand, but still have the control mechanisms on the other.

Furthermore, the structure is formed to enable the different departments to Function as efficiently as possible. The *Sales* department is divided *geographically*, whereas the *Operations* department is divided both through *product* and *geography*. Lastly, the *Innovations* department is divided according to *product* and *technology*. Each department has a structure that benefits its functions and the ability to innovate the most, according to the clients' needs and internal business understanding.

<u>Goals:</u>
When goals are being set according to the company strategy, the end result is well-defined, and all stakeholders can be aligned in the same direction. If the goal of Agrovision, as

mentioned before (section 5.5.1. General Analysis), is to be the "Google of Agriculture" then it is obvious that ML is part of this goal. However, as this project only deals with ML, the applicable main goal is to have a functioning ML analytics system integrated into Agrovision. All the actions and the ownership of this goal, are specified in this document, making it very clear and straightforward to align all the stakeholders towards it.

Culture:

When approaching a complicated matter such as culture, it is preferable to use an existing model to help define the different aspects and put them in the appropriate location. The model chosen here is called "Cultural Typology" by Robert E. Quinn and Kim S. Cameron. It defines four different cultures of organisations. By identifying the organisation's correct culture, its values, norms, and behavioural patterns can be recognised and analysed. As mentioned in section 5.5.1., Agrovision is an organisation with an *outside-in* approach, which means it is heavily influenced and orientated towards external market forces.

There are two cultures which are externally oriented. Out of the two, *Adhocracy* is the most suited to Agrovision's culture. It was discussed before that Agrovision's flat structure is in such a way which allows and promotes innovation and flexibility. In the past, and no doubt in the future as well, Agrovision's strategy has been to grow through buying other companies and integrating them. Lastly, when a product is delivered to the client, it is considered a successful endeavour.

Additionally, as organisations rarely match exactly into one culture or another, some elements of *Hierarchy* culture in Agrovision (as mentioned in 5.5.1., Agrovision having some elements of inside-out approach). Work is defined by procedures to make it as efficient and organised as possible. There is a long-term focus on stability which is also considered as successful.

Summary:

Both the cultures encapsulate Agrovision's culture. These two cultures stem from Agrovision being an established company on the one hand, but had to adapt to being innovative and flexible on the other, in order to keep up with changing market demands. Agrovision has structured itself according to these cultures, to be able to innovate and be flexible but have control measures put in the right places for efficiency and stability. The culture, combined with the goals and strengths & weaknesses gives a prominent picture of Agrovision's internal mechanisms relating to this project. In the next section, the external and internal analysis come together to form a strategy.

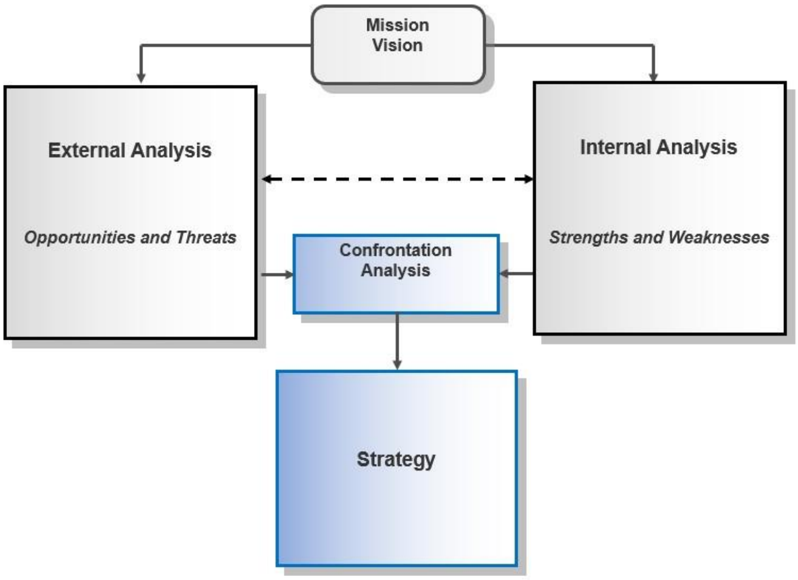

Figure 20 – the model for formulating the change strategy – continuing with Strategy

5.5.4. Strategy

According to Michael D. Watkins from Harvard Business Review, the definition of strategy is (Watkins, 2007):
"A business strategy is a set of guiding principles that, when communicated and adopted in the organization, generates a desired pattern of decision making". By this definition, the strategy is how people throughout the organisation should take decisions and allocate resources to achieve well-defined objectives. To reach such strategy, models were utilised that derived information from the internal and external analysis.

Customer Value Strategy:
In order to further define the different strategies for change which Agrovision can employ, the three customer value strategies by Treacy & Wiersema are utilised:

a) Product Leadership - This strategy embodies the goal of Agrovision to be the "Google of Agriculture."
b) Operational Excellence – leading the agricultural market in price and convenience, making its operations lean and efficient. Though not as high on the Agrovision priorities, this strategy is still one which they strive for – delivering the best product at the best economic value.
c) Customer Intimacy - Agrovision relies heavily on its extensive market knowledge, customer feedback and bespoke reporting.

As can be seen, there is an element of all three strategies deployed by Agrovision, even if one has more emphasis than the other. This means Agrovision is attempting at overall excellence and market leadership, which does correspond with 'being like Google'.

Summary:

If these three strategies are plotted on the Cultural Typology (as depicted in figure 5), it is clearly visible that Agrovision's strategy is concentrated on the external market, which was mentioned before, but is being revalidated once again. Moreover, the Adhocracy culture and Product Leadership strategy are placed together and are the most dominant features in Agrovision. Once again, this reaffirms the goal of being like Google by the need to innovate, be flexible and have a strong emphasis on growth. Regarding only the project, the *strategy* was to get *familiar* with and *offer ML analytics* to Agrovision's customers. This goes hand in hand with the organisation's culture and strategy. The next section discusses the desired future situation according to all the previous findings.

5.5.5. Future Situation

Once all the necessary analyses have been done, there is a need to describe the future situation which will exist once the change process is complete. If the complete strategy of Agrovision had to be changed, this step would have taken some further analysis. However, because it is only a relatively minor change, many different tools do not have to be applied. For example, the Balanced Scorecard is usually done to assist in change management because it can give a complete picture of the organisation's various aspects. Nevertheless, the balanced scorecard is a time consuming and lengthy process, so it is not needed for this relatively small change. Furthermore, Agrovision's strategy or structure will have minor changes, so there is no need to re-evaluate these criteria. The factors which do need to be defined and analysed are stated below.

Key Success Factors:

KSF's are created to form a common language for the forthcoming change. They create commitment by the organisation, define the feasibility and what support is needed for the change. The definition of Key Success Factor (KSF) according to Cambridge Dictionary is (Cambridge Dictionary, n.d.):

"one of the most important things that a company or organization must do well in order for its business or work to be successful".

If the same definition is applied to this project, it would be – "the most important things Agrovision must do well for the project's objective to be successful". The objective of the project is defined in the main question:

What are the best AI solutions for AgroVision to implement in their DWH for better analysis and prediction of the data?

By the above definitions and objective, the KSF's of this project are:

a) Knowledge of AI or ML analytics.
b) Data understanding.
c) Appropriate software architecture.

Improvement Actions:
For each KSF, there needs to be an Improvement Action, or what needs to be done to realise the KSF. If the KSF's are derived from the main question, the improvement actions are derived from the sub-questions. In other words, all the activities carried out and specified during this project are done to realise these KSF's.

Summary:
What properties should Agrovision possess to achieve the project's objective are defined by the KSF. Because of that, the KSF's define the desired future situation of Agrovision. The improvement actions are the means to get to the future situation. Once the previous analyses were done and the future situation defined, the change's implementation plan can be formulated.

5.5.6. Implementation

There is more than one possible way of implementing a change plan. Two of the most common ways are Project Management and Process Management. By definition, project management is temporary and aimed at specific results and should be concluded within a limited time. Process management is used when there are very complex projects, often being inter-organisational. It usually starts with scenario analysis and the changes are gradually realised.

This project matches the definition of Project Management because it is temporary and there is a very specific objective. This project is divided into five phases:

a) Initiation
b) Research
c) Determination
d) Design
e) Implementation

An After-Care phase conventionally follows these phases, but this is outside the scope of this project (the phases of this project are all defined in section 4.8). Compiling the change plan was done in the design phase and the implementation of the change plan would have been part of the Implementation phase. The chosen model for the change implementation is Kurt Lewin's 3-stage model of change (Study.com, n.d.):

Unfreezing:
The term unfreezing stems from the fact that the starting situation (frozen) before the change is flawed and needs to be unfrozen to change or move. This is the initial stage of the change process within which the *motive* for change is determined, the final objectives are defined and the relevant stakeholders are made aware of the impending change.

In reality, this stage started before the beginning of this project, when Agrovision realised they need to incorporate ML analytics into their value proposition. This project was created to make the desired change. The early stages of the project were part of the unfreezing process. The central part of this stage provided the different stakeholders with the need for change, the anticipated future situation, the change's primary goal, and how the change will occur. The unfreezing was done during the project's initiation, research, and design phases. The different steps taken are described in sub-questions one, two, three and six.

Moving:
This is the *implementation* stage of the process. Here the incentives for the change need to be communicated as well as making agreements in order to mitigate the resistance to the change. Other actions that can be taken to alleviate the resistance indicate the advantages of the change (to the organisation and personal) and demonstrate the results that can demonstrate these advantages.

The implementation would have occurred during the implementation phase and would have been done according to the change plan. The implementation did not take place in practice, as described in section 5.6.

Refreezing:
This is the stage where the new change is being *reaffirmed* and the new situation is wholly *embedded* in the organisation, hence the name refreezing. At this stage, feedback can be given and rewards can be offered for assimilating the change. Successes should be communicated and failures corrected. At the end of this stage, the change is entirely accepted and the process is finished.

The refreezing stage was not in the project's scope as it concerns the After-Care phase. As such, it was not carried out.

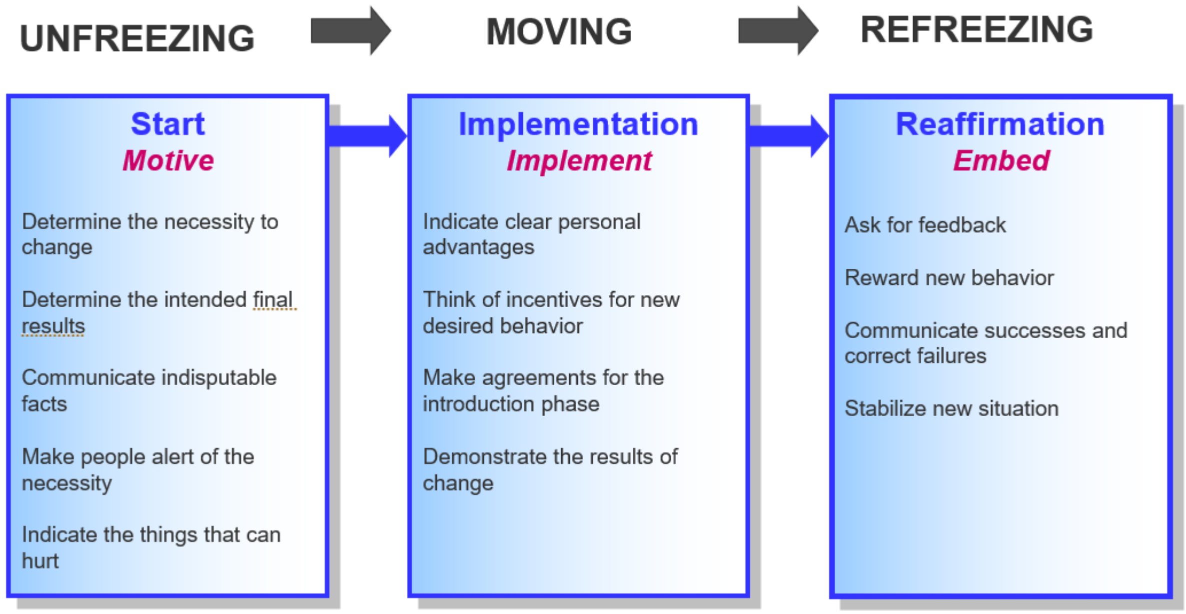

Figure 21 – change implementation process

5.5.7. Change For Structure and Personnel

As the change is relatively small, Agrovision's personnel's consequences are small as well. When talking to Jacob Rasmussen, Managing Director of Agrovision, he mentioned there would be no immediate change in the company structure because of the new ML solution. He stated that there is planning to create an ML analytics team in roughly two years when the analytics personnel will be proficient enough with the ML work and assistance will not be needed from IBM 2020 (Rasmussen J., Personal Communications, 13 November 2020). The current analytic personnel will handle the extra ML analytics work for the foreseeable future.

5.5.8. Summary

With any change in the organisation's strategy or processes, there needs to be a change plan to negate or completely eliminate the resistance to the change. Even if the change is relatively small, as in this project, a change plan is very beneficial to ensure all stakeholders are aligned to the same objectives, see the benefits of these objectives and are willing to go through a period of uncertainty for these objectives. The next step is implementation, which would have been done according to the implementation and change plans.

5.6. Results - Sub-Question Seven

Unfortunately, as mentioned before, the implementation process did not occur, except for the small part determining the use-cases. Nevertheless, given more time for the project, the implementation could have been completed because all the research and work before it was complete. The most suitable solution was chosen, according to the requirements, the implementation plan was made and the change management plan was also specified. These steps will enable to carry out the implementation when Agrovision decide to implement the ML analytics solution.

5.6.1. Start of Implementation Process

The first part of the implementation plan was to submit use-cases to IBM in order for them to understand the business needs and be able to investigate which data exists in order to analyse these use-cases.

5.6.1.1. Use Cases

Four different experts at Agrovision were contacted to contemplate use-cases which they think can add the most business value to Agrovision. The four people were chosen because of their expert domain knowledge of pig-husbandry (Browers J., De Weerdt J., Søgaard C., Wisselink A., Personal Communications, 5 November 2020). These use-cases were formulated in a one-sentence form and were all assembled as a complete list of 19 (see appendix 2).

<u>IBM Suggestion:</u>
IBM's suggestion was to choose between three to seven use-cases for the model building. If there are too few, it can happen that all are not attainable rendering the project unsuccessful. If there are too many, it will be hard to determine which has the most business value. For that reason, a meeting was held with the same expert to determine which of the use-case holds the most business value. The other criterion considered was the availability of data. During that meeting, it was decided the three use-cases which to provide IBM with (Use Cases to IBM (meeting), Personal Communications, 17 November 2020):

1. Forecasting the number of piglets to be born.
2. Where is there a need to replace a sow (an adult female pig that has given birth)?
3. Predict coming litter according to the weight of sow or back-fat (amount of fat a pig has on its back).

End of Physical Steps:
These three use-cases were sent to IBM. Unfortunately, this was the last physical step which took place during the project (the reasons are specified in section 5.6.2.). These use-cases were the basis for continuing with the implementation according to the implementation plan.

5.6.2. Reasons for Delays

The reasons for the implementation not taking place are mainly delays. As shown in the Gantt Chart below, the different stages took a longer time than initially anticipated. The determination stage took longer than anticipated because the decision team needed more information to reach a conclusion. The designing phase has taken longer than expected because there was a need to get authorisation from Agrovision's management for sharing the data with IBM, which was also delayed. Being the last part of the project, the implementation did not occur due to these delays.

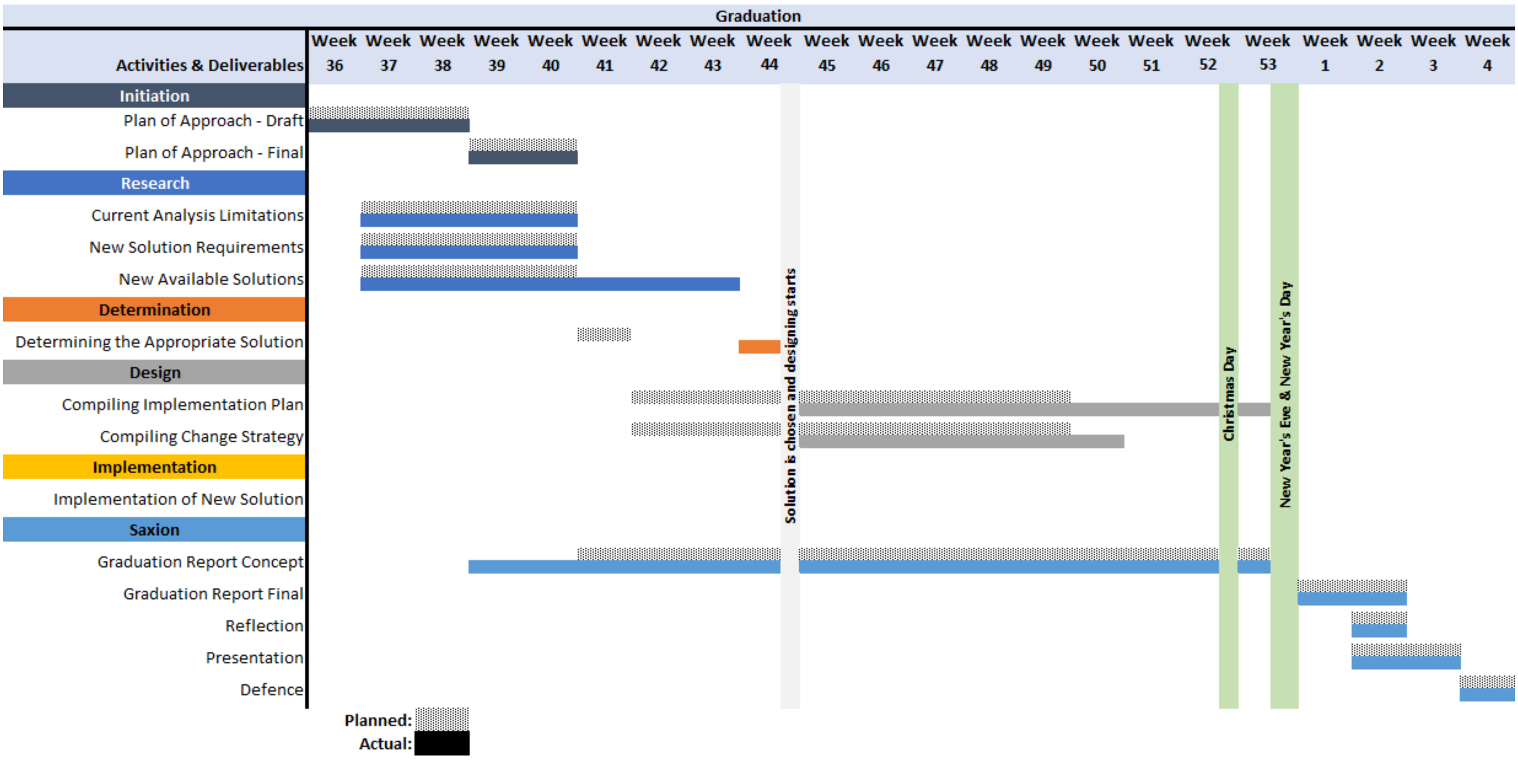

Figure 22 – Gantt chart planned vs actual

5.6.3. Summary

It was regrettable that the implementation did not occur, but very early during the project, it was ascertained that the iterative part of the implementation would have taken much longer than the project's time scope. There are many criteria to take into account during this process and there is a need for a great deal of data to be available. From conversations with other organisations went through this process, it would have taken between a few months to several years to complete. Nevertheless, the process can still continue after this project has ended, should Agrovision choose to do so.

6. Conclusion

This thesis's main question covers many different topics, from choosing the best solution to implementing it to using it to analyse the data. This is no coincidence, as this project's objective was meant to give a complete solution to Agrovision's problem of not having ML analytics as part of its value proposition. Even if the project's implementation did not take place, its foundations have been systematically and meticulously placed using a firm methodology.

The ICT Research Methodology has proven itself in each step of the project. It was able to encompass every activity utilised, thus providing a robust framework for the project. Moreover, its validation capabilities were proven invaluable, as every finding was justified and verified. Even though other methods were used during the research, such as the cultural typology by Quinn and Cameron or the Confrontation Matrix, they were always used within the ICT Research framework and as such were an integral part of it. As a product of such methodology, the research outcome was of the highest quality.

As stated during the project, the necessity for ML analytics within Agrovision was apparent to all stakeholders. Nonetheless, the most suitable solution and how to implement it was a complete unknown. This research not only found the appropriate solution, but also described in great detail, in the implementation and change plan, the different steps that must occur for its integration into Agrovision.

When going back to the main question, the answers are resoundingly positive. The best solution was found according to all aspect considered and the plans were depicted for realising it. Should Agrovision decide to continue this project, it has these fundamentals to rely upon for the best possible implementation and later increasing its value proposition.

7. Recommendations

At the beginning of the project, there was one fact which was clear to all stakeholders – there was a need to add ML analysis to Agrovision's value proposition. This was the starting point which eventually led to creating this assignment. This need to incorporate ML is not always so evident to other enterprises as some do not see the added business value which can be gained.

However, this project only achieved a certain part of its final objective because implementing the ML solution was not possible. Agrovision should aspire to continue this project as the necessity of ML is so clear-cut. Moreover, Agrovision can decide whether to maintain the same line this project has taken with the same suggested solution, or it can choose a different solution which might be appropriate for other requirements than the ones which were debated here.

Even if Agrovision chooses a different solution, many of the steps which were described here will still be applicable. The software architecture or the connectivity might change if the solution exists in a different configuration than Watson Studio, but Agrovision can still utilise the implementation plan excluding these aspects. The same applies to the change plan. Some small details will not be applicable in the new situation, but most of it should be consulted and utilised for a problem-free solution integration.

Whichever direction Agrovision will decide to continue this project, the most important recommendation is that ML analytics will be integrated and implemented because both Agrovision and their clients would like to benefit from the added business value and competitive edge.

8. Discussion

The primary problem Agrovision had was that no ML analytics was offered as part of its value proposition. This created a twofold problem for Agrovision. The clients of Agrovision who are not benefiting from ML analytics can fall behind to farmers who do utilise it. This can have further effect when these clients of Agrovision decide to work with software that includes ML analytics, meaning they stop working with Agrovision.

The reasoning for the Project:

The above problem gave rise to this project, during which a suitable ML solution was obtained for Agrovision and a clear implementation plan was devised to integrate the solution into Agrovision's software architecture, including how to prepare the data, create the cube for analysis, generate the ML model and verify the results.

Need for ML Analytics:

As the research proceeded further, it was realised that the need for ML analytics was not a trivial one. The world population continues to grow, which reduces the amount of land that can be cultivated, but the need to increase food production is only mounting. There is a need to extract more yield from every piece of land. For that reason, farmers and the companies who support them have turned to ML analytics.

Advantages for Farmers:

Unlike more conventional analytics methods, ML analytics has the ability to search through much more considerable amount of data and draw conclusions based on patterns that emerge in the data. It uses sophisticated algorithms to discover these patterns and is also able to learn from itself, using the same algorithms, to improve these pattern recognitions. These patterns can reveal developments that are currently happening, such as diseases evolving in the crop, and crucially make predictions of future events, such as the amount of milk produced. These ML abilities can give the farmers insight in real-time or even before time, which can help make the correct business decisions according to the overall strategy, thus reducing cost and increasing productivity.

Advantages for the Agribusinesses:

For the Agribusiness organisations which either support or benefit from the farmers' products, ML analytics has great value also. If a farmer can commit to a distributor of a certain amount of produce, the distributor can prepare itself well ahead of time for that amount. This will negate the risks which are involved and in turn will reduce the costs. It will also make the distributer more reliable, which will increase its market value.

These processes which ML analytics drive forward are already taking shape with organisations who employ this practice. They benefit from distinct advantages over their competitors by producing more at a higher reliability and at a lower cost.

Project Limitations:
As mentioned before, the most significant limitation of the project was time. Even if the initial time predictions for the processes before creating and implementing the model were correct, the model's actual creation would have taken a considerable amount of time, which is much more than the five months scope of this project. ML analytics' iterative nature makes it a long and extensive process, which in many occurrences is stretched beyond the eight weeks allocated to it initially. Once taking into account the biological factor, this process would probably have been even longer. Unfortunately, this limitation was only apprehended during the project. When Agrovision continues this project, the time frame for creating the model will be adjusted to match the factors that influence it.

The research during this project has revealed much more than was expected. Of course, there were very distinct answers, such as the appropriate ML solution for Agrovision, but many other elements were only realized as the project unfolded. Elements such as the time frame for creating the ML model, the biological factor, the use cases which need to be researched and submitted to IBM, the complexity of the change management plan and the like. These elements might have delayed the research to some extent, but were also a reminder of why this research was needed – to uncover all the small steps that are part of integrating an ML solution into Agrovision.

9. Bibliography & Appendices

Afimilk (n.d.). Retrieved 29 October 2020, from https://www.afimilk.com/

Allflex (n.d.). Retrieved on 29 October 2020, from https://www.allflex.global/

Bonestroo, W.J., Meesters, M., Niels, R., Schagen, J.D., Henneke, L., Turnhout, K. van (2018): *ICT Research Methods. HBO-i, Amsterdam*. Retrieved 10 September 2020, from http://www.ictresearchmethods.nl/

Cambridge Dictionary. (n.d.). *Meaning of critical success factor in English*. In dictonary.cambridge.org. Retrieved 02 December 2020, from https://dictionary.cambridge.org/dictionary/english/critical-success-factor

De Vlieger, Rick (13 April 2016). *Confrontation Matrix*. Retrieved 26 November 2020, from https://www.calltheone.com/en/consultants/confrontation-matrix

Dinsmore, T. (14 November 2018). *How to choose your AI vendor*. DataRobot. Retrieved 07 October 2020, from https://www.datarobot.com/blog/how-to-choose-your-ai-vendor/

Grit, R., Julsing, M. (2015). *How to do research (1st edition)*. Retrieved 06 November 2020

Grit, R. (2011). *Project Management, A Practical Approach (3rd edition)*. Retrieved 05 October 2020

IBM Watson Studio (n.d.). Retried 16 December 2020, from https://www.ibm.com/nl-en/cloud/watson-studio

Marr, B. (19 February 2016). *A short history of machine learning – every manager should read.* Forbes Magazine. Retrieved 23 September 2020, from https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#1e983f215e78

Marr, B. (05 June 2020). *The ten best data analytics and BI platforms and tools in 2020.* Forbes Magazine. Retrieved 29 September 2020, from https://www.forbes.com/sites/bernardmarr/2020/06/05/the-10-best-data-analytics-and-bi-platforms-and-tools-in-2020/#69d264c9560d

Morgan, J. (08 July 2015). *The 5 Types Of Organizational Structures: Part 2, 'Flatter' Organizations*. Forbes Magazine. Retrieved on 12 November 2020, from https://www.forbes.com/sites/jacobmorgan/2015/07/08/the-5-types-of-organizational-structures-part-2-flatter-organizations/?sh=2505490c6dac

Oko A. (2020). *Plan of Approach V2.0* [Unpublished Project Plan]. Saxion University of Applied Sciences. Retrieved 07 October 2020.

Over ons - Agrovision (n.d.). Retrieved 30 October 2020, from https://www.agrovision.com/nl/over-agrovision/

Quinn, R.E., Cameron, K.S. (n.d.). *Robert S. Quinn and Kim S. Cameron's Cultural Typology*. Quinn Association. Retried 23 November 2020, from https://www.quinnassociation.com/en/culture_typology

Sharda, R. Delen, D. Turban, E. (2018). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective (4th Edition).* Retrieved 02 November 2020.

Study.com (n.d.). *Lewin's 3-Stage Model of Change: Unfreezing, Changing & Refreezing*. Retrieved 03 December 2020, from https://study.com/academy/lesson/lewins-3-stage-model-of-change-unfreezing-changing-refreezing.html

Treacy M., Wiersema F. (1993). *Customer Intimacy and Other Value Disciplines*. Harvard Business Review. Retrieved 01 December 2020, from https://hbr.org/1993/01/customer-intimacy-and-other-value-disciplines?cm_sp=Article-_-Links-_-Comment

United Nation. *World population prospects 2019* (n.d.). Retrieved on 24 December 2020, From: https://population.un.org/wpp/Graphs/DemographicProfiles/Line/900

Voortman, H. (2017). *Strategic Change Management* [PowerPoint slides]. BlackBoard@Saxion. From https://leren.saxion.nl/webapps/portal/execute/tabs/tabAction?tab_tab_group_id=_1695_1

Watkins, D. Michael (10 September 2017). *Demystifying Strategy: The What, Who, How and Why*. Harvard Business Review. Retrieved 26 November 2020, from https://hbr.org/2007/09/demystifying-strategy-the what#:~:text=Here's%20my%20definition%3A%20A%20business,desired%20pattern%20of%20decision%20making

Microsoft Azure (n.d.). *What is Azure?.* Retrieved 13 January 2021, from https://azure.microsoft.com/en-us/overview/what-is-azure/#most-popular-questions

9.1. Personal Communications

- *Bakker A.* – Software Developer, Innovations, Agrovision
- *Beijer J.* – BI Consultant, Innovations, Agrovision
- *Browers J.* – Export Manager, Sales Marketing & PMT, Agrovision
- *Rasmussen J.* – Managing Director, Agrovision
- *Rooks W.* – BI Consultant, Innovations, Agrovision
- *Said Fredsted P.* – Customer Service Employee, Operations, Agrovision
- *Søgaard C.* – Product Manager, Sales Marketing & PMT, Agrovision
- *Veldhuis B.* – Manager Innovations, Innovations, Agrovision
- Weerdt De J. – Product Manager, Agrovision
- *Wesselink M.* – Teacher/Researcher, Academy Creative Technology, Saxion University of Applied Sciences
- *Wisselink A.* – BI Specialist, Innovations, Agrovision

9.2. Glossary

- ADEX – Agrovision Data Exchange
- AI – Artificial Intelligence
- AMT – Architectural Management Team
- Back Fat – the amount of fat a pig has on its back (used to determine the total amount of fat or the weight of the animal)
- BI – Business Intelligence
- Boar – male pig
- DB – Database
- DWH – Data Warehouse
- EA – Enterprise Architecture
- IoT – Internet of Things
- KPI – Key Performance Indicator
- KSF – Key Success Factors
- ML – Machine Learning
- OLAP – OnLine Analytical Processing
- Sow – a female pig
- SSAS – SQL Server Analytics Service

9.3. Appendix 1 – Table of Nine Solutions

Name	Advantages	Disadvantages	Cost[1]
IBM Watson Studio	• Complete ML solution • Can connect to Azure DB and embedded on website • Can be programmed with Python/R • Graphic interface (zero programming) • Will set up a test environment for free • Provide training (for a fee	• Completely new to Agrovision • Large learning curve	
Qlik Sense	• Currently considered one of the best BI tools • Easy connection to DB • Easy embedding into website • Provide on-premise training (for a fee) • Well documented	• Completely new to Agrovision • Large learning curve • No 'out of the box' ML algorithms – need to program with Python/R • Similar to Power BI	
Oracle Machine Learning	• Comprehensive solution • Autonomous database • Faster than Azure (their claim) • Can be programmed with Python/R	• Need to maintain two DBs • Completely new to Agrovision • Large learning curve	
KNIME	• Graphic interface • Working in the Azure cloud • Connects to Power BI for reporting • Can be programmed with Python/R • Free for testing (one user)	• Smaller company – fewer resources? • No office in the Netherlands	
Python/R	• Completely customizable	• Completely new to Agrovision • Large learning curve	

[1] Though these costs were presented to Agrovision personnel, they are not presented here for privacy reasons

	• Can be built for Agrovision specific needs • Many existing libraries • Widely used for ML	• Need extra personnel	
Python/R in Power BI	• Can be built for Agrovision specific needs • Many existing libraries • Existing Power BI knowledge	• Completely new to Agrovision • Large learning curve • Need extra personnel	
Azure ML Studio	• Using the same environment as the rest of Agrovision • Can be easily integrated into existing Agrovision products • Can be connected to Power BI • Graphic interface (zero programming) • Can be programmed with Python/R	• Completely new to Agrovision • Need extra personnel	

9.4. Appendix 2 – Initial List of Use-Cases

	Use-Case	Name
1	Predict daily gain of finishers.	Jan Browers
2	Profit margins of future selling.	
3	Forecasting number piglets to be born.	
4	Disease predictions.	
5	Which variables (management decision, climate etc.) can change meat quality?	
6	Which variables (management decision, climate etc.) can change fertility & reproduction?	
7	In heat predictions.	Arnold Wisselink
8	Optimum sow serving time after weaning (in hours)	
9	Low activity alarm – why?	
10	Low feeding alarm – why?	
11	In autumn birth is less (some by 5% some by 20%). Why?	
12	How does birth weight affect slaughter weight?	
13	When a certain group will be ready for slaughter?	Camilla Søgaard
14	When to buy feed?	
15	What kind of feed proves better?	
16	How many sows/inseminations to get X KG of meat?	Jelle De Weerdt
17	When is there a need to replace a sow?	
18	Predict coming litter according to weight of sow.	
19	Predict which farm will stop using Agrovision.	