AI for communication and marketing

Churn prediction model

Chiara Barbieri, 517096

Index

Business context

The chief marketing officer of a retail e-commerce platform is concerned about a growing rate of customers churning, which negatively effects the average order value (how much is spent on each order) and the overall customer lifetime value (customer future value).

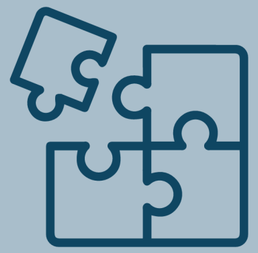
To reduce this trend, a churn prediction model, namely a model capable of predicting whether a certain customer is at high risk of churn, so at risk of not purchasing anymore from the company, has to be created and employed. This model will allow to boost loyalty and profitability, since it leads to aimed data-driven retention strategies, such as personalized offers or targeted communications.

Project Objectives

Employ explorative data analysis

Analyze the data to understand:

- The distribution of the features
- The correlation between features to understand the relationships between them



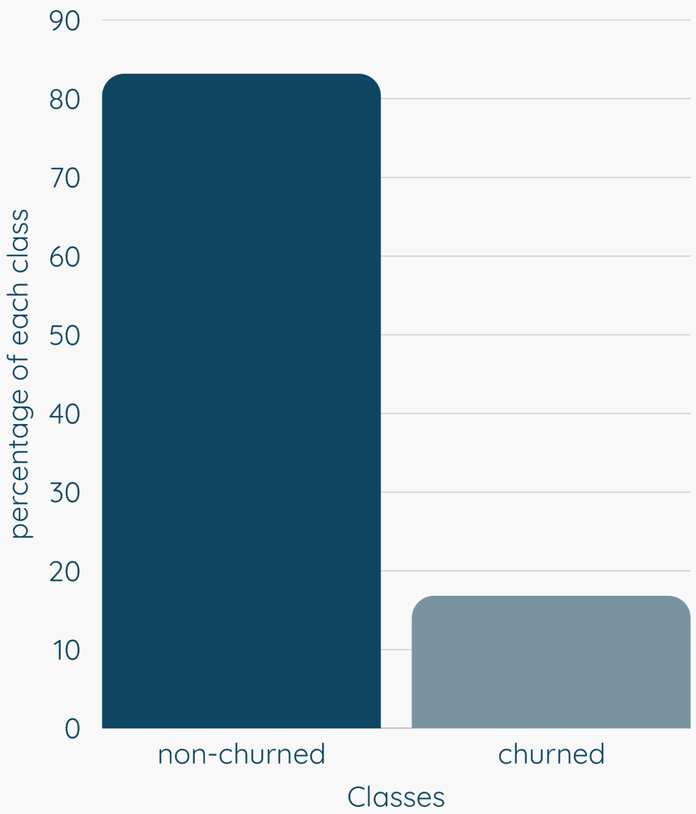
Develop a churn prediction model

This is to identify high-risk of churn customers, more precisely predict in which class (low, medium and high risk of churn) the customers belong to

Give a data-driven business insight

Given the different churn classes, report the key discoveries made through data and the different marketing strategies that could be used for the retention of customers

Dataset description



Data in the dataset

- Type of data: transactional data and customer behavioural data
- The features represent different aspects of the experience of the customer with the company

Target

- The target is the feature “Churn”, this feature can have 2 possible values:
  - 0: non-churned customer (83.17% of customers)
  - 1: churned customer (16.83% of customers)

Sample size

- Number of customers included in the analysis: 5630
  - After preprocessing we have 5626
- Number of features included in the dataset: 20
  - ‘Customers ID” is not used for training
  - ‘Churn’ is the label

Key variables

Tenure
Amount of time the customer has been loyal to the company

Warehouse To Home
Distance between the warehouse and the customer's home

Satisfaction Score
Satisfaction in the products and the services of the company

OrderCount
Total number of orders placed during the previous month

Complain
Whether complains were raised in the last month

Order Amount Hike From last Year
Percentage increases in orders with respect to the previous year

CouponUsed
Total number of coupon used during the last month

Cashback Amount
Amount of cash refound received by a customer after a purchase

Days Since Last Order
Days that have passed since the last order made by the customer

Analytical methodology

Explorative analysis and variable selection

The main steps taken are:

- Dealing with null values
- Understanding the distribution of the attributes and making the appropriate changes
- Definition of the correlation matrix to understand the relationships between variables
- Plotting the repurchase curve

Data balancing

Technique used: SMOTE (Synthetic Minority Over-Sampling)

This technique was chosen because it creates new data for the minority class without removing samples from the majority class.

Model configuration and optimization

- Model used: Gradient Boosting (XGBoosting)
- Optimization technique: Bayesian optimization using cross validation and the hyperparameter space

Churn model Pipeline

1. Familiarization with the dataset
2. Explorative analysis & variable selection
3. Data splitting
4. Stratified over sampling for class balance on the train set
5. Model configuration & optimization
6. Model evaluation
7. Definition of churn risk classes
8. Business insight

Analytical methodology

Model evaluation

Validation of the model done using the following methods:

- roc_auc score and the ROC curve plot
- recall score, precision score, accuracy score, f1 score
- Confusion matrix and classification report

Definition of churn risk classes

The churn risk classes were created using the probability of churn found by the model.

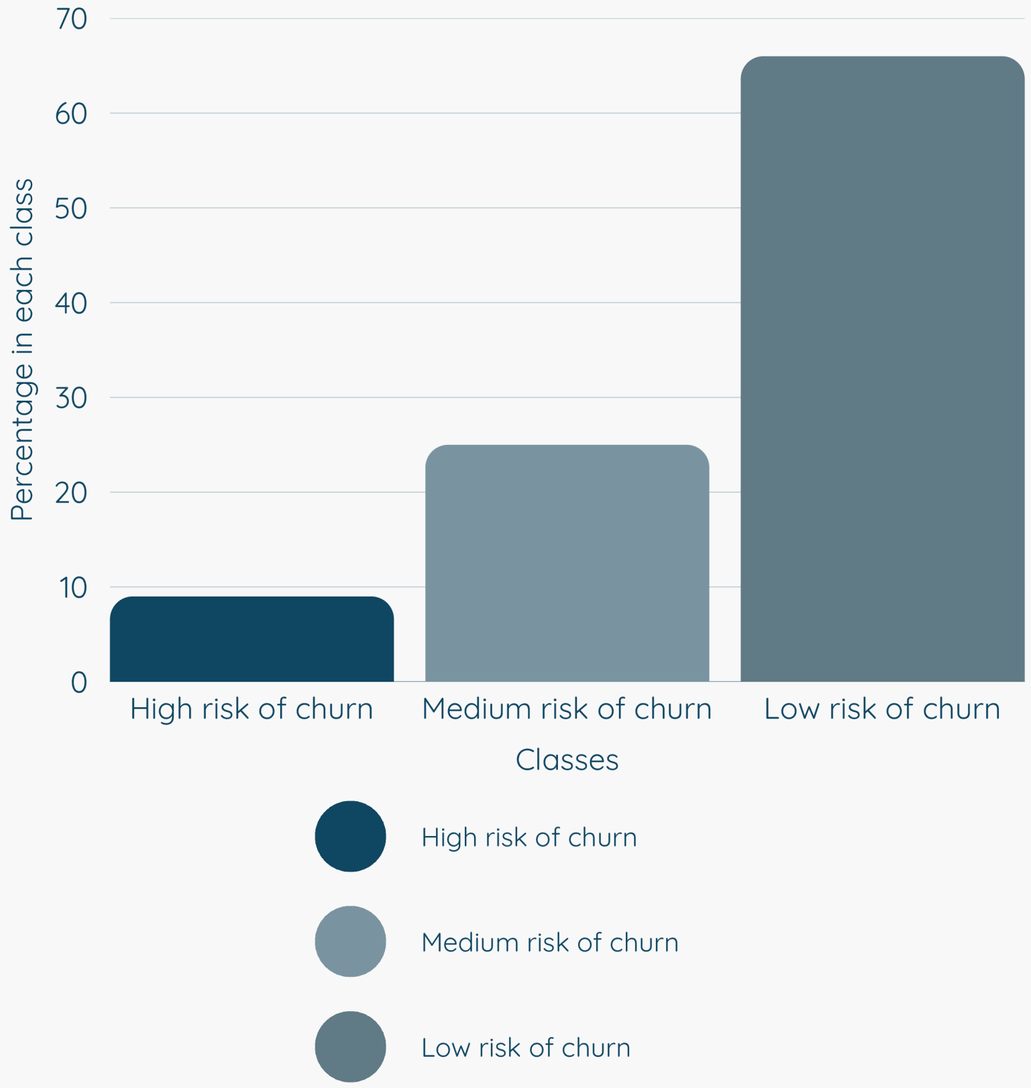
- If churn probability is more or equal to 0.7: **High-risk of churn class**
- If churn probability is more or equal to 0.3 and below 0.7: **Medium-risk of churn class**
- If churn probability is less then 0.3: **Low-risk of churn class**

Moreover, also the importance of the feature for the prediction of churn is defined

Churn model Pipeline

1. Familiarization with the dataset
2. Explorative analysis & variable selection
3. Data splitting
4. Stratified over sampling for class balance on the train set
5. Model configuration & optimization
6. Model evaluation
7. Definition of churn risk classes
8. Business insight

Assessment of the business insights



The distribution of the classes shows that 34% of the customers are at risk of churn. The 4 factors that most influence the customer's churning are:

- The cashback amount
- The distance between the warehouse and their house
- The days since the last purchase
- The % of more goods ordered during the previous year

Finally from the correlation matrix we remember that:

- High correlation between the number of complaints and the likelihood of churn
- The churn probability decreases as tenure increases
- The number of coupons used, the number of orders done and the days that have passed since the last order are all directly proportional

Practical recommendations

Improve customer service

Resolve complaints faster and follow up proactively to address root causes. This is done for boosting satisfaction and trust.

Design targeted retention campaigns

Deliver personalized offers depending on the churn class

Optimize the coupon strategy and the cashback amount

Use smaller and frequent rewards to encourage steady purchasing, so encourage a higher shopping frequency

Enhance delivery experience

Find techniques to shorten delivery times & add real-time tracking of the package and be transparent about waiting times

Strenghten loyalty programs

Give to the customers rewards proportional to the tenure and purchasing history of the customer, done to incetivize loyalty

Thank you