

Customer Churn Prediction - DS Roadmap

Name: Khalid Mohamad Aladin Hafez Omar
Position: Senior Data Scientist
Date: 16 July 2023

Introduction and Company Objective

Introduction:

OLIST is an e-commerce company similar to Amazon; it is a marketplace where sellers and buyers come together. For the immediate short term, the company leadership has come up with the following four business goals:

Company Strategic Goals and Objectives:

1. Increase the revenue
2. Increase the efficiency of services
3. Improving customer experience
4. Increase the number of active customers

Objective:

Create proposals of DS/Analytics projects to help achieve company strategic goals.

Data Science Projects Proposed - 1

Projects	Functional Area	Project Description
Project 1: Delivery Date Prediction	Marketing and Logistics	The logistics team at Olist uses heuristics to provide an estimated delivery date for the orders placed. It is very conservative about the delivery dates. CMO. He found that on average, the estimated time to deliver products that is given to customers is twice that of the actual delivery time. Such a high expected delivery time is driving away Olist's customers. So, the CMO is looking to use ML to get a far more accurate expected delivery date.
Project 2: Sentiment Analysis	Marketing & Sales	The Olist Chief Marketing Officer sought to comprehend customer experience through post-delivery reviews and pinpoint areas for improvement. Recognizing the potential of NLP for sentiment analysis and topic modeling, they were mindful of its limited sophistication in Portuguese, the language of the customer reviews.
Project 3: Customer Churn	Marketing	Customer churn is a vital metric for an e-commerce company's CMO. Olist aims to develop customer churn models to identify 'at-risk' customers and devise tailored retention strategies. This will yield insights into the factors influencing churn, enhancing their retention efforts.

Data Science Projects Proposed - 2

Project 4: Customer Acquisition Cost Optimisation	Marketing & Finance	Olist's Marketing team conducts several promotional campaigns for customer acquisition. Nonetheless, the CFO is concerned about excessive spending due to substantial discounts and benefits, leading to inflated customer acquisition costs. To address this, the CFO aims to implement a process that evaluates campaign effectiveness relative to customer lifetime value.
Project 5: Fraud Detection	Finance	Fraud poses a significant challenge in the e-commerce industry, leading to substantial financial losses. It encompasses various types, such as merchant identity fraud, advanced fee and wire transfer scams, chargeback fraud, and more. To safeguard the organization, the CFO aims to leverage analytics to detect fraudulent transactions effectively.
Project 6: Price Optimisation	Sales	Pricing is a critical aspect of e-commerce, significantly influencing revenue, sales, profit, and demand. OLIST's sales team aims to develop a data-driven price optimization algorithm, considering factors like location, customer attitude, and competitor pricing. This algorithm will enhance sales and revenue by responding to price changes based on predicted customer segmentation.

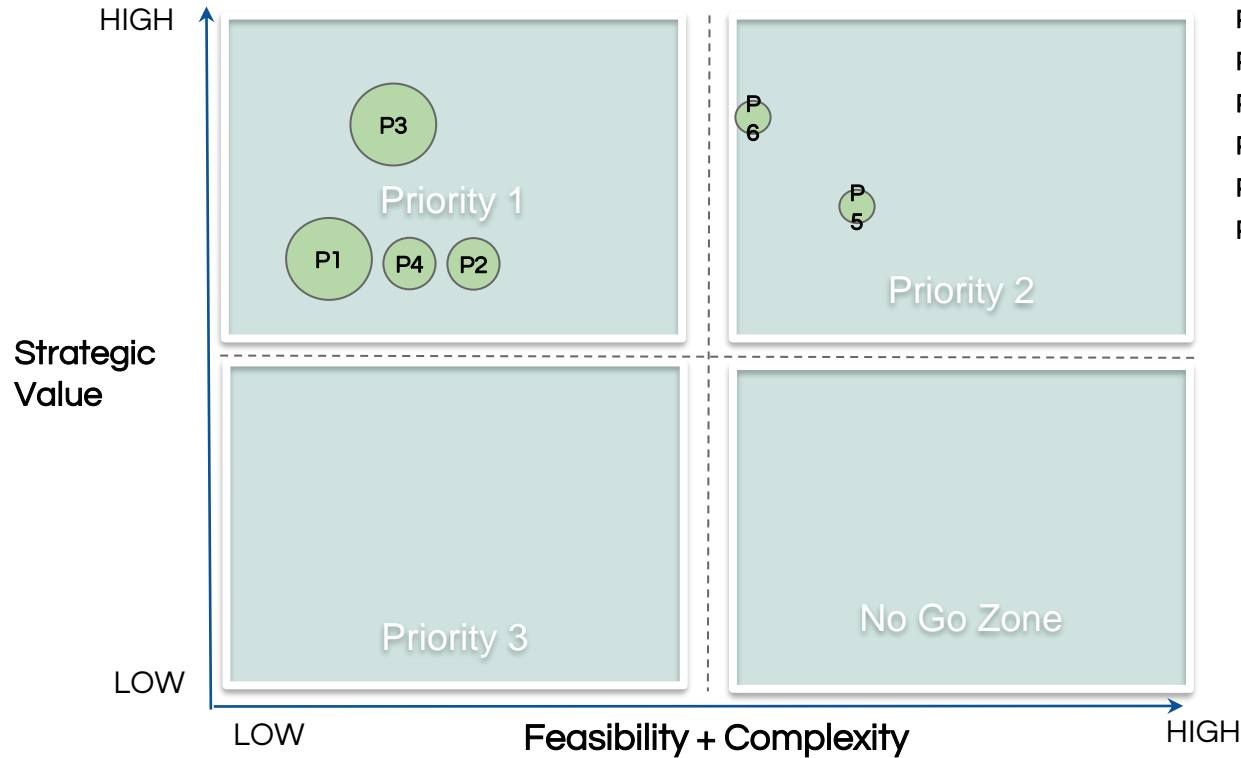
Summary of Projects Prioritization

			Raw Scores							Normalised Scores					
Use Case	Maximum Possible Score	Minimum Score	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	P6	
Feasibility Score	5	0	4.1	3.5	3.7	3.63	3.9	3.27	0.82	0.7	0.74	0.726	0.78	0.654	
Complexity Rating	20	0	16.5	16.83	13.83	14	9.83	12.5	0.825	0.725	0.841	5	0.7	0.625	
Strategic Value	20	0	14	14	18	10	16	18	0.7	0.7	0.9	0.5	0.8	0.9	
Business Value (Enter in numericals the business value generated)	3867	0	3746.05	2764.11	3867	3003	631	1800	0.97	0.71	1.00	0.78	0.16	0.47	
									3.31	2.84	3.48	2.70	2.23	2.64	

Highest-Priority Descending order of Data Science Projects Follow the Money = Business Value then Strategic Value

Order	Project	Higher Value Data more Feasible Data feasibility	Infrastructure feasibility	Complexity	Strategic Value	Business Value
		1=High; 5=Low	1=High; 5=Low	1=High; 5=Low	1=Low; 5=High	1=Small; 5=Large
First	Project 3: Customer Churn	3.4	4	4.2	4.5	5
Second	Project 1: Delivery Date Prediction	4.2	4	4.125	3.5	5
Third	Project 2: Sentiment Analysis	3	4	3.625	3.5	3
Fourth	Project 4: Customer Acquisition Cost Optimisation	3.6	3.67	3.5	2.5	3
Fifth	Project 6: Price Optimisation	3.2	3.33	3.125	4.5	2
Sixth	Project 5: Fraud Detection	3.8	4	2.45	4	1

Data Science Opportunity Matrix : Modeling each of the six projects in terms of feasibility, complexity, strategic and business value impact.



Project 1: Delivery Date Prediction

Project 2: Sentiment Analysis

Project 3: Customer Churn

Project 4: Customer Acquisition Cost Optimisation

Project 5: Fraud Detection

Project 6: Price Optimisation

Data Science Road Map with the first four data science projects chosen for implementation.

<u>Order</u>	<u>Project</u>	<u>Order Justification</u>
1	Project 3: Customer Churn	This project demonstrates both significant business and strategic value, aligning with the organization's vision and mission, and possessing ample resources for effective execution, earning it the highest priority.
2	Project 1: Delivery Date Prediction	This project's substantial business and strategic value, alignment with the organization's vision and mission, and ample resources for effective execution grant it the second highest priority based on the prioritization study.
3	Project 2: Sentiment Analysis	This project demonstrates moderate business value, aligns with three business goals, and is feasible in terms of data and resources, resulting in its third priority based on the conducted studies.
4	Project 4: Customer Acquisition Cost Optimisation	This project exhibits moderate business value, aligns with three business goals, but is less feasible in terms of data and resources compared to the Sentiment Analysis project. Its priority was determined based on thorough studies.

Customer Churn Project Impact on Company Objective

Business Process

- Implementing a machine learning solution automates the churn prediction process, saving time and effort compared to manual analysis. This allows the sales and marketing team to focus on executing effective retention strategies and enhancing the overall customer experience while increasing revenues by focusing only on customers with high probability to churn thus reducing marketing expenses
- Streamlined processes
- Data-driven decision-making..

Monetary:

- OLIST can proactively implement targeted retention strategies, reducing churn rates and improving customer loyalty.
- Reducing churn improves profitability by increasing customer lifetime value. Acquiring new customers can be expensive, while retaining existing customers can be more cost-effective.
- Competitive Advantage being customer centric.
- Saving costs on retention expenses and increasing revenues and customer satisfaction.
 - Example Cost of retention of 1 customer = 10 BRL (assumed) thus saving 80% than money spent on acquisition = 50
Savings = $50 - 10 = 40$ BRL **Total Savings** = $40 * 100$ customer per month assumed = **4000 BRL/ Month savings**
 - Cost of sending 10 BRL discount to customers not churning if done on randomly on 50% of customers $96000 * 0.5 = 48000$ customers * 20 = 9,600,000 RBL on 760 days thus 1263/ day RBL for unoptimized retention while optimized retention will allow you to send to only 10% (assumed) thus $10 * 9600 = 96000$ $96000 / 760 = 126$ BRL/day Savings = $1263 - 126 = 1137$ BRL/day

DS Solution Approach

Predictive Model

1. Use relevant customer data, including purchase history, engagement metrics as, review score, demographic information (location), city, customer_id, product category name, customer_id and customer_unique_id..
2. Clean and preprocess the data, handling missing values, outliers, and ensuring data quality.
3. Extract meaningful features from the data that capture customer behavior, preferences, and interaction patterns frequency of purchase and assume that a customer churns if he didn't purchase 6 months after his first order. Segment customers on RFM after extracting features as frequency, resenancy and monetary and extract the data to be used from model training.
4. Train and optimize a Gradient Boosting model using the prepared dataset to predict customer churn probabilities.
5. Evaluate the model's performance using appropriate metrics as accuracy and confusion matrix, and fine-tune the model if necessary.
6. Deploy the model into the production environment, integrate it with the existing systems, and continuously monitor its performance, updating the model periodically as needed.

Changes in Business Process:

No change will take place in marketing strategy but the marketing and customer service team will now save time and money on retention strategies and ways to improve services along as increasing customers lifetime value (CLV).

Steps of Developing Proof of Concept P.O.C

- a. Gather and preprocess relevant data from the OLIST dataset, including customer attributes, purchase history, engagement metrics, and demographic information. Clean and format the data to ensure its accuracy and suitability for analysis.
- b. Choose appropriate machine learning algorithms, such as logistic regression, XGBoost, or LightGBM, based on the problem's nature. Utilize the prepared data to train the selected models to predict customer churn.
- c. Split the dataset into training and testing sets to assess the model's performance. Measure key metrics like accuracy, precision, recall, and F1-score to evaluate how well the model can identify churn patterns and predict customer behavior.
- d. Analyze the POC's outcomes and compare them against the predefined success criteria. Determine if the developed model effectively identifies significant churn factors and suggests personalized retention actions in line with the business goals.
- e. Based on the POC results, fine-tune the model and incorporate feedback from domain experts to enhance its accuracy and reliability.

Through this process, a successful POC will provide evidence of the DS solution's viability and its potential to address OLIST's strategic goals, helping in making informed decisions before full-scale implementation.

Technical Infrastructure Needed to Support the Data Science Organization

Data Requirements	What data should be included in the Data Strategy?	<ul style="list-style-type: none">● Relevant customer data, including purchase history, engagement metrics (e.g., review scores), demographic information (e.g., city), customer_id, product category name, customer_unique_id.● We will use some unstructured data to convert them to structured data for model training.
Data Governance	Data Availability	<ul style="list-style-type: none">● The required customer data is available in the OLIST dataset, containing approximately 100k orders from 2016 to 2017.● Marketing Funnel Dataset is also available for the same period and orders.
	Usability	<ul style="list-style-type: none">● The data is well-structured and accessible, facilitating its utilization for analytics and machine learning projects.
	Integrity	<ul style="list-style-type: none">● The data source is integrate but depending on the application, the dataset's integrity is crucial to ensure accurate and reliable results. Steps will be taken to validate, clean, and preprocess the data as needed.

Technical Infrastructure Needed to Support the Data Science Organization

Skills and Capacity	Data literacy skills and organizational capacity	<ul style="list-style-type: none">• Our team possesses the necessary data science and analytics expertise to effectively analyze and interpret the data for the proposed projects but as the model complexity increases the team might not be able to move with as the team is small, and resources are limited.
Support for Machine Learning	Machine learning	<ul style="list-style-type: none">• We will leverage various machine learning algorithms, such as Logistic Regression, KNeighborsClassifier, SVC (<u>requires high GPU resources</u>), DecisionTreeClassifier, RandomForestClassifier, LGBMClassifier, XGBoost, or LightGBM, to address different aspects of the projects, including customer churn prediction and personalized retention actions. For batch processing we might not need fast computation but we will need CPU/GPU support until the model is trained for fast implementation and continuous monitoring AWS or Azure could be an temporary solution.

Success Metrics:

1. Churn Prediction Accuracy: Metrics such as accuracy, precision, recall, and F1 score. A higher accuracy indicates a more reliable model for identifying at-risk customers.
2. Churn Reduction Rate: This metric reflects the effectiveness of the implemented strategies in reducing churn and improving customer loyalty.
3. Customer Lifetime Value (CLV) Improvement: Monitor changes in CLV for customers who were identified as at-risk but successfully retained. An increase in CLV indicates that the implemented retention initiatives have positively impacted customer behavior and long-term value.

Conclusion

In conclusion, OLIST's strategic goals and objectives are aligned with the proposed data science projects, aimed at increasing revenue, improving efficiency, enhancing customer experience, and growing the number of active customers. Among the six proposed projects, Customer Churn, Delivery Date Prediction, Sentiment Analysis, and Customer Acquisition Cost Optimization stand out as the highest priority projects due to their substantial business value, strategic importance, and feasibility in terms of data and resources.