Data Processes

Project Plan

Date: 09.01.2022

Group Members:
Azeez Abdikarim, Bela Bönte, Eleonora Renz, Eduardo Freyre

Task:
To design a project plan in which a business goal must be detailed and consequently, a plan designed.

Disclaimer:
This document has been produced as a report for the final project on the course Data Processes. The input data has been generated artificially, and thus, the conclusions obtained during the project are by all means unsupported.

Table of Contents

Document History

Version	Date	Issued by	Description
V1	2021-12-24	Renz, Eleonora	First Draft
V2	2022-1-4	Freyre, Eduardo	Second Draft
V3	2022-1-7	Bönte, Bela	Third Draft
V4	2022-1-8	Abdikarim, Azeez	Fourth Draft

List of Figures

List of Table

1. Executive Summary

1.1 Summary

The planning and implementation of a business goal is a complex undertaking that can always lead to complications. To minimize the risks of delays or budget overruns, extensive planning is therefore necessary. This project report includes a detailed planning of the work steps to be completed, the risks to be weighed and a budget planning to achieve the targeted project goal. The introduction of the project report deals with the necessity and applicability of the evaluated data.

2. Introduction

2.1 Motivation

The Omicron variant is much more contagious than the Delta variant, which has prevailed up to now. In many parts of the world, Omicron has already caught up with or replaced the Delta variant. For this reason, measures have already been tightened in some European countries, such as the Netherlands or Germany, to prevent overloading of the healthcare system [3]. Despite this, there may be a shortage of beds that cannot be solved by redistributing patients. In this case, a so-called triage occurs.

“The sorting out and classification of patients or casualties to determine priority of need and proper place of treatment.”[2]

To make this decision effectively, many factors are important, but the prospect of survival is most crucial. In this situation, a data-driven decision-making process would help doctors assess a patient’s prospect for survival. This product is not intended to supplant a doctor’s opinion, but rather to augment their decision making process. While an individual doctor is limited in experience to the COVID cases they’ve interacted with, a data-driven can provide recommendations informed by the experience of thousands of COVID patients.

2.2 Purpose and scope of the project report

The goal of this project report is to identify the order of tasks that must be completed to bring this product to production, as well as to acknowledge the risks associated with a medical triaging model. Special attention is paid to how this project will be implemented. A concise timeline is necessary because of how quickly the Omicron variant is spreading, and the risk that more dangerous variants may develop. Each section of the project report covers a different aspect of the development. So, the report defines a time frame for the entire project team, lists budget constraints, and assesses risks.

3. Goals

3.1 Project Goal

Context:
With the increasing number of Omicron cases, and the already high burden on the healthcare system, triaging may have to be applied. This would place an incredible psychological mental burden on the attending medical practitioners [1]. To take some responsibility away from doctors, a computer-based approach that provides information about a patient's chances of survival could be used to support making decisions. Therefore, the goal is an application for decision support, based on a model which can, given the initial conditions of a patient, produce a statistically accurate estimate of the survival chance if they were admitted to hospital. This methodology could be sold to governments or hospitals worldwide facing the same problem.

Table 1: Table of project goals

Project objectives	Description	Indicator of success
Increase ICU survival rate	Triaging patients allows doctors to selectively treat patients with the highest likelihood of survival. If the model is effectively triaging patients, it can be expected that patients with an extremely low chance of survival are not occupying ICU beds. Since these deaths will not occur in the hospital, survival rates within the ICU will increase.	Survival Rate of Patients that stay at least one day in the ICU
Increase Net Promoter Score (NPS)	NPS measures the likelihood that customers recommend the product to others. The metric would be compiled by responses from doctors and nurses, who would evaluate how the product improves their ability to allocate resources.	Doctors and Nurses
Increase turnover	To increase sales, various factors must be optimized. More customers must be acquired (Marketing) and a suitable pricing model must be developed.	Turnover
Increase profit	Profit can be increased by scaling the project. In this context, that look like getting more hospitals to utilize this data-driven triaging method.	Profit

Refinance costs	In order to refinance development costs, stable hosting of the application, all costs should be covered with after 4-6 months of in-service	Credit Debit

3.2 Data Science Goal

Context:
The technical goal of this data science company is to use advanced computerized predictions that will predict, with a high degree of accuracy, whether or not an admitted patient will survive.

Table 2: Table of quantifiable data science goals

Data mining objective	Description	Indicator of success
Filter relevant data	Each patient has a wide range of data. However, these are of varying interest. In order to statistically predict whether a person will survive, the most influential data attributes must therefore be determined.	Performance of the model (Accuracy, precision, recall)
Model with high degree of accuracy (specially recall)	Since physicians are supposed to use the data later to simplify decisions, the system must work reliably and with high significance.	Difference between hospitals with statistical assistance program and those without
Minimize False Negatives	In this context, we want to minimize the chance that the model predicts a patient will survive when they are admitted, but they actually end up dying. This is a waste of hospital resources.	High accuracy and recall in predictions
Minimize False Positives	The greatest harm a triaging model could cause, is suggesting a doctor refuse to treat a patient because they are likely to die, when in actuality the patient would survive if they were admitted.	Low false positive rate

4. Work plan

4.1 Work Packages

Table 3: Timeline for project tasks

Task Name	Assigned To	Start Date	End Data	Days	Status
Planning of project		15.11.2021	30.11.2021	15	Done
Clarification of data to analyze (Data Exploration)		01.12.2021	07.12.2021	7	Done
Exploration of field of interest (e.g Kaplan-Meier)		01.12.2021	07.12.2021	7	Done
Data analysis and graphs		08.12.2021	20.12.2021	12	Done
Data preprocessing		20.12.2021	01.01.2022	12	Done
Find suitable models (train and test)		02.01.2022	07.01.2022	6	Done
Evaluate models		02.01.2022	07.01.2022	6	Pending
Maintain and Monitor Models		07.01.2022	-	Ongoing	Ongoing

4.2 Deliverables

Table 4: Descriptions of each project task's deliverables

Task Name	Assigned To	Deliverables
Planning of project		Complete breakdown of work packages, risk analysis and budget planning.
Clarification of analyzable data (Data Exploration)		First review of the existing data and their meaning. Clarify how the data was collected and how it generally relates to the solution to the problem.
Exploration of field of interest (e.g. Kaplan-Meier)		Looking at already known patterns in medicine from other contexts. Suggestions for the incorporation into the existing problem definition.
Data analysis and graphs		Clarify the relationships of the data to each other and investigate for missing data. With the help of graphs, specifics within the data set are visualized for the entire team.
Data preprocessing		The insights gained from the analysis, such as missing data, different scales, and erroneous results, are used to put the data into a format that allows accurate AI-powered predictions to be made
Find suitable models (train and test)		In view of the problem, different ML methods are to be evaluated. Afterwards, these have to be tested and adjusted in order to achieve the desired result.
Evaluate models		Since the model has only been trained with a comparatively small data set, data must be collected from new cases. These data are immediately compared with the predictions of the model and, if necessary, incorporated into the model.

4.3 Milestones

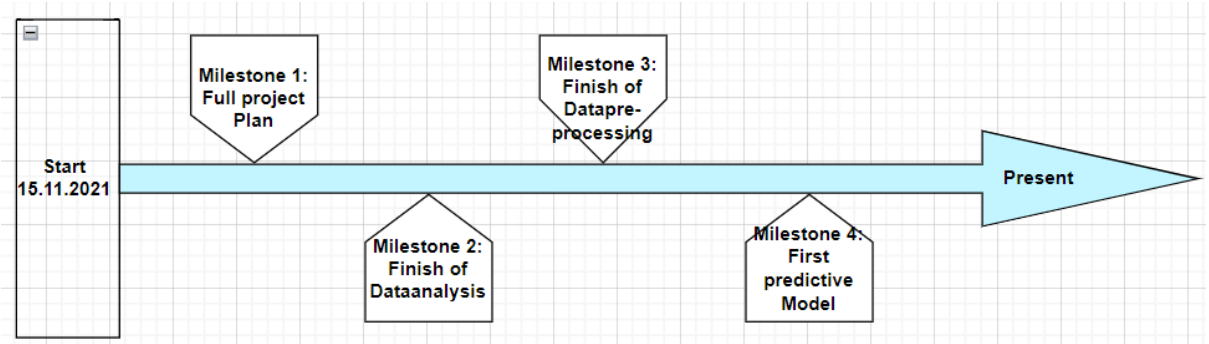

Figure 1: Milestones on time axis

Table 5: Milestones of project plane

Milestone	Description	Approximated time
Full project plan	Division of the task and comprehensive insight into the problem are completed.	30.11.2021
Finish of data analysis	All aspects of the review and analysis of the existing data is completed.	20.12.2021
Finish of data processing	The data set has been freed from erroneous data and only the data required for problem solving is left	01.01.2022
First predictive model	The first model with significant explanatory power was created.	07.01.2022

4.4 Visualization (Pert and Gantt Diagrams)

Notice:
Since the diagrams were created in the first step (Planning of the project), this step is not taken into account in the following diagrams.

Pert Chart:

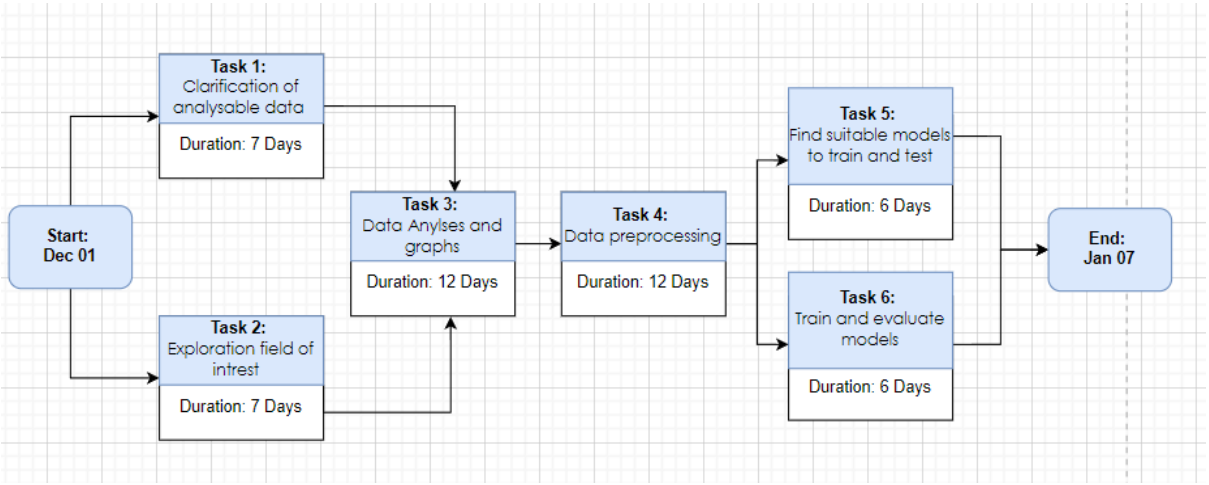

Figure 2: Pert Diagram based on internal work plan

Gantt Chart
The gantt chart shows the sequence of individual tasks in a temporal context.

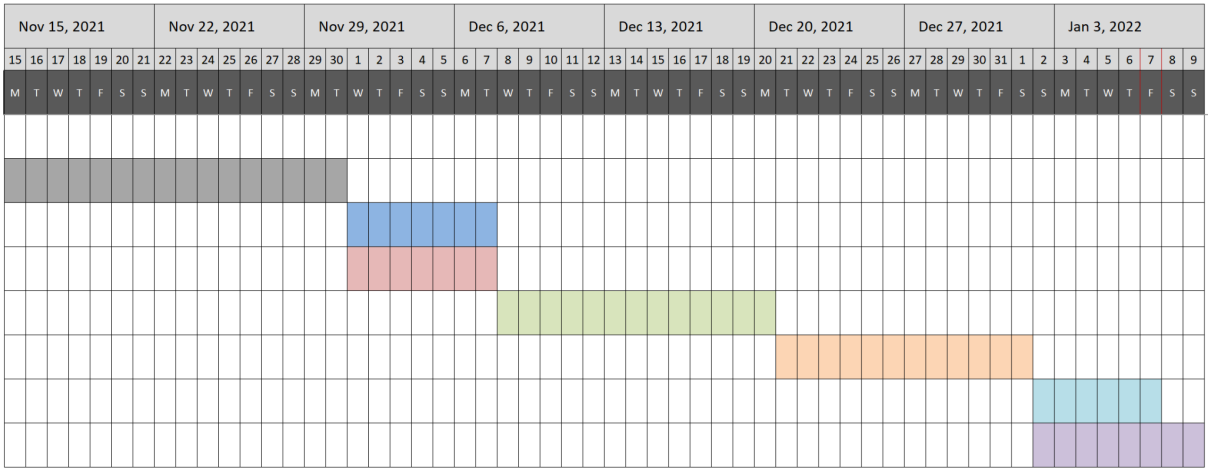

Figure 3: Gantt Chart of project Plan

The colors in the gantt chart symbolize the following tasks:

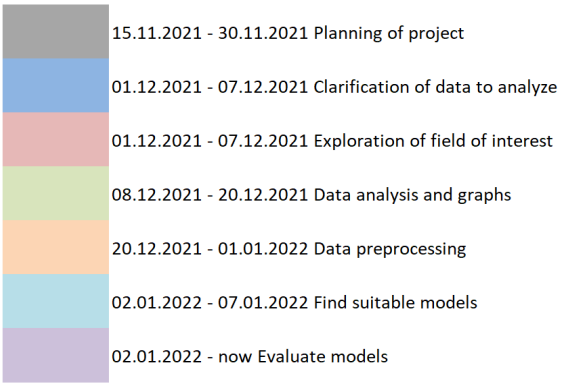

Figure 4: Tasks shown in the gantt chart

5. Risk Plan

5.1 Identifying Sources of Risk

There are a number of risks that should be acknowledged while developing a tool that will help triage incoming COVID-19 patients. Specifying these risks early will help to minimize any potential negative impact to customers as well as maintain overall project efficacy. The main category of risk that this project is exposed to is algorithmic risk. Some of these risks present themselves during the development phase of the algorithm, while emergent risks may only become noticeable once the algorithm has operated in production for some time.

Table 6: Project Risk Plan

Risk	Probability	Contingency Measure
Stale Training Data Since COVID is an ever evolving virus, it's possible that the profile of patients with the highest chance of survival changes.	Medium - but the probability increases the longer the algorithm exists in operation without being retrained.	To keep pace with the evolution of the coronavirus, twice a month the hospital's data team should share the latest patient data with the algorithm's engineering team so that the triaging model can be retrained.
Reputational Risk Families, doctors, and patients may object to the fundamental idea of triaging, especially the fact that a computer system is being used to help in these decisions	High	It will be important to effectively communicate that this triaging model is not making decisions on behalf of doctors, but rather adding an additional opinion for them to consider when deciding what level of care to administer
Correlation Risk Results of the triaging model may be correlated to specific protected attributes (ex. race or gender), however this can only be identified once the algorithm has experience in production. While these features may not be explicitly utilized to train the model, they could be proxied by a combination of multiple data features.	Low	Correlation risk will be assessed twice a month. If it is discovered that the model's results correlate to an individual's protected features, then the model should be withdrawn from production to give analysts time to re-evaluate the set of training data and design a new algorithm that doesn't proxy protected features.

5.2 Monitoring Plan

The monitoring plan for this project calls for the model to be reviewed twice a month. The project plan specifies that the modeling infrastructure and automated reports will be created during the 8 week algorithm development timeline. These reports can be reviewed by a data analyst, while a data scientist can use the findings to decide what updates need to be made to the model. It is requested that new data will be sourced from the hospitals twice a month, so that these new observations can be used to update the model's training data. This is important, so that the model stays up to date with the latest coronavirus strains.

6. Budget

The budget for this project will be split between computer infrastructure, model development, and monitoring costs. This budget covers both the short term costs involved with designing the triaging model, as well as the long term costs associated with maintaining and monitoring the model's accuracy. A majority of the budget will be used in the short term to develop the model and its monitoring tools, while the remaining budget will be used to fund a 3 month monitoring period.

A budget of 40 thousand euros will be used to demonstrate how funds for this project would be allocated. A larger budget could be used to hire more engineers, which would speed up the model's development timeline, as well extend the month to month monitoring. Since it is impossible to estimate when hospitals will no longer be overwhelmed by COVID patients, this budget is flexible, and provides a clear estimate for how much additional months of modeling/monitoring would cost.

Table 7: Proposed project budget

Budget Proposal (40.000 euros over 3 months)			
Short Term (8 weeks)	Computer Infrastructure	● Cloud Storage ● Cloud Compute	● 2.000 euros ● 2.000 euros
	Develop Monitoring Infrastructure	● 4x Data Scientists (40 hours x 8 weeks x 25 euros/hour)	● 32.000 euros
	Model Development Cost		**36.000 euros**
Long Term (Month to Month)	Computer Infrastructure	● Cloud Storage ● Cloud Compute	● 1.000 euros ● 1.000 euros
	Model Monitoring	● 1 Data Analyst (32 hours x 20 euros/hour)	● 640 Euros
	Model Updates	● 1 Data Scientist/Engineer (32 hours x 25 euros/hour)	● 800 Euros
	Monthly Operating and Monitoring Cost		**3.440 euros**

7. Conclusions and Future Steps

This project plan details a two-month timeline for developing a model to triage COVID patients, as well as a month-to-month plan for monitoring and maintaining the efficacy of the model. The goal of the project is to provide doctors with a tool to efficiently triage incoming patients, so that hospitals can achieve the highest possible survival rate during a time of incredible stress on the healthcare system. While it is vital that doctors make the ultimate decision about the care a patient receives, the aim of this triaging model is to identify which patients may be too sick to save, so that resources can be allocated to patients with the strongest chance of survival. Due to the novelty of the COVID virus, doctor's have, at most, two years experience with COVID patients. Providing them with intelligence, generated by a model trained on thousands of COVID patient outcomes, will augment a doctor's ability to make effective decisions.

If this initial model proves to be effective (measured by NPS scores by doctors along with the model's recall score), numerous steps can be taken to develop the product further. Currently, the model is trained off of ~2000 historical patients. Acquiring additional records of COVID patient outcomes would provide the triaging model with a larger training data set, which would likely improve model performance. The triaging model could also be improved in future iterations by modifying the model's output so that it no longer predicts a binary classification (Exitus vs Non-Exitus), but rather the probability that a patient may PERISH. Delivering the model's output as a percentage would communicate the confidence the model has in its prediction, and would therefore be more interpretable by doctors.

References

[1] Bard, T. and Bursztajn, H., 2022. *Triage Trauma and Moral Distress*. [online] Psychiatric Times. Available at: <https://www.psychiatrictimes.com/view/triage-trauma-moral-distress> [Accessed 6 January 2022].

[2] Medical Dictionary Online. 2022. *Triage - Medical Dictionary online-medical-dictionary.org*. [online] Available at: <https://www.online-medical-dictionary.org/definitions-t/triage.html> [Accessed 6 January 2022].

[3] Uniyal, P., 2022. As Omicron overwhelms hospitals, we must talk about triage. *Hindustan Times*, [online] Available at: <https://www.hindustantimes.com/lifestyle/health/as-omicron-overwhelms-hospitals-we-must-talk-about-triage-101641370093362.html> [Accessed 6 January 2022].