UNIVERSIDAD TORCUATO DI TELLA

Master in Management + Analytics

Customer Churn Detection and Marketing Retention Strategies in the Online Food Delivery Business

TESIS

Luis Ezequiel Muñoz

April 2022

Tutor: Elena Fumagalli

ABSTRACT

The purpose of this thesis is to analyze the behavior of customers within the Online Food Delivery industry, through which it is proposed to develop a prediction model that allows detecting, based on valuable active customers, those who will leave the services of Alpha Corporation in the near future.

Firstly, valuable customers are defined as those consumers who have made at least 8 orders in the last 12 months. In this way, considering the historical behavior of said users, as well as applying Feature Engineering techniques, a first approach is proposed based on the implementation of a **Random Forest** algorithm and, later, a boosting algorithm: **XGBoost**.

Once the performance of each of the models developed is analyzed, and potential churners are identified, different **marketing suggestions** are proposed in order to retain said customers. Retention strategies will be based on how Alpha Corporation works, as well as on the output of the predictive model. Other development alternatives will also be discussed: a **clustering model** based on potential churners or an **unstructured data model to analyze the emotions** of those users according to the NPS surveys. The aim of these proposals is to complement the prediction to design more specific retention marketing strategies.

<u>KEYWORDS</u>: machine learning, boost, xgboost, random forest, feature importance, marketing, retention, churn prediction.

RESUMEN

El presente trabajo tiene la finalidad de analizar el comportamiento de usuarios dentro de la industria de deliveries online, mediante el cual se propone desarrollar un modelo de predicción que permita detectar sobre la base de usuarios activos valorables aquellos que abandonarán los servicios de Alpha Corporation en un próximo futuro. En primera instancia, se definen como usuarios valorables aquellos consumidores que han realizado por lo menos 8 órdenes en los últimos 12 meses. De esta manera, considerando el comportamiento histórico de dichos usuarios, como así también aplicando técnicas de Feature Engineering, se propone una primera aproximación basada en la implementación de un algoritmo de **Random Forest** y, posteriormente, un algoritmo de boosting: **XGBoost**.

De esta manera, una vez que se analiza la performance de cada uno de los modelos desarrollados y se logran identificar aquellos potenciales churners, se propondrán diferentes sugerencias de marketing con la finalidad de retener a dichos usuarios. Las estrategias de retención se basarán en cómo funciona Alpha Corporation acompañado por el output del modelo predictivo. También se sugerirá otras alternativas de desarrollo que complementen la predicción para diseñar estrategias más puntuales: un **modelo de clusterización** sobre la base de potenciales churners o un **modelo de datos no estructurados para analizar las emociones** de aquellos usuarios según las encuestas NPS.

Palabras claves: machine learning, boosting, xgboost, random forest, feature importance, marketing, retention, churn prediction.

ACKNOWLEDGMENT

First and foremost I am extremely grateful to my thesis tutor, PhD. Elena Fumagalli for her invaluable advice, continuous support, and patience during this work. Her immense knowledge and plentiful experience have encouraged me in all the time of my academic research, as well as her guidance and advice carried me through all stages of my writing project.

My gratitude also goes to Universidad Torcuato Di Tella as well as its professors who provided me with the necessary tools and knowledge to make this work possible. Also to my colleagues from the Master in Management + Analytics (MiM) who allow me to share and discuss experiences that not only enriched me to do my best from the professional and academic aspect, but also from a personal one.

Finally, I would like to express my deepest gratitude to my parents, Nuchi and Sergio; my grandma, Teti; my sisters, Noel and Gusi. Without their tremendous understanding, support and encouragement in the past few years, it would not be possible to complete this thesis and be where I currently am, professionally and personally.

AGRADECIMIENTOS

En primer lugar, estoy muy agradecido con mi tutora de tesis, PhD. Elena Fumagalli por sus invaluables consejos, apoyo continuo y paciencia durante este trabajo. Su inmenso conocimiento y su abundante experiencia me han impulsado en todo el tiempo de mi investigación académica, como así también su orientación y consejos que me acompañaron a través de todas las etapas de elaboración de la tesis.

Mi más profundo agradecimiento también a la Universidad Torcuato Di Tella así como a sus profesores que me brindaron las herramientas y los conocimientos necesarios para hacer posible este trabajo. También a mis compañeros del Máster en Management + Analytics (MiM) que me permitieron compartir y discutir experiencias que no solo me enriquecieron para dar lo mejor de mí desde el aspecto profesional y académico, sino también desde el personal.

Finalmente, quisiera expresar mi eterno agradecimiento a mis padres, Nuchi y Sergio; mi abuela, Teti; mis hermanas, Noel y Gusi. Sin su enorme comprensión, apoyo y aliento en este último tiempo, no habría sido posible completar esta tesis y encontrarme donde estoy actualmente, profesional y personalmente.

CHAPTER 1 - INTRODUCTION

CHAPTER 2 - DATA

CHAPTER 3 - THEORETICAL FRAMEWORK

CHAPTER 4 - EXPLORATORY ANALYSIS

CHAPTER 5 - FEATURE ENGINEERING AND FEATURE SELECTION

CHAPTER 6 - MODEL DEVELOPMENT

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

As we all know, smartphone technology has played a big role in revolutionizing the food delivery service, turning a simple phone call into a completely huge online experience when ordering. Technology has also prompted companies to understand and get to know their customers better: not only by being aware of their personal preferences but also trying to explain the reasons why certain users stop using their delivery services. *Convenience* should be the best word to describe why people choose this type of service as it is the biggest appeal to consumers. Only a few personal information disclosure steps are required when it comes to placing an order, followed by some clicks on the mobile device you are using, choosing your payment method (whether cash or debit) and, by the time you have set the table, your delivery order will be waiting for you by the door.

The demand of **Online Food Delivery Service** (OFDS) has been growing incredibly fast in the past few years, reaching its highest peak this year as a result of the global COVID-19 pandemic. From a business perspective, restaurant owners would grab opportunities that are viewed as a new way of revenue generation. As mentioned above, consumer preference is the main motivating factor for business owners to engage with online delivery and, the idea of counting with a service like this makes them aware of how powerful and helpful this tool is for their business to remain competitive in the market.

The foodservice market size in Argentina was estimated to grow 21 percent from 2018 to 2022, a percentage three times higher than the estimated one for Brazil and Chile over that same period of time. In Colombia and Peru, the market size of this industry could increase up to six percent, while foodservice in Mexico could register a market size growth of four percent during the same period. For 2020, it was estimated that the foodservice industry in Latin America would generate 264 billion U.S. dollars[1]. However, as is known, the world was turned upside down by the outbreak of COVID-19. With people confined in their homes, food delivery has been even more critical in supporting the foodservice industry. Both Argentina and Brazil, the top online food delivery markets in Latin America, saw an increase in the use of delivery platforms.

In the present research, we will focus on the Argentinian market, and we will analyze a business issue that affects one of the top OFDS market players which, from now on, we will refer to as "Alpha Corporation'' to preserve confidentiality. After the COVID-19 outbreak in Argentina, the number of Alpha Corporations' app downloads grew by 126 percent compared to the same period of 2019.[2]

Even though the company under analysis is strengthening its position as top market player, counting with almost 90% of restaurants and stores provided with its own logistic network, market competition has become tougher and more intense in the past few years, especially since stay-at-home orders were issued and the lockdown

---

[1] Based on Statista Dossier, *Online food delivery in Latin America.* (2020)

[2] These results were obtained by an outsourced consultant hired by the company under-analysis. The whole document is not allowed to be shared.

began. For that reason, customer retention now became Alpha Corporation's number one concern. Alpha now needs to focus on keeping its customers happy and improving its retention rate. If customers are unhappy, they will not just stop using the service, but they will leave the company for one of its competitors and that is the last thing the business needs in the current context. Thus, Alpha Company should be very careful of customers moving to the competitors now during the pandemic when usage frequency is high because there is a lot of academic evidence that *"habits" stick* and its customers will not switch back once they get used to their competitors' services. Besides, there is a lot of academic research showing that gaining customers' back when they leave is way harder than getting new ones. [3] [4]

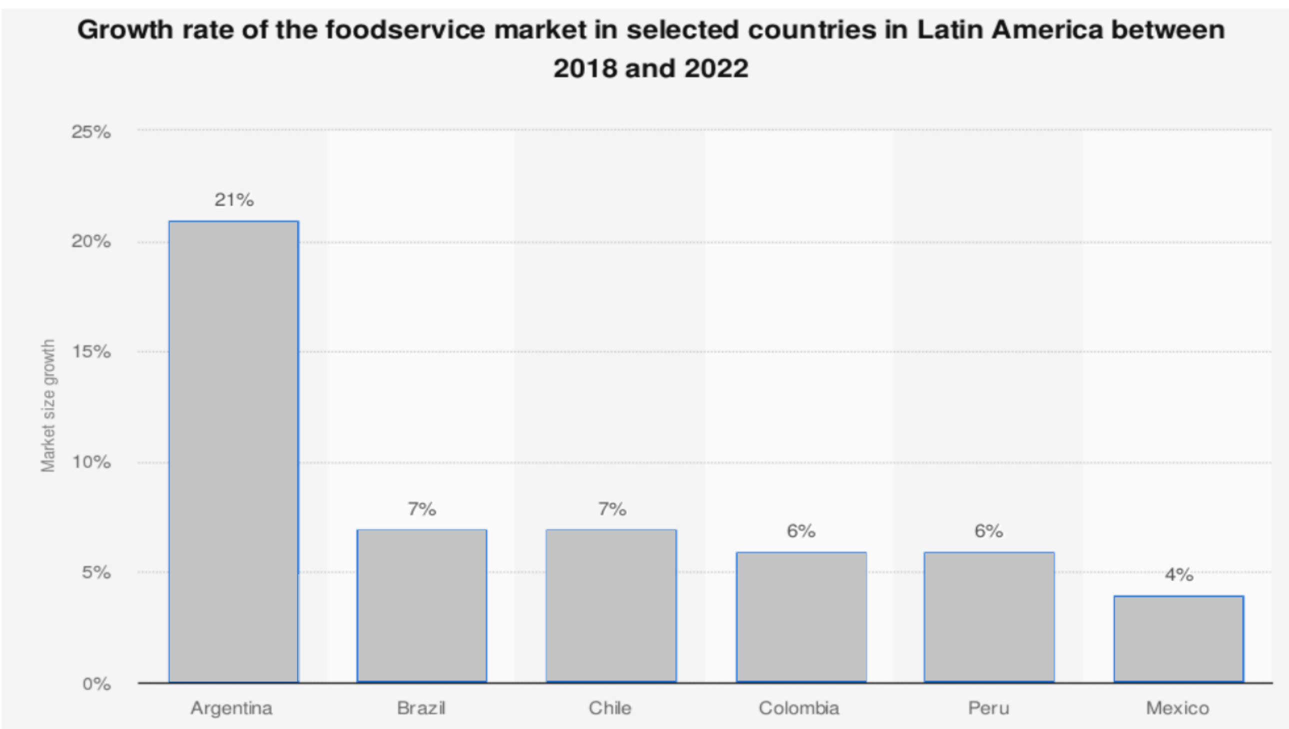


Fig 1.1 *Latin America: Foodservice market size growth 2018/22 - Forecasting*

[3] Frederick Newell, Martha Rogers. *Loyalty.Com: Customer Relationship Management in the New Era of Internet Marketing* (2002)

[4] Rajat Paharia. *Loyalty 3.0: How to revolutionize Customer and Employee Engagement with Big Data* (2013)

1.2 PROBLEM & OBJECTIVES

Based on what was mentioned above, the problem to be solved is not only related to the idea of **detecting customers who intend to leave** Alpha Corporation, but also trying to **understand their motivations to exit the relationship** by clustering them into different groups. The solution will require Machine Learning concepts as well as Marketing and Consumer Behavior notions.

The first part of the thesis will be dedicated to the **analytical side** and will consist of developing the churn prediction model and, once the customers most likely to exit the relationship are identified, the second part of the thesis will be presented which is related to the **managerial side**, and it will be dedicated to developing and analyzing the effectiveness of different marketing and retention strategies targeted to the potential churners that were identified in the analytical part.

1.3 EXPECTED OUTPUT

Hence, the **expected output** will consist of:

- A predictive model that will be mainly concerned with forecasting how Alpha Corporation's customers will behave in the future (whether they will exit the relationship or not) by analyzing their past behavior.

- Once the model is executed and flags customers who are more likely to churn, a variety of marketing and retention strategies will be discussed, determining which ones would be better suited to try and retain those potential churners.

- Some other suggestions are also proposed that go beyond the scope of this thesis, but that can be complemented to obtain a more straightforward output based on the initial predictive model developed.

- Last but not least, if the company agrees to collaborate, an experiment will test the efficacy of the retention strategies proposed above (i.e., treatment), comparing the effects of a treated cluster member with one that has been randomly assigned to the control group.

1.4 METHODOLOGY

To begin with, it is necessary to define what ***valuable users*** and ***churn*** mean to Alpha Corporation. Specifically, Alpha considers users to be *valuable* if they have ordered at least 8 times since they created their account. Therefore, for the purpose of our research question, *valuable customers who intend to leave* are those who placed orders recently but do not order again within the following 60 days (2 months). As such, we plan to analyze the historical data related to each valuable user to try and predict whether he/she will leave or not within the following 2 months. In order to do so, Alpha Corporation granted us permission to analyze the information contained in the full company database. We will have access to a dataset comprising 77 variables and more than 80.000 observations. Based on our research question, we provide an overview of some variables we can use for the analyses in the table below.

However, we will not limit our inquiry to the variables that are theoretically relevant. In order to make use of the majority of the information contained in the dataset, we intend to also extract features from raw data via data mining techniques (i.e., feature engineering).

Once a more comprehensive set of variables is selected, we will run our predictive model and analyze the output to define different typologies of customers that are likely to leave the company. After the potential churners are identified we will qualitatively analyze them to provide a series of marketing strategies to retain those users.

Finally, if feasible, we would like to conduct an experiment (e.g., A/B test) by randomly assigning potential churners to either the treatment or control group. The strategies we propose in the previous step will then be implemented with the treatment group, thus enabling us to test whether they will prevent churn in that specific cluster by means of comparing churn rates between the control and treatment groups. If our strategy is effective, the churn rate for the treatment group will be lower than that of the control group.

CHAPTER 2

DATA

All data used for the development of this thesis has been provided by Alpha Corporation under a non-disclosure agreement, in which it was requested not to mention the company's real name as well as modifying some sensitive inputs related to its clients and business strategies.

For the purpose of our research, it was decided to work with the following datasets in order to identify those features that will allow us to predict those valuable customers, who have placed more than eight orders within the last year, but do not order again in the following 60 days.

2.1 DATASETS

2.1.1 LOGISTIC DATASET

This dataset contains information related to the management of the flow of orders between the point of origin (customer placing an order) and the point of consumption (rider arrives at customer's place). Some features that allow us to check whether they meet the requirements of valuable customers or not, are shown below.

- ***Logistic Delivery on Last Order [boolean]:*** Whether the customer's last order was delivered by Alpha's own logistic (1) or not (0).

- ***Avg Delivery Time [numeric - minutes]:*** Average time it took to deliver each order from the moment the rider picked it up at the restaurant to the moment it was finally delivered.

- ***Avg Order Delay [numeric - minutes]:*** Average time difference between promised delivery time and actual delivery time.

- ***Avg Restaurant Delay [numeric - minutes]:*** Average time difference between the scheduled pickup time and the actual pickup time.

- ***Avg Accepting Time [numeric - minutes]:*** Average time lag between the customer placing an order and the restaurant confirming it.

- ***Avg Distance [numeric - meters]:*** Average distance in meters from the pickup spot to the delivery point.

- ***% Order Delay [numeric]:*** Percentage of total orders whose actual delivery time was longer than the promised one, over the total orders placed by the customer.

- ***% Restaurant Delay [numeric]:*** Percentage of total orders whose actual pickup time was longer than the scheduled one, over the total orders placed by the customer.

- ***% Accepting Delay [numeric]:*** Percentage of total orders whose time between the customer placed an order and the restaurant confirmed it is longer than 5 minutes.

2.1.2. FINANCIAL DATASET

This dataset contains information corresponding to the financial details of each order, indicating how well the company is doing regarding generating revenue and profits. Some features that allow us to verify what was mentioned before are shown below.

- ***% Orders Online Payment [numeric]:*** Percentage of confirmed orders with online payment over confirmed orders by customer.

- ***Avg Check-out size [numeric - local currency]:*** Average Order Value, net of discounts. Total paid by a customer divided by total confirmed orders.

- ***Avg Check-out size with discounts [numeric - local currency]:*** Average Order Value, including discounts. Total orders amount placed by a customer over total confirmed orders.

- ***Total Alpha Subsidy [numeric - local currency]:*** Total discount applied to an order that is paid by the company.

- ***% Subsidized Orders [numeric]:*** Percentage of confirmed orders with a discount applied over total confirmed orders.

2.1.3 SESSIONS DATASET

This dataset contains information regarding different metrics that allow the company to monitor and evaluate its customers' behavior. Some of the analyzed features are presented below.

- ***Sessions [numeric]:*** Total number of times a user logs in the app or visits the website. We consider the last 30 sessions for each customer.

- ***Avg Sessions Difference [numeric]:*** Average days between one session and the following one (within the last 30 days).

- ***Sessions Stddev [numeric]:*** Sessions Standard Deviation.

- ***Sessions Slope [numeric]:*** Describes both the direction and the steepness of the sessions' evolution over time.

- ***Days without Session [numeric]:*** Days without logging in or visiting the app since the query was actually run.

- ***Avg Session Time [numeric]:*** Average time spent by the customer in the app or visiting the website.

- ***Avg Session Time when No Transaction [numeric]:*** Average time spent by the customer in the app or visiting the website when no order was placed.

- ***Avg Session Time when Transaction [numeric]:*** Average time spent by the customer in the app or visiting the website when an order was placed.

- ***% Sessions Restaurant [numeric]:*** Percentage of sessions in which the customer visited only the category “restaurants”.

- ***% Sessions Others [numeric]:*** Percentage of sessions in which the customer visited a different category other than "restaurants".

- ***Avg Hits [numeric]:*** Indicates the average amount of clics made by an user within the app or website.

2.1.4 HISTORICAL O-REGISTRY DATASET

This dataset contains information regarding the amount of orders placed by users during each month. We would like to check whether the amount of previous orders helps us predict whether certain users do not order again in the following two months. For that reason, two features have been chosen to be part of the predictive model.

- ***orders_m_n1 [numeric]:*** Total amount of orders placed by the user a month prior the query is run.

- ***orders_m_n2 [numeric]:*** Total amount of orders placed by the user two months prior to the query is run.

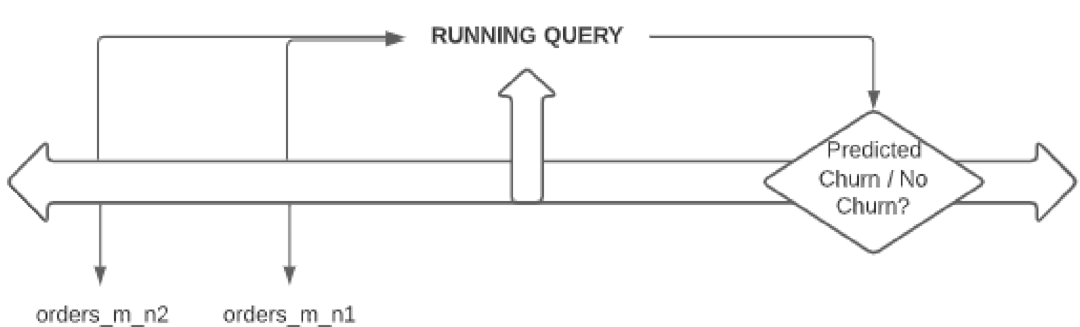


Fig 2.1 *Simplified Data Diagram for Prediction*

2.1.5 CATEGORY DATASET

This dataset reflects different aspects related to an order: date, time tier or vertical. Some features used to predict *Churn* are listed below.

- ***% Weekdays Orders [numeric]:*** Total orders placed from Monday to Thursday over total confirmed orders.
- ***% Weekends Orders [numeric]:*** Total orders placed from Friday to Sunday over total confirmed orders.
- ***% Orders Morning [numeric]:*** Total orders placed from 7 am to 11 am over total confirmed orders.
- ***% Orders Lunch [numeric]:*** Total orders placed from 11 am to 4 pm over total confirmed orders.
- ***% Orders Afternoon [numeric]:*** Total orders placed from 4 pm to 9 pm over total confirmed orders.
- ***% Orders Noon [numeric]:*** Total orders placed from 9 pm to 7 am over total confirmed orders.
- ***% Orders Other Categories [numeric]:*** Orders from a different category than "restaurants" over the total amount of confirmed orders.
- ***% Orders Coffee [numeric]:*** Orders from category "Coffee" over the total amount of confirmed orders.
- ***% Orders K [numeric]:*** Orders from a certain type of restaurants which have an agreement with Alpha Company.

2.1.6 DATASETS SUMMARY

Below is a table that summarizes the variables mentioned above, as well as which department they belong to.

BELONGS TO	*VARIABLE NAME*	*TYPE*
LOGISTICS	*Logistic Delivery on Last Order*	*Boolean*
	Avg Delivery Time	*Numeric*
	Avg Order Delay	*Numeric*
	Avg Restaurant Delay	*Numeric*
	Avg Accepting Time	*Numeric*
	Avg Distance	*Numeric*
	% Order Delay	*Numeric*
	% Restaurant Delay	*Numeric*
	% Accepting Delay	*Numeric*
FINANCIALS	*% Orders Online Payment*	*Numeric*
	Avg Check-out size	*Numeric*
	Avg Check-out size with discounts	*Numeric*
	Total Alpha Subsidy	*Numeric*
	% Subsidized Orders	*Numeric*
HISTORICAL O-REGISTRY	*orders_m_n1*	*Numeric*
	orders_m_n2	*Numeric*

SESSIONS	***Sessions***	***Boolean***
	Avg Sessions Difference	***Numeric***
	Sessions Stddev	***Numeric***
	Sessions Slope	***Numeric***
	Days without Session	***Numeric***
	Avg Session Time	***Numeric***
	Avg Session Time when No Trans.	***Numeric***
	Session Time when Transaction	***Numeric***
	% Sessions Restaurant	***Numeric***
	% Sessions Others	***Numeric***
	Avg Hits	***Numeric***
CATEGORY	***% Weekdays Orders***	***Numeric***
	% Weekend Orders	***Numeric***
	% Orders Morning	***Numeric***
	% Orders Lunch	***Numeric***
	% Orders Afternoon	***Numeric***
	% Orders Noon	***Numeric***
	% Orders Other Categories	***Numeric***
	% Orders Coffee	***Numeric***
	% Orders K	***Numeric***

Table 2.1 *Simplified Data Diagram for Prediction*

2.2 KPI CALCULATION

- ***Order Late [numeric - percentage]:*** Represents the percentage of the total orders that have been delivered in a time greater than the one promised through the app or the website.

$$Order\ Late\ = \frac{Orders\ whose\ delivery\ time\ was\ longer\ than\ the\ promised\ one}{Total\ confirmed\ orders\ by\ customers}$$

- ***Restaurant Late [numeric - percentage]:*** Represents the percentage of the total orders that have been picked up at the restaurant/vendor in a time greater than the one promised through the app or the website.

$$Restaurant\ Late\ = \frac{Orders\ whose\ actual\ pickup\ time\ was\ longer\ than\ the\ scheduled\ one}{Total\ confirmed\ orders\ by\ customers}$$

- ***Accepting Late [numeric - percentage]:*** Represents the percentage of the total orders that have been accepted by the restaurant/vendor in a time greater than five (5) minutes through the app or the website.

$$Accepting\ Late\ = \frac{Orders\ whose\ time\ between\ it\ was\ placed\ and\ confirmed\ is\ longer\ than\ 5\ minutes}{Total\ confirmed\ orders\ by\ customers}$$

- ***Online Payment Rate [numeric - percentage]:*** Represents the percentage of the total orders that have been paid online with debit, credit or Alpha's coins over the total orders confirmed by customers through the app or the website.

$$Online\ Payment\ Rate = \frac{Orders\ paid\ online\ (debit, credit\ or\ Alpha's\ coins)}{Total\ confirmed\ orders\ by\ customers}$$

- ***Subsidized Rate [numeric - percentage]:*** Represents the percentage of the total orders that have been subsidized by Alpha Corporation through the app or the website.

$$Subsidized\ Rate = \frac{Placed\ orders\ with\ a\ discount\ (voucher\ or\ subsidy)}{Total\ confirmed\ orders\ by\ customers}$$

- ***Sessions Restaurant Rate [numeric - percentage]:*** Represents the percentage of the total orders that have been placed within the category *Restaurant* through the app or the website.

$$Sessions\ Rest.\ Rate = \frac{Orders\ placed\ in\ "Restaurant"\ category}{Total\ confirmed\ orders\ by\ customers}$$

- ***Sessions Other Rate [numeric - percentage]:*** Represents the percentage of the total orders that have been placed within a category other than *Restaurant* through the app or the website. Some examples: pharmacy, cafeterias, supermarkets, etc.

$$Sessions\ Other\ Rate\ =\ 1\ -\ Sessions\ Rest.\ Rate$$

- ***K Rate [numeric - percentage]:*** Represents the percentage of the total orders that have been placed at a K restaurant through the app or the website.

$$K\ Rate\ =\ \frac{Orders\ placed\ at\ a\ "K"\ restaurant}{Total\ confirmed\ orders\ by\ customers}$$ [5]

- ***Tier Time [numeric - percentage]:*** Represents the percentage of the total orders that have been placed during a certain period of time (morning, lunch, afternoon, noon) through the app or the website.

$$Tier\ Time\ =\ \frac{Orders\ placed\ during\ morning\ /\ lunch\ /\ afternoon\ /\ noon}{Total\ confirmed\ orders\ by\ customers}$$

- ***Avg Hits [numeric - percentage]:*** The total number of clicks on the website and the app, over the number of total sessions..

$$Avg\ Hits\ =\ \frac{Total\ clicks\ on\ website\ /\ app}{Total\ sessions}$$

[5] "K" restaurants are those who have an exclusive agreement/contract with Alpha Corporation.

CHAPTER 3

THEORETICAL FRAMEWORK

The current chapter consists of four sections. The first one introduces basic concepts related to Customer Relationship Management (CRM) whereas the second segment explains how different Machine Learning and Data Mining techniques have contributed to its development nowadays. Meanwhile, section number three explains the impact and importance of analyzing and monitoring customer churn rates in the OFDS industry. Finally, the last section of this chapter describes the Machine Learning models and techniques most commonly used to analyze customer churn in the OFDS industry, some of which, will be used in the present work.

3.1 CRM: BRIEF INTRODUCTION

The essence of customer relationship management (CRM) is understanding customer needs and leveraging that knowledge to improve a company's long-term profitability. When successfully deployed, CRM can have a dramatic effect on bottom-line performance.[6]

When it comes to Customer Relationship Management there are three building blocks that coexist indistinctly. The first building block is an understanding of **customer decision-making**: almost 85 percent of all buying decisions have an emotional

[6] Anne Stringfellow, Winter Nie, David Bowen. *CRM: Profiting from understanding customer needs.* (2004)

component, thus according to Tehrani[7] understanding customers' emotional needs is vital for predicting and influencing their purchasing behavior. In fact, there is a huge difference between knowing *about* customers and actually knowing *the* customers.

The second building block is **customer information**. As discussed by Hennig-Thurau and Hansen[8] the information presently collected for CRM may be divided into three categories. *Personal information*, such as customer demographics, is useful in basic customer segmentation and for selecting advertising media. *Customer history* is a record of purchase transactions and such non-purchase transactions as complaints and service records. *Profitability information* expands on customer history by permitting the estimation of customer lifetime value. The most important category, customers' deep-seated needs, is often ignored, but such information can provide crucial insights into what drives customers' decision-making processes.

Finally, the third building block is **information-processing capability**. CRM systems need to integrate information from multiple sources and across different functions. Data must be organized by customers so that decisions can be made at the single customer or customer segment level. Fast processing allows for the information to be used in real-time (e.g., point-of-contact decision-making). For example, Alpha Company believes that the micro-segmentation of its customers is essential for identifying and keeping its most valuable customers. When a customer reaches the contact center complaining over a delivery, the full history of the account appears on the computer screen, and information on general customer characteristics and spending tendencies is available to guide the representative on how to "retain" that customer: offering a voucher, a discount, replacing the order again or offer a refund.

---

[7] Tehrani, Nadji. *The essence of CRM success: Customer Interaction Solutions*. (2002)

[8] Thorsten Hennig-Thurau. *Relationship Marketing*. (2000)

Some other authors like Olafsson, Li and Wu[9] believe that a valuable customer is usually dynamic and that the relationship evolves and changes over time. Thus, a critical role of CRM is to understand this relationship by studying the customer life-cycle (lifetime) which refers to various stages of the relationship between customer and business. This allows us to design models which predict more accurately those valuable customers who may stop placing orders in the near future.

The life cycle of customers is the base to maintain the business and consists of four main steps which are shown below. [10]

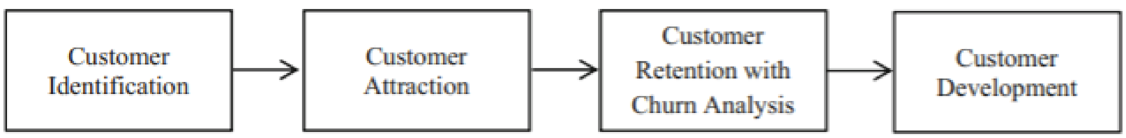


Fig 3.1 *CRM cycle steps.*

1. **<u>Customer Identification</u>**: This is the first step of the CRM framework. The objective of this step is to identify your target or prospect, a customer that is associated with classification or segmentation.

2. **<u>Customer Attraction:</u>** In this step, identified target customers are prioritized by providing them resources in order to attract them.

3. **<u>Customer Retention:</u>** Focus of the activities and actions realized to fulfill the customer expectations and satisfaction for maintaining the customer. Various loyalty programs, promotional schemes and complaints management are key elements to retain customers.

[9] Olafsson, Li, Wu. *Operations research and data mining.* (2008)

[10] Sabbeh SF. *Machine-learning techniques for customer retention.* (2018)

4. **Customer Development:** This step is related to enhancing transaction amount, business value, and specific customer effectiveness. The foremost features of customer development are customer lifetime, value analysis, and up/cross-selling.

All through our research we will be focusing on the third step of the CRM Cycle: *Customer Retention* as the objective of this thesis is to conduct a Customer Churn Analysis. All of this is closely related to loyalty and customer retention rates because an adequate study of churn and customer retention can help companies increase their ratio of valuable lifetime customers.

3.2 ML CONTRIBUTIONS TO CRM

Nowadays, lack of data is no longer a problem, but the inability to extract useful information from the dataset is. Due to this amazingly increase in the amount of data, many companies (including Alpha Corporation) have developed a greater dependency on statistical methods as a way of extracting useful information from different sources. These statistical methods provide a structured and organized way of looking at data and have been the base for developing new techniques and tools to transform data into useful information and knowledge: manually and automatically.

As mentioned above, these advances in Information Technology (IT) have made dramatic improvements in information-processing capabilities while data integration problems have been addressed through the use of data warehouses. Machine Learning techniques are used to improve the customer relationship. The research in Artificial

Intelligence (AI), Machine Learning (ML), Deep Learning, and Data Mining (DM) presented various excellent market algorithms and techniques to process customers' needs and discover diverse approaches and techniques for creating an operative CRM to fulfill those needs and requirements.[11]

Data is the main source of Machine Learning and Data Mining: on one side, DM discovers useful patterns and knowledge in the data, whereas ML comprises the algorithm that improves the automation without explicit program through experience based on data. As mentioned above, the basic goal of CRM is to help companies to understand customers' needs, streamline the processes and improve the profitability. A well-defined CRM system can be based on a customer segmentation strategy, identifying those who may be lost through the following months.

The application of ML and DM tools in CRM is an emerging trend in the global economy, since most companies -including Alpha- try to analyze and understand their consumer behaviors and characteristics for developing a competitive CRM strategy. The broad application of these techniques falls over two major categories:

- **Descriptive:** Increases the understanding of historical data and its content. Some techniques might be related to association analysis or data visualization
- **Predictive or Perspective:** Focuses on forecasting and devising. Orientates the decision process. Some examples might be related to classification models, regression or forecasting.

[11] Ngai EW, Xiu L, Chau DC. *Application of data mining techniques in customer relationship management* (2009)

3.3 IMPACT AND IMPORTANCE OF CUSTOMER CHURN IN THE OFDS INDUSTRY

Tanneedi[12] pointed out that customer churn has become a dreadful problem for the **Online Food Delivery Service** (OFDS) industry since customers never have a second thought to leave if they do not exactly get what they are expecting. The study emphasized that Big Data Analytics with Machine Learning are considered effective as means for identifying churn.

On the other hand, Almana[13] mentioned that the most complicated issue that this type of industry faces is customer churn. To effectively deal with the churn prediction challenge, the current study has used Machine Learning algorithms along with Data Mining tools. An effort in retaining existing customers could result in a considerable growth in profits as well as revenues. The necessity of retaining old customers calls for rigorous prediction of customer churn algorithms which are both precise and understandable. The study figured out certain factors that may impact customers to churn: customer loyalty is another factor which can be defined by quality of service and customer service delivered by Alpha company. Issues such as delivery late, canceled orders or mistaken orders might impact customers to switch to the competitors.

Today users will not think twice before switching if they do not get what they strive for, which is referred to as churning in Alpha. Customer churning is strongly associated with customer satisfaction. As the expenditure of gaining a new customer is

[12] Tanneedi, Naren Naga Pavan. *Customer Churn Prediction Using Big Data Analytics* (2016)

[13] Almana, Amal. *A Survey On Data Mining Techniques* (2014)

comparatively far higher than the expenditure of keeping an old one, OFDS have now changed their center of attention from acquisition of a customer to retention of itself.

Thus, Zineldin[14] pointed out that as a result of the fast development of this sector, the delivery providers are in position to progress towards extension of the subscriber base. To address the demand of retention in the competitive market, the retention of on-hand customers has become a complicated task and emphasized that it is important for Delivery Companies to have a churn prediction model in order to prevent their users from moving to other operator services. Consequently, the underlying principle of this study is to develop the customer churn prediction model. Machine learning can possibly be the sort of tool which could help delivery companies in churn prediction models. Machine learning is a kind of artificial intelligence tool which gives the capability to let computers learn the algorithm instinctively without human contribution.

The current study makes an effort to predict Alpha's customer churn employing Big Data analytics. Statistical analyses and machine learning applications such as Decision Trees or applications like XGBoost have been used over all the datasets mentioned above. The training data is used to develop classifiers employing machine learning techniques. Collecting knowledge from Alpha's database could contribute to predicting the level of involvement of the customers as to whether they are likely to leave the company or not.

[14] Mosad Zineldin. *The royalty of loyalty: CRM, quality and retention* (2017)

3.4 APPLIED ML TECHNIQUES

In simple words, ML is about predicting the future based on historical data, using a remarkable technique that gets better with time as it learns about the customer's profile, its retention and establishing a lifetime value. The role of ML techniques is excellent in CRM due to its predictive analytics capabilities - while the conventional perspective of CRM offers only a simple perception about users' data patterns based on the past or even present data, ML tools allows you to stay one step ahead: developing a predictive model that learns continuously and gets better with time.

3.4.1 SUPERVISED LEARNING

Supervised machine learning algorithms are designed to learn by example. The name *supervised learning* originates from the idea that training this type of algorithm is like having a teacher supervise the whole process.

When training a supervised learning algorithm, the training data will consist of inputs paired with the correct outputs. During training, the algorithm will search for patterns in the data that correlate with the desired outputs. After training, a supervised learning algorithm will take in new unseen inputs and will determine which label the new inputs will be classified as based on prior training data. The objective of a supervised learning model is to predict the correct label for newly presented input data. Therefore, supervised learning can be split into two subcategories: **Classification** and **regression**.

3.4.1.1 CLASSIFICATION

During training, a classification algorithm will be given data points with an assigned category. The job of a classification algorithm is to then take an input value and assign it a class, or category, that it fits into based on the training data provided.

In this thesis, the model development will be based on classification as it aims to determine if a customer will churn or not. With two classes to choose from (churn, or not churn), this problem is called a binary classification problem. The algorithm will be given training data with customer historical behavior that are both churn and not churn. The model will find the features within the data that correlate to either class and create the mapping function: $Y = f(x)$. Then, when provided with an unseen customer, the model will use this function to determine whether the user will churn (or not) in the near future.

Thus, classification issues can be solved with numerous algorithms. Whichever algorithm is selected to use depends on the data and the situation. Some popular classification algorithms are presented below.[15]

- Linear Classifiers
- Support Vector Machines
- Decision Trees
- K-Nearest Neighbor
- Random Forest

[15] Bramer, Max. *Principles of data mining. Vol. 180.* (2007)

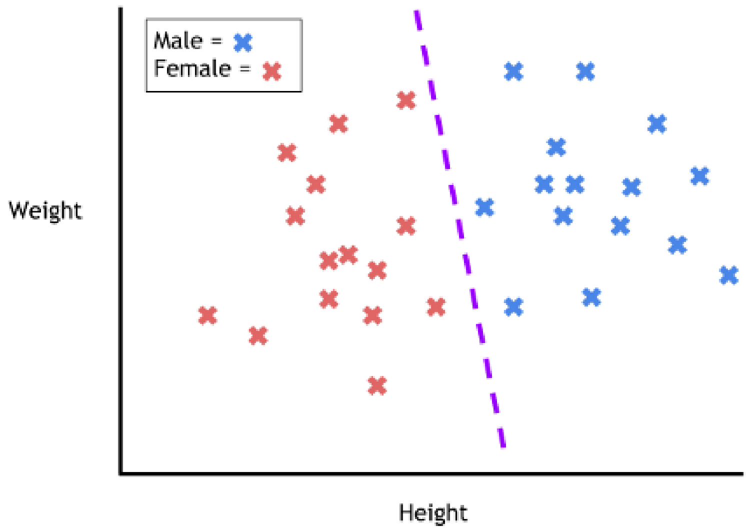


Fig 3.2 Classification example based on two labels: male and female

3.4.1.2 REGRESSION

Regression is a predictive statistical process where the model attempts to find the important relationship between dependent and independent variables. The goal of a regression algorithm is to predict a continuous number such as sales, income, and test scores.

For simple regression problems such as this, the model's predictions are represented by the line of best fit. For models using two features, a plane will be used. Finally, for a model using more than two features, a hyperplane is required.

Although a regression is not required for the development of the model of this thesis, the following case can be analyzed to understand the operation of the second aspect in supervised learning. In the case of determining a student's test grade based on how many hours they studied the week of the test, the following data is plotted with a line of best fit.

There is a clear positive correlation between hours studied (independent variable) and the student's final test score (dependent variable). A line of best fit can be drawn through the data points to show the model's predictions when given a new input. In this case, for example, it can be predicted how well a student would do with five hours of studying. by using the line of best fit to predict the test score based on other students' performances. [16]

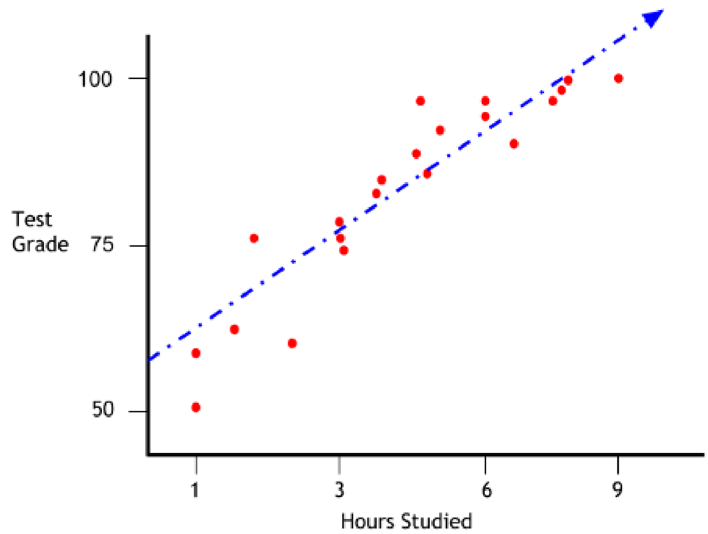


***Fig 3.3** Regression example based on hours studied and grade*

There are many types of regression algorithms. The three most common are listed below:

- Linear Regression
- Logistic Regression
- Polynomial Regression

[16] Aidan Wilsom. *A Brief Introduction to Supervised Learning* (2019) - Article available here

3.4.2 UNSUPERVISED LEARNING

In those cases when data does not have any special designated attribute it is called *unlabelled*, and techniques that are applied in these cases are known as *unsupervised learning*. The aim is simply to extract the most information possible from the data available.

Sometimes the aim of the research is to use a training set to find any relationship that exists amongst the values of variables, and this is commonly known as *association rules*. There are a lot of possible association rules derivable from any type of dataset, and most of them have little or no value at all, so it is important to bring some additional information indicating how reliable those association rules are.

On the other hand, there are some *clustering algorithms* that examine data and find groups that are similar. Given a set of data points, it is possible to apply clustering algorithms to classify each data point into a specific group, and those data points that are part of the same group should have similar characteristics and properties, while data points in different groups should have highly dissimilar properties and features.

At the end of this thesis, some suggestions will be proposed which involve a clustering model development to gain some valuable insights from the churn data by observing in which groups the data points fall into when applying clustering algorithms and, based on those groups, being capable of designing different marketing retention strategies for Alpha Corporation.

CHAPTER 4

EXPLORATORY ANALYSIS

How to choose the most suitable algorithms for your data set and how to define the feature variables that can potentially be used for the ML model? The **Exploratory Data Analysis (EDA)** answers this question: it is an approach for visualizing, understanding and summarizing the important characteristics of a certain dataset.

EDA is used by data scientists to analyze and investigate datasets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions. It is primarily used to observe what data can reveal beyond the formal modeling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques considered for the data analysis are appropriate. [17]

The main purpose of EDA is to get to know the data before making any biased assumptions: identifying obvious errors in data, understanding patterns, detecting correlations, outliers and interesting relations among the variables that might be useful for training the ultimate model. This step ensures the results are valid and applicable to any desired business outcomes and goals, and once it is complete, features can be used for more sophisticated data analysis and modeling.

[17] IBM Cloud Education. *Understanding Exploratory Data Analysis.* (2020)

On the other hand, one of the most important tools in exploratory analysis for detecting data quality is developing what is called ***quality report***, which describes the main characteristics of each feature using standard statistical measures. For that reason, while working on this thesis, after analyzing its quality report, some data quality issues such as missing values, large number of outliers or inappropriate data types were identified and immediately solved within the SQL query.

4.1 UNDERSTANDING DATA

Once the data quality issues are solved, it is time to understand the dataset and gain some insights from past behaviors in order to improve our prediction over their impact on future outcomes. The main goal of data understanding is to gain general insights about the data that will potentially be helpful for the further steps in the analysis process, but data understanding should not be driven exclusively by the goals and methods to be applied in later steps. Although these requirements should be kept in mind while initially exploring data, one should approach the data from a neutral point of view: never trust any data as long as you have not carried out some simple plausibility checks.

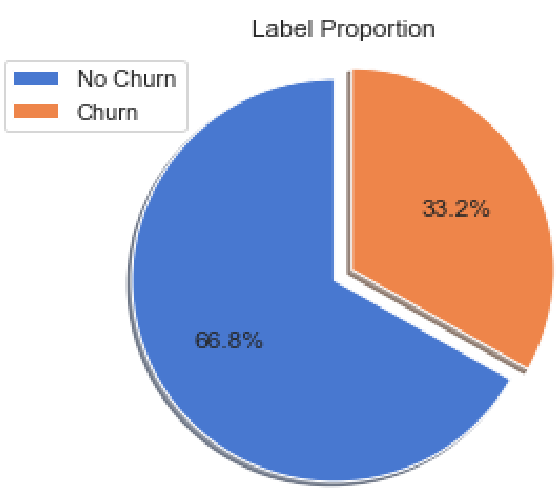
4.1.1 DATASET STRUCTURE

As mentioned previously, all data used for the development of this thesis has been provided by Alpha Corporation within different databases which were merged after an exhaustive study and coding with SQL language.

After the process of merging, the final dataset contains **85.084 observations** as well as **46 features (36 + 10 KPIs),** where each row is a particular user (identified by its respective *user_id* which was modified to preserve the company's privacy). Some features are integers, while some others are booleans and plain text.

Out of the **85.084 users** under analysis, **28.238** have **churned** in the following two months whereas **56.846 are still customers**. In other words, **33.20%** of the whole dataset has churned within the following two months based on the concept of churn defined by Alpha Company: *"valuable customers who intend to leave are those who placed at least 8 orders in the last 12 months but do not order again within the following 60 days (2 months after the analysis)"*.

The following chart illustrates the Churn/No Churn proportion.




```
False    56846
True     28238
Name: TARGET, dtype: int64
```

Fig 4.1 *Churn vs No Churn proportion*

4.1.2 CHURN RATE BASED ON TIME & ORDERS

In order to get some insights from data, the evolution of the Churn Rate will be analyzed through this subsection, considering some variables that involve time and are strictly related to the orders. Some examples that will be analyzed are the quantity of days without placing an order as well as the amount of orders placed two months ago (*orders_m_n2*) or the previous month (*orders_m_n1)*.

For that reason, a plotting function has been developed in order to represent the Churn Rate in relation to a certain variable under analysis. Thus, this will allow us to obtain some insights that may be significant to explain to the company which business aspects are leading its users to churn the service and, subsequently, be able to design retention marketing strategies for those users. Plots are presented below.

Regarding the variable *days_without_orders*, it can be perceived that there is a remarkable and positive relation between the churn rate and the number of days without placing an order. This means, in other words, that the greater the number of days that elapse since an order is not placed, the greater the probability of churn by customers.

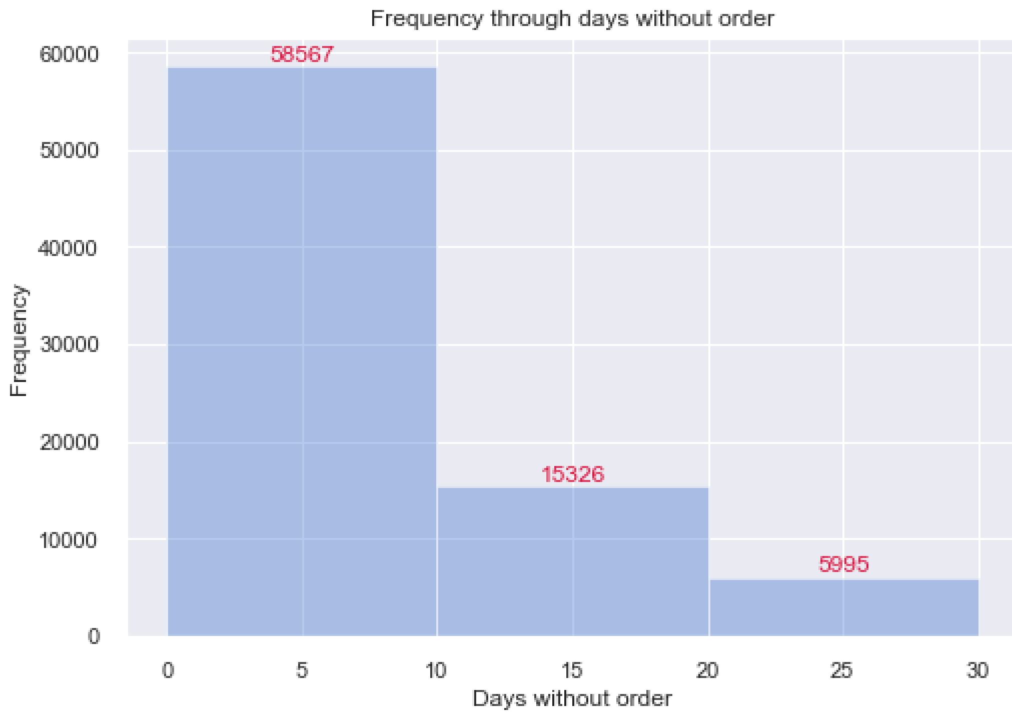


Fig 4.2 *Frequency through days without placing an order (plot max. limited 30)*

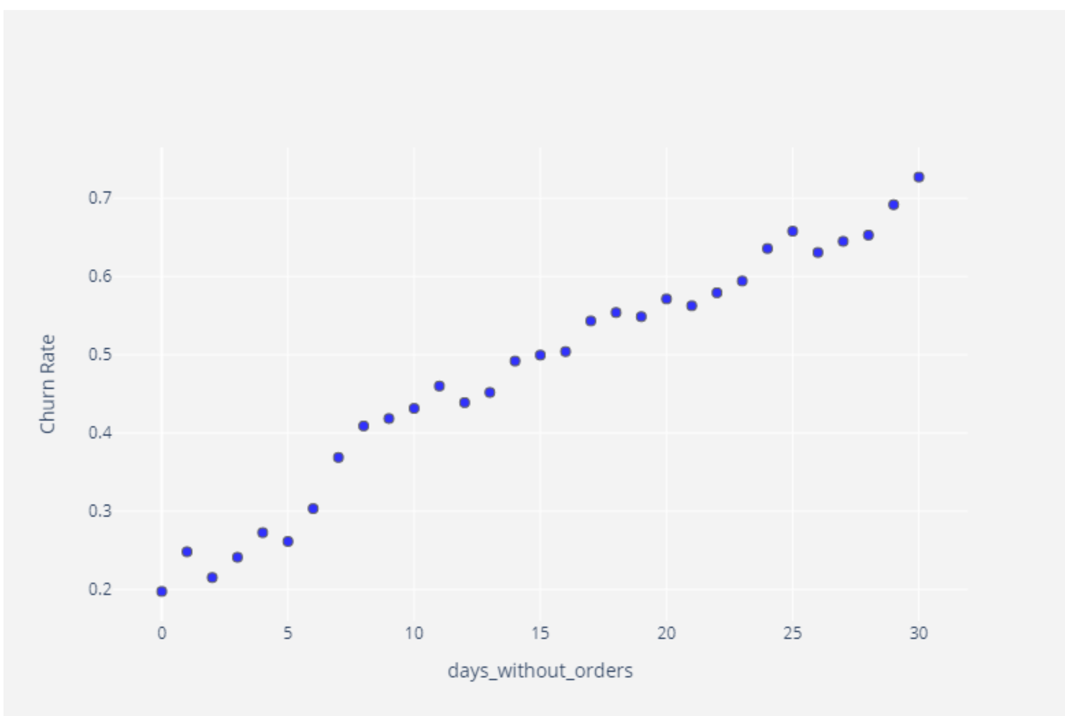


Fig 4.3 *Churn rate based on days without placing an order (plot max. limited 30)*

On the other hand, the variable *orders_m_n2* indicates the total quantity of orders that were placed by the customer two months previous to the analysis. A downward trend is observed: the higher the number of orders placed two months prior analysis is, the lower churn rate.

The same situation occurs when it comes to the variable *orders_m_n1* which represents the number of orders placed the months prior to the analysis by each customer. Therefore, it can be deduced therefore that those users who maintain a more active purchase history in the app are less likely to churn in the near future.

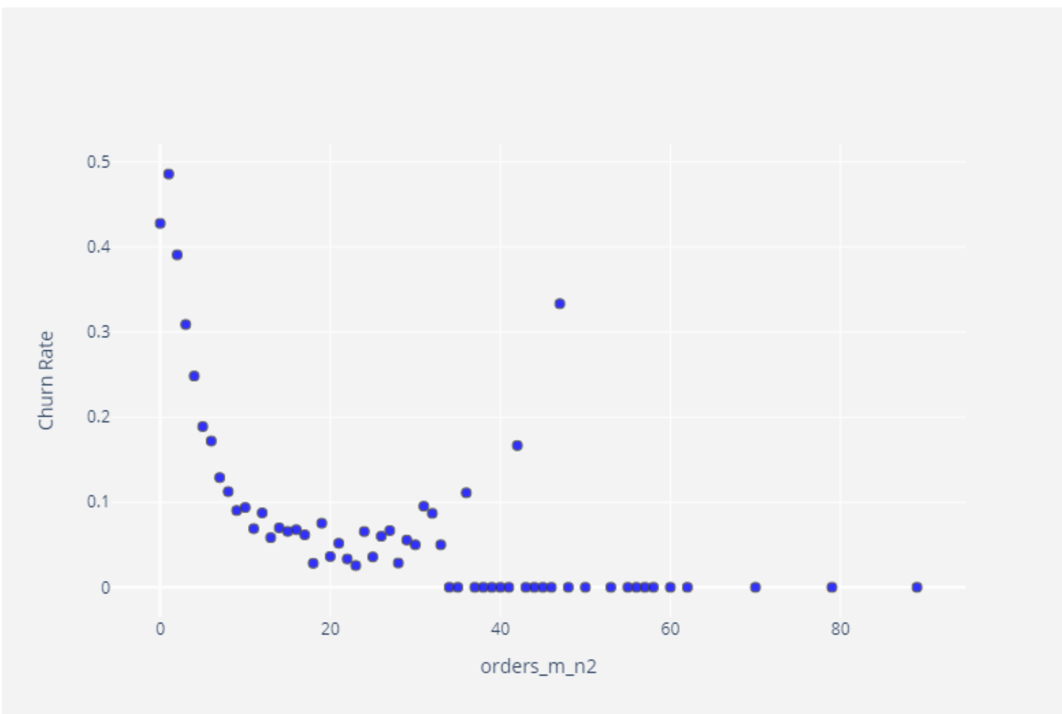


Fig 4.4 *Churn rate based on total orders placed two months prior to the analysis.*

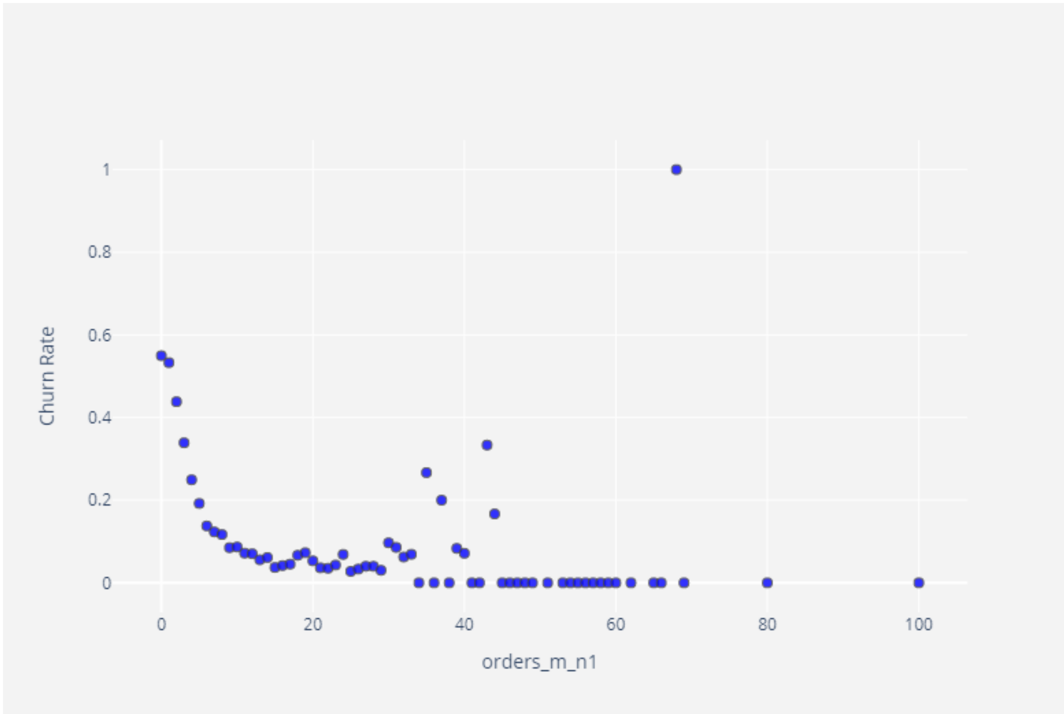


Fig 4.5 *Churn rate based on total orders placed one month prior to analysis.*

All the previous analysis considers both: orders with vouchers (any kind of discount) as well as those that can be called *pure orders* or *neutral orders* (referring to those that do not have any type of discount). For this reason, it is of great importance to verify whether the evolution of non-discount orders maintains a similar trend or not.

In this case, the following variables will we used: *orders_m_n2_without_vouchers* and *orders_m_n1_without_vouchers* which represent the total of pure or neutral orders -that do not present any type of discount in their price- and have been placed by each customer two or one month prior the analysis respectively.

Again, it can be observed the same inverse trend is maintained in both cases: by increasing the number of orders in months prior to the analysis, the Churn Rate of users decreases significantly. Although there are some outlier values, the inverse trend can be verified nevertheless. Both plots are presented below.

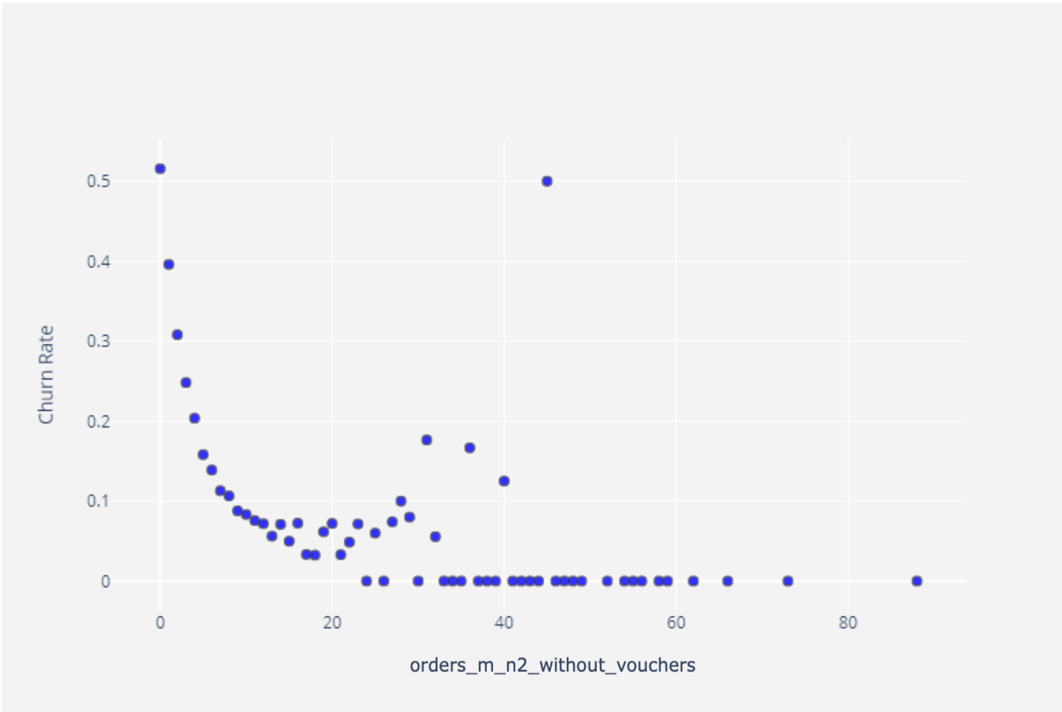


Fig 4.6 *Churn rate is based on total pure/neutral orders placed two months prior to analysis.*

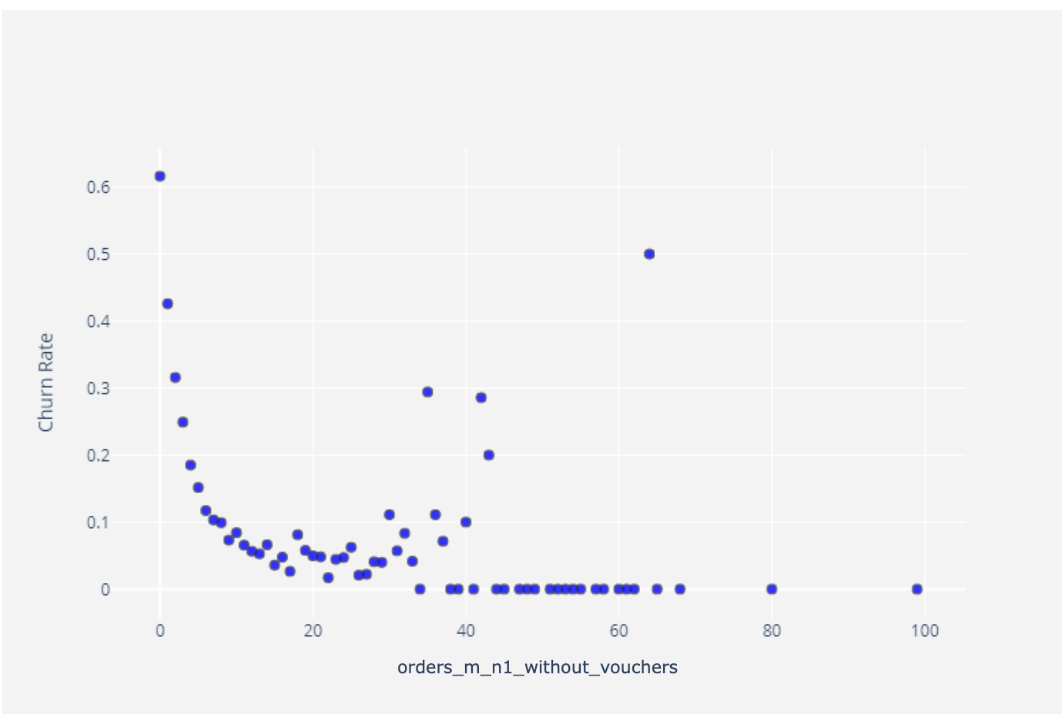


Fig 4.7 *Churn rate is based on total pure/neutral orders placed one month prior to analysis.*

4.1.3 CHURN RATE BASED ON LOGISTICS

As mentioned in Section 2, one of the most fundamental aspects of the industry under analysis is its logistics as they provide the delivery service to third parties also through their own logistics. For that reason, it is extremely useful to verify if there is any relation/trend between the churn rate and the variables related to the logistics department of Alpha Company.

Thus, the *log_orders* variable is used, which represents the number of orders that have been placed by each customer and been delivered thanks to Alpha's own logistics. It is important to emphasize that the company offers two alternatives:

1. **Logistic Orders**: Those orders that have been delivered by the company's own logistics.
2. **Market Orders**: Those orders that have been delivered by the restaurant/vendor's own logistics. In this case it is not possible to get the logistics metrics since the delivery is in charge of the vendor itself.

From the plot presented below, it can be seen that there is a notable downward trend in the Churn Rate as the number of logistics orders increases. One possible explanation is that customers highly value tracking the delivery from the app and the website, as well as having a greater reliability in the delivery and its respective logistics statistics that are also provided (e.g. expected delivery time). On the contrary, they might be disappointed when their orders are delivered by Alpha's vendors logistics as they have no chance to monitor their delivery. However, this outcome is not as straightforward as it seems to be: another reason might be related to the total orders, where the bigger the number of orders, the lower the churn rate is. This side effect can be easily addressed by plotting the churn rate against the share of logistic orders (logistic orders/ total orders).

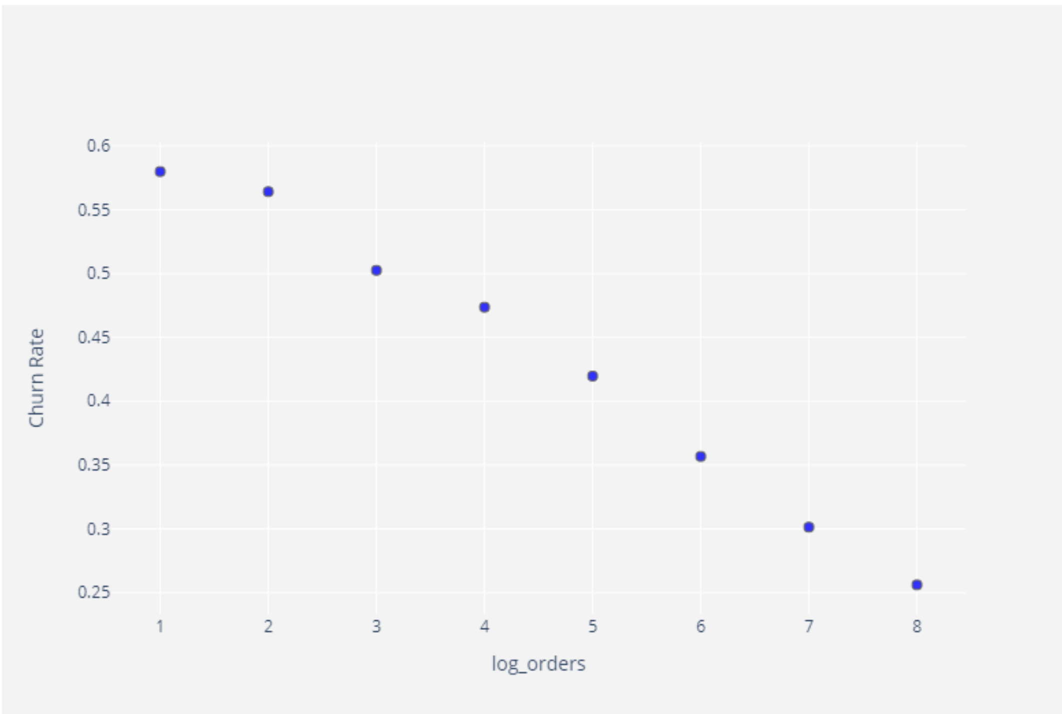


Fig 4.8 *Churn rate based on logistic orders.*

4.1.4 CHURN RATE BASED ON SESSIONS

One of the most powerful features of Alpha Corporation is its ability to identify and track users based on their session tagging. In order to begin with, a session is defined as a period of time wherein a user interacts with Alpha's app or website and usually triggered by opening the app, a session records the length and frequency usage to show developers, marketers and product managers how much time users spend within an app and website. Sessions can be analyzed in a way that reveals how users truly interact with it, allowing us to identify opportunities and problems as well as optimizing our model under development.

In the case of analyzing the variable *days_without_sessions*, which represents the total number of days without signing in the app or visiting the website, a notable direct and positive relation between days without sessions and churn rate is confirmed. This means, in other words, that those users who have less interaction with the website or app have a greater tendency to churn. The graphic is presented below.

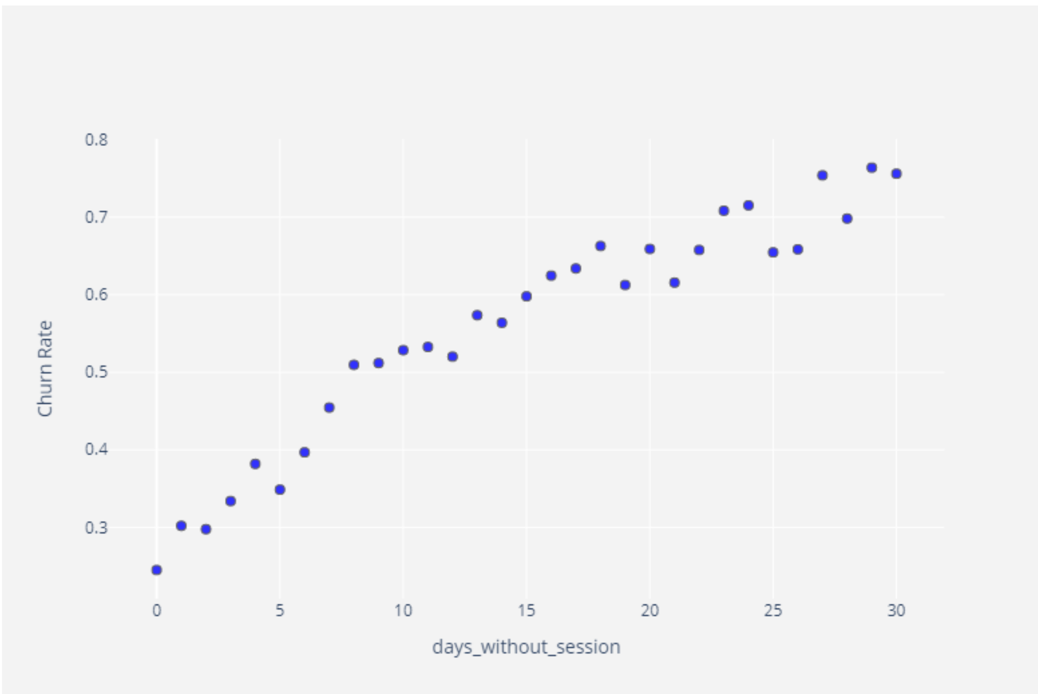


Fig 4.9 *Churn rate based on days without sessions.*

4.1.5 CHURN RATE BASED ON ORDER AMOUNT

Orders total amount is a measurable financial value that indicates how well the company is doing regarding generating revenue and profits. Based on its distribution, it can be observed that, although the Churn and No Churn areas are overlapping, the lower the average order amount ($), the greater the churn probability of said customer. These monetary values have been modified to preserve the confidentiality of the company.

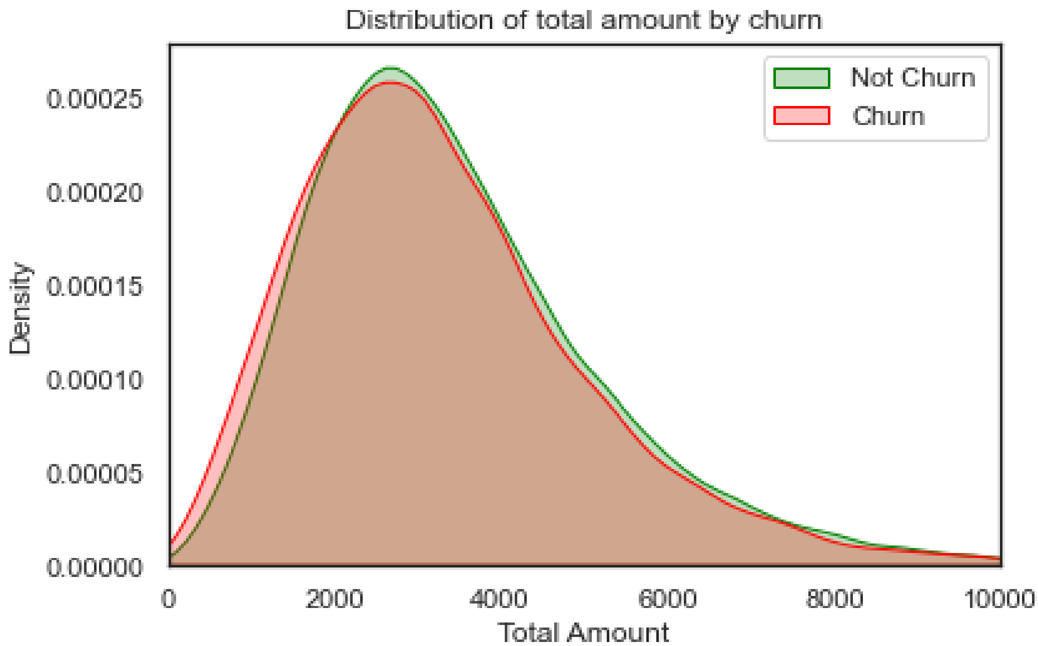


Fig 4.10 *Churn vs No Churn based on total amount (rescaled for confidentiality)*

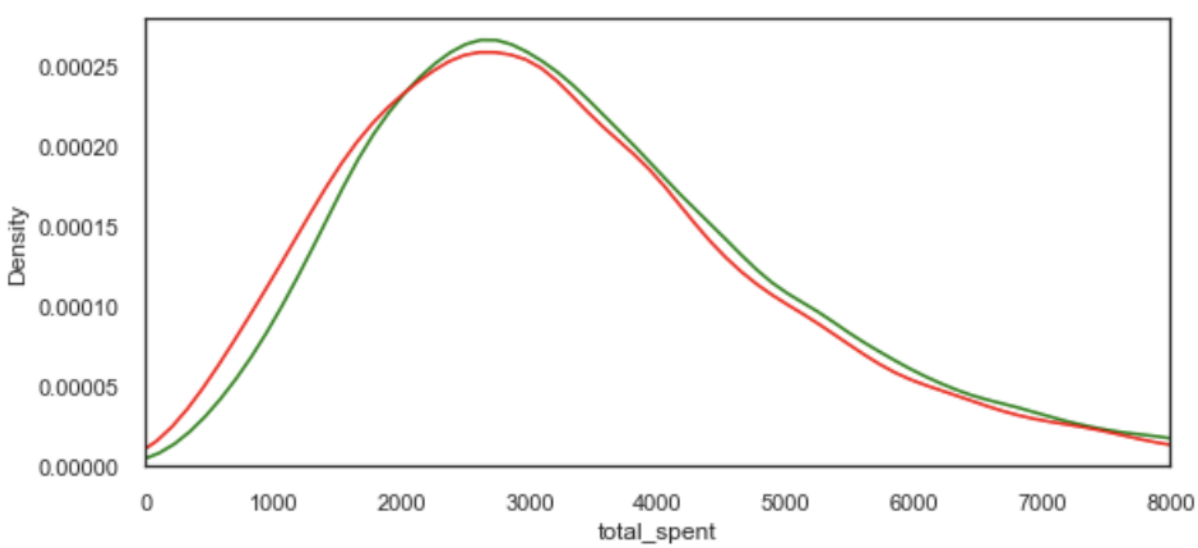


Fig 4.11 *Churn vs No Churn based on total spent, no discount applied (rescaled for confidentiality)*

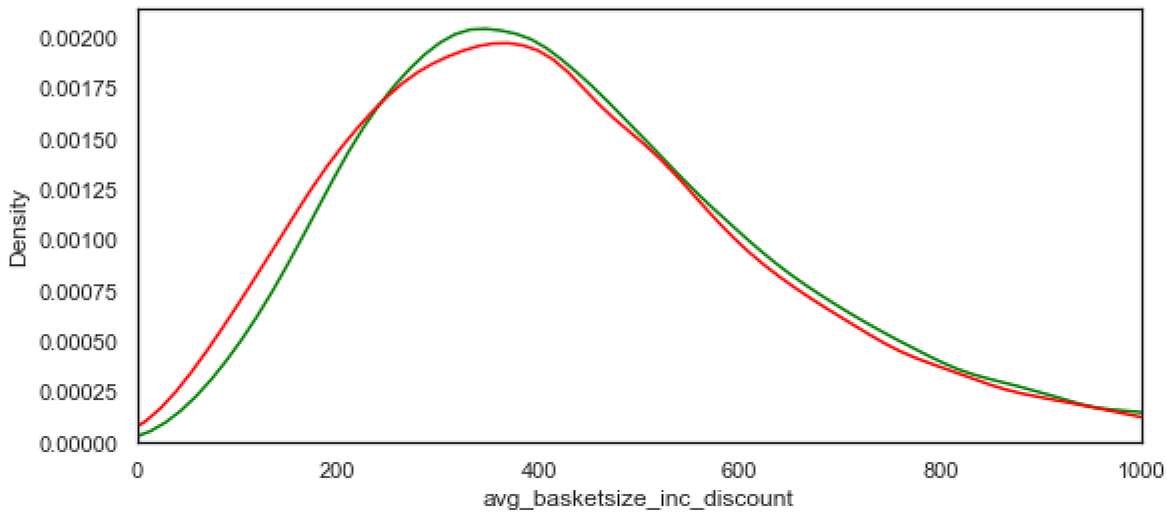


Fig 4.12 *Churn vs No Churn based on total spent, discounts applied (rescaled for confidentiality)*

4.1.6 HEAT MAP & VARIABLE CORRELATION

A Cluster Heat Map is a rectangular tiling of a data matrix with cluster trees appended to its margins. Within a relatively compact display area, it facilitates inspection of row, column, and joint cluster structure. Moderately large data matrices (several thousand rows/columns) can be displayed effectively.

From the graph presented below, it can be observed that the delivery time is highly and positively correlated to the accepting time: thus, the more it takes for a restaurant to accept an order, the more it takes to deliver it.

In the same way, the higher the total amount of the order, the more items are included in that order, or they require more time to be elaborated. In such a way, the average delivery time that elapses from placing an order to actually receiving it, would also be increased.

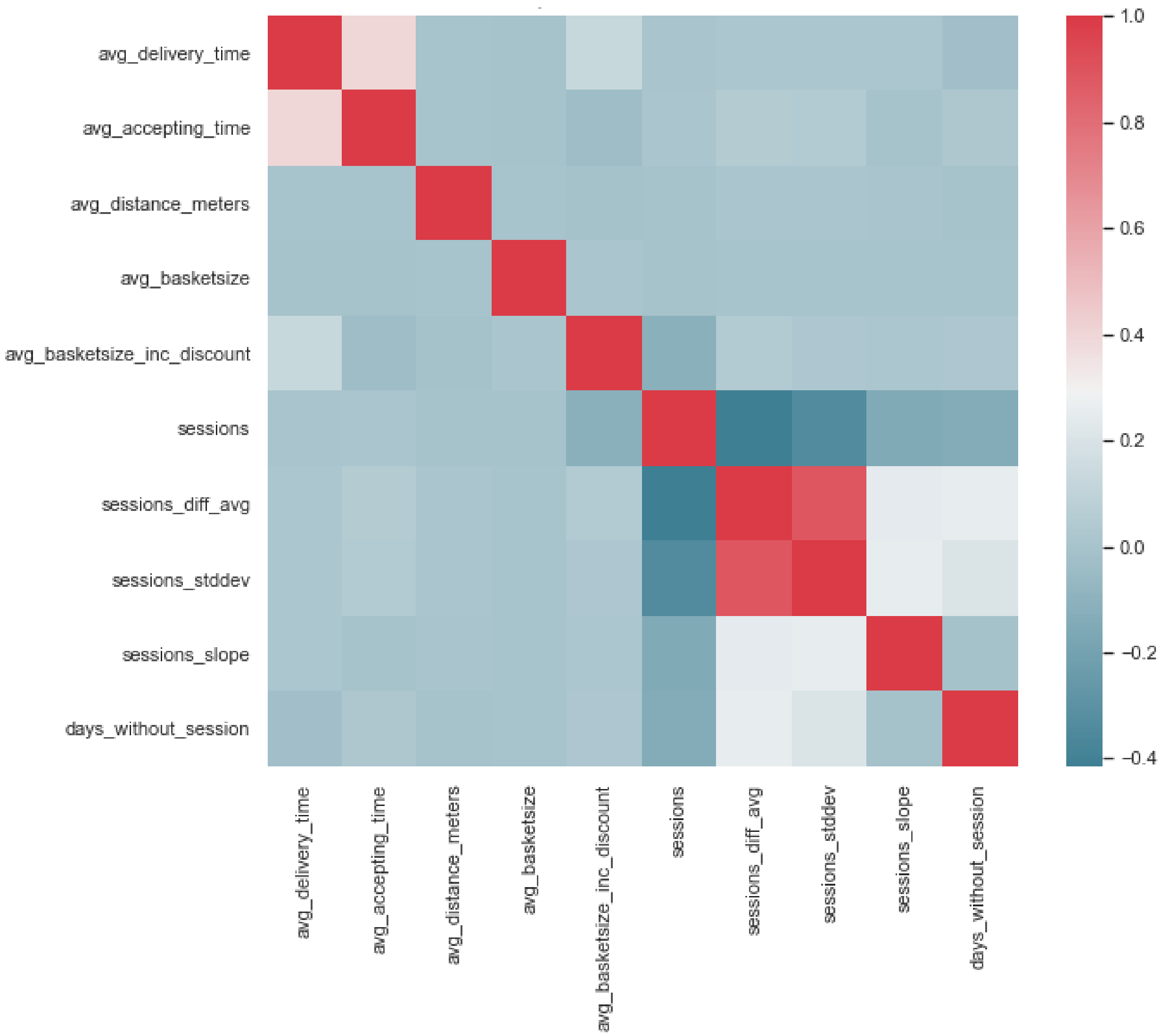


Fig 4.13 *Heat map and some correlated features*

CHAPTER 5

FEATURE ENGINEERING AND FEATURE SELECTION

In order to improve the performance of any machine learning model, it is important to focus more on the data itself instead of continuously developing new algorithms. This is exactly the aim of feature engineering. It can be defined as the clever engineering of data hereby exploiting the intrinsic bias of the machine learning technique to our benefit, ideally both in terms of accuracy and interpretability at the same time. Oftentimes it will be applied in combination with simple machine learning techniques such as regression models or decision trees to boost their performance (whilst maintaining the interpretability property which is so often needed in analytical modeling), but it may also improve complex techniques such as XGBoost and neural networks. Feature engineering aims at designing smart features in one of two possible ways: either by adjusting existing features using various transformations or by extracting or creating new meaningful features (a process often called *featurization*) from different sources (e.g., transactional data, network data, time series data or text data).[18]

It is clear from the definition that the feature engineering steps may encompass many methodologies to reach its goals. Some feature engineering techniques rely mainly on domain knowledge, others on intuition or are obtained through data analysis. Those techniques are typically applied after gathering and cleaning the input

---

[18] Verdonck, Tim. Baesens, Bart. *Special issue on Feature Engineering* (2021)

data. In the cleaning step, one typically deals with missing values, errors, outliers and duplicates.

Throughout this chapter, the different Feature Engineering techniques that were applied will be presented. The importance of each of them will also be mentioned, as well as their respective implementation.

5.1 MISSING VALUES

Nowadays, with ever-increasing data velocity and volume, missing data has become a common phenomenon. In many applications of Machine Learning considerable part of data seems to be missing, however, despite the frequent occurrence of missing data, most machine learning algorithms handle this problem.[19]

Generally, there are three approaches for dealing with missing features in machine learning: **omitting samples-features**, **imputing null values with some default value** of other samples or **applying standard learning algorithms** to deal with missing features (e.g. ELM algorithms). Considering that the null values were less than 1% of the total base due, mainly, to errors in data processing, the first approach was implemented: **omitting samples**. This step simply omits the samples containing missing features and applies standard learning algorithms in the remaining samples. The advantage of applying this approach is that it is simple and makes computation inexpensive. Notably, the key point of omitting is keeping as much as possible useful information while omitting, but sometimes it is difficult to do so as inevitably it will

[19] Hang Gao, Xin-Wang Liu, Yu-Xing Peng. *Sample-Based Machine Learning with Missing Data* (2015)

omit some useful information. When there is massive information retained after being partly omitted, this approach can be a better choice, being the reason why it was chosen. Otherwise, in the situation of much useful information being omitted while few being retained, this kind of approach affects learning precision seriously.

Below, part of the code is also presented, where it can be observed that the omission of null values represents less than 1% of the total base. It was also checked there were no infinite values, which sometimes occur when calculating certain KPIs (e.g. dividing some features by the order total amount which, in some cases, is equal to 0 because the entire order is subsidized). On the other hand, the explanation of the approach that was carried out is also graphically presented.

```
# In case of infinite values.
data.replace([np.inf, -np.inf], np.nan, inplace=True)
```

```
# NA values are dropped. Not significat impact as they are less than 1% of the total data base.
data.dropna(inplace=True)
```

```
data.shape
```

Fig 5.1 *Missing values and infinite values - Python Code*

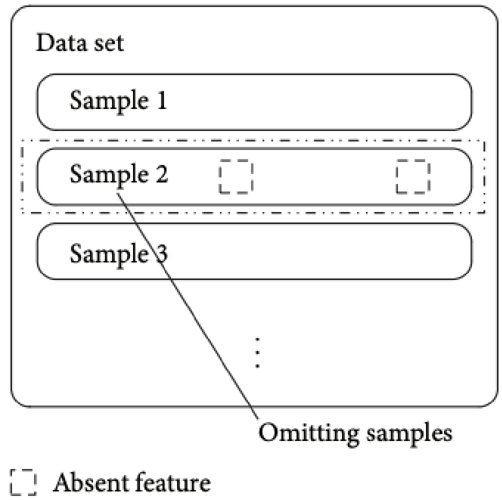


***Fig* 5.2** *Omitting null samples explanation*

5.2 LOCAL OUTLIER FACTOR

Local Outlier Factor (LOF) is a well-known unsupervised data mining technique which can locate the anomalous points of the given dataset with free specific data distribution assumptions.

LOF indicates the degree of how isolated a sample is with respect to its surrounding neighbors. If sample X is not an outlier, the LOF value would be approximately equal to 1. However, if sample X is an outlier, the LOF value would be larger than 1. Therefore, LOF can identify whether a sample is an outlier without any data distribution assumption.[20]

Considering the LOF execution time and the low impact it had on final results (absence of numerous outliers), it was discarded from the final code.

[20] Lei Wang, Xiaogang Deng, Yuping Cao. *Multimode complex based on local outlier factor method* (2017)

5.3 SMOTE

Imbalanced data is a common problem in classification. This phenomenon is growing in importance since it appears in most real domains. It has special relevance to highly imbalanced data-sets (when the ratio difference between classes is high). Many techniques have been developed to tackle the problem of imbalanced training sets in supervised learning. Such techniques have been divided into two large groups: those at the **algorithm level** and those at the **data level**. Data level groups that have been emphasized are those that try to balance the training sets by reducing the larger class through the elimination of samples or increasing the smaller one by constructing new samples, known as undersampling and oversampling, respectively.

The imbalanced data-set problem in classification domains occurs when the number of instances that represent one class is much larger than the other classes. The minority class is usually more interesting from the point of view of the learning task. For this particular case, potential churners are the ones who matter the most and represent the minority class, with a distribution close to **70% - 30%**.

With this approach, the positive class is oversampled by taking each minority class sample (churn users) and introducing synthetic examples along the line segments joining any/all of the *k* minority class nearest neighbors. Depending upon the amount of oversampling required, neighbors from the k-nearest neighbors are randomly chosen.[21]

As mentioned above, it can be observed that the data set presents a certain imbalance (not critical as in some other cases) but that it is advisable to solve it before running the predictive model. In this way, SMOTE was applied to the train data while

[21] Enislay Ramentol, Rafael Bello, Francisco Herrera. *Smote: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE* (2011)

oversampling the minority class and, thus, balancing both categories: from **17.666 churn users to a total of 49.909**, being equal to the total number of non-churn customers.

Getting training and test set

```
X_train, X_test, y_train, y_test = train_test_split(
    data.drop(labels=['TARGET'], axis=1),
    data['TARGET'],
    test_size=0.15,
    random_state=0)

X_train.shape, X_test.shape

X_train = X_train.fillna(X_train.mean())
X_test = X_test.fillna(X_test.mean())
```

Applying SMOTE for imbalanced classes

```
smt = SMOTE(random_state=0)
X_train_smote, y_train_smote = smt.fit_resample(X_train, y_train)
```

```
from collections import Counter
print("Before SMOTE:", Counter(y_train))
print("After SMOTE:", Counter(y_train_smote))
```

```
Before SMOTE: Counter({0: 49909, 1: 17666})
After SMOTE: Counter({0: 49909, 1: 49909})
```

Fig 5.3 *Applying SMOTE - Python Code*

5.4 CORRELATED FEATURES

As previously mentioned, feature selection is the process of reducing the number of input variables when developing a predictive model. It is desirable to reduce the number of them to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model. Statistical-based feature selection methods involve evaluating the relationship between each input variable and the target variable applying statistics and selecting those that have the strongest relationship with the target variable.

Empirical evidence from the feature selection literature shows that, along with irrelevant features, redundant information should be eliminated as well. A feature is said to be redundant if one or more of the other features are highly correlated with it.

In this case, we proceeded to discard certain attributes that were denoted with perfect correlations. Then, following the same logic applied in the analysis of variability of the attributes, it was decided to rule out features that exceeded a threshold of 0.85, discarding four of them.

```
# Treshold fixed
threshold = 0.85

# Features buffer
corr_features = set()
corr_matrix = X_train.corr()

for i in range(len(corr_matrix.columns)):
    for j in range(i):
        if abs(corr_matrix.iloc[i, j]) > threshold:
            colname = corr_matrix.columns[i]
            corr_features.add(colname)

print('Number of Correlated Features: ', len(corr_features) )
print('Correlated Features:', corr_features)
```

```
Number of Correlated Features: 4
Correlated Features: {'last_order_total_amount_with_dc', 'sessions_stddev', 'total_spent', 'stddev_intermittence'}
```

```
# Drop correlated features
X_train.drop(labels=corr_features, axis=1, inplace=True)
X_test.drop(labels=corr_features, axis=1, inplace=True)

X_train.shape, X_test.shape
```

```
((99818, 32), (11925, 32))
```

```
X_train.head()
```

	Unnamed: 0	user_id	avg_intermittence	slope_intermittence	days_without_ord	log_ord	avg_vendor_late	avg_rider_late	avg_order_late	avg_delivery_time	...
0	52352	19589973	3.125000	0.196429	0	8	1.500000	-1.375000	-6.428571	30.125000	...
1	23417	21857625	3.833333	-0.372024	3	3	5.000000	0.333333	-1.333333	27.333333	...
2	42229	4471105	18.267857	1.541667	4	8	2.750000	-5.125000	1.000000	24.625000	...
3	64108	16018779	26.702381	7.598214	19	8	10.500000	-9.375000	0.125000	35.625000	...
4	9280	15126547	14.994048	6.172619	12	6	1.833333	-4.333333	-3.600000	21.000000	...

Fig 5.4 *Correlated features - Python Code*

5.5 DUPLICATED & CONSTANT FEATURES

Constant features are the type of variables that contain only one value for all the outputs in the dataset. Constant features provide no information that can help in classification of the record at hand. They are redundant data available in the dataset as their presence has no effect on the target, therefore, it is advisable to remove all the constant features from the dataset.

```
# Drop correlated features
X_train.drop(labels=corr_features, axis=1, inplace=True)
X_test.drop(labels=corr_features, axis=1, inplace=True)

X_train.shape, X_test.shape
```

```
((99818, 36), (11925, 36))
```

```
print("X_train", X_train.shape)
print("Y_train", y_train.shape)
```

```
X_train (99818, 36)
Y_train (99818,)
```

Fig 5.5 *Dropping constant features - Python Code*

On the other hand, quasi-constant features are those, as their name suggests, that are almost constant. Previously, a threshold of 0 has been placed for constant features, thus for quasi-constant features, this can be adjusted arbitrarily. The procedure is almost the same as the previous one. It is recommended to examine the quasi-constant features in the already reduced dataset.

In this case, instead of defining a threshold parameter equal to zero, we now will define a 0.01 threshold, which means that if the variance of the values in a certain column is less than 0.01, then that feature will be removed.

```
sel = VarianceThreshold(
    threshold=0.01)  # Treshold
sel.fit(X_train)
```

```
VarianceThreshold(threshold=0.01)
```

```
# Get features
sum(sel.get_support())
```

```
36
```

```
# Fields with variance lower than 0.01:
low_var_features = [x for x in X_train.columns if x not in X_train.columns[sel.get_support()]]
print(len(low_var_features))
print(low_var_features)
```

```
0
[]
```

```
# Remove them:
X_train.drop(labels=low_var_features, axis=1, inplace=True)
X_test.drop(labels=low_var_features, axis=1, inplace=True)

X_train.shape, X_test.shape
```

```
((99818, 36), (11925, 36))
```

```
print("X_train", X_train.shape)
print("Y_train", y_train.shape)
```

```
X_train (99818, 36)
Y_train (99818,)
```

Fig 5.6 *Dropping quasi-constant features - Python Code*

Last but not least, the problem of identifying duplicate features has been extensively studied in the database field in the context of data integration. In this sense, identifying and fusing duplicate features could potentially bring gains in terms of prediction quality and computational costs by reducing data sparsity and dimensionality.

```
# Iterate over columns
duplicated_feat = []
for i in range(len(X_train.columns)):
    col_i = X_train.columns[i]

    # List of duplicated values
    duplicated_i = [col_j for col_j in X_train.columns[i+1:] if X_train[col_i].equals(X_train[col_j])]

    # Print duplicated
    if len(duplicated_i):
        print(f"{col_i}:{duplicated_i}")

    # Save final list
    duplicated_feat.extend(duplicated_i)
```

```
# List of duplicated values:
dup_feat_unique = set(duplicated_feat)
print(len(dup_feat_unique))
```

```
0
```

Fig 5.7 *Dropping duplicated features - Python Code*

CHAPTER 6

MODEL DEVELOPMENT

As mentioned previously, within the artificial intelligence agenda, machine learning has become a mature field over the last several decades. Most machine learning algorithms are designed to minimize a loss function or maximize the probability of the observations in the training set. Depending on different scenarios, there are certain costs associated with the model development process, such as: cost of acquiring data, cost of cleaning and feature engineering, computation cost related to the time it takes to train a model or cost of incorrect classification.

For a machine learning model to work well in practice, it is important to optimize the model fitting performance and, at the same time, minimize the cost associated with the learning process. In other words, this means the machine learning algorithm which is being developed requires to be sensitive to the cost it is dealing with.[22]

The aim of developing a machine learning model for this business case is to predict whether a customer will churn or not, thus it can reduce the misclassification costs of classifying a non-churner as a churner (Type I error) or a churner as a non-churner (Type II error). In other words, when an incentive which can be sent to a real churner is sent to a non-churner, it means either that a real churner has not received an incentive to be persuaded to stay (increasing in their risk of churn) and a non-churner has received an incentive which they were not supposed to receive (loss of marketing budget for Alpha company).

[22] Balaji Krishnapuram, Shipeng Yu, Bharat Rao. *Cost-Sensitive Machine Learning* (2011)

Before going into details of the first model, it is necessary to present the following table, which will be useful to understand the business dynamic.

	Benefits granted	**No Benefits granted**
Churn	The customer will churn. Benefits were granted to retain the user.	The customer will churn. No marketing benefits were granted. The user will not be retained.
No Churn	The customer will not churn. Benefits were granted, which should have been granted to churn users (cost).	The customer will not churn. No benefits were granted. No cost associated in this case.

Table 6.1 *Model developing and business dynamic table*

6.1 TEST & MODEL VALIDATION

As the name suggests, **model validation** is the procedure for ensuring that the trained model produces adequate results for their data, in accordance with both quantitative and qualitative goals. In preparation for usage, model validation assures the efficacy and correctness of a model. The model might perform badly if it is not validated and, in case it has not been adequately verified, then it may not be able to adjust to new circumstances, and it can be overfitted to accept and utilize new inputs appropriately.[23] This process is carried out after the model training and the aim is to test

[23] Neoklis Polyzotis, Martin Zinkevic, Eric Breck, Steven Whang. *Data validation for Machine Learning (2019)*

the speculation capacity of a developed model. For that reason, picking the correct validation method is likewise critical to guarantee exactness and biasness.

In this particular case, **K-fold Cross Validation** has been applied to carry out the validation process. This allows evaluating the model by training numerous models on subsets of the available input data and evaluating them on the matching subset of the data. Data is split into a training set that calibrates the statistical model, and an independent test set that is used to estimate the model's predictive performance.[24] This is repeated many times with different splits of the sample data into K-Folds (5 folds), eliminating the possible selection biases when contrasting two models where one may be overfitting.

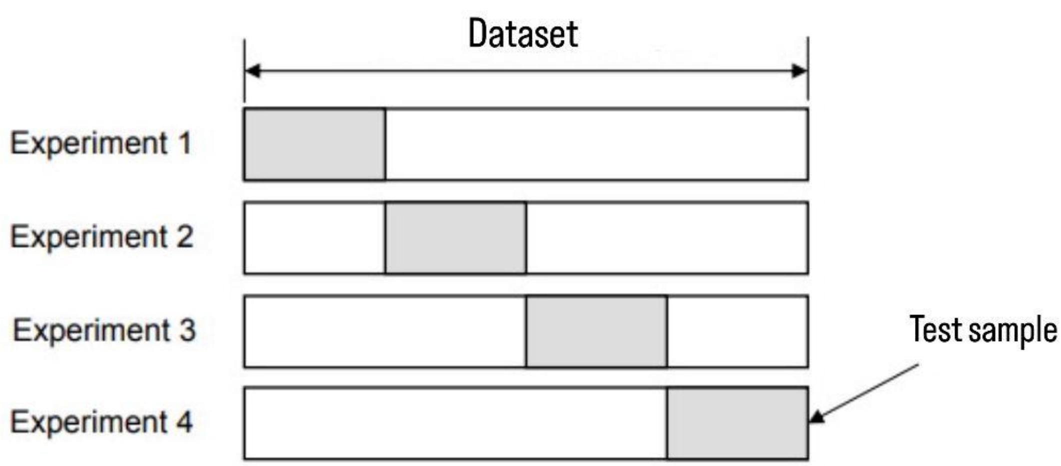


Fig 6.1 *K-Fold Cross Validation*

On the other hand, as mentioned previously, the **train-test split** is a technique for evaluating the performance of a machine learning algorithm. The procedure involves

[24] Thomas Keevers. *Cross-validation for model validation (2019)*

taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the **training dataset**. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This last one is referred to as the **test dataset**.

For this particular case, the split percentages were established by **Train: 80%** and **Test 20%** making sure the share (%) meets the project objectives with considerations that include:

- Computational cost in training the model.
- Computational cost in evaluating the model.
- Training dataset representativeness.
- Testing dataset representativeness.

6.2 RANDOM FOREST

6.2.1 DEFINITION

Random forests (RF) are becoming increasingly popular in many scientific fields because they can cope with complex interactions and even highly correlated predictor variables. In random forests and the related method of bagging, an ensemble of classification trees is created by means of drawing several bootstrap samples or subsamples from the original training data and fitting a single classification tree to each sample. Due to the random variation in the samples and the instability of the single classification trees, the ensemble will consist of a diverse set of trees.[25]

Besides, in random forest models, another source of diversity is introduced when the set of predictor variables to select from is randomly restricted in each split, producing even more diverse trees. In addition to the smoothing of hard decision boundaries, the random selection of splitting variables in random forests allows predictor variables that were otherwise outplayed by their competitors to enter the ensemble.

In other words, the RF is an *ensemble learning technique* consisting of the aggregation of a large number of decision trees, resulting in a reduction of variance compared to the single decision trees. RF is usually considered a black-box algorithm, as gaining insight on its prediction rule is hard due to the large number of trees. One of the most common approaches to extract from the random forest interpretable information on the contribution of different variables consists in the computation of the so-called variable importance measures outlined in the *Variable importance measures* section.[26]

6.2.2 FEATURE IMPORTANCE

Determining feature importance is one of the key steps of the machine learning model development pipeline. The outcome of the feature importance stage is a set of features along with the measure of their importance. Once the importance of features is

[25] Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, Achim Zeileis. *Conditional variable importance for random forests* (2008)
[26] Raphael Couronné, Philipp Probst, Anne-Laure Boulesteix. *Random forest versus logistic regression* (2018)

determined, the features can be selected appropriately. Feature importance scores play an important role in a predictive modeling project: provides insight into the data and the model, as well as defining the basis for dimensionality reduction and feature selection that can improve the efficiency and effectiveness of a predictive model on the problem.

Feature importance scores can provide **insight into the dataset**. The relative scores can highlight which features may be most relevant to the target, and conversely, which features are the least relevant. This may be interpreted by a domain expert and could be used as the basis for gathering more or different data.

Feature importance scores can provide **insight into the model**. Most importance scores are calculated by a predictive model that has been fitted on the dataset. Inspecting the importance score provides insight into that specific model and which features are the most important and least important to the model when making a prediction. This is a type of model interpretation that can be performed for those models that support it.

Feature importance can be used to **improve a predictive model**. This can be achieved by using the importance scores to select those features to delete (lowest scores) or those features to keep (highest scores). This is a type of feature selection and can simplify the problem that is being modeled, speed up the modeling process (deleting features is called dimensionality reduction), and in some cases, improve the performance of the model.

In this way, a function was developed to obtain the most important features and those that show an Area under the Curve (AUC) greater than (> 0.50) are plotted.

```python
# Decision Tree Classifier
roc_values = []
for feature in X_train.columns:
    clf = DecisionTreeClassifier()
    clf.fit(X_train[feature].to_frame(), y_train)
    y_scored = clf.predict_proba(X_test[feature].to_frame())[:, 1]
    roc_values.append(roc_auc_score(y_test, y_scored))
```

```python
# Feature Importance Plot with ROC-AUC > 0.5
roc_values = pd.Series(roc_values)
roc_values.index = X_train.columns
roc_values.sort_values(ascending=False).plot.bar(figsize=(22, 6))
selected_feat = roc_values[roc_values>0.51]
len(selected_feat), X_train.shape[1]
```

```
(22, 32)
```

Fig 6.2 A *Feature importance - Python Code*

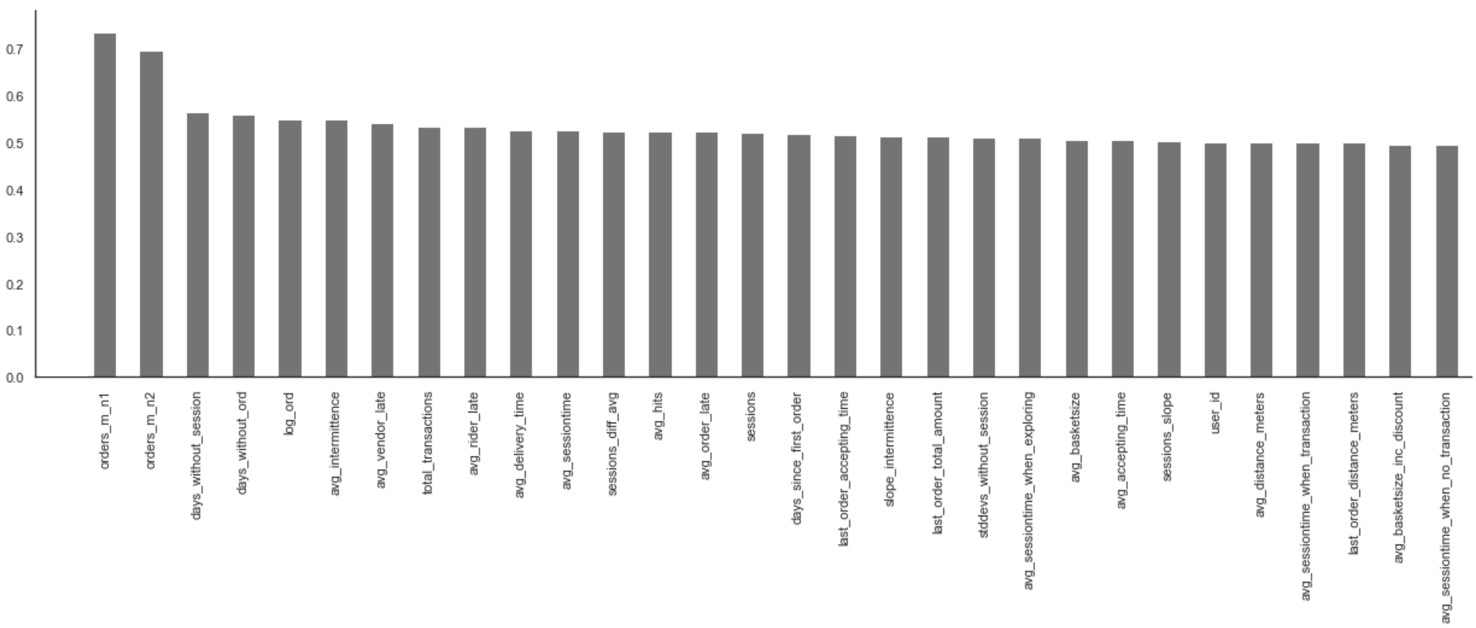


Fig 6.2 B *Feature importance output*

From the feature importance distribution graph, it can be observed that most important variables are linked to the **purchase behavior of previous months** (one and two months), as well as other aspects related to the number of **days without sessions** in the app/website and the **average duration** of those sessions, the **number of logistics**

orders (delivered by Alpha's own logistic), as well as those **logistics KPIs**: such as delay of accepting restaurants, delay in the delivery shipment or the percentage with an actual delivery time greater than that the agreed one.

6.2.3 CLASSIFICATION METRIC

AUC stands for **Area Under the Curve**, so to speak about ROC AUC score it is ROC curve first. The Receiver Operating Characteristic curve, mostly known as ROC Curve, is a chart that visualizes the tradeoff between true positive rate (TPR) and false positive rate (FPR). Therefore, for every threshold, TPR and FPR are calculated and plotted on one chart. Thus, the higher TPR and the lower FPR is for each threshold the better and so classifiers that have curves that are more top-left-side are better.

Several points in ROC space are important to note. The lower left point (0, 0) represents the strategy of never issuing a positive classification; such a classifier commits no false positive errors but also gains no true positives. The opposite strategy of unconditionally issuing positive classifications is represented by the upper right point (1, 1). The point (0, 1) represents perfect classification. Further developed models will be compared based on the metrics mentioned above. [27]

6.2.4 HYPERPARAMETERS TUNING

This section presents the most important parameters for Random Forest and their common default values as implemented in Python considered in this thesis. The best way to think about hyperparameters is like the settings of an algorithm that can be

[27] Tom Fawcett. *ROC Graphs: Notes and Practical Considerations for Researchers* (2004)

adjusted to optimize performance; while **model parameters** are learned during training — such as the slope and intercept in a linear regression — **hyperparameters** must be set by data scientists before training.

Hyperparameter tuning relies more on experimental results than theory, and thus the best method to determine the optimal settings is to try many combinations to evaluate the performance of each model. However, evaluating each model only on the training set can lead to one of the most fundamental problems in machine learning: **overfitting**. If the model is optimized for the training data, then it will score very well on the training set, but will not be able to generalize to new data, such as in a test set. When a model performs highly on the training set but poorly on the test set, this is known as overfitting, or essentially creating a model that knows the training set very well but cannot be applied to new problems.

6.2.5 RF EXPERIMENTS IMPLEMENTED

Once the code function of the model was defined, as well as the hyperparameters, we proceeded to run tests implementing the different steps regarding Feature Engineering and Feature Selection which were mentioned previously. In this sense, it will be possible to observe how the model performance has improved as these methods were applied.

At first, two code functions were defined to compare the performance: On one side a Random Forest function was developed, whereas on the other side a Linear Regression function was developed as well. However, the regression output turned out to be too low compared to the Random Forest performance.

```python
# Score calculation Random Forest
def run_randomForests(X_train, X_test, y_train, y_test):
    rf = RandomForestClassifier(n_estimators=200, max_depth=5, random_state=0)
    rf.fit(X_train, y_train)
    print('Train set')
    y_pred = rf.predict_proba(X_train)[:,1]
    print(f"RandomForest ROC-AUC {roc_auc_score(y_train, y_pred)}")
    print('Test set')
    y_pred = rf.predict_proba(X_test)[:,1]
    print(f"RandomForest ROC-AUC {roc_auc_score(y_test, y_pred)}")
```

```python
# Function for Logistic Regression
def run_logistic(X_train, X_test, y_train, y_test):
    logit = LogisticRegression(random_state=0)
    logit.fit(X_train, y_train)
    print('Train set')
    y_pred = logit.predict_proba(X_train)[:,1]
    print(f"Logistic Regresion ROC-AUC: {roc_auc_score(y_train, y_pred)}")
    print('Test set')
    y_pred = logit.predict_proba(X_test)[:,1]
    print(f"Logistic Regresion ROC-AUC: {roc_auc_score(y_test, y_pred)}")
```

Fig 6.3 *RF and Logistic Regression functions*

Subsequently, we proceeded to run a Random Forest model on the original dataset, then the dataset with basic filtering and finally eliminating the highly correlated features plus embedded methods. In this way, an increase in the performance of the test set was observed as different filtering techniques were applied.

- Initial original dataset: **ROC-AUC** of **0.8567.**
- Applying basic filtering: **ROC-AUC 0.8569**.
- Eliminating highly correlated features: **ROC-AUC 0.8598**
- Finally, applying embedded methods: **ROC-AUC 0.8623**

```python
In [112]: # original
run_randomForests(X_train_original.drop(labels=['user_id'], axis=1),
                  X_test_original.drop(labels=['user_id'], axis=1),
                  y_train, y_test)

Train set
RandomForest ROC-AUC 0.8934623053796704
Test set
RandomForest ROC-AUC 0.8567660454569804
```

Fig 6.4 *Original dataset RF output - Python Code*

```
# original
run_randomForests(X_train_original.drop(labels=['user_id'], axis=1),
                  X_test_original.drop(labels=['user_id'], axis=1),
                  y_train, y_test)
```

```
Train set
RandomForest ROC-AUC 0.8934623053796704
Test set
RandomForest ROC-AUC 0.8567660454569804
```

```
# filter methods - basic
run_randomForests(X_train_basic.drop(labels=['user_id'], axis=1),
                  X_test_basic.drop(labels=['user_id'], axis=1),
                  y_train, y_test)
```

```
Train set
RandomForest ROC-AUC 0.8796230125643274
Test set
RandomForest ROC-AUC 0.8569273455674988
```

```
# filter methods - correlacion
run_randomForests(X_train_corr.drop(labels=['user_id'], axis=1),
                  X_test_corr.drop(labels=['user_id'], axis=1),
                  y_train, y_test)
```

```
Train set
RandomForest ROC-AUC 0.8813854381010415
Test set
RandomForest ROC-AUC 0.859865784866373
```

```
# embedded methods - Random forests
run_randomForests(X_train_rf,
                  X_test_rf,
                  y_train, y_test)
```

```
Train set
RandomForest ROC-AUC 0.8745119251928176
Test set
RandomForest ROC-AUC 0.8623018538838437
```

Fig 6.5 *Implemented RF output - Python Code*

.

6.3 XGBOOST

6.3.1 DEFINITION

XGBoost is an algorithm that has recently been dominating applied Machine Learning for structured or tabular data. It is an implementation of gradient-boosted decision trees designed for speed and performance. It stands for eXtreme Gradient

Boosting, being an implementation of gradient boosting machines. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, as in this particular case, decision-tree based algorithms overperform most of the times. [28]

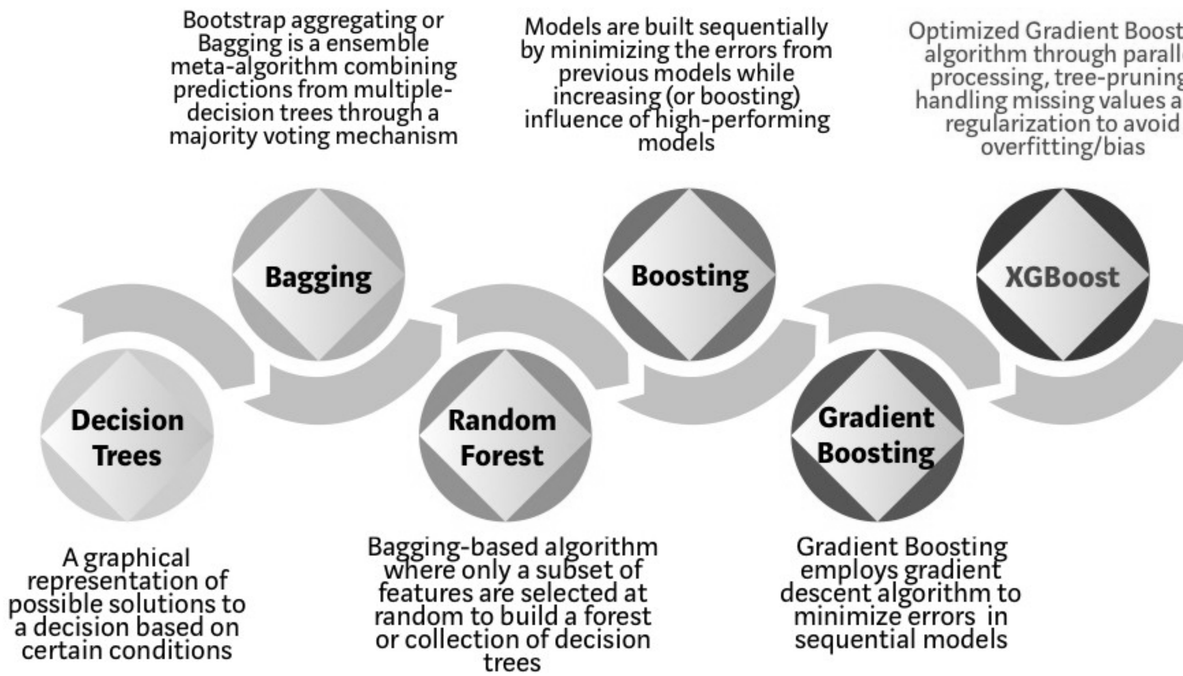


Fig 6.6 *Evolution of XGBoost Algorithm from Decision Trees*

6.3.2 HOW GRADIENT BOOSTING WORKS

Gradient Boosting involves three elements:

- A loss function to be optimized.
- A weak learner to make predictions.

[28] Jason Browniee. *XGBoost with Python: Gradient Boosted Trees with XGBoost and scikit-le* (2019)

- An additive model to add weak learners to minimize the loss function.

6.3.2.1 LOSS FUNCTION

The loss function used depends on the type of problem being solved. It must be differentiable, but many standard loss functions are supported. For example, regression may use a squared error and classification may use logarithmic loss. A benefit of the gradient boosting framework is that a new boosting algorithm does not have to be derived for each loss function that may want to be used, instead, it is a generic enough framework that any differentiable loss function can be used.

6.3.2.2 WEAK LEARNER

Decision trees are used as the weak learner in gradient boosting. Specifically, regression trees are used that output real values for splits and whose output can be added together, allowing subsequent model outputs to be added and correct the residuals in the predictions. Trees are constructed in a greedy manner, choosing the best split points based on purity scores like Gini or to minimize the loss.

6.3.2.3 ADDITIVE MODEL

Decision trees are added one at a time, and existing trees in the model are not changed. A gradient descent procedure is used to minimize a set of parameters, such as the coefficients in a regression equation or weights in a neural network. After calculating error or loss, the weights are updated to minimize that error.

After calculating the loss, to perform the gradient descent procedure, a tree must be added to the model in order to reduce the loss. Generally, this approach is called functional gradient descent or gradient descent with functions. The output of the new tree is then added to the output of the existing sequence of trees in an effort to correct or improve the final output of the model. A fixed number of trees are added or training stops once loss reaches an acceptable level or no longer improves on an external validation dataset.

6.3.3 WHY XGBOOST FOR THIS PROBLEM?

This current business case is actually related to tabular data, in which tree learners excel while the "*most attractive*" deep learning models tend to underperform.

XGBoost is the first and most known model of the most recent generation of tree booster ensembles, which gave superior speed and accuracy over older models such as AdaBoost. This means that while it is not automatically the best choice for tabular data, XGBoost makes sense as a primary attempt, as if it fails to give good performance it is a likely indicator that boosting is not an optimal approach to the problem.

Besides, more generally, some other reasons why it was decided to develop a XGBoost model are listed below:

- It is **relatively fast**. Modern libraries like XGBoost come equipped with several speed enhancements, making it possible to train a well-performing model in a short amount of time.

- XGBoost and similar libraries can also be trained on GPUs (some assembly required), making **training** even **faster** on larger datasets.

- It is an **ensemble learning algorithm**, which combines the predictions of multiple base learners (usually, each one being a fairly weak performer on its own) to generate one overall prediction for each input. This allows it to learn more complex relationships between the features and targets in the training set.

- Based on what was mentioned before, it comes equipped with **several performance tuning hyper-parameters** (some vary by library), making it a highly versatile learner. This can be a bit overwhelming at first, and it makes the search for a good set of parameters more difficult. But often, one can achieve good performance by tuning just the learning rate, number and depth of decision trees, stopping early while leaving other parameters at their default settings.

- It is **easy to implement**, since it typically performs quite well even with minimal tuning.

Finally, Deep Learning (DL) gets a lot of the attention these days, and understandably so, given all the recent advances in its applications to computer vision, natural language processing, and transfer learning. However, DL models also require a lot of overhead in terms of setup, training, and deployment. And sometimes, it is simply just not the best tool for the job based on how the data is structured, similarly to what happens with Alpha Corporation data. For this reason, we choose to develop a Gradient Boosting model for this thesis.

6.3.4 HYPERPARAMETERS TUNING

As previously mentioned, hyperparameter tuning is choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a model argument whose value is set before the learning process begins.

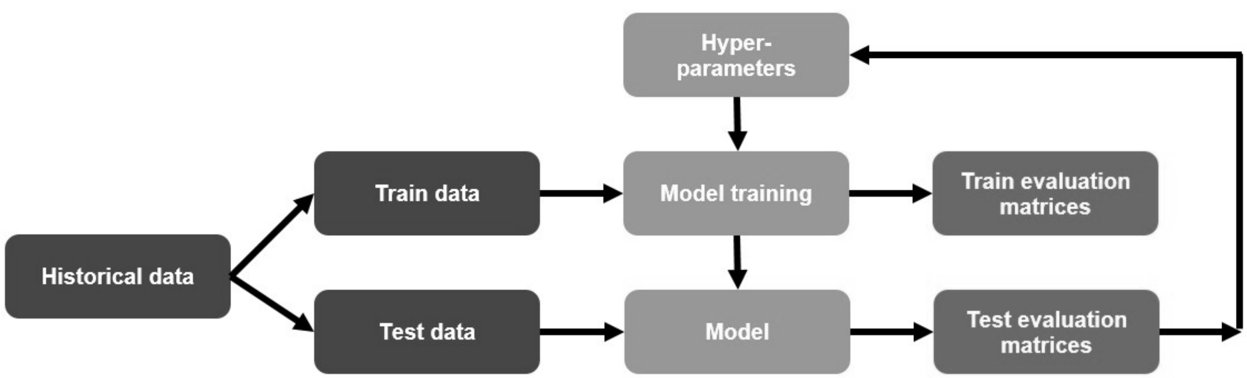


Fig 6.7 *Tuning hyperparameters and ML model development*

Below, a couple of hyperparameters will be mentioned. All of them were tuned while applying a *grid search technique*.

6.3.4.1 GRID SEARCH

When a machine learning model is trained, it is usually unclear which hyperparameters lead to good results. While there are estimates and rules of thumb, there is often no way to avoid trying out hyperparameters in experiments. However, machine learning models often have several hyperparameters that affect the model's performance in a nonlinear way.

Based on how the data is structured, grid search has been used to automate the process of searching for optimal model hyperparameters. The grid search algorithm exhaustively generates models from parameter permutations of a grid of parameter values.

The concept behind the grid search technique is quite simple and straightforward: it is an exhaustive technique that tests all permutations of a parameter grid. The number of model variants results from the parameter grid and the specified parameters. The grid search algorithm requires the following information to be provided:

- The hyperparameters that we want to configure (e.g. tree depth)
- For each hyperparameter a range of values (e.g., [50, 100, 150])
- A performance metric so that the algorithm knows how to measure performance (e.g., accuracy for a classification model)

A sample parameter grid is shown below. In this case a specific range was set up [*16, 32, and 64]* for *n_estimators* and a range of *[8, 16, and 32]* for *max_depth*. Then, the search grid will test a total of 9 different parameter configurations.

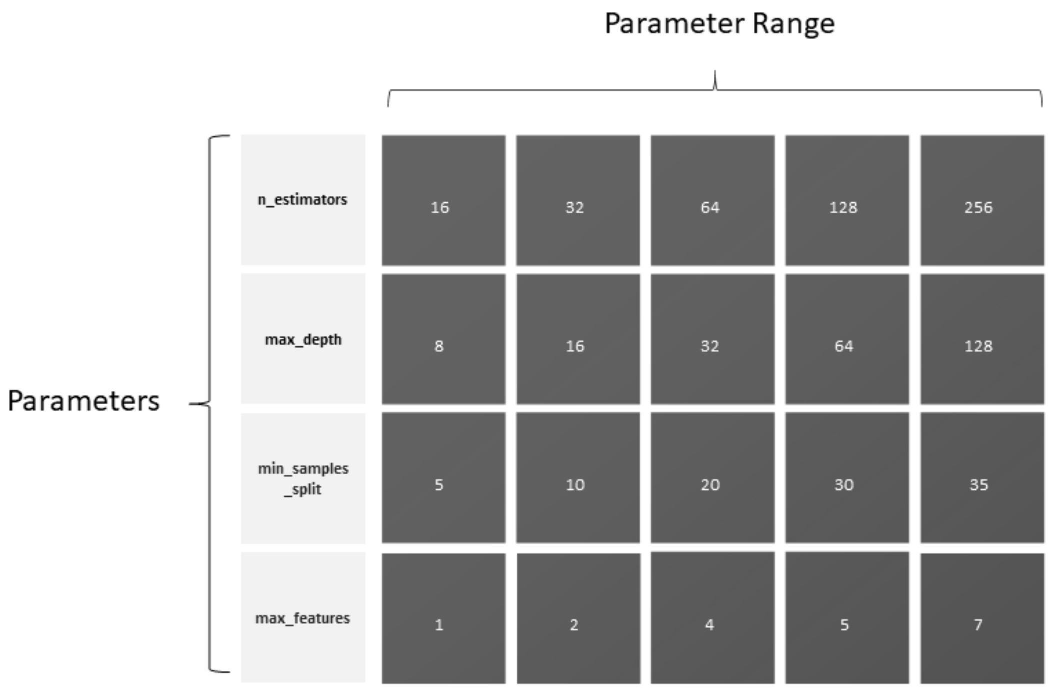


Fig 6.8 *Exemplary parameter grid for the tuning of a XGBoost model with four hyperparameters*

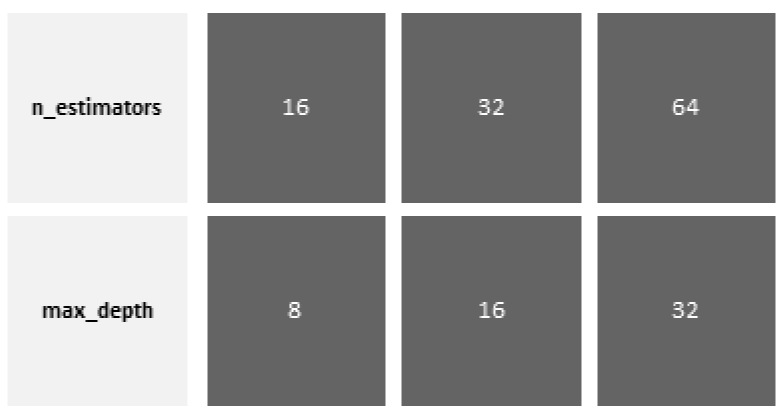


Fig 6.9 *A parameter grid with two hyperparameters, 9 different parameter configurations tested.*

The advantage of the grid search is that the algorithm automatically identifies the optimal parameter configuration from the parameter grid. However, the number of possible configurations increases exponentially with the number of values in the parameter grid. So in practice, it is essential to define a sparse parameter grid or run the algorithm several times with different parameter ranges.

6.3.4.2 MAX DEPTH

The *max_depth* hyperparameter is defined as the longest path between the root node and the leaf node. Using this parameter, it is possible to limit up to what depth is requested for every tree in the random forest to grow.

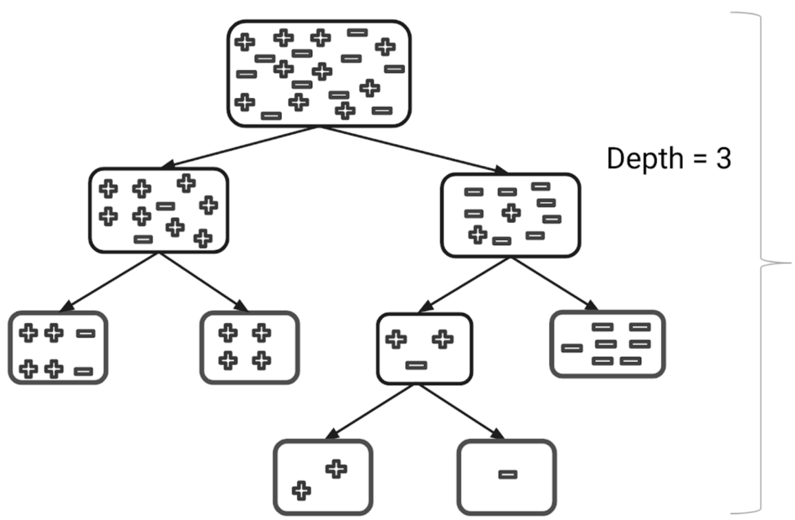


Fig 6.10 *Tree depth: Graphic definition*

On the one hand, while the max depth of the decision tree increases, the performance of the model over the training set increases continuously. On the other hand, as the *max_depth* value increases, the performance over the test set increases initially but after a certain point, it starts to decrease rapidly. In other words, the tree starts to overfit the training set and therefore is not able to generalize over the unseen points in the test set.

6.3.4.3 LEARNING RATE

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is challenging, as a value too small may result in a long training process that could get stuck, and a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process.

The learning rate may be the most important hyperparameter when training a Random Forest model. Therefore, it is vital to know how to investigate the effects of the learning rate on model performance and to build an intuition about the dynamics of the learning rate on model behavior. As mentioned previously, if the learning rate is set too low, training will progress very slowly as you are making very tiny updates to the weights in your network. However, if your learning rate is set too high, it can cause undesirable divergent behavior in your loss function.

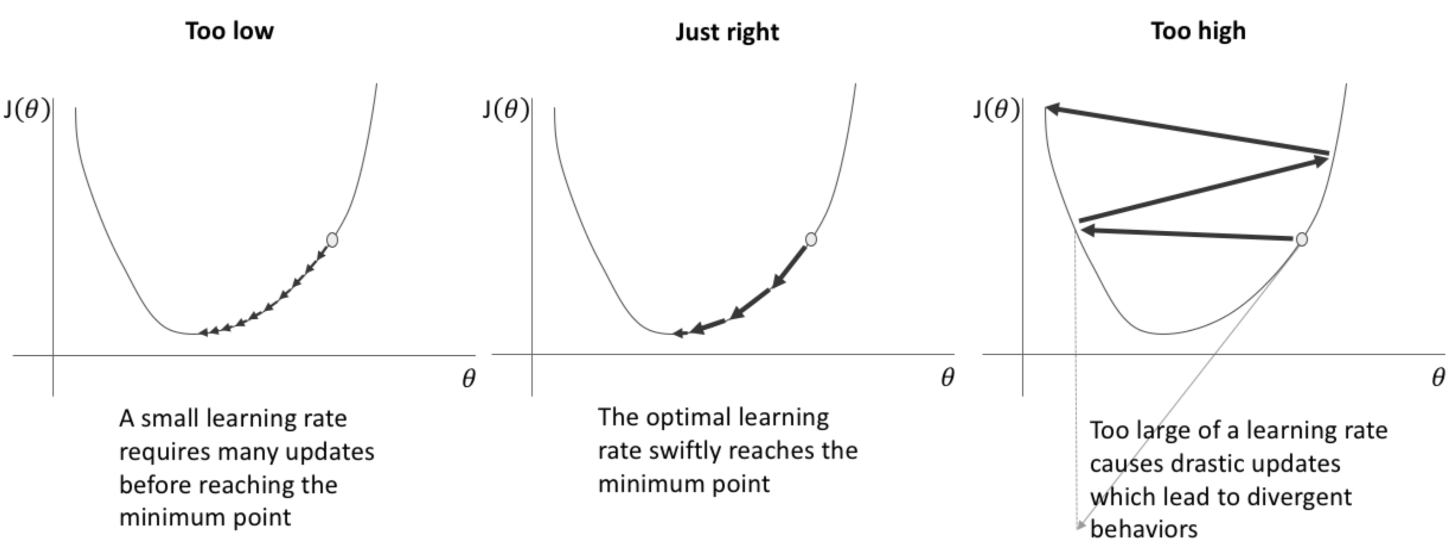


__Fig 6.11__ Learning Rate: Graphic definition

6.3.4.4 MINIMUM CHILD WEIGHT

The minimum weight or number of samples (if all samples have a weight of 1) required in order to create a new node in the tree. A smaller *min_child_weight* allows the algorithm to create children that correspond to fewer samples, thus allowing for more complex trees, but again, more likely to overfit.

Thus, this parameter and maximum depth can be used to control the complexity of the trees. It is important to tune them together in order to find a good trade-off between model bias and variance.

6.3.4.5 GAMMA

Gamma is a regularization parameter. In contrast with *min_child_weight* and *max_depth* that regularizes usage *within tree* information.

The range of that parameter is [0, Infinite]. It represents how much the loss has to be reduced when considering a split, in order for that split to happen. To summarize: the higher Gamma is, the higher the regularization. Default value is 0 (no regularization). Gamma values around 20 are extremely high, and should be used only when using high depth.

6.3.4.6 HYPERPARAMETERS IMPLEMENTATION

Once the hyperparameters have been explained, we proceed to develop a function to find those that approach the best performance.

```python
from sklearn.model_selection import RandomizedSearchCV, GridSearchCV

clasificador = xgb.XGBClassifier()
```

```python
# Hyperparameter Optimization

params = {
    'learning_rate'      : [0.03, 0.05, 0.07, 0.10, 0.15],
    'max_depth'          : [2, 3, 4],
    'min_child_weight'   : [1, 3, 5, 7, 10],
    'gamma'              : [0.0, 0.1, 0.2, 0.3, 0.4],
    'colsample_bytree'   : [0.3, 0.4, 0.5]

}
```

Fig 6.12 Hyperparameter optimization - Python Code

6.3.5 XGBOOST RESULTS

Having implemented their respective functions, the XGBOOST model was run using the previously mentioned tuned hyperparameters, obtaining a performance of **0,9032** in training while **0,8595** in testing, which is a bit lower than Random Forest - embedded methods.

```python
### Fitting best model
xgb_train = xgb.DMatrix(X_train, label=y_train)
clf = xgb.XGBClassifier(
            max_depth = 4,
            n_estimators = 300,
            subsample = 0.4,
            learning_rate = 0.03,
            nthread = 4,
            colsample_bytree = 0.5,
            colsample_bylevel = 0.5,
            min_child_weight = 3,
            seed = 0)

## Cross Validation running
cv = xgb.cv(clf.get_xgb_params(), xgb_train,
        num_boost_round=600,
        early_stopping_rounds=50,
        nfold=5,
        metrics=['auc'],
        seed=0)

# Best parameters
clf.set_params(n_estimators=cv.shape[0])
clf.fit(X_train, y_train, eval_metric='auc')

# AUC for training and testing set - Running
auc_train = roc_auc_score(y_train, clf.predict_proba(X_train)[:,1])
y_proba_xgb = clf.predict_proba(X_test)[:,1]
auc_test = roc_auc_score(y_test, y_proba_xgb)
print(f"TRAIN: {auc_train}, TEST: {auc_test}")
```

```
TRAIN: 0.9032250883775361, TEST: 0.8595043668644503
```

Fig 6.13 XGBoost output - Python Code

6.4 COMPARED RESULTS & INSIGHTS

As mentioned previously, the area under the curve is a good way to estimate the accuracy of the model. An excellent model possesses an AUC near to the 1 which indicates that it has a good measure of separability whereas an AUC is 0.5, which means the model has no class separation capacity present whatsoever.

Considering that AUC-ROC was the metric applied for this business case, the performance of the XGBoost and Random Forest models can be observed through the following plot (having applied data cleaning and data engineering). Although both models show favorable results in detecting users who will not place orders, the first model developed by Random Forest has shown slightly better performance than XGBoost.

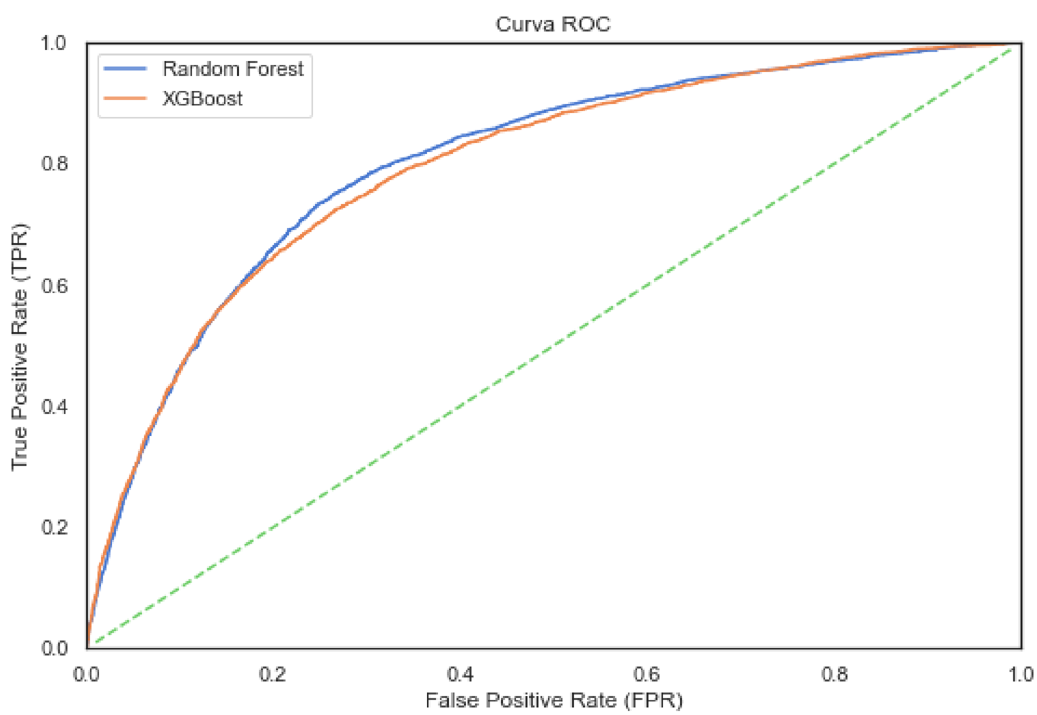


Fig 6.14 *ROC Results - Comparing RF and XGBoost output*

RANDOM FOREST PERFORMANCE:	ROC-AUC Training **0.8745**
(Embedded Methods)	ROC-AUC Testing **0.8623**
XGBOOST PERFORMANCE:	ROC-AUC Training **0.9032**
	ROC-AUC Testing **0,8595**

Customers more likely to churn: Based on the analysis carried out previously, as well as the results obtained from the predictive model, it can be said that the behavior of customers during previous months are great indicators if the user will leave or not. In this way if, for example, the model runs in May to predict whether the customer will churn in July or not, the behavior based on the number of orders in the month of May, April and March is a strong churn indicator. This explains the high correlation between the previous-months-behavior variables and the Churn, based on the graph of importance of variables presented above. For this reason, customer loyalty to make monthly orders plays a relevant role and will be analyzed in depth in the following section.

In addition, improving Alpha's own logistics offer would also have a positive impact: as it can be seen in results presented in *Chapter 4, p. 49-50*, where customers value having the chance to track their orders in real time as well as less delivery time, coverage in case of erroneous products, etc. It can be seen that, in the case of predictive churn customers, there is a significant correlation with the logistics of the orders.

On the other hand, Alpha Corporation is capable of tracking their customers based on their session tagging. Sessions can be analyzed in a way that reveals how users truly interact with the company, allowing them to identify opportunities and problems. In this case, the number of sessions is also one of the outcomes from this model: higher number of churn users present the lowest number of sessions during the last couple of months.

Likewise, those customers who do not place orders in other verticals rather than restaurants present a higher risk of churning. It can be seen that diversification in purchases: restaurants, coffee, breakfast/snacks, pharmacies, markets, pets, beverages, etc. has a lower probability of churn.

In summary, training the model based on historical data allows it to learn and detect certain patterns,which improve the predictive outcome: when AUC is approximately 0.5, the model has no discrimination capacity to distinguish between positive class and negative class, however, in this case, the model reached a testing performance level of 0,8623 which means there is a 86,23% chance that the model will be able to distinguish between positive class (churn) and negative class (no churn). Based on what was mentioned above, the behaviors that churn users generally display are summarized in the table below. This behavior analysis can also be carried out in depth with other models that are presented in the following sections as future research directions.

SOME COMMON CHURN BEHAVIORS
Previous months' behavior: *Have customers stopped placing orders or the number of orders has been decreasing in the last couple of months?* Based on results presented in *Chapter 4 and 6,* there is a high correlation between previous months behavior and the churn rate, thus customer loyalty plays an important role in the analysis. The decrease in orders during the last couple of months is related to a higher churn rate, this is the reason why the feature importance distribution graph presented in *Chapter 6* shows that one of the most important variables is linked to the purchase behavior of previous months.
Sessions history: *Have customers been using the app lately? If so, how many times? Has the number of sessions been improving in the last couple of months? How has the CVR (conversion rate) been performing lately?* The behavior of user sessions is also an indicator of relevant magnitude within the churn predictive model. By reducing the number of sessions, or presenting drops in the CVR index, churn rate tends to be higher. Based on the feature importance graph presented in *Chapter 6*, as well as the distribution graphs presented in *Chapter 4*, features like *days_without_sessions* in the app/website and the *average_duration* of those sessions play a relevant role in the prediction model development and understanding the behavior of those churn users.
Own Logistics Orders: *Considering customers' orders, what percentage were delivered by Alpha's own logistics? Has the percentage of logistics orders been increasing lately?* Users who have placed orders without Alpha's own logistics present a higher probability of churn, which is also a behavior to consider in this section (*p. 49-50, Chapter 4*).
Single vertical customers: *Do customers tend to place orders from only one vertical (e.g. placing orders in restaurants only)? Do they diversify their behavior afterwards by trying and ordering from new verticals?* Based on outcomes, single-vertical-customers are more likely to churn than those who place orders in multiple verticals such as pharmacies, markets, coffee shops, pets, etc.

Table 6.2 *Common churn behaviors summary*

CHAPTER 7

MARKETING IMPLICATIONS

7.1 RETENTION MARKETING APPROACH

Retention marketing is a form of marketing that aims to maximize the value of customers the company already has. It is different from the traditional concept of *acquisition-focused* marketing as the main aim is not to acquire new users, but to encourage repeat purchases or placed orders from already existing customers.[29] For Alpha's particular case, a B2C e-commerce company in the Online Food Delivery Industry, this means encouraging customers to return to their website or app and keep on placing repeated, high-value orders, with no intention of migrating to their competitors.

Thus, customer retention with regard to customer satisfaction, is becoming increasingly significant, because nowadays, customers have the opportunity to choose from such a large range of products and services, and because the competition within various markets is becoming ever fiercer. Following the adage that one has to invest six times as much in a new customer as in an already existing one, companies are discovering new potential in the building up and maintenance of a stable customer base. This is the source of commercial success and long-term growth.

---

[29] Rizal Ahmad, Francis Buttle. *Customer Retention: a potentially potent marketing management strategy* (2010)

From the point of view of the customer, two possibilities need to be differentiated with regard to customer satisfaction and their retention. These possibilities are, on one hand, **being loyal**, and on the other hand, **being constrained**.

Loyalty is present in voluntary commitment. This is the case when customers remain with a company even when they have the opportunity to change at any time. Rational, economic, and emotional reasons can be cited. In the area of emotional reasons, customer satisfaction and retention is of particular significance.[30] It must be given the highest priority, because customers will only remain loyal under Alpha's existing conditions, if they are so satisfied with the company's service that they do not want to change.

On the other hand, **a constrained situation** is where customers have no factual or legal possibility to change their provider, or if this can only be done with great difficulty or at great expense.[31] They are involuntarily bound to a company. However, this is not the particular case of Alpha Corporation as they do have pretty strong competitors within the market and switching between them represents no cost at all – just downloading a new app and making sure they know how to use it.

---

[30] Gerhard Raab, Riad A. Ajami, G. Jason Goddard. *Customer Relationship Management: A Global Perspective* (2016)
[31] Gerhard Raab, Riad A. Ajami, G. Jason Goddard. *Customer Relationship Management: A Global Perspective* (2016)

7.1.1 WHY IS IT IMPORTANT?

- **Builds Customer Loyalty**: Based on a previous study carried on internally in the company, loyal customers place orders more regularly and in greater quantities (higher revenues). They know the business can provide the delivery service they want. As a valued customer, they feel the company is more likely to listen to their requests and provide superior service than other competitors. Retained customers are more likely to place additional orders if we pay attention to what verticals (type of products) they are more interested in. Certain strategies may consist of providing discounts for loyal customers to encourage them to reorder, such as a 10 percent discount after a certain number of orders or free delivery over a certain basket/ticket amount.

- **Reduce Marketing Costs:** Customers are familiar with the delivery services, which reduces the marketing expenses, such as creating brand awareness and advertising (e.g. riders wearing an Alpha's jacket or facemask when delivering).

- **Word of mouth advertisement and referral program:** If you retain valuable customers, it is more likely that they will tell their close ones about the business. Based on several papers that analyze brand evangelism, such as the one written by Matzler, Pichler and Hemetsberger[32], it can be said that customers feel good telling people about the quality service they receive. On the other hand, word-of-mouth advertising also gives Alpha business credibility. Customers respect the opinion of people close to them and give prospective confidence that

[32] K Matzler, EA Pichler, A Hemetsberger. *Who is spreading the word? The positive influence of extraversion on consumer passion and brand evangelism* (2007)

the business has the skills to perform their deliveries on time with no issues at all.

- **Less Price Sensitivity:** Retained customers are not as price-conscious[33]; they are aware that Alpha Corporation provides an excellent delivery service and they are willing to pay for it.

- **Provide Valuable Feedback:** Retained customers usually provide valuable feedback.[34] [35] As they place frequent orders, they are aware of areas that could be improved. This can lead to new opportunities that may have been overlooked in the past.

7.2 CHURN MANAGEMENT & MARKETING

The concept of Customer Churn is applied in the contemporary marketing field and should be highly considered by B2C e-commerce companies such as Alpha Corporation. Nowadays, due to improved access to information, customers are more transient, and it is easier and less costly for them to switch between competitors. Thus, companies are aware of this and are interested in identifying potential churners in order to attempt to prevent defection by targeting such customers with incentives.[36]

---

33 Daniel P.Hampson, ShiyangGong. *How consumer confidence affects price conscious behavior: The roles of financial vulnerability and locus of control* (2021)
34 John Goodman, Steve Newman. *Understand Customer Behavior and complaints* (2003)
35 Young Namkung, Soo Cheong (Shawn) Jang, Soo KeunChoi. Cus*tomer complaints in restaurants: Do they differ by service stages and loyalty levels?* (2011)
36 Niccolò Gordini, Valerio Veglio. *Customer Churn Prediction and Marketing Retention Strategies* (2016).

Therefore, with the purpose of retaining customers, it is crucial to build a churn prediction model that is as accurate as possible in order to minimize the customer risk of churn. As mentioned initially, customer churn indicates the propensity of customers to cease doing business with the company in a given time period, and building up from scratch a customer churn prediction model consists of, in other words, building a machine learning model that ranks the customers from most likely to leave the company (churn flag) to least likely to leave the company (no churn flag).

E-commerce context has provided many new opportunities to consumers, in this particular case, to place an order almost instantly through their phone. The rapid expansion of internet, e-commerce and social media has made the study of delivery food consumer behavior on e-commerce a fundamental research agenda. Thus, it is likely to develop marketing strategies through trust-building mechanisms and affecting customers' intention to place an online order or churn. In fact, this rapid growth raises important research questions about the levels of loyalty and churn management in the web/app environment. It reflects the compelling advantages that e-commerce and social media offer over conventional brick-and-mortar stores (calling a restaurant to place an order), including easier interconnectivity and participation on the web, easier share information, greater flexibility, enhanced market outreach, lower cost structures, faster transactions, broader product lines, greater convenience, and customization. All these advancements have developed online food delivery commerce into a vibrant and lucrative e-commerce channel, highlighting this is an important point as customer involvement through the Internet is a key factor in the development of new marketing strategies.

However, this situation also comes with its own set of challenges: the importance of retention for companies becomes even clearer in the food delivery industry, where customers place larger and more orders with higher transactional values. As a consequence of not having virtual switching costs, competitors in the world are only a few *clicks* away and consumers are able to compare and contrast competing benefits and services with minimal expenditure of personal time or effort. Therefore, the food delivery business feeds the ease of moving from one company to another.

7.2.1 CHURN MANAGEMENT GENERALIZATIONS

From the results of the present thesis, it is possible to highlight some major characteristics of churn management that can be generalized to other companies operating in Alpha's industry and to all food delivery customers:

1. The prediction of churn and non-churn is a typical binary classification issue.
2. The data problem could be balanced or imbalanced, depending on the company performance – one of the most problematic aspects is that certain customers create fake users that increases the level of churn (a situation that is specifically analyzed by the fraud department). In this particular case, even though the data is not perfectly balanced, it is quite so: range of 70% - 30% approximately.
3. Large learning applications will inevitably have some types of noise in the data, thus the task of integrating them may turn out to be quite complicated.

7.3 PROACTIVE RETENTION STRATEGIES

Therefore, marketing programs can consist of three main marketing strategies, listed below:

1. **Subscription Management Strategy** (SMS): Develop a subscription program *Alpha Club* which, based on a fixed monthly price, guarantees the customer different benefits such as free delivery, vouchers, discounts, etc. The main purpose is to provide retention incentives to renew another period for customers who are most likely to churn when their subscription period is close to expiration, and thus to have them extend their service contract.[37] [38] Some possible retention incentives within this strategy could be to lower the subscription price or to offer more benefits.
2. **Long-lasting Management Strategy** (LMS): Focuses on soliciting current long-lasting customers for premium services at marginal increases of service fees (e.g. no charging delivery fee based on distance or cash-back for certain time frames and verticals). Long-lasting, overage customers may benefit from this offer because they enjoy upgraded services at only a marginal cost increase and avoid unexpected and expensive overage charges. Possible customer retention incentives within this strategy consist of offering a better service free of charge, such as improved delivery speed or a preference line when contacting the customer help center.

---

[37] Tony Chen, Ken Fenyo, Sylvia Yang, Jessica Zhang.*Thinking inside the subscription box: New research on e-commerce consumer*s (2018)

[38] Study carried out by an outsourced consultant hired by Alpha Company to get insights in the case of developing a subscription model. Even though results are not allowed to be shared due to confidentiality, one of the outcomes confirmed this.

3. **Complaints Management Strategy** (CMS). Focuses on targeting the customers who are not satisfied with the service and make many complaint calls/chats to the service center, causing high operating costs. Therefore, the objective of CMS is to make these customers satisfied with the customer care service providing highly responsive services and financial incentives, to reduce complaint calls accordingly, and keep them ordering with Alpha's own logistics. Possible customer retention incentives within this strategy are: 1) **obtaining feedback** on customer satisfaction through NPS questionnaires, or 2) **offering a specialized employee available** for the customer to satisfy their particular needs. An innovative solution could consist in presenting the calculated churn prediction score as one of the parameters on the screen of the employee of the customer care center when a customer calls in with a complaint. This score would enable the employee to take the appropriate action, for example, determining the order of answering incoming phone calls or chats. Customers with the highest churn probability would be answered first, whilst the others would remain on hold. Again, a customer with a high churn probability (flagged as potential churner), who calls in very unsatisfied, can be offered some incentives to make up for the disservice they experienced.

Customer Name	Status	Prediction	Action
Christa Sanders	Active	Active	✓ No Action
Lucas Garcia	Active	Churned	⚠ At-Risk!
Holly Tucker	Inactive	Churned	✓ No Action
Rebecca Presland	Active	Active	✓ No Action

Fig 7.1 *Labeling sample shown to improve contact center performance (innovative proposal)*

7.3.1 EARLY STAGE RETENTION

The biggest drop in retention usually happens during the first couple of months. Based on previous analysis carried on with the company, the biggest drop mostly happens during the first and second month of usage. In the graph below, it can be observed how the number of new monthly users decreases over time (churn), and it is evident that the main drop occurs in the first and second month of acquisition. Due to confidentiality, the number of users has been rescaled, but the trend remains significant for the reader to ascertain the main concern of users churning.

September	104,351	28,226	20,869	19,960	18,292	16,974
October	109,096	29,367	24,027	21,845	20,019	
November	88,152	26,686	20,415	18,217		
December	84,881	25,481	19,671			
January	87,816	26,147				

Fig 7.2 *Customer acquisitions and drop over time (read from left to right, rescaled for confidentiality)*

Depending on the industry, the time frame might be a little different—anywhere from the first visit, to the first week, or even the first couple of months. In this sense, a cohort analysis (grouped by acquisition cohort) is necessary to be carried out in order to understand when users drop out and define the time frame that constitutes early-stage retention. But regardless of how the time frame is defined, improving early-stage retention is crucial to minimizing churn in the long run.[39]

Considering the Online Food Delivery context and how customers behave in Alpha Corporation considering internal data, short-term retention refers to the **first six months** since the first order was placed. Thus, once the customers who churn are

[39] Artun Omer, Levin Dominique, *Predictive marketing: easy ways every marketer can use customer analytics and big data* (2015)

identified with the prediction model, it is possible to track whether they are in their early stage, and if so, some strategies to reduce early-stage churn can be implemented. The following early-stage proposed solutions are generic, which means that they do not necessarily require the prediction model developed initially. These strategies will improve the retention level in the long term, regardless of whether we have or have not developed a Machine Learning Model. However, when mentioning the medium-long term strategies afterwards, the output from the predictive model developed initially will be necessary because implementing the marketing retention strategies to all customers would not be profitable (extremely high costs), thus having a churn prediction model would reduce the number of users reached and with more favorable results.

1. **Improve the user onboarding experience:** Reducing early-stage churn requires shortening time to value and getting more customers to experience the service core value as early as possible. This means improving the user onboarding experience by identifying a single workflow that demonstrates the service core value, and then design the onboarding to guide users through it.[40] The main goal is to keep the ordering process simple — do not overwhelm users with unnecessary details. Using tooltips and hotspots to guide users towards key actions within Alpha's app. Optimize the user experience (UX) to reiterate Alpha service value and benefits, because if users do not find value with the delivery service quickly and efficiently, there is a very good chance that they will abandon the service in favor of another competitor. In this way, based on the outcome of the predictive model, and considering that customers who do not diversify

[40] Katryna Baiboni. *Proactive strategies for reducing customer churn* - Article (2017)

orders through different verticals tend to have a higher probability of churn, then the development of an onboarding program in which the user is *"taught"* and encouraged to order in these verticals would add value to Alpha Company. Thus, an acquisition plan based on daily steps (roadmap) can be developed which seeks to build customer loyalty by presenting all the ordering alternatives that are available: not only in restaurants, but also in coffee shops, markets, pharmacies, etc.

Moreover, it is also important to mention how **progressive disclosure** and **cognitive load**[41] impacts the customer's onboarding experience. In simple words, progressive disclosure aims at making the interface easier to use at first, while complex features are gradually revealed later. During the onboarding, it is critical to show only the core features of the service, and as users get familiar, unveil new options. It keeps the interface simple for new users and progressively brings power to advanced users. On the other hand, cognitive load refers to the way information or tasks are presented to a user, thus the more information the heavier the load. For that reason, the main goal for Alpha Company is that new customers move from completing simple actions to executing more complex ones, lowering the chances that users will feel overwhelmed.

Considering the example presented below, the first situation *Fig 7.3a* shows a large number of offers and alternatives that lead to great confusion for a new user, in instance they might be wondering where they can choose the order categories or simply feel overwhelmed by the amount of information presented to them. This situation can be tackled by offering progressive disclosure and reducing the cognitive load *Fig 7.3b* where Alpha Company offers simple actions

[41] *Growth Design Community*. Available in: https://growth.design/

to start with: *What kind of category would you like to order?* Afterwards some new features can be presented gradually: restaurants, verticals, offers, delivery types, etc.

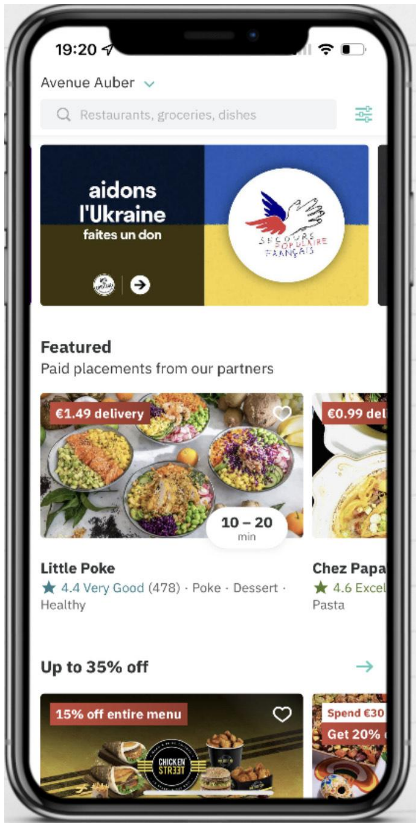


Fig 7.3a *Random company's current homepage*

Fig 7.3b *Implementing a first launching site to explain progressive disclosure and cognitive load concepts*

2. **Send a personalized welcome email and push notification through the app after some time with no sessions:** A personalized, friendly onboarding welcome email as well as push notifications can help build trust with new users and encourage them to place their first order. The welcome should remind users of

the aim of Alpha and how the service can help them, explaining different ways of paying for an order, how long it may take to receive an order, and different verticals where they can place an order. It is important to show them that the company cares about their experience. From the data presented above, the churn probability is highly correlated with the customers' behavior in the previous months: whether they have placed orders or not. For this reason, developing a system that allows push-notifications to be sent to customers whose reorder rate is decreasing during early stages, could also have a significant impact on retention.

Over this topic, it can also be mentioned the relevant role that **visual priming** plays when it comes to a personalized and enjoyable early-stage experience. McNamara defines visual priming[42] as an improvement in performance in a perceptual or cognitive task, relative to an appropriate baseline, produced by context or prior experience, which consists of subtle visuals that influence how customers respond.[43] Thus, presenting a friendly-looking food background when opening the app, lets customers wonder about what their next order will be, increasing the chances of a positive experience. In this particular case, visual priming can also be improved by tracking the time and season of each session, e.g. presenting coffee, smoothies and croissants emojis on a sunny background during a summer-breakfast session. On the other hand, in case of re-ordering, the system can track previous behavior based on their *customer_id*, e.g. a client who usually orders fast food may see the following screen when opening the app during dinner time:

[42] Timothy McNamara. *Priming: Perspectives from Memory and Word Recognition* (2005)
[43] *Growth Design Community*. Available in: https://growth.design/

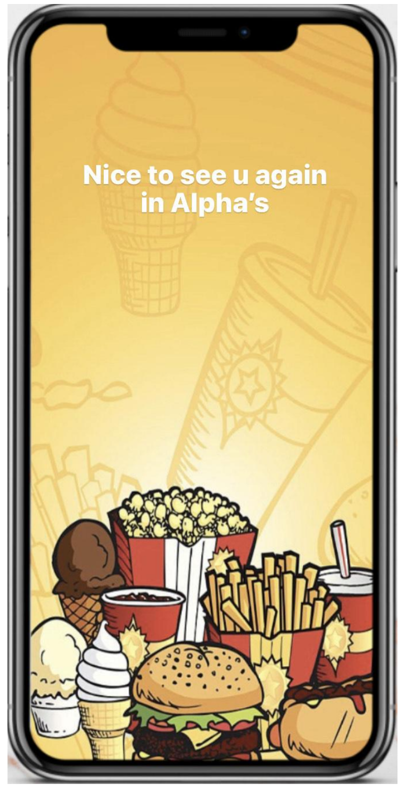


***Fig 7.4** Visual Priming to provide an enjoyable and personalized experience*

3. **Test different ways for brand building:** Video, written content, tooltips, static images can all help build brand awareness through customers. A good idea of reaching this is through on-street-branding such as riders wearing Alpha clothes while delivering or wearing Alpha-covid masks which makes a huge social impact, not only by catching new users' attention but also reinforcing those who already have been captured, making people feel that the company cares not only for their customers but also employees. Reinforcing the company status through social media (e.g. LinkedIn) by taking part in social movements of diversity and inclusion could also be a key to increase brand awareness and a great way to retain customers.

7.3.2 MID & LONG-TERM RETENTION

Once the customers experience the core value of Alpha's delivery services during the early stage, it is time to turn the attention towards building customer loyalty and continued engagement.

Again, customer attrition is best dealt with proactively. Just because a user has started using Alpha's app does not mean they will keep using it. Increasing mid and long-term retention means getting customers to create habits around services provided by the company so that they continue experiencing the core value it provides.

Some ways to reduce mid and long-term churn consist of:[44] [45]

1. **Overcome feature blindness and increase new feature adoption:** As users become more familiar with the service, they build habits around certain features, and tend to ignore other areas of it. They might try out each feature a couple of times before forgetting about them altogether, limiting how they use the app to a single, narrow manner (e.g. ordering only one vertical, like restaurants, while ignoring the rest of them - like cafeterias or supermarkets). This has been tracked by some internal projects carried on by the User Analytics Team.

   It is a phenomenon known as *feature blindness*, and while it may be fine for some users, it can spell disaster when users ignore core functionality. Getting users to build habits around the whole service is crucial, but it can be difficult to break them out of their routine and encourage new behaviors. To combat feature blindness, it is necessary to address the need as it arises, instead of when users are focused on more familiar activities.

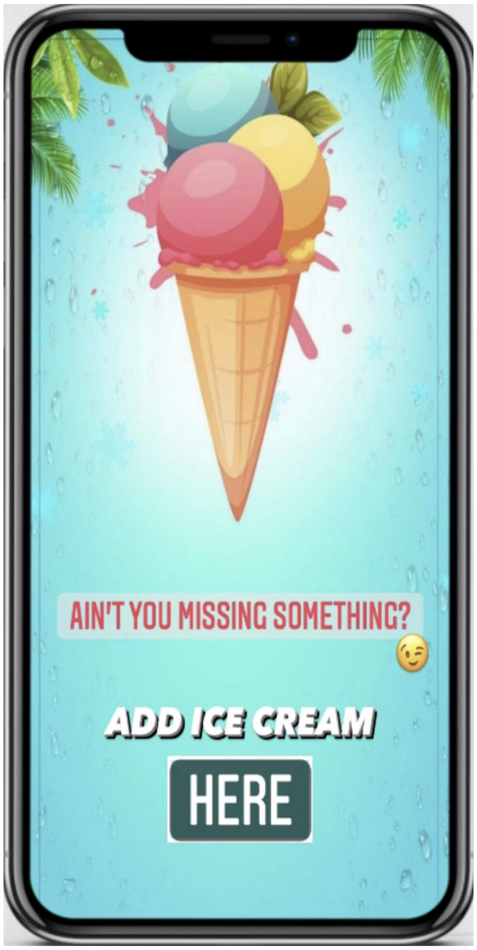
[44] Katryna Baiboni. Proactive strategies for reducing customer churn - Article (2017)
[45] Ryan, Hoover. *Hooked: How to build habit-forming products* (2014)

Temptation Bundling could be one way of solving this by offering complementary products of different verticals along with the main order. For example, a customer placed a dinner order during summer in vertical = restaurant; based on time and cuisine data, before checking out, an ice cream offer is presented (vertical = desserts), becoming a reminder for customers about different categories.

Besides, **Social Proof** confirms that customers adapt their behaviors based on what others do, meaning that when they are unsure or when the situation is ambiguous, they are most likely to look and accept the actions of others as correct. Thus, the greater the number of people, the more appropriate the action seems.[46] Therefore, Alpha Company could implement a *People also ordered [list]* interface with different vertical offers before the check-out view based on other customers' behaviors.

Some sample interfaces are presented below: Temptation Bundling Principle on *Fig 7.5a*, based on time + cuisine data, offering complementary products of different verticals related to the main order placed by the user; whereas on the right *Fig 7.5b*, the Social Proof Principle is applied by offering different verticals products based on previous customers' behaviors.

[46] Growth Design Community. Available in: https://growth.design/



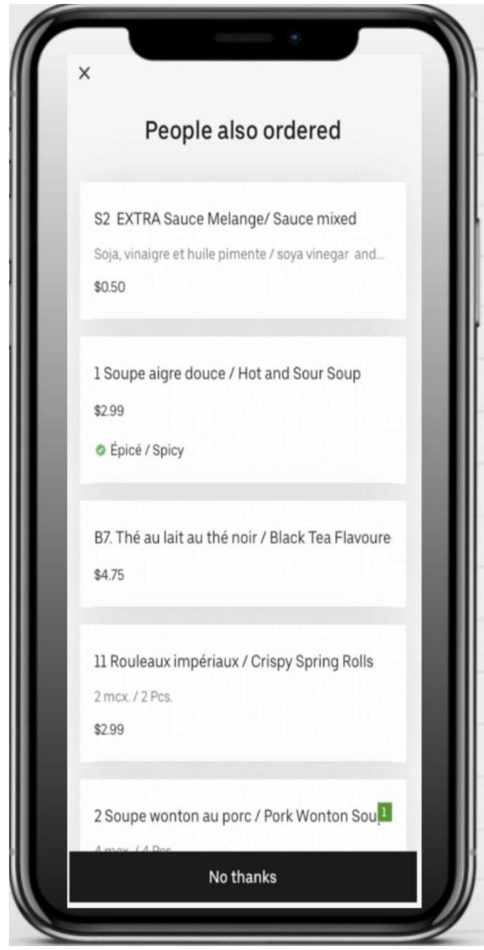


Fig 7.5a *Temptation Bundling applied when ordering, to overcome feature blindness*

Fig 7.5b *Social Proof Principles applied when ordering, to overcome feature blindness*

2. **Offer special discounts based on the predictive Machine Learning model developed:** When it comes to re-engaging potential churners, emails and push notifications are often the best bet. A well-written retention email, as well as push notification offering any kind of discount based on the customer behavior, can help keep potential churner users from leaving, reminding them of the value Alpha Corporation provides and reducing the chances they will abandon the app

forever. For example: Customer X used to place fast-food orders during dinner, he is predicted as a potential churner by the model, then an email/push notification should look like this.

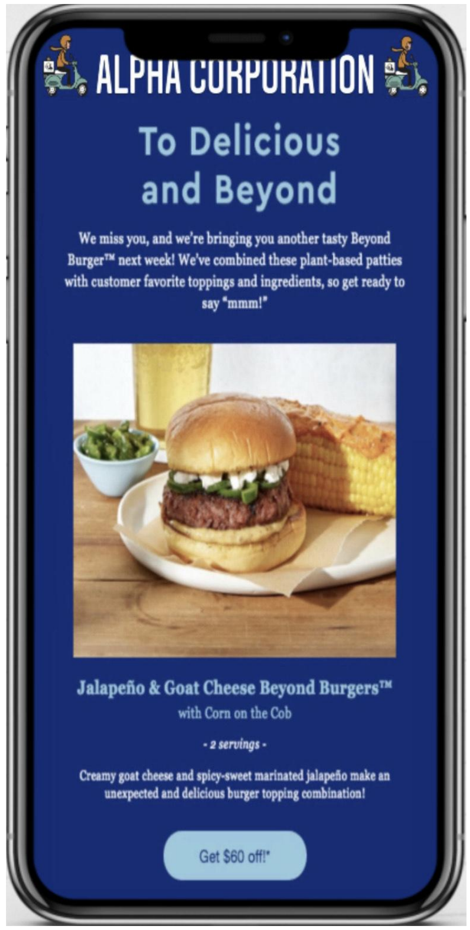


Fig 7.6 *Email with offers based on ML Predictions*

With the current data, orders placed during breakfast, lunch, snack, and dinner time can be tracked. However, to complement the churn analysis (and with the information that is currently available) one of the possible future studies consists on the implementation of a clustering model, which is proposed in the following section (improving, thus, the Principle of Personalized Approach).

3. **Develop a subscription program for potential churners and design/implement an attractive launching page to subscribe:** At its basic form, a subscription business model is one that charges customers a recurring fee for certain benefits. But on a deeper level, ecommerce subscriptions are about strong customer relationships. Subscriptions turn customers, who already see the value a company provides, into loyal followers who become reliable sources of recurring revenue. In fact, the longer a customer uses your product or service, the more valuable they become to the company. Plus, higher customer retention rates mean lower acquisition costs in the long term.[47]

   Based on the predictive model output, it is possible to distinguish those customers who are likely to churn in the near future. For that reason, launching a subscription program for those customers could also be a great way of retaining them. By paying a fixed price monthly, those potential churners will get exclusive and unlimited discounts (e.g. starting at 20% from over 3,000 restaurants), a certain amount of free delivery treats from all shops and restaurants, priority customer support, etc.

   All benefits mentioned previously become a way to reduce churn, in fact, based on outcomes presented in *Chapter 4*, it appears that one common behavior shared among them is lower engagement during the previous months of churning, thus offering these benefits through subscriptions could improve the number of orders placed by potential churners while keeping profitable economic KPIs.

[47] Chuck Longanecker. *Why you should use a subscription business model?* - Article available here: https://www.entrepreneur.com/article/243573

Keeping this in mind, and based on certain principles such as **the impact of transparency** and **providing exit points**, certain subscription pages could be implemented to retain potential churn customers. Some examples will be discussed below and how those principles may affect customer retention rate.

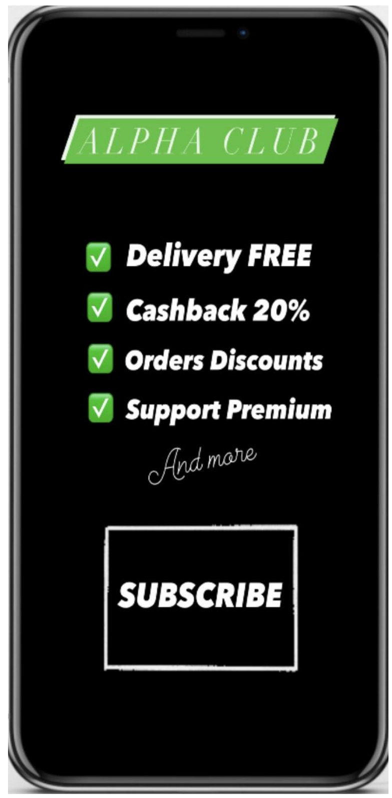


Fig 7.7 *First Subscription interface*

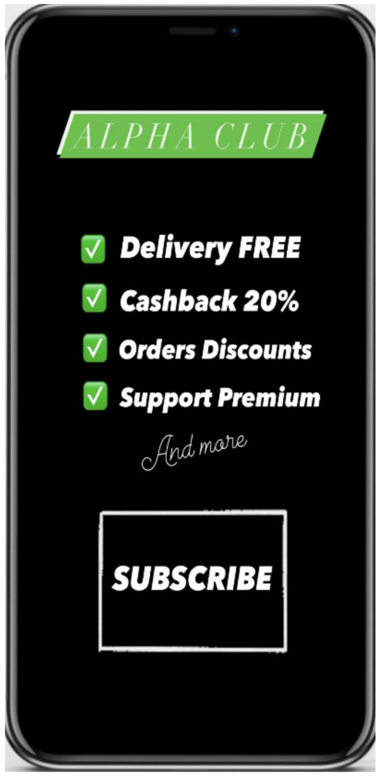
First of all, the launching page presented above follows the typical paywall pattern: it grabs your attention with a simple design and contrast, presenting all benefits and the reasons to subscribe to Alpha's Club program and right below a call-to-action button: *subscribe*. At first sight, this launching page seems to fulfill its purpose: based on the ML predictive outcome, presenting this offer to potential churn users could somehow tackle the issue of low rate retention. However, this can also be improved based on **principles of**

transparency, trial, reactance and exit points, increasing even more the chances of customer retention. Therefore, some of the issues Alpha would face if implementing this subscription page are listed below.

First, you can barely see the **exit button** on the top left of the screen and, while minimizing distracting navigation elements can increase the retention rate, pushing too far makes customers feel trapped and obliged to subscribe, which leads to a total opposite outcome. Considering the **reactance principle**[48] users are less likely to adopt a behavior when they feel forced, thus when taking away too much of a person's behavioral freedom, it backfires and increases resistance to persuasion, reducing drastically the retention rate.

Secondly, research by Jaycee Day[49] reveals that there are two main concerns when reaching the subscription paywall: 1) a trial period to test how the subscription works. 2) forgetting to cancel the trial and being charged. For that reason, adding a **trial option** and addressing the customer's fear of forgetting to cancel a subscription by creating a roadmap and trial reminder close to the trial expiry date, fulfills the **principle of transparency**, reducing customers complaints and increasing customer retention rate.

[48] Growth Design Community. Available in: https://growth.design/
[49] Henry Priest. 101 Cognitive Biases (2020)



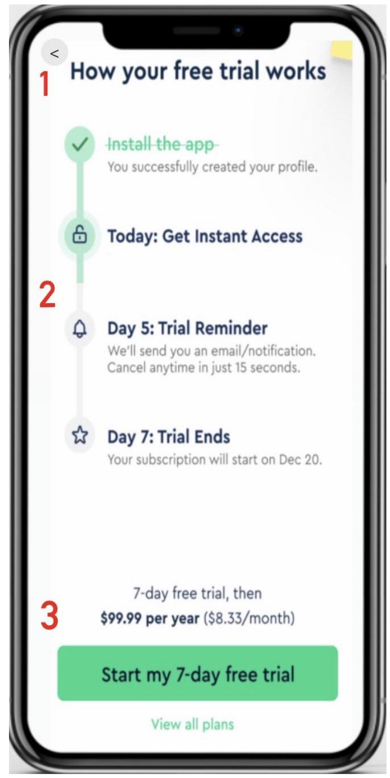


Fig 7.8a *Subscription interface comparison* ***Fig 7.8b*** *Considering principles mentioned above*

To summarize, the development of a subscription program is a solution to increase the retention rate based on the outcome of the predictive model. Likewise, the implementation of an appropriate launching subscription page is also relevant: the impact of exit points *Fig 7.8b(1)* based on the reactance principle, as well as a trial period and roadmap/reminders to avoid the main fears of users while respecting the principle of transparency *Fig 7.8b(2)*, and finally, the price and types of subscription plans offered *Fig 7.8b(3)*.

4. **Measure customer satisfaction and collect feedback & NPS Machine Learning Model development for the potential churners:** Maintaining open lines of communication with customers is key to reducing churn. The more they use Alpha service, the more important it becomes to understand what they value.

Sending a carefully timed survey can help to get to know the users and their needs. NPS Score from placement orders and contact center, as well as developing a non-structured data Machine Learning model exclusively for NPS analysis (which will be discussed in *Chapter 8*) are great tools to reduce churn while understanding what the most common reasons are.

5. **Propose a Roadmap Milestone for potential churners:** When recognizing those potential churners, a roadmap milestone can be proposed, with different objectives to be achieved and their respective benefits. For example: reaching a certain amount of orders, or placing an order at some point of the day (such as lunch or dinner), or placing an order in a vertical different from restaurants (such as pharmacy or coffee shops). This will allow greater engagement with customers who are expected to leave the company, by presenting different aspects of the service provided by Alpha. This is highly related to the **Goal Gradient Effect** which indicates that motivation increases as users get closer to their goal, thus the closer users are to reaching a milestone, the faster they work towards reaching it. It is also related to the **Variable Reward Principle** which demonstrates that people especially enjoy unexpected rewards. In fact, based on previous internal company studies and based on the results of the data analysis presented in *Section 4.1.2*, the higher the percentage of shared orders between categories, the lower the churn rate.

6. **Default address and tracking location**: Based on the principle of personalization, using data to craft a better experience is key to user retention.[50] Alpha's customers are evolving and growing with the platform every time they

[50] Growth Design Community. Available in: https://growth.design/

use it. So, it is important to make sure the service leverages past behaviors to remove repetitive and unnecessary steps. For example, allowing customers to save a default address whenever they order, so setting the address when reordering is no longer necessary (it saves the customer a lot of psych, especially if they are starving). However, in case of tracking location and not matching with the default address, vendors available within the current location will be listed and when placing an order the customer will be asked to modify the address if they want to (disclaimer: this should be one of the latest steps as we need to respect the progressive disclosure principle mentioned in point number 1).

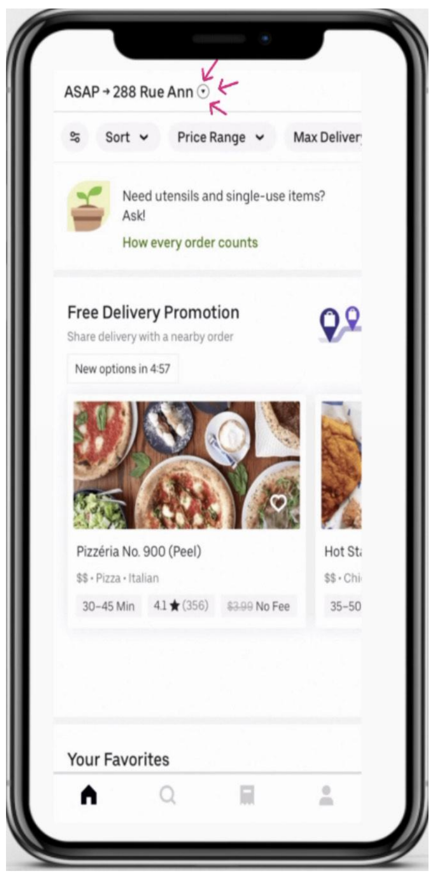
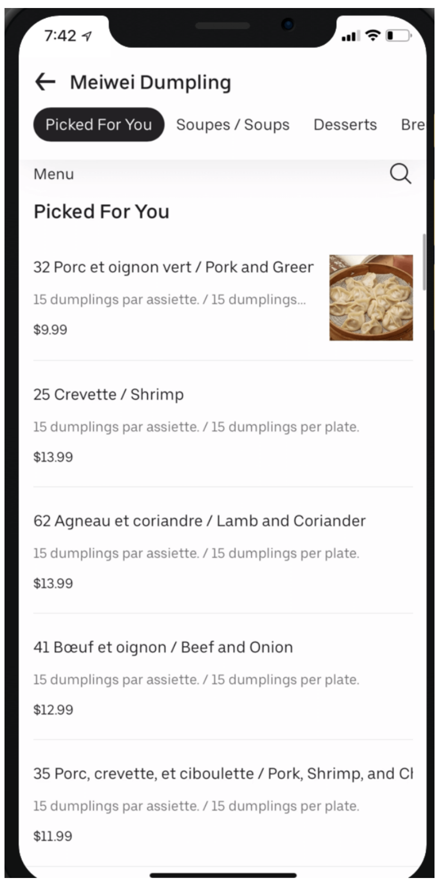


Fig 7.9 *Default address when reordering (on top left, pink arrows) - Personalization principle*

7. **Use Sensory Appeal:** This principle states that people are engaged by things that appeal to multiple senses. The senses amplify one another when they are mixed. It is really powerful because users do not perceive them as marketing messages.[51]

In this particular case, considering a study carried out by the Product Analytics Team, a large number of sessions leave the app when viewing the menu, and this churn rate is increased in those vendors which do not present images of their products. Thus, adding juicy visuals to engage different senses next to each product could be also a solution to improve the retention rate.



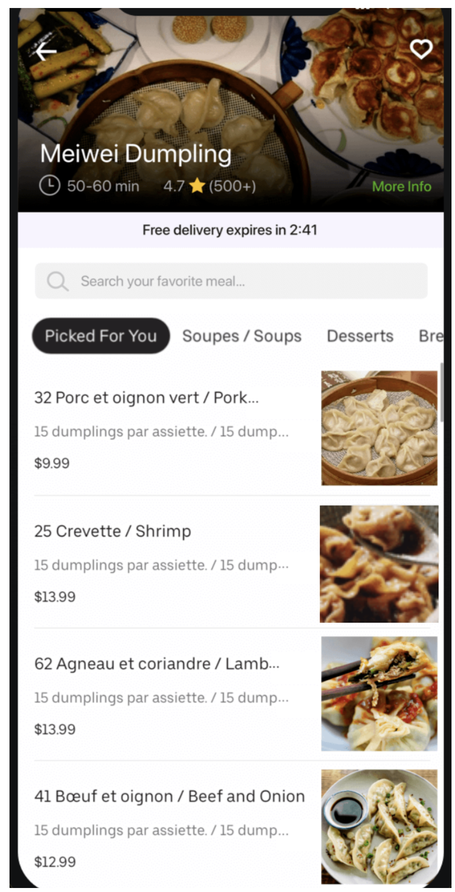


Fig 7.10a interface without images.

Fig 7.10b interface designed considering the Sensory Appeal Principle

[51] *Harvard Business Review*, The Science of Sensory Marketing (2015)

8. **Once the monthly subscription program is rolled out, encourage customers to upgrade to annual or longer-term contracts if possible:** Based on point number 3, once the monthly subscription program is rolled out, it is a good idea to engage those users who are already subscribed to go for an annual contract, as for long-term retention, nothing beats an annual contract.

   Annual subscriptions give more time in which to demonstrate value, and a longer time frame makes it much more likely that the service will become a part of the customers' routine, thus a higher percentage of annual contracts experience lower churn rates. The **Principle of Choice and Decision Fatigue**[52], which states that choosing something inherently means giving up something else and that people prefer not to be constantly making decisions, play a relevant role here as customers on an annual contract only have one purchasing decision to make each year instead of monthly bases, and they have access to a price offer that would lose in case of opting for a monthly subscription.[53]

[52] Sheena Iyengar. The art of choosing (2010)
[53] David VanHoose. *eCommerce Economics* (2011)

7.4 SUMMARY OF RETENTION STRATEGIES

SITUATION / ISSUE	*PROPOSED STRATEGIES*	*PRINCIPLES & REACTIONS*
Improve user onboarding experience	- First pages with easy steps and complex features presented at the end. - Avoiding cognitive load information on interfaces.	• Progressive Disclosure • Cognitive Load • Step-by-step experience
Personalized emails and push-up notifications when no sessions	- Developing a friendly-looking background, personalized based on previous behaviors, increasing positive experience	• Visual Priming
Brand awareness	- Video, images and written content from the Brand Team. - Riders wearing Alpha clothes while delivering. - Riders wearing Alpha masks during Covid times. - Reinforce company status through social movements of diversity and inclusion.	• Brand building and awareness • Social proof • Noble Edge Effect
Overcome feature blindness	- Offering complementary products of different verticals, allowing customers to know about their existence. - Offering additional products based on other customers behavior: *"Other customers also ordered this: [list]"*	• Temptation Bundling • Feature blindness • Social Proof
Special discounts based on the Predictive Churn Model	- Once the potential churn customer is flagged, offer them discounts based on previous behaviors	• Service re-engagement • Personalization

Develop Subscription Program	- Implementing subscriptions plans with different benefits. - Design an interface with exit points. - Allow a trial version with a reminder through the roadmap of the subscription trial.	• Transparency • Reactance • Mid-Long term engagement • Commitment and Consistency
Tracking NPS Results	- Maintaining open lines of communication with customers.	• Customer support care
Roadmap Milestone for potential churners	- Propose different goals and rewards customers could get from achieving those goals.	• Goal Gradient Effect • Variable Reward
Default address and location tracker	- Allow customers to set a default address. - When the default address does not match with the track location, then list vendors available in current location and ask to modify the address afterwards	• Remove unnecessary and repetitive steps. • Transparency • Progressive disclosure
Use sensory appeal	- Ask vendors to publish their menus with photos of each single product.	• Sensory Appeal
Offer different time-frame subscription plans (e.g. annual plans)	- Offer more than monthly plans and design an appealing interface for subscription.	• Principle of Choice • Decision Fatigue

***Table 7.1** Retention Strategies summary*

CHAPTER 8

CONCLUSION

From a marketing perspective, developing a model that precisely predicts customer churn could have many managerial and financial implications for B2C Online Food Delivery companies in order to minimize the probability of churn.

In particular, this thesis suggests three main managerial implications for Alpha Corporation. Firstly, the **correct classification of a customer as churner or non-churner** can reduce the misclassification costs: either classifying a non-churner as a churner (Type I error) or a churner as a non-churner (Type II error). Failing to identify a churner has significantly more negative managerial and financial consequences on the company's retention strategies than classifying a non-churner as a churner. In fact, in the former case, the company loses direct contact with the client and therefore it cannot engage its customer contact strategy. Basically, when an incentive which can be sent to a real churner is sent to a non-churner, it means either that a real churner has not received an incentive to be persuaded to stay (increasing in their risk of churn) and a non-churner has received an incentive which they were not supposed to receive (loss of marketing budget for Alpha Corporation). Thus, Alpha does not even have the opportunity to induce potential churners to stay by sending them marketing incentives, and therefore, it loses the customers and forgoes profits, opportunities and so forth.

Companies seeking to adopt such a retention strategy should be equipped with models that can accurately identify customers who are likely to churn in a given future time period. This becomes even more critical in B2C e-commerce contexts. The model developed in this thesis helps to improve the churn prediction rate (correctly classifying customers into churners or non-churners based on their likelihood to leave the company) and thus reducing misclassification costs, which makes it possible for marketing managers to come up with a tailor-made retention strategy that focuses on the customers at real risk and save the scarce marketing resources. Therefore, it can help Alpha Corporation to more accurately identify churners and thereby develop more efficient, effective and targeted retention strategies. As a consequence, the Random Forest and XGBoost models can help this company not only to reduce costs by sending fewer incentives to non-churners, but also to increase revenues by contacting potential churners with the highest profit potential. This means that more valuable churners are more likely to change their mind and stay loyal to the company services, leading to an increase in retention rate and delivering more revenue/profit.

Secondly, the thesis indicates that **it is not necessary to simply rely on traditional techniques like Linear Regression or Neural Network** for customer churn prediction. Models can be very wasteful if churn predictions are inaccurate (first case) or if it takes a long time to develop and train the model (second case). The results in this thesis show that Random Forest (along with embedded methods) and XGBoost models can provide useful information in the context of churn prediction and management. Using any of the developed and presented models, Alpha Corporation must be able to analyze customer data, understand in an effective and efficient manner their needs, and identify the riskiest customer segments in terms of churn to improve their retention strategy and avoid losses. The high predictive accuracy rate of Random Forest represents, in fact, a strategic asset to better manage the customer relationship, and it could lead to incremental profits for the company in case of implementing this method. Even a small improvement in the accuracy of churn prediction models can be associated with substantial financial gains for the company, especially those operating in the Online Food Delivery Industry, where customer churn has been shown to be a major cost item due to the high level of competition. For instance, the high accuracy rate of both models could help marketing managers deal with both prediction and targeting, as well as identify those variables that may affect the most (features importance).

Thirdly, this thesis also **suggests managerial recommendations** for optimizing the decision-making process in a customer churn prediction context. In fact, while applying Random Forest or XGBoost, Alpha Corporation can accurately identificate customers who are truly at risk to churn, focus their effort on these customers and potentially save money. In particular, the company can better serve their current customer base by building and maintaining relationships, rather than expanding efforts to recruit new customers, which are often characterized by a high attrition rate. In fact, long-lasting customers spend more with these kinds of companies, provide new referrals through positive word of mouth if satisfied, and are less costly to serve, because the company has more knowledge about them. Moreover, long-term customers are less prone to competitive marketing actions. Thus, losing long-lasting customers leads to opportunity costs, whilst also it increases the necessary costs to attract new customers. From a marketing manager's point of view, the aforementioned findings imply paying attention to their current customers flagged as churners.

Once customers with the highest probability of churning are identified thanks to the predictive model, marketing data analytic teams can **go beyond** and **develop a clustering model** based on the churn database which had been predicted. Thus, some push notifications can be sent to contact them and offer some customized micromarketing messages with discounts, vouchers and benefits before they switch to another company. For example, the marketing manager may organize a campaign that consists of sending out push notifications to offer a certain discount based on customers behaviors. Thus, marketing managers can decide whether to develop tailor-made marketing programs to incentive customers to remain with the firm, programs which will presumably reduce the likelihood of churn.

Another suggestion proposed in this thesis, and which acts in a complementary way with the predictive model, is to **develop a model based on unstructured data** that allows the **emotion of potential churners to be analyzed**. In this way, it would be possible to analyze the surveys carried out (with their respective NPS score) and thus discover what aspects produced dissatisfaction in the clients, leading them to abandon Alpha's services.

In other words, the solution path can be summarized in the following four aspects: **define, prepare/predict, retrain and strategize/act**[54]

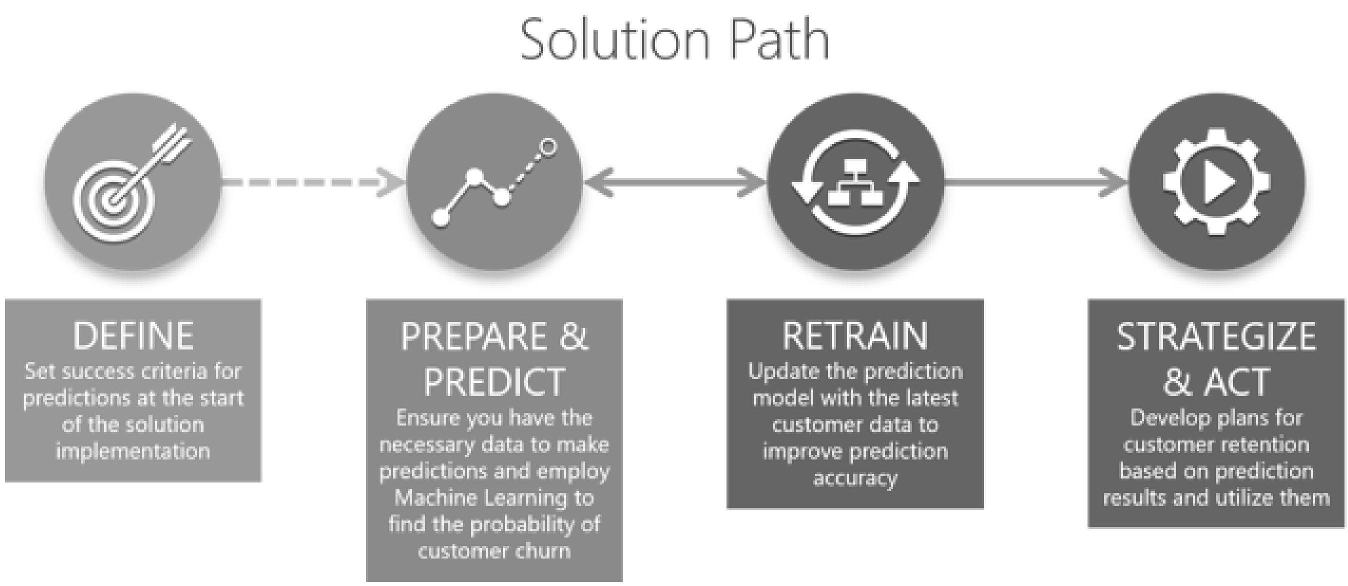


***Fig 8.1** Solution path summary*

[54] Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow (2019), 2nd Edition.

To sum up, the model proposed in this thesis aims to predict customer churn in the Online Food Delivery context, helping this B2C e-commerce company not only to reduce costs by sending fewer incentives to non-churners, but also to increase revenues by contacting potential churners with the highest profit potential. This means that valuable churners are more likely to change their mind and stay loyal to Alpha Corporation, leading to an increase in retention rate and delivering more revenue/profit. In this way, once the potential churners have been identified, different retention marketing strategies are established: Offer benefits such as **vouchers**, **discounts or free shipping** to said users; **design a milestone roadmap** where objectives are set and, by meeting them, benefits are obtained; **implement a monthly subscription plan** that would be offered to potential churners, allowing them to pay a reduced and fixed monthly price, allowing access to certain benefits; implement personalized attention for said users, being able to have **priority access when contacting the call center**; send **push notifications** or emails with unique offers, among others.

Finally, suggestions are also proposed that go beyond the scope of this work, but that can be complemented to obtain a more straightforward output: implementation of an **artificial intelligence model based on the clustering** of potential churners according to their purchasing behavior, as well as, an **analysis of customers' emotions based on unstructured data from NPS surveys** to recognize the main causes related to users churn.

8.2 FUTURE RESEARCH DIRECTIONS

Although the scope of this thesis is limited only to the development of a prediction model of churn users and, subsequently, to the proposal of user retention strategies from a Marketing and Business Analytics point of view, there are a wide variety of solutions that could complement them in obtaining an output of greater importance at the business level.

As a first approximation, the development of the prediction model and retention strategies such as the development of a subscription plan for said potential churners or the implementation of a milestone roadmap can constitute the first link in the retention chain: knowing who are the users who are likely to churn in the near future, and what actions can be taken to retain those users.

8.2.1 CLUSTERING MODEL DEVELOPMENT

Customer clustering or segmentation is the process of dividing an organization's users into groups or **clusters** that reflect **similarity amongst customers** in that particular group. The goal of such a process of clustering is to decide how to relate to users in each of these clusters to maximise the benefits those clients bring to the business and make for more engaging transactions with them. In the context of customer segmentation, cluster analysis is the use of mathematical modelling to achieve such goals. These homogenous groups of customers are known as *customer archetypes* or *personas*.

In this case, once the churn prediction model has been developed, you can dig deeper to gain more insights on different customer retention marketing strategies. That is why a first approach is linked to the development of an artificial intelligence model that, considering the basis of potential churn, emphasizes clustering them into different groups based on the behavior they present. This means that, for example, they can be grouped based on the frequency of their orders during the week or weekend, or type of placement orders like vertical types of consumption, etc.

In this way, the predictive model will allow an approach to solve the unknown of: *who are the users to whom the marketing campaigns should be targeted*. While a second approach, the implementation of this clustering system, will allow knowing more accurately the consumption behaviors of those users and to propose more focused strategies for each one of the segments.

So the question here is: *how would the clustering model work?* Once it is implemented, it runs over the **potential churn database**, being able to detect patterns and dividing them into different segments. According to the behavior they share in common, the impacts of retention campaigns can be designed. Thus, if a cluster group is made up of users who place orders at night, a voucher will be triggered with a discount on dinner restaurants. In order to encourage other Alpha Company services, one could also choose to provide benefits for other types of verticals, encouraging the user who normally orders in a restaurant during dinner to place an order in cafeterias or supermarkets.

From a technical perspective, **unsupervised learning** models are usually applied when data scientists do not have prior knowledge about the categories and labels in the data. Any information pertaining to how the data can be grouped is not present. Thus, it is just data with several rows and columns with numerical values, but no information about what rows belong to what group, if any.

8.2.1.1 K-MEANS ALGORITHM SUGGESTION

Considering that the predictive churn model is already trained and the output is available, one suggestion to go deeper with the analysis consists of developing a **KMeans model**, as it is a popular unsupervised clustering algorithm designed to group data into clusters and label data points. It is widely used in applications such as market segmentation, document clustering, image segmentation, and image compression, etc.

The algorithm is based on the distance between the points in *N-dimensional* space and relies on finding centroids for each of the numbers of clusters we want to identify in the data. It shuffles and randomly chooses as many points as the number of clusters required to be identified and assigns them as cluster centers. Afterward, iteratively, it calculates the distance between these cluster centers (or centroids) and data points to group the data points into clusters. The images below show the visual grouping of data points in a 2-dimensional space for better understanding.

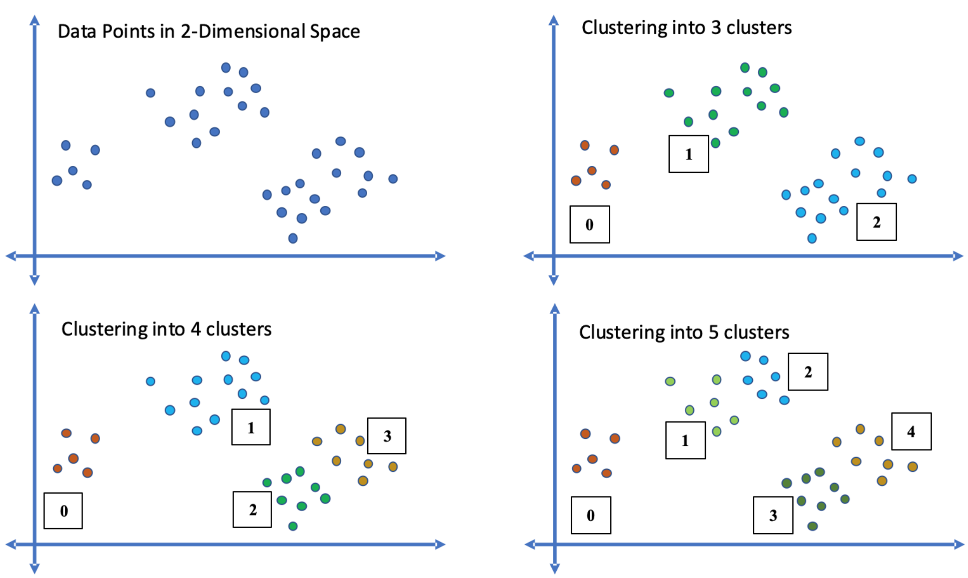


***Fig 8.2** A set of points group into an increasing number of clusters*

Thus, in case of developing this model, the overall scope of the analysis can be summarized as:

1. **Creation of an unsupervised learning model** using the available customer churn data (model developed through Chapter 6) to label the data into multiple clusters.
2. **Analyze the clusters** to determine if they align with any customer churn behavior.

8.2.2 UNSTRUCTURED MODEL DEVELOPMENT

As mentioned initially, one way to retain churners is to analyze their loyalty and how satisfied they feel about the service provided by Alpha company. The **Net Promoter Score (NPS)** is a metric that is associated with customer loyalty and is widely used across industries. The NPS is measured through regular market research surveys, which are conducted at a frequency ranging from monthly to an annual basis. In such surveys, NPS is derived based on a question in which customers are asked to rate the likelihood of recommending Alpha's services to their friends or colleagues. Apart from the NPS question, such a survey also includes questions related to the customer satisfaction score on experience attributes: key areas of the interaction between the customer and the company that may influence the NPS (e.g., service attributes like: delivery costs, service quality, billing and touchpoint experiences, such as website, call center, mobile app, etc). Moreover, the common practice in CX surveys is to include responses from a sample of customers related to all competing companies in the market, thus giving crucial information about Alpha Corporation's position against competition. The results obtained from the NPS surveys allow primarily for the monitoring of the

NPS and CX attribute satisfaction score trends for all market players. Such an analysis reveals areas of strong and weak performance, which trigger companies to take corrective actions.

All the feedback surveys carried on by Alpha Corporation provide the opportunity to express, communicate and share users' opinions, thoughts, views and perspectives—on diverse issues, matters, and topics—through text and scores. Analyzing and studying these responses from feedback surveys may indicate emotional states and the reasons behind those emotions. However, the massive volume of this data makes this analysis very difficult. Artificial Intelligence can help to find feelings, personal traits, views, and their effects on feedback surveys in an automated manner.

8.2.2.1 TEXT-BASED EMOTIONAL PREDICT MODEL

Text-based emotion detection using artificial intelligence employs Natural Language Processing (NLP), which combines techniques in linguistics and computations to help computers comprehend and produce texts in the form of human languages. Wider applications in NLP consist of machine translation, speech recognition systems, sentiment analysis, text classification, questions and answering, chatbots, text summarization, and so on. Emotion detection is an extended derivative of sentiment analysis and it is taking out finer-grained emotions like anger, happiness, sadness, anxiety, depression, etc., and applying this feedback to future decision-making.

Therefore, once the results of those potential churners have been obtained, a model can be developed to analyze the emotion of said users regarding the services provided by the company. It will be possible to detect, based on the scores and surveys that they have responded, if there are certain essential aspects of which improvements could be made, or if the reason for leaving the company is linked to dissatisfaction (negative emotion). It will be possible to analyze the words they use, their expressions, emojis and scores. This suggestion of text-based emotion detection using artificial intelligence methods can be useful in many fields, not only for the Online Food Delivery industry, but also human–computer interaction, education, data mining, psychology, E-learning, software engineering, website customization, information filtering systems, gaming, etc.

The focus of this sentiment analysis suggestion is to derive information from human language for interpreting views and satisfaction to assign polarities like positive, negative, or neutral based on customers responses on surveys (and not only considering the NPS score). Thus, developing this emotion detection model based on text from NPS surveys could be addressed by two approaches:

1. **Explicit detection**: Explicit means clearly stated words or emotion-bearing words like "*happy*" are used in the written text to express the emotions. Explicit detection is identifying and classifying written text into emotion classes with the help of emotion-bearing words. Explicit detection is used where more specified key phrases are used to express emotions, such as the keyword-based approach of emotion detection.

2. **Implicit detection:** In contrast, implicit means identifying and categorizing text into emotion classes without emotion-bearing words is referred to as implicit emotion detection.

To summarize, customers provide their feedback about Alpha's services with NPS surveys, where they express their thoughts and opinions. This feedback is used by businesses to understand customer requirements, in order to reduce the churn rate by increasing customer support on those features that are considered important for potential churners. Therefore, it becomes necessary to understand the impact of these aspects on customers' satisfaction. Thus, from a technical perspective, deep learning and ensemble techniques need to be used for further analysis on this topic, allowing to improve robustness and understanding of churners' feelings/satisfaction to respond accordingly.

BIBLIOGRAPHY

- Olafsson, Li, Wu. *Operations research and data mining.* (2008)
- Sabbeh SF. *Machine-learning techniques for customer retention*. (2018)
- Ngai EW, Xiu L, Chau DC. *Application of data mining techniques in customer relationship management* (2009)
- Tanneedi, Naren Naga Pavan. *Customer Churn Prediction Using Big Data Analytics* (2016).
- Almana, Amal. *A Survey On Data Mining Techniques* (2014)
- Bramer, Max. *Principles of data mining*. Vol. 180. (2007)
- Aidan Wilsom. *A Brief Introduction to Supervised Learning* (2019)
- IBM Cloud Education. *Understanding Exploratory Data Analysis*. (2020)
- Verdonck, Tim. Baesens, Bart. *Special issue on Feature Engineering* (2021)
- Hang Gao, Xin-Wang Liu, Yu-Xing Peng. *Sample-Based Machine Learning with Missing Data* (2015)
- Lei Wang, Xiaogang Deng, Yuping Cao. *Multimode complex based on local outlier factor method* (2017)
- Enislay Ramentol, Rafael Bello, Francisco Herrera. *Smote: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets* (2011)
- Balaji Krishnapuram, Shipeng Yu, Bharat Rao. *Cost-Sensitive Machine Learning* (2011)
- Carolin Strobl, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, Achim Zeileis. *Conditional variable importance for random forests* (2008)

- Raphael Couronné, Philipp Probst, Anne-Laure Boulesteix. *Random forest versus logistic regression* (2018)

- Tom Fawcett. *ROC Graphs: Notes and Practical Considerations for Researchers* (2004)

- Jason Browniee. *XGBoost with Python: Gradient Boosted Trees with XGBoost and scikit-learn* (2019)

- Niccolò Gordini, Valerio Veglio. *Customer Churn Prediction and Marketing Retention Strategies* (2016).

- Daniel P.Hampson, ShiyangGong. *How consumer confidence affects price conscious behavior: The roles of financial vulnerability and locus of control* (2021)

- John Goodman, Steve Newman. *Understand Customer Behavior and complaints* (2003)

- Young Namkung, Soo Cheong (Shawn) Jang, Soo KeunChoi. Cus*tomer complaints in restaurants: Do they differ by service stages and loyalty levels?* (2011)

- Andrew Hershy. *Is Random Forest better than Logistic Regression*? (2017) Article available in: https://towardsdatascience.com/is-random-forest-better-than-logistic-regression-a-comparison-7a0f068963e4

- *XGBoost: Gamma* (2018). Article available in: https://xzz201920.medium.com/xgboost-gamma-1b0f3b00c70e

- *Vishal Morde. XGBoost Algorithm* (2019). Available in: https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

- W. Verbeke, B. Baesens, C. *Bravo, Profit-Driven Business Analytics* (2018)

- S. Barocas, M. Hardt, A. Narayanan, *Fairness and machine learning, Limitations and Opportunities.*

- N. V. Chawla, K. W. Bowyer, L. O.Hall, W. P. Kegelmeyer, *SMOTE: synthetic minority over- sampling technique Journal of artificial intelligence research* (2002).

- Aurelien Geron. *Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow* (2019), 2nd Edition.

- Growth Design Community. Available in: https://growth.design/

- Christoph Molnar. *Interpretable Machine Learning, a Guide for Making Black Box Models Explainable* (2020). Article available in: https://christophm.github.io/interpretable-ml-book/

- Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. *Why should I trust you? Explaining the predictions of any classifier* (2016).

- Sweta Chakraborti. *Customer Clustering for better Customer Engagement* (2021)

- Vicent Giner-Bosch. *Assessing Classification methods for churn prediction by composite indicators* (2012).

- Sheena Iyengar. *The art of choosing* (2010).

- The Psychology of Choice: How to make easier decisions? Article available in: https://blog.hubspot.com/sales/the-psychology-of-choice#:~:text=Choosing%20something%20inherently%20means%20giving,don't%20grab%20it%20today.

- Akhil Sharma. *Unsupervised Learning using KMeans Clustering* (2021). Article available in: https://medium.com/data-science-on-customer-churn-data/unsupervised-learning-using-kmeans-clustering-629b26e1ce1e

- Katryna Balboni. *Proactive strategies for reducing customer churn* (2020).

- Gerhard Raab, Riad A. Ajami, G. Jason Goddard. *Customer Relationship Management: A Global Perspective* (2016)

- Ryan Hoover. *Hooked: How to build habit-forming products* (2014)

- Vineeth Balasubramanian, Shen-Shyang Ho, Vladimir Vovk. *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations and Applications* (2014).

- Jason Brownlee. *Understand the impact of Learning Rate on Neural Network Performance.* (2020). Article available in: https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/#:~:text=The%20learning%20rate%20is%20a,the%20model%20weights%20are%20updated.&text=Momentum%20can%20accelerate%20training%20and,to%20converge%20the%20optimization%20process.

- Jason Brownlee. *What is the difference between a parameter and a hyperparameter?* (2017). Article available in: https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/

- Cambridge Spark. *Hyperparameters tuning in XGBoost* (2017). Article available in: https://blog.cambridgespark.com/hyperparameter-tuning-in-xgboost-4ff9100a3b2f

- Ralph Kimball, Margy Ross. *The Data Warehouse Toolkit* (2013). Third Edition.

- Artun Omer, Levin Dominique, *Predictive marketing: easy ways every marketer can use customer analytics and big data* (2015)