POLITECNICO DI TORINO

College of Engineering and Management

Master of Science in Engineering and Management

Master of Science Thesis

Machine Learning Study to Improve Surgical Case Duration Prediction

POLITECNICO DI TORINO

Supervisor:

Domenico Augusto Maisano

Candidate:

Raed Akkawi

April 2022

Abstract

An accurate estimate of the duration of an intervention is essential to optimize the utilization of the operating room, it plays a fundamental role in reducing the cost of the operating room (OR). The approaches mostly used by hospitals are based on historical averages based on a specific surgeon or specific type of procedure obtained from the electronic medical record (EMR) scheduling systems. However, the low predictive accuracy of the EMR leads to negative impacts on patients and hospitals, such as rescheduling of surgeries and cancellations which costs a lot of money.

Our aim in this study is to improve the prediction of surgeries duration using advanced machine learning algorithms to construct a predictive model. Firstly, we obtained a large data set containing 66,857 surgery cases undergone in Rivoli, Pinerolo, Susa, and Venaria from 2016 and on. After exploring the data by detecting the outliers and plotting, we proceeded to clean the data and analyse it, then we trained the model. We computed historic averages of each surgery and they were used as a baseline model for comparison with the model we created.

After constructing the model, it was implemented on an application to be more user-friendly and be used by anyone by inserting different variables as input and getting the predicted time of the surgery in addition to a graph showing occupation time of the operation theatre of the chosen type of surgery in the past according to the data we have. The machine-learning algorithm showed higher predictive capability than electronic medical record (EMR), which is a notable advancement towards statistical modeling of case-time duration in all surgical departments, enabling improved operating room efficiency, cost, and scheduling.

Acknowledgments

First, I would like to thank my academic supervisor, Professor Domenico Maisano, for his constant guidance and support throughout this entire project, which without I would not have been able to make this work possible.

I am also grateful to the local health authority ASLTO3 in which I carried out this investigation, for allowing me to conduct this study and providing all the data needed. Particularly, I appreciate the contribution of Dottore Francesco Calabrese for his cooperation, endless support, and of course friendship, and Dottore Davide Minniti for being my company tutor.

This work wouldn't be possible without the unmeasurable support received from my family. Special thanks to my mother, for inspiring me and teaching me the most important values in life.

Finally, I would like to thank my friends for uncountable stimulating discussions and the countless, highly needed motivation.

Summary

List of Figures

List of Tables

Chapter 1- Introduction

This project documents the efforts to improve efficiency of operation rooms in hospitals of Pinerolo, Rivoli and Susa which are under the management of the local health authority ASL TO3 where the work was developed.

First, we obtained a large data set which included all surgeries undergone in all 3 hospitals since 2016. Knowing that any over-usage or under-usage have a big impact on the profitability and the effectiveness of operation rooms, we cleaned these data to examine it and analyse it as an attempt to find if there is any way to improve organization and scheduling of surgery rooms. Through the analysis we established a linear correlation between surgery time and occupation time of the operation room. The interventions were divided based on different criteria as per department, discipline, hospital and for each criteria the mean was compared with others to identify abnormalities.

After having finished the analysis of the data we obtained, we asked for more variables within the data we already had where we can find some more information on the patients such as age, height, weight... thus, we can build a predictive model that can ameliorate and improve estimation of time of each surgery which can improve the scheduling process of the surgery room.

Using these data, we trained a model with XGBoost package in R, that resulted a more accurate estimation than using historical average of the specific surgery and resulted lower percentage of overage and underage.

After constructing the model, we implemented it on an application to make it usable, where the user insert all the variables related to the surgery like where it will be held and type … in addition to variables related to the patients.

This application returns the predicted time of surgery with the average time of this surgery, with an illustration of all surgeries done in the 3 hospitals since 2016 till now. According to the doctor who was leading the project the application was returning plausible numbers.

Finally, the report suggests to standardized and collect more data related to the operation and procedure of surgeons to construct a model with more accurate prediction.

1.1. Abbreviations and Acronyms

ASL Azienda Sanitaria Locale

OR Operation Room

PACU post-anesthesia care unit

EHR Electronic health record

The first case on-time starts (FCOTS)

Δtime is the difference between usage

time of OR and the time of surgery

MAE Mean Absolute error

1.2. Company

Azienda Sanitaria Locale (ASL) is the local health authority in Italy responsible for planning and organizing the health and medical assistance to people that live in its territorial area supplying diagnosis and treatments by public and/or private providers. The ASLs are divided into districts (Distretti) and each health district is responsible to plan the territorial medical assistance and the coordination among health structures on their territory in addition to some health services (drug addiction, mental health, service for handicapped people...). In case a hospital doesn't have the conditions required by national law to be hospital authority or *Azienda Ospedaliera* which is a trust that manages one or more hospitals, it is managed directly by the ASL in charge of the territory where the hospital is placed. This thesis describes an improvement project started by ASL TO3 which is the health authority of an area of Torino, this project focused on analysing the time of each surgery undergone in hospitals of Rivoli and Pinerolo and Susa, then trying to build a model that predicts the time of each surgery and therefore to improve the estimation of case-time duration relative to current standards.

1.3. Operation Rooms

Operation rooms (OR) or also called operation theatres are a significant source of revenue and overhead knowing that around 60% of hospitalized patients will eventually require surgical intervention (T. Gordon, S Paul, A Lyles, J Fountain 1988; Pelkokorpi 2011). Therefore, OR utilization is a critical factor in assessing the productivity and efficiency of a hospital, and having a more effective scheduling or operation management, in general, can provide significant benefits for hospitals. The main times for each patient when OR is ready starts when the patient enters OR, anaesthesia induction, surgery starts, surgery ends, patient leaves OR to the recovery room or post-anaesthesia care unit (PACU) where anaesthesiologists monitor the patient's condition and assure the transition to an awakened state from the anesthetized state. However, it's important to differentiate between the time in-out which is the time of the whole process starting from the moment the patient enters the surgery room till he exits, and the time of the surgery which is the time needed specifically for the undergone surgery. Many recent studies focused on the improvement of the First case on-time starts (FCOTS) in which it is shown the importance of starting the procedure on time and how it affects the ability of OR to stay on schedule for the rest of the day (Allen, Taaffe, Neilley, Busby, 2019), but in this study, we will focus more on improving the capability to predict the case time duration using machine learning. Historically, the prediction was done either by the surgeon based on his personal experience or by an electronic health record (EHR) based on historical means for case duration. More about inefficient utilization of OR impact on the hospital will be discussed in chapter 2.

Chapter 2- Problem Background

The largest hospital cost category is the operating room (33%) (Macario, Vitez, Dunn, McDonald, MD, 1995) with a mean cost ranging widely from 36$ to 37$ per minute with 20$ to 21$ of direct costs, wages, and benefits of 13$ and 14$ of the total costs and the remaining small portion of 2.5$ to 3.5$ of the total cost for unbilled supplies, including packs, gloves, and sutures (Kaitlynn Ely,2018), and these costs increased and are still increasing with time (see Fig 1)(Childer MD, Maggard-Gibbons MD, 2018).

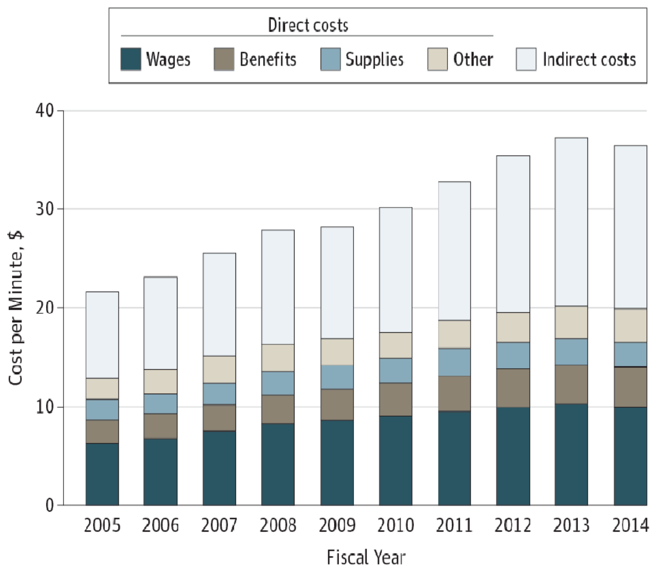

Figure 1 Direct costs of operation room per minute from 2005 to 2014.

Scheduling of surgeries has relied on project cases estimated by surgeons themselves. However, some studies have demonstrated limited accuracy where 100 cases were analysed and surgeons estimated correctly only 26% of the time and underestimated 32% of the time, and overestimated 42% of the time (Daniel M Laskin, A Omar Abubaker, A Strauss, 2013). When scheduling decisions are made several objectives must be in mind like completing all procedures by the end of the day in addition to having the patients released to the recovery units at a regular rate to guarantee the availability of recovery beds for the next patient. Therefore apart from the cost, the accuracy of prediction impact all the objectives set, for example, if a scenario is considered where some surgeries finish later than predicted time and others earlier, in this scenario a point in which an unexpected spike in demand for recovery beds and the nursing staff won't be able to accommodate due to the limited number of bed, in that case, patients will have to wait for beds in the OR which will delay next surgeries, and in some cases where these delays are acute and next surgeries will have to be rescheduled, which decrease the patient's satisfaction and increase the costs on the hospital.

This scenario emphasizes the importance of the prediction process and shows how essential it is to deliver high-quality care and to be efficient. Thus, inaccurate predictions not only increase costs but also reduce revenue (Master, Zhou, Miller, Scheinker, Bambos, Glynn, 2017). Historical averaging predicts that the duration of the scheduled surgery is equal to the average of the duration of the specific surgery undergone in the past, this method means generalizing and it does not take into consideration several variations from one patient to another like gender, age, weight, height...

Another method for case time prediction is to rely on experts' opinions where surgeons can provide an estimate of how much time is needed for a given procedure, this estimation can be inaccurate as that surgeons do not bear the overtime costs running past the end of the day so they tend to try to maximize the number of scheduled surgeries which therefore maximize their compensation, and sometimes surgeons underestimate the time required for a procedure(Master, Zhou, Miller, Scheinker, Bambos, Glynn, 2017).

Chapter 3- Problem Break Down

Data were collected from the OrmaWeb corporate platform by ASL TO3. The period from 1st January 2016 until the end of June 2021 has been selected. It was decided to include both the operating room usage time and the time of the surgery in the analysis, knowing that the operating room usage time includes the minute the patient enters and wait for his anaesthesia, time of surgery after having the anaesthetic induction and the time between the end of anaesthesia after the surgery ends till he exits the room.

3.1. Data Collection and Cleaning

All operating blocks present were initially selected. The data thus collected in Excel format were processed using R software to improve its quality. R is a programming language for statistical computing, and it was used for cleaning and plotting the data, and to construct the predictive model later. We proceeded to:

- Re-fill empty database cells with null values due to unavailable or missing data.

-Rename some types of interventions that are indicated with similar if not identical names but resulted due to excess characters or grammatical errors that have led to an increase in the type of surgeries despite being identical.

-Rename some items to reduce the redundancy of information.

The initial data contained 66856 cases that were reduced to 57423 after removing the cardiologic surgeries performed in off-site locations like Venaria hospital. The average room occupancy time is 74.73 minutes, but as can be seen from the box plot (see Fig2), each discipline has its distribution of this parameter. General surgery, O.R.L, Gynaecology, Urology, and Orthopaedics have a wide range of room occupancy that depends on the number of different interventions that are performed. The same is true for the average intervention time which, as will be examined later, is the main determinant of the occupation time of the room, which is 41.73 minutes.

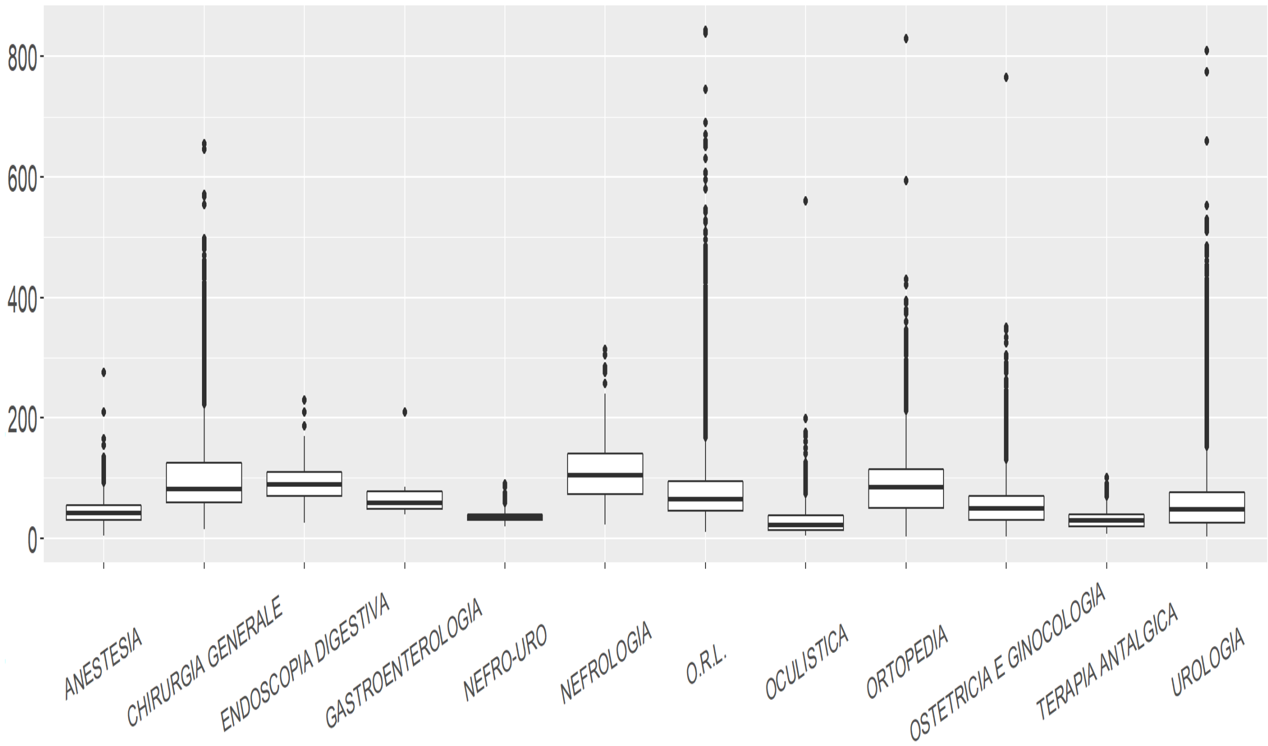

Figure 2 Boxplot showing the time of occupation of each department.

The room occupancy time distributed over the 3 hospitals is respectively:

-96.23 minutes for Rivoli hospital.

-72.45 minutes for Pinerolo hospital.

-50.33 minutes for Susa hospital.

The average operative time is instead of:

-55.08 minutes for Rivoli hospital.

-39.23 minutes for Pinerolo hospital.

-29.78 minutes for Susa hospital.

The difference between hospital units in terms of occupancy of OR is explained by considering the different disciplines that operate in each hospital knowing that each discipline has its distribution as shown previously in Figure 2.

Except for missing values and/or errors, the time of occupation of the room and the time of the intervention is generally reported correctly and reliably according to the doctor that led this project.

The analysis that was carried out has 2 purposes which are:

-establish if there is a linear correlation between the surgery time and the occupation time of the room.

-define which surgeries have a greater delta between operation room usage time and the surgery time.

For this analysis, all interventions were chosen to be grouped in 4 groups following priority of each as follows: Elective, urgent, and emergent, in addition to a group of urgent surgeries in the afternoon, at night or on holidays.

From an evaluation carried out through box plotting it emerged that within the database, despite an initial screening of the data considered out of scale, there were still hundreds of outliers, but difficult to identify manually. The graphical analysis of the relationship between the duration of the intervention and the occupancy time of the operating room is shown in figure 3 below using ggplot in R (See appendix 1).

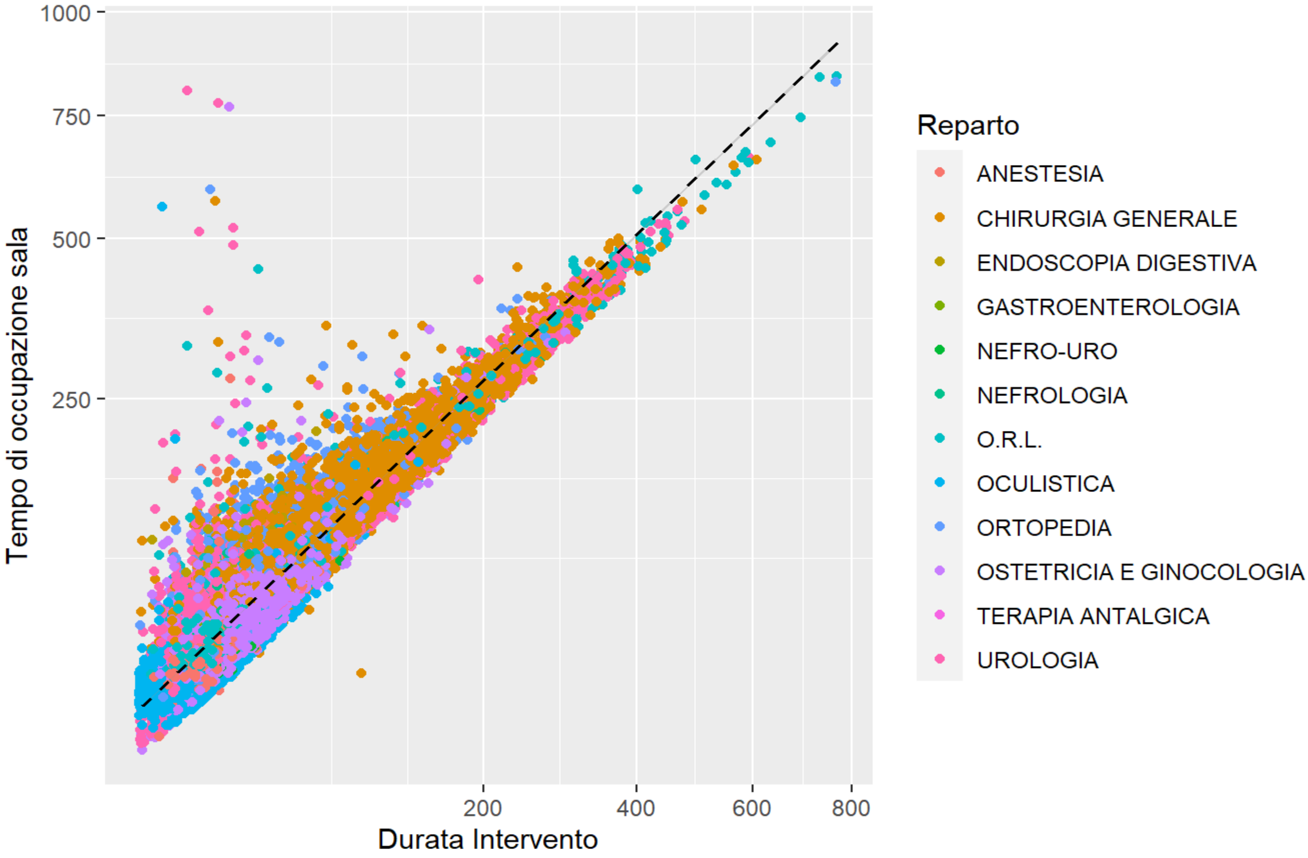

Figure 3 Graph showing the time of occupation of the surgery rooms with respect t time of surgery for different types of surgery.

Without the exclusion of the outliers, it is noted the presence of a linear correlation between the time taken for the surgery and the time of occupation of OR, but the same graph highlights a wide variability of the time of occupation of the room compared to the actual duration of the surgery. The total time of occupation of the room is therefore certainly correlated to the duration of the intervention, everything that is carried out inside the operating room before or after the intervention represents another determining factor. The Δtime understood as the difference between the time of occupation of the room and the operating time, in other words the gap time, is linked to pre-and post-operative procedures, many of which are of anaesthetic competence, and which usually have a variable duration that depends on the type of procedure.

The literature describes the average timing of pre-and post-operative procedures, but these are too influenced by the clinical and organizational context in which they are measured (types of intervention performed, techniques used, disciplines involved, etc..). Taking this into account, it is possible to evaluate the Δtime and how it influences the time of OR then identify which interventions present the highest Δtime. The graphical analysis of the relationship between the duration of intervention and the occupancy time of the operating room in 3 hospitals is shown below where on the y axis it is represented the occupation time of OR and on the x-axis the difference between the time of OR and the time of surgery which is Δtime.

Figure 4 Surgeries Δtime with respect to time of OR in 3 different hospitals.

The graph shows all the surgeries carried out; each colour represents a discipline. The closer the single surgery gets to the dotted line, the longer the room is occupied without surgery being taken place.

A similar graph is shown but where the subdivision that is proposed is by priorities mentioned previously.

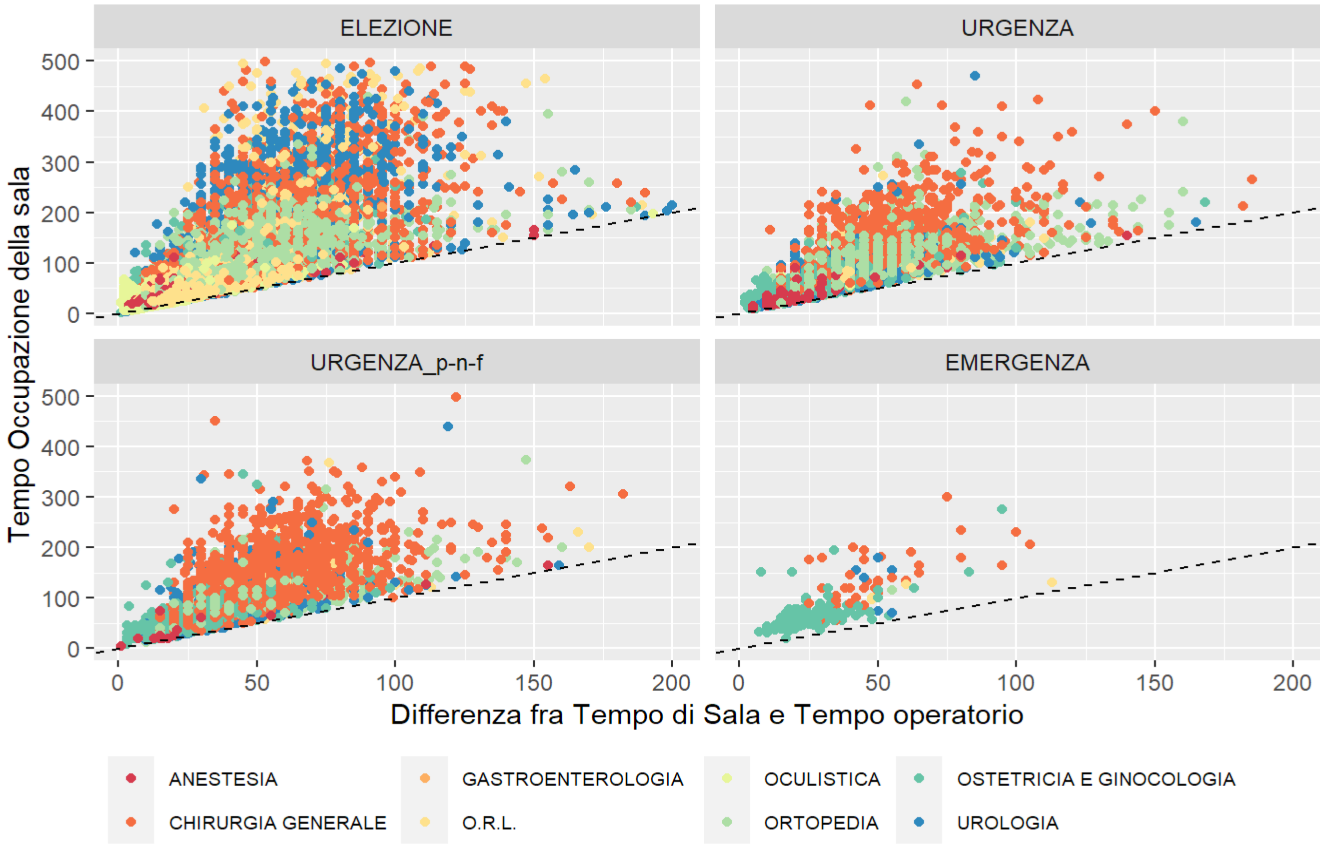

Figure 5 Difference of time of OR and time of surgery with respect to time of OR for different priorities.

It is thus highlighted that there are some interventions for which the duration of the room occupation is mainly influenced by the pre-and post-operative time (interventions close to the dotted lines). What can be stated through the graphical analysis is that some disciplines have greater distribution of room occupation time, and it doesn't show a big difference between elective and urgent interventions.

However, it is difficult to define what is a correct benchmark for the Δtime and average of a single centre.

From literature, it is possible to define a theoretical limit of around 150-200 minutes, which does not represent the average, but a theoretical upper extreme and which includes both the pre-and post-operative time.

It was decided to define a function capable of identifying the anomalous values trying to modulate them in such a way as to include all the interventions in which Δtime is excessive compared to the overall time of occupation of room and bearing in mind that not all interventions require the same pre-operative and post-operative time.

Consequently, we programmed a function (see Appendix 2) capable of identifying these anomalous values. The function considers as an anomalous value any intervention that has a Δtime that differs more than 2 times the standard deviation from all other surgeries that have the same procedure and priority. The anomalous interventions are then selected based on room occupation time, compared to the difference between room occupancy and duration of intervention. The sample of outliers selected is equal to 592 interventions of which:

-442 in Election.

-81 in urgency.

-68 Urgency in the afternoon, night, or holiday.

-1 in Emergency.

In total, anomalous interventions represent 1.03% of all interventions, that were not included in the initial database. Analysing the types of intervention based on priority, it is possible that some selected interventions do not represent real outliers, but that their duration has been influenced by the clinical context in which they took place. The interventions identified as outliers are distributed as follows:

-330 in Pinerolo.

-223 in Rivoli.

-39 in Susa.

A list of some of these interventions whose delta time has been identified as anomalous as shown in the following table:

Priority	Hospital	Distribution	Data	Type of surgery	Room occupation time	Duration of intervention	Delta time
URGENZA_p-n-f	Rivoli	UROLOGIA	2016-04-29	CATETERIZZAZIONE URETERALE	810	7	803
URGENZA_p-n-f	Rivoli	UROLOGIA	2018-11-02	CATETERIZZAZIONE URETERALE	775	15	760
ELEZIONE	Rivoli	OSTETRICIA E GINOCOLOGIA	2018-01-14	TAGLIO CESAREO	765	19	746
ELEZIONE	Pinerolo	ORTOPEDIA	2019-07-16	ARTROSCOPIA DEL GINOCCHIO	593	13	580
ELEZIONE	Pinerolo	OCULISTICA	2020-02-13	INIEZIONE DI SOSTITUTI VITREALI	560	3	557
ELEZIONE	Susa	CHIRURGIA GENERALE	2020-02-12	ASPORTAZIONE DI VENE DELL'ARTO INFERIORE	570	15	555
URGENZA_p-n-f	Rivoli	UROLOGIA	2020-09-21	CATETERIZZAZIONE URETERALE	515	10	505
ELEZIONE	Rivoli	UROLOGIA	2019-11-05	ESTRAZIONE ENDOSCOPICA DALL'URETERE E PELVI RENALE DI: COAGULO DI SANGUE, CALCOLO CORPO ESTRANEO	515	20	495
ELEZIONE	Rivoli	UROLOGIA	2017-10-03	CISTECTOMIA RADICALE	485	20	465
ELEZIONE	Pinerolo	O.R.L.	2021-03-04	LARINGECTOMIA RADICALE	448	30	418
URGENZA_p-n-f	Rivoli	UROLOGIA	2018-03-26	CATETERIZZAZIONE URETERALE	380	13	367
ELEZIONE	Rivoli	UROLOGIA	2016-03-17	PIELOPLASTICHE	340	25	315
URGENCY	Pinerolo	O.R.L.	2019-04-20	RIDUZIONE CHIUSA DI FRATTURA NASALE NON A CIELO APERTO	320	7	313
ELEZIONE	Rivoli	CHIRURGIA GENERALE	2020-05-28	ALTRI INTERVENTI SULLA CUTE E SUL TESSUTO SOTTOCUTANEO	325	15	310

ELEZIONE	Susa	ORTOPEDIA	2017-03-28	ACROMIOPLASTICA SPALLA	335	35	300
ELECTION	Pinerolo	UROLOGY	2017-09-19	ALTRI INTERVENTI SULLA VESCICA	313	24	289
ELECTION	Pinerolo	ORTHOPEDICS	2021-06-23	RIDUZIONE CRUENTA DI FRATTURA DI TIBIA E FIBULA, CON FISSAZIONE INTERNA	327	40	287

Table 1 Sample of the outliers that show an unusual difference between the time of occupation and time of surgery

Table above shows some interventions where surgery duration was so short, yet the delta time is very high with respect to it.

However, an evaluation of each single surgery is needed to affirm if a surgery is an outlier discriminating based on the type of surgery, for example, cystoscopy lasting 600 minutes is an anomalous value knowing usually it lasts usually 5 to 15 minutes (according to Mayoclinic) and it cannot be said the same for Laryngectomy that can last from 5 to 20 hours (Erika Roth, reviewed medically by Judith Marcin M.D. 2018).

The identified outliers are shown in the following figure:

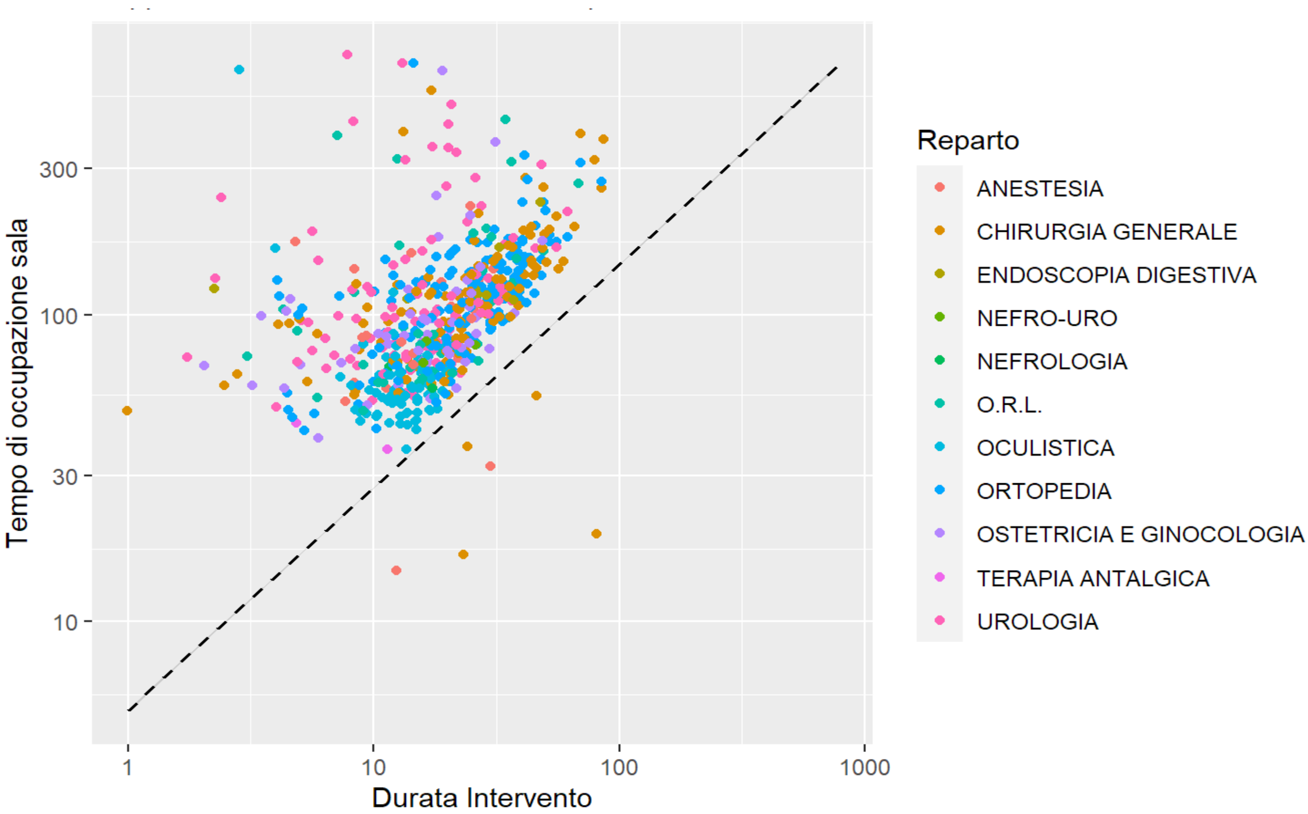

Figure 6 Representation of outliers and relation between the occupation of operation room with respect to time of surgery for each department.

The graph highlights that within the database several surgeries present a high delta between occupation time and surgery time. However, no type of analysis can specify whether the discrepancy between room time and intervention time is linked to organizational problems or simply it was registered in a wrong or later time than real one. Regarding the surgeries that are placed on the line of the logistic regression, these are generally surgeries that lasted much longer than identical surgeries and are carried out in the same settings of priority. More will be discussed in the following section.

3.2. Data Analysis

The following table is a summary of the average room occupancy of the operating room for different medical departments for the data excluding outliers.

Department	Average Occupation time (min)
ANESTHESIA	43.79
CHIRURGIA GENERALE	103.03
ENDOSCOPIA DIGESTIVA	91.44
GASTROENTEROLOGIA	71.53
NEFRO-URO	36.17
NEFROLOGIA	110.20
O.R.L.	90.80
OCULISTICA	26.78
ORTOPEDIA	87.57
OSTETRICIA E GINOCOLOGIA	57.49
TERAPIA ANTALGICA	31.73
UROLOGIA	66.75

Table 2: Summary of average Occupation time for each department

The average Δtime or gap time for all surgeries that were considered after excluding outliers is around 32.5 Minutes, while for the three hospitals it is shown in the following table **in minutes** and is divided by priority.

Presidio	Election	Urgency	Urgency_PNF	Emergency
Pinerolo	28.45	46.54	49.19	30.02
Rivoli	39.76	40.81	41.14	36.46
Susa	17.53	37.22	34.26	30.00

Table 3: Review of the Δtime divided by unit and priority

The difference of Δtime among hospitals is explained by considering the disciplines that operate in the various facilities and which each have a particular distribution of Δtime, and the same can be said for time of occupation of OR.

However, the average Δtime for the 3 hospitals is respectively of:

-38.55 minutes for Pinerolo.

-39.54 minutes for Rivoli.

-29.75 for Susa.

All the new data (excluding outliers) are represented in the following representation.

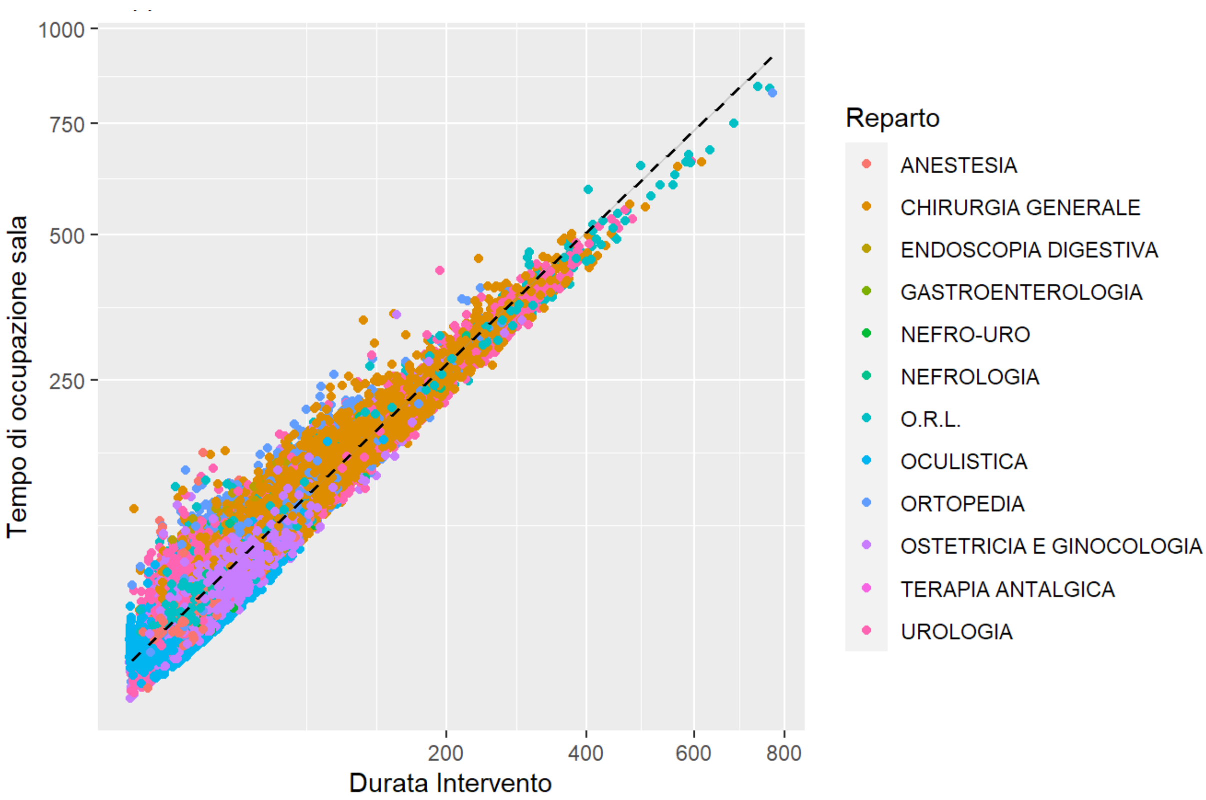

Figure 7 Representation of Data without outliers showing the time of intervention with respect to time of occupation of the operation room.

From the graph, it also emerges that a model based on the duration of intervention can explain the occupation time of the operating room for operations that do not exceed 6 hours (300 minutes). For interventions of longer duration, it is possible to observe that the duration of surgeries tends to correlate less well with the overall room time.

The same can be said for surgeries of limited duration, where Δtime represents the most relevant component of the time of occupation room. It can therefore be assumed that generally, the main determinant of occupancy time of the room is the time necessary to carry out the surgery and that the process of selecting the values considered consistent within the clinical organizational aspects, thus excluding the outliers, improves the linear model. Assuming that there is a linear correlation between room time and surgery time, it is possible to define a linear regression model and verify if the removal of the outliers improves the Δtime.

The linear regression was carried out mainly to numerically define the relationship between the time of occupation of the room and the time of intervention, evaluating if by removing the outliers, the relationship becomes more significant, and it's represented in the table below using stargazer package in R software (see appendix 3).

Comparison of model regression with and without outliers

	Dependent variable:	
	Time of occupation of the Surgery room	
	Without outliers	With outliers
TIME OF OPERATION	1.25***	1.24***
	(0.001)	(0.002)
Constant	21.62***	22.57***
	(0.091)	(0.108)
Observations	56,41	57,005
R^2	**0.931**	**0.902**
Adjusted R^2	0.931	0.902
Residual Std. Error	16.051 (df = 56411)	19.331 (df = 57003)
F Statistic	762,502.400*** (df = 1; 56411)	522,364.000*** (df = 1; 57003)
Note:	*p<0.1;* ***p<0.05;*** p<0.01	

Table 4:Comparison of a regression model with and without outliers

From the analysis of the linear model, the R^2 improves, passing from the value of 0.902 to a value of 0.931, which means removing the outliers improves the linear model.

The efficiency of scheduling surgical interventions is dependent on the over-use and under-use of the room with respect to the scheduled time, and these times can be determined by various factors, but in the context of an election, they can be determined by the type of intervention that is carried out. When you are faced with an intervention that has well-standardized execution times, the possibility of making an accurate forecast and scheduling that intervention in such a way as not to have downtime or extend the opening time of the surgery room is greater. The diametrically opposite is true for interventions that have variable timing and therefore do not allow for that intervention to be included in an agenda without increasing the risk of under or overuse. An index of organizational relevance is therefore proposed which is the Coefficient of variation (see appendix 4), reported as the possibility that a future intervention has the same duration and the same time of occupation of the room as the past interventions (forecast certainty). This index is more realistic the higher the number of interventions of the same type that have been carried out. In this analysis taking into consideration, the median is better than the average because almost no intervention has a normal distribution, it is better to rely on an index less influenced by extreme values. The following table shows the numerousness, median, room occupation, and the index S which indicates the skewness of each surgery. Note that the more negative is the skewness the greater the chance of finishing earlier than the median, and the more positive, the greater the possibility of ending after the median. An example of some of the surgeries done in Pinerolo hospital is shown knowing that there are hundreds of types of surgeries and the impossibility to show all types in all hospitals.

Type of intervention	Hospital	Coefficient of variation	Numb of surgeries	Median (min)	median_Room (min)	INDEX _S
ERNIORRAFIA OMBELICALE	B.O. AGNELLI	1.006	106	35.0	69.5	1.56
INIEZIONE DI SOSTANZE TERAPEUTICHE NELL'ARTICOLAZIONE O NEL LEGAMENTO	B.O. AGNELLI	0.94	17	2.0	10.0	1.52
ASPORTAZIONE RADICALE DI ALTRI LINFONODI	B.O. AGNELLI	0.88	8	35.0	67.5	1.33
ALTRI INTERVENTI SULLA VESCICA	B.O. AGNELLI	0.82	25	15.0	44.0	1.78
RIMOZIONE DI DISPOSITIVO IMPIANTATO DA SCAPOLA, CLAVICOLA E TORACE (COSTE E STERNO)	B.O. AGNELLI	0.82	9	15.0	45.0	0.95
BIOPSIA DEL PENE	B.O. AGNELLI	0.81	7	13.0	29.0	1.30
RIDUZIONE INCRUENTA DI FRATTURA DI TIBIA E FIBULA SENZA FISSAZIONE INTERNA	B.O. RIVOLI	0.78	4	59.0	92.5	0.07
ASPORTAZIONE DI LINFONODI CERVICALI PROFONDI	B.O. AGNELLI	0.78	26	53.5	107.0	1.20
ASPORTAZIONE O DEMOLIZIONE DI LESIONE O TESSUTO DELLA LINGUA	B.O. AGNELLI	0.76	13	16.0	57.0	1.34
ESTRAZIONE ENDOSCOPICA DALL'URETERE E PELVI RENALE DI: COAGULO DI SANGUE, CALCOLO, CORPO ESTRANEO	B.O. AGNELLI	0.75	481	20.0	55.0	0.79
RIDUZIONE CON FISSATORE ESTERNO	B.O. SUSA	0.74	5	125.0	160.0	-0.17
LITOTRISSIA CON ULTRASUONI O ELETTROIDRAULICA	B.O. AGNELLI	0.74	66	20.0	31.0	1.51
GLOSSECTOMIA PARZIALE	B.O. AGNELLI	0.71	10	111.0	172.0	0.56
RIMOZIONE DI DISPOSITIVO IMPIANTATO DALL'OMERO	B.O. AGNELLI	0.707	40	9.0	32.5	0.94
CHIUSURA FAV	B.O. AGNELLI	0.70	7	75.0	114.0	0.88
ARTROCENTESI	B.O. AGNELLI	0.69	6	15.0	44.0	-0.05
REVISIONE DI ANASTOMOSI DELL'INTESTINO CRASSO	B.O. AGNELLI	0.69	9	78.0	132.0	0.73
TENORRAFIA FLESSORI	B.O. AGNELLI	0.68	6	20.0	55.0	0.77
EXERESI CONDILOMI	B.O. AGNELLI	0.66	4	12.5	27.5	0.72
ALTRI INTERVENTI SULLA CAVITA+ ORALE	B.O. AGNELLI	0.66	8	17.0	54.5	1.75
RIDUZIONE INCRUENTA DI LUSSAZIONE DELL'ANCA	B.O. RIVOLI	0.64	5	10.0	25.0	1.29
RIPARAZIONE DEL GINOCCHIO	B.O. AGNELLI	0.64	20	33.5	72.5	0.68
RIPARAZIONE DELLA LARINGE	B.O. AGNELLI	0.63	4	34.0	68.0	0.96
ASPORTAZIONE LOCALE DI LESIONE O TESSUTO DEL RETTO	B.O. AGNELLI	0.62	17	25.0	65.0	1.01
RIMOZIONE DI DISPOSITIVO IMPIANTATO DA CARPO E METACARPO	B.O. AGNELLI	0.61	5	27.0	40.0	0.72

ALTRI INTERVENTI SULLO SCROTO E SULLA TUNICA VAGINALE	B.O. AGNELLI	0.61	13	41.0	66.0	0.41

Table 5 Important indexes for some types of surgeries

It was identified 318 types of interventions that were performed less than 4 times in the observation period. On this type of intervention, it was decided not to proceed with a timing distribution analysis as any result on such a limited sample would not be generalizable. The number of interventions carried out less than 4 times is 543. Most of these were carried out between the hospitals of Rivoli and Pinerolo. The duration of this type of intervention is highly variable and the low number does not allow a further evaluation of surgery times. The following graph represents these surgeries:

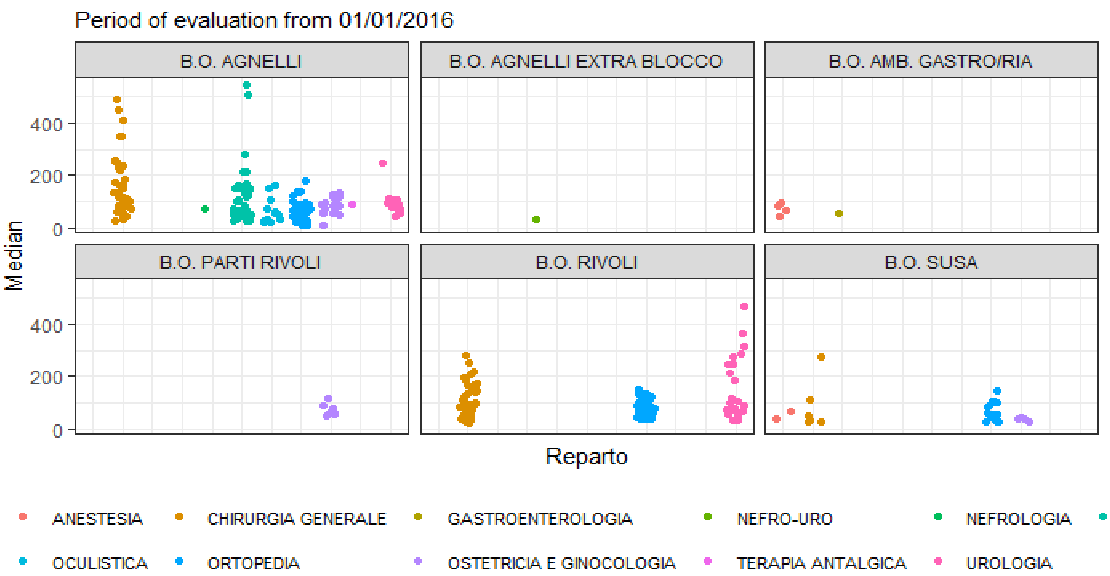

Figure 8 Representation of surgeries undergone less than 4 times

Finally, the surgical activity in the whole period that is under consideration is examined in the following graph that shows the trend of surgical activities or in other words the number of interventions undergone in the 3 hospitals during this period with respect to dates.

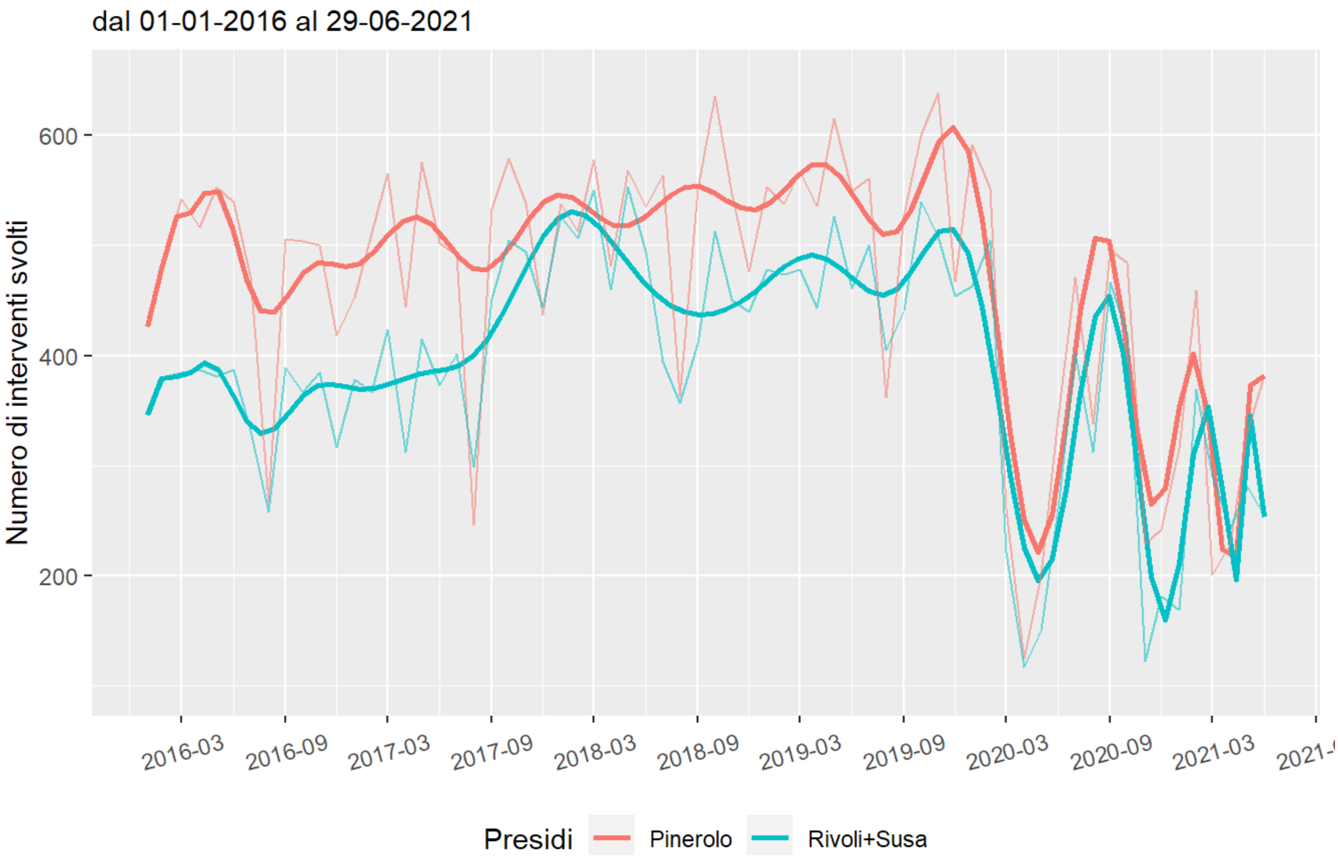

Figure 9 Trend of surgeries in hospitals.

.

It is noted that in the period preceding the pandemic, the number of interventions carried out was relatively stable and that a greater number of interventions took place in Pinerolo.

The graph highlights the effect of the three pandemic waves at the beginning of 2020, the end of 2020, and the beginning of 2021, which led to a reduction in surgical activity in all three centres.

Chapter 4- Building the Model

After obtaining approval, we got new data that included more details related to the procedure and the patient's information (age, gender, height, weight...), and the name of the surgeons that have undergone each surgery.

The dataset was comprised of data for around 7 years from January 2014 till August 2021 with a total of 92913 surgeries undergone in the same hospitals.

4.1. Cleaning New Data

The dataset we received had some missing values in addition to some information that doesn't affect our model like code of procedure or name of nurses... Initially, the dataset was composed of 92 variables. After cleaning and eliminating unnecessary information, and removing Cardiology and Venaria hospital, the data set is composed of 79693 surgery and 30 variables that are represented in figure 10.

Figure 10 Representation of time of surgery with respect to time of occupation of the operation room

As done in the previous chapter, the same function to identify outliers to ameliorate accuracy was defined (see appendix 2) and that was proven to ameliorate accuracy with linear regression test. The results are illustrated in the following figure.

Figure 11 Representation of Data without outliers

Having identified the outliers, the obtained data were used to create a dataset that was used to construct the model using Xgboost (see appendix 5) package in R software.

This dataset followed exclusion criteria that consisted of keeping only useful data and variables, therefore only elective surgeries after January 2016 were taken into consideration which resulted in a dataset of 51663 surgeries with 13 variables composed of the surgical unit, specialty, scheduled surgery, hospitalization regime, type of anesthesia, name of first surgeons, and the diagnosis. In addition to some variables related to the patient like age, sex, weight, height, and ASA score, which is a subjective score assessed of a patient's overall health that has a range from 1 to 5. Where 1 is for a patient that is completely healthy and fit and 5 is a moribund patient who is not expected to survive without the operation (Committee on Economics on American Society of anesthesiologists, 2014). In addition to the total time which was called time in-out and which is the total time of entrance till the exit of the patient to the OR.

4.2. Constructing The Model and Results

Due to its importance in scheduling OR, the total minutes from patient's entry till room exit was defined as case-time duration. The data were randomly divided into a training data set (41,330 cases; 80% of data) and the rest testing data set (10,333 cases; 20% of data) where the machine learning model was developed on the 80% training data and tested with the 20% dataset (see appendix 6), Missing variables among some interventions were marked as NA.

Knowing that Xgboost cannot handle categorical features by itself and only accepts numerical values, categorical variables were converted into binary representation for each category which is called dummies, in each dummy variable label 1 represents the existence of the level in the variable, while 0 represents its non-existence. Having chosen some parameters for the model we proceeded in constructing the model (see appendix 7). The evaluation metric of the model was chosen to be mean absolute error (consider appendix 8).

The best results after constructing the model are shown by R as follows:

```
##### xgb.Booster
raw: 5.8 Mb
niter: 2340
best_ntreelimit : 2241
best_iteration : 2241
best_score : 15.98075
best_msg : [2241]
train-mae:11.909       test-mae:15.980
```

Table 6: Results XGBoost

Where the training set showed an MAE of 11.90 minutes and testing dataset an MAE of 15.98 minutes.

However, this model's prediction was compared to the prediction of a model that uses the average of the specific type of intervention within our data. The key metric to measure model performance and make the comparison was illustrated through distribution of prediction error in testing dataset in term of predicted Δtime which is (the total time of OR – time predicted) and Δtime which is (the total time of OR – mean of the specific intervention) and the results are shown in fig (12-13) with a red box that denotes the -10 to 10% tolerance threshold for within cases.

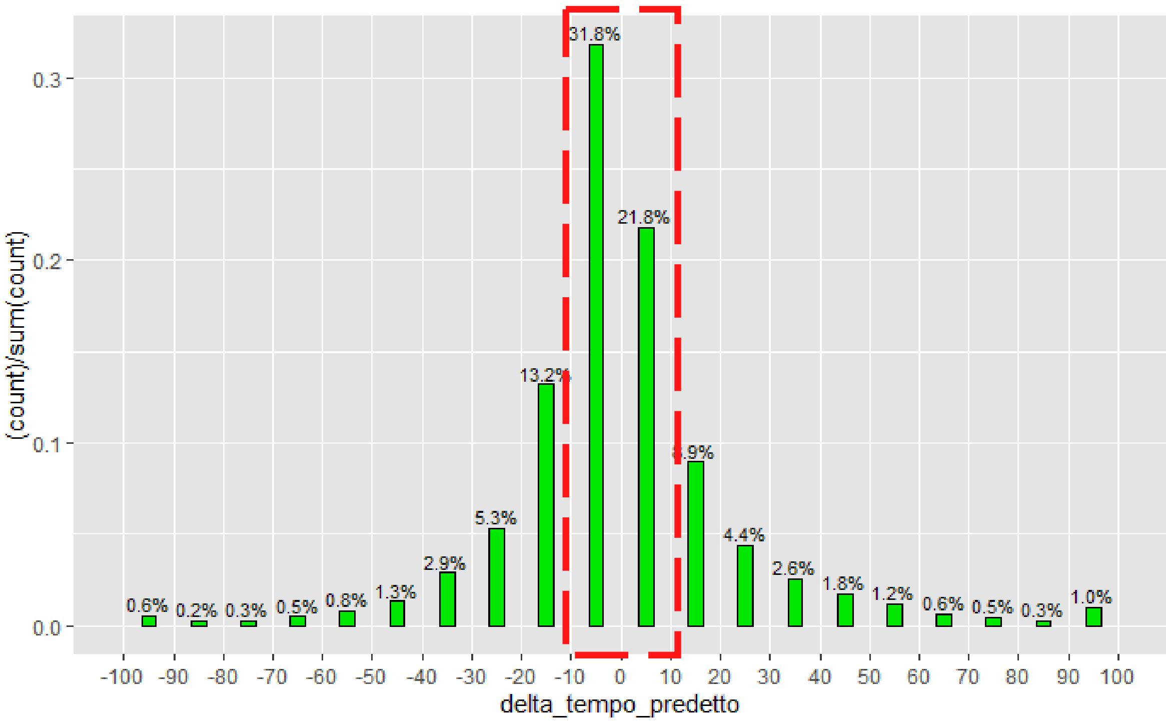

Figure 12 Distribution of Prediction Error in Testing Dataset Using XGBoost.

Where on the Y-axis it is presented the frequency of cases and on the X-axis the error (10% bins). Positive error represents overestimation while negative error represents underestimation.

On the other hand, the same graph showing the performance of the model was plotted for the prediction of time using the model that uses the mean of each intervention and showed the following results in figure 13.

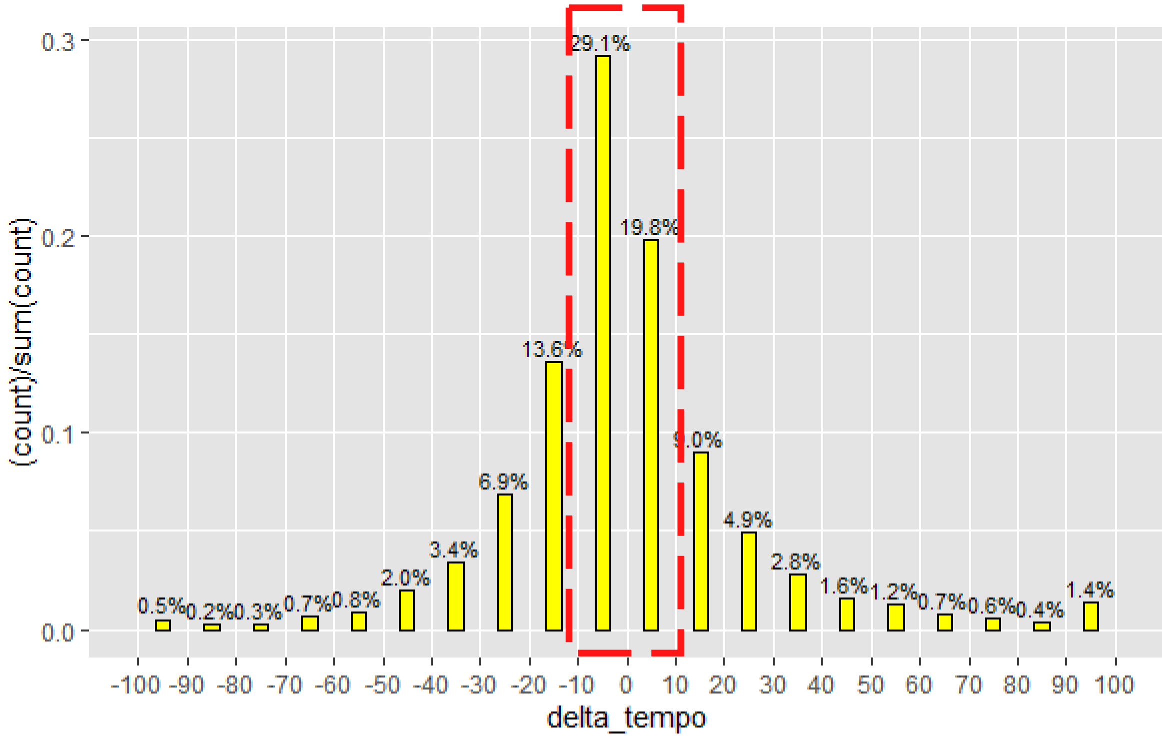

Figure 13 Distribution of Prediction Error in Testing Dataset Using the mean

Within the -10% and 10% tolerance threshold the XGBoost model have better predictive capability with higher accuracy, lower percentage of underage and percentage of overage. Where XGBoost could predict 56.3% of Δtime within the 10% tolerance, where the model based on the mean predicts 48.9% withing the 10% tolerance.

To make the model usable by anyone or more user-friendly, we implemented it on an application using R shiny apps , where the application takes the input that is related to the type of surgery and the surgeon who will do it, in addition to data of the patient, and as an output gives the predicted time and the average time of the specific surgery. In addition to a graph that shows all the interventions undergone during the period of the study (from 2016 to 2021) in all 3 hospitals as shown in figure below.

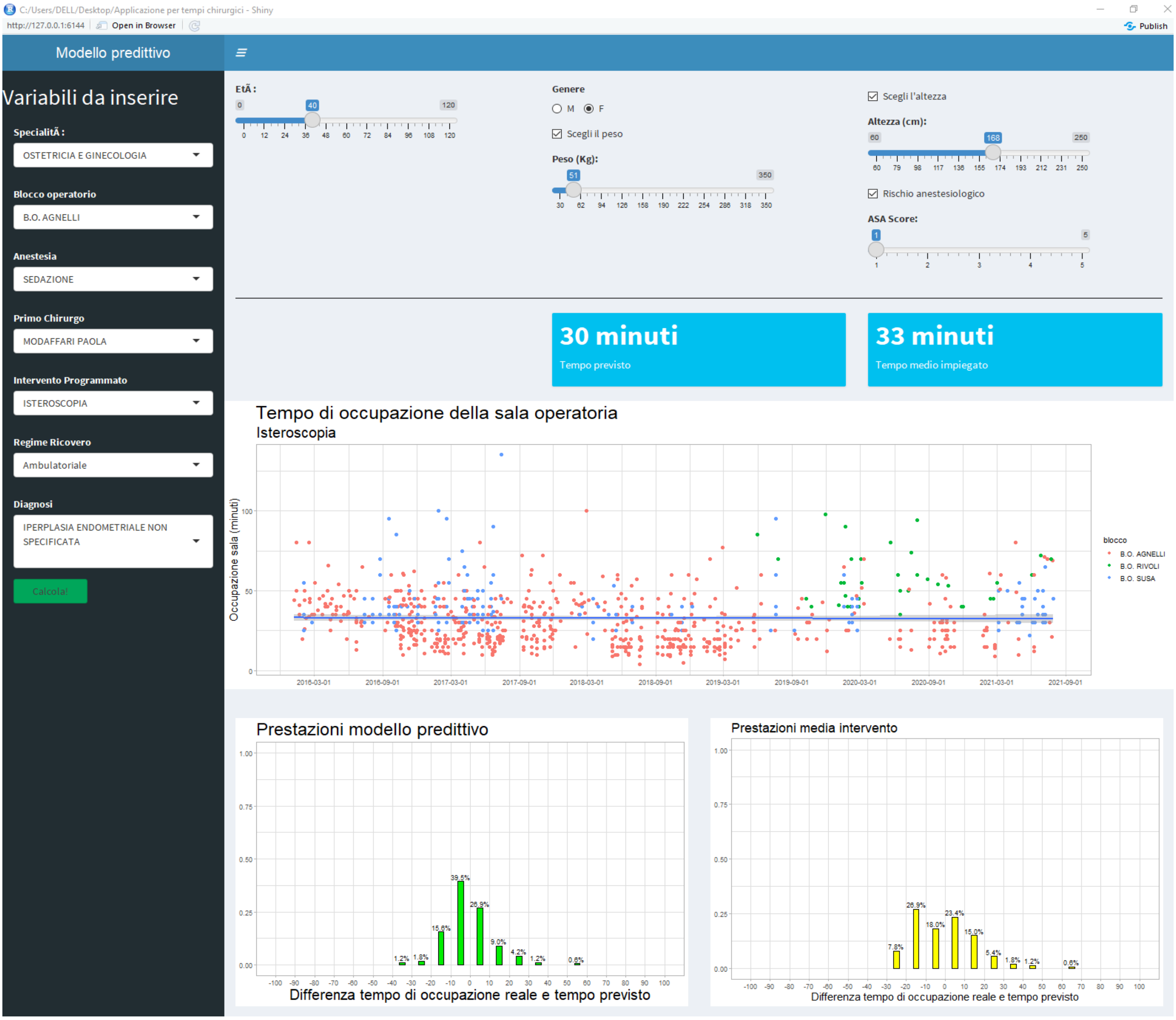

Figure 14: Application example

In the bottom of this presentation, we see both performances of the predictive model using XGBoost and the model of the mean for the same intervention, which illustrates the distribution of percentage of error in predicting the time specifically for this type of surgery, where on the x- axis it shows the difference between real time of occupation and time predicted, knowing that the performance of the model differs for each surgery depending on different factors like availability of data, coefficient of variation, skewness...

Chapter 5- Conclusion

From the analysis we established several points, first each discipline has its own distribution of time for room occupancy and obviously the same for the intervention time itself, knowing the relation between the two is linear. In addition, the occupation time of some interventions is mainly influenced by the pre- and post-operative time. However, no type of analysis other than an evaluation of the single intervention can specify whether the discrepancy between room time and intervention is linked to organizational problems or simply a problem with time or data entry. According to data we have, each type of surgery has its skewness and coefficient of variation, which affects the forecast certainty of each surgery.

Subsequently, after proceeding with constructing the predictive model and comparing it with a model that uses the mean of the intervention, XGBoost model showed superior results in predictive performance, and showed less percentage of error, which is in other words less possibility of overestimation and underestimation of surgical case. Within the 10% threshold, XGBoost model had a superior accuracy of 7.4% compared to the model that uses the mean.

While the estimation of case duration is improved, this does not necessarily mean that the utilization of OR overall will improve.

The fact that in the large dataset we obtained, there were a lot of missing values among the data for some interventions that affected a lot the accuracy of the model.

Moreover, the different procedures and human factors introduce a large amount of uncertainty, from literature it is proposed that the surgeons' procedures are the ones with most notable effect on the time of surgery.

In conclusion, more standardized cases may see higher benefits of machine learning approach knowing that its simpler for training. We propose extracting additional information from operation and procedure of surgeons' data to be used as variables for ML algorithm training that can help constructing a model with higher accuracy.

Appendix 1- Code to plot the data with ggplot

All graphs were plotted using ggplot 2 in R which is an open-source data visualization package for the statistical programming language R.

```
ggplot(data=dati, mapping=aes(x=Minuti_INTERVENTO,
                              y=Minuti_INOUT,
                              colour=Reparto))+
  geom_point(position = position_jitter(w=0.1, h=0.1))+
  stat_smooth(data=Dati, method="lm", color="black", size=0.7 , linetype = "d
ashed")+
  scale_x_log10()+
  scale_y_log10()+
  xlab("Durata intervento")+`
  ylab("Tempo di occupazione sala")
```

Appendix 2- Code for the function OutFinder to find outliers

```
OutFinder2 <- function(x, na.rm = TRUE, ...) {
  M <- mean(x, na.rm = TRUE)+2*sd(x, na.rm = TRUE)
  y <- x
  y[x > M]<-NA
  y
}



Outliers<- dati %>%
  filter(Minuti_INOUT>30) %>%
  filter(Minuti_INTERVENTO<Minuti_INOUT | is.na(Minuti_INOUT) | is.na(Minuti_
INTERVENTO)) %>%
  group_by(Priorità, TipoIntervento) %>%
  mutate(Outlier = OutFinder2(deltatempo)) %>%
  filter(is.na(Outlier)) %>%
  filter(Minuti_INOUT/deltatempo<1.1 & Minuti_INOUT <60|
           Minuti_INOUT/deltatempo<1.2 & Minuti_INOUT <120|
           Minuti_INOUT/deltatempo<1.35 & Minuti_INOUT <240|
           Minuti_INOUT/deltatempo<1.45 & Minuti_INOUT <480|
           Minuti_INOUT/deltatempo<1.60 & Minuti_INOUT>480)

datisenzaoutliers<-anti_join(dati, Outliers)

a<-datisenzaoutliers %>%
  filter(deltatempo>150 & Minuti_INOUT/deltatempo<1.5|
           deltatempo<1)

Outliers<-bind_rows(Outliers, a)

datisenzaoutliers<-anti_join(dati, Outliers)
```

Appendix 3- Stargazer summary statistics table code

```
stargazer(modello_lineare, modello_lineare_outlier, type = "html",
          title="Comparison of model regression with and without outliers", a
lign=TRUE,
          dep.var.labels=c("Time of occupation of Surgery room"),
          covariate.labels=c("TIME OF OPERATION"))
```

R^2 or coefficient of determination is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. in other words, R^2 shows how well the data fit in the regression model (Corporate Finance institutes). R^2 ranges from 0 to 1, and its interpretation can mean how well the regression model fits the observed data, if R^2 is equal to 80% that reveals that 80% of the data fit the regression model, therefore a higher R^2 indicates a better fit for the model.

Appendix 4- Coefficient of variation

The coefficient of variation measures the dispersion of data points around the mean, it is useful to compare the degree of variation among series of data. It is calculated by dividing the standard deviation by the mean. The higher the coefficient of variation, the greater the level of dispersion of data around the mean which means the harder to predict and vice versa.

Appendix 5- XGBOOST

XGBOOST is an optimized distributed gradient boosting library that is designed to be highly flexible, efficient, and portable (XGBoost Documentation).

XGBoost stands for 'Extreme Gradient Boosting' where the term Gradient boosting originated from the paper *Greedy Function Approximation: A Gradient Boosting Machine, by Friedman.*

Briefly, XGBoost is used in supervised learning problems, where a training data set is used to predict a target variable. It is a decision tree-based ensemble Machine learning algorithm that uses a gradient boosting framework, it is the evolution of decision trees as shown in following figure.

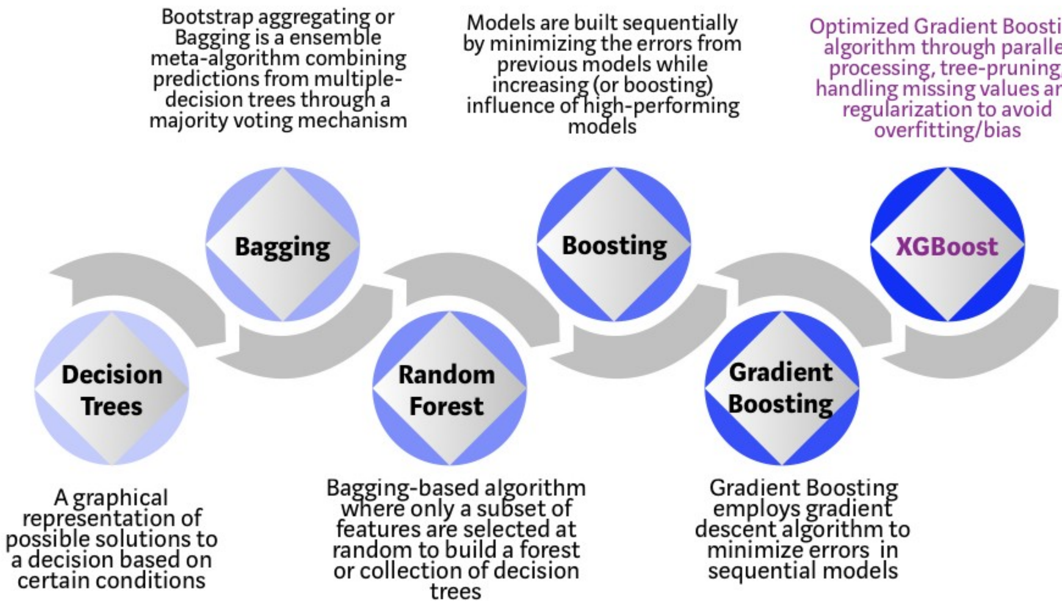

Figure 15:Evolution till XGBoost

There are many advantages of XGBoost like:

- Highly Flexible
- Uses the power of parallel processing.
- It is faster than Gradient Boosting.
- It supports regularization.
- It is designed to handle missing data with its in-build features.

XGBoost uses decision trees as base learners; combining many weak learners to make a strong learner. As a result, it is referred to as an ensemble learning method since it uses the output of many models in the final prediction as illustrated in Figure 16

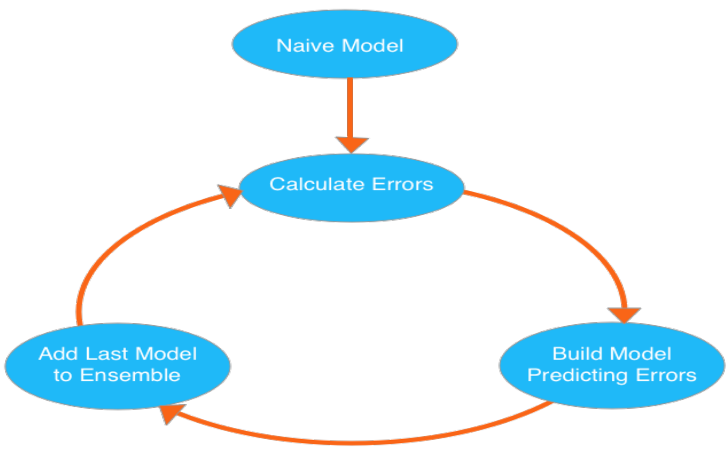

Figure 16: How XGBoost works.

XGBoost or Extreme Gradient Boosting can be used into various use cases such as ranking, classification, regression, and user-defined prediction problems.

Appendix 6- Splitting Data into training and testing dataset

```
set.seed(12345)


sample <- sample.int(n = nrow(xgb_Dati), size = floor(.80*nrow(xgb_Dati)), repl
ace = F)


train <- xgb_Dati[sample, ]


test  <- xgb_Dati[-sample, ]
```

Appendix 7- Constructing the model

```
params <- list(eta = 0.1,
               min_child_weight = 3,
               max_depth = 6,
               gamma = 1,
               eval_metric = "mae")

w <- list(train = dtrain, test = dtest)

xgb_modello1 <- xgb.train(data = dtrain,
          booster = "gbtree",
          params = params,
          watchlist = w,
          early_stopping_rounds = 100,
          nrounds = 5000
          )
```

As we can see a series of parameters were set for the XGBoost model where:

*Eta means the learning rate which is step size shrinkage used to prevent overfitting. Eta shrinks the weight of feature making the process of boosting more conservative

*min_child_weight is the minimum sum of instance weight(hessian) needed in a child. If the tree partition step results in a leaf node with the sum of instance weight less than min_child_weight the building process will give up further partitioning. In other words, it’s like saying to the model stop trying to split once the sample size in a node goes below a given threshold.

*max_depth is the maximum depth of a tree.

*gamma is the minimum loss reduction required to make a further partition on a leaf node of the tree.

*eval_metric is an evaluation metric for validation data (14).

Appendix 8- Mean Absolute Error

Mean Absolute error (MAE) in statistics is a measure of error between paired observations expressing the same phenomenon, Examples of Y versus X include comparisons of predicted versus observed where MAE is calculated as:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Where y is the prediction, x is the true value and n total number of data points.

References

1) T.Gordon, S Paul, A Lyles, J Fountain. Surgical unit time utilization review 1988: resource utilization and management implications: https://pubmed.ncbi.nlm.nih.gov/3049900/

2) Antti Peltokorpi 2011, How do strategic decisions and operative practices affect operating room productivity? https://pubmed.ncbi.nlm.nih.gov/21814829/

3) Robert W Allen, Kevin M Taaffe, Vivian Neilley, Erik Busby 2019: a First Case On-Time Starts Measured by Incision On-Time and No Grace Period: A Case Study of Operating Room Management: https://pubmed.ncbi.nlm.nih.gov/30845060/

4) Alex Macario, MD, MBA; Terry S.Vitez, MD; Brian Dunn, BA; Tom McDonald, MD 1995. Where are the costs in perioperative care?: Analysis of Hospital costs and charges for inpatient surgical Care: https://pubs.asahq.org/anesthesiology/article/83/6/1138/49/Where-Are-the-Costs-in-Perioperative-Care-Analysis

5) Kaitlynn Ely 2018. What are the implications of the costs of Operating room time? https://www.ajmc.com/view/what-are-the-implications-of-the-costs-of-operating-room-time

6) Christopher P.Childer, MD; Melinda Maggard-Gibbons, MD, MSHS 2018: Understanding costs of care in the operating room https://jamanetwork.com/journals/jamasurgery/fullarticle/2673385

7) Daniel M Laskin, A Omar Abubaker, Robert A Strauss 2013. Accuracy of predicting the duration of surgical operation https://pubmed.ncbi.nlm.nih.gov/23351763/

8) Neal Master, Zhengyuan Zhou, Daniel Miller, David Scheinker, Nicolas Bambos and Peter Glynn 2017. Improving predictions of pediatric

https://link.springer.com/article/10.1007/s41060-017-0055-0 surgical durations with supervised learning :

9) Mayoclinic https://www.mayoclinic.org/tests-procedures/cystoscopy/about/pac-20393694

10) Erican Roth, reviewed medically by Judith Marcin, M.D. 2018 Laryngectomy: Purpose, Procedure and recovery https://www.healthline.com/health/laryngectomy

11) Corporate Finance instituteshttps://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/

12) Committee on Economics https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system

13) Xgboost guide https://xgboost.readthedocs.io/en/stable/parameter.html