E-COMMERCE CUSTOMER CHURN PREDICTION

TEAM MEMBERS

- Wendy Oira - **Team Leader**
- Edward Sang - **ATL**
- Emmanuel Titi
- Kevin Kiprono
- Ruth Nguli
- Jasmine Wakini

BUSINESS UNDERSTANDING

Introduction

Customer churn is when a proportion of your customers decide to stop supporting your brand. This could mean a lot of different things, depending on your business.

For example, customer churn could be when customers choose not to renew their subscription or stop purchasing items from your online store. Because retailers have to find a way to replace this lost business, customer churn presents a long-term problem for many brands, especially if they have a niche audience or similar offerings to a competitor.

Problem Statement

The e-commerce companies must either retain the existing customers or draw new customers. But the latter is difficult to achieve as there are less chances of that happening. Henceforth, this leaves us with the major option of retaining existing customers and in order to do so, different marketing strategies need to be considered by the companies. The predictive models can enable the businesses to recognize the customers who are going to churn and come up with ways to re-engage with their customers and prevent them from churning.

Main objective

To predict whether a customer will churn or not based on a customer's profile and activities.

Specific Objectives

1. To identify whether a preferred mode of payment for a specific customer is a determinant of churning.
2. To determine whether a customer who previously made a complaint will churn.
3. To determine if a customers access to the mobile application and website affects churning
4. To determine if the distance of the customers home from the warehouse will cause churning ,in this case, will a longer distance cause churning.

Research Questions

1. Which preferred mode of payment for a specific customer is a determinant of churning?
2. Which of the customers who previously made a complaint will churn?
3. Will customers access to the mobile application and website affects churning?
4. Will the distance of the customers home from the warehouse will cause churning ,in this case, will a longer distance cause churning?

Business Impact

The project will help Jumia, Jiji and Kilimall come up with key customers who are likely to churn so that they can come up with ways to re-engage with their customers and prevent them from churning.

DATA UNDERSTANDING

Data Description

We got our data from Kaggle
The data set we're using is E commerce customer churn analysis and prediction

The data set belongs to a leading online E-Commerce company. An online retail (E commerce) company wants to know the customers who are going to churn, so accordingly they can approach customers to offer some promos.

The dataset contains 5,630 records and 20 columns

Explore Data

Data Glossary:

- CustomerID - Unique customer ID
- Churn - Churn Flag
- Tenure - Tenure of customer in organization
- PreferredLoginDevice - Preferred login device of customer
- CityTier - City tier
- WarehouseToHome - Distance in between warehouse to home of customer
- PreferredPaymentMode - Preferred payment method of customer
- Gender - Gender of customer
- HourSpendOnApp - Number of hours spend on mobile application or website
- NumberOfDeviceRegistered - Total number of deceives is registered on particular customer
- PreferedOrderCat - Preferred order category of customer in last month
- SatisfactionScore - Satisfactory score of customer on service
- MaritalStatus - Marital status of customer

- NumberOfAddress - Total number of added added on particular customer
- Complain - Any complaint has been raised in last month
- OrderAmountHikeFromlastYear - Percentage increases in order from last year
- CouponUsed - Total number of coupon has been used in last month
- OrderCount - Total number of orders has been places in last month
- DaySinceLastOrder - Day Since last order by customer
- CashbackAmount - Average cashback in last month

Our data is made up of float, object and integer data types.

Verify Data Quality

The dataset contains necessary variables for our research questions

DATA PREPARATION

Loading the Data

Loaded the dataset on the colab notebook

Data Cleaning

We have missing values in the Tenure, WarehouseToHome, HourSpendOnApp, OrderAmountHikeFromlastYear, CouponUsed,OrderCount and DaySinceLastOrder variables .
For HourSpendOnApp we replaced the means with mean
For the rest of the columns we used mode
No duplicate values.
Outliers were present in our data frame but we did not delete them because they are a true representation of real life situations.

MODELLING

Evaluate Results

We used 4 models for our analysis;

1. Random Forest which gave us an accuracy of 98.86%
2. Gradient Boost which gave us an accuracy of 99.29%
3. For Support Vector Machine we looked at 3 kernels those are linear, polynomial and rbf in which polynomial performed best with an accuracy of 98.36%

To challenge our solution we used another model that is K-Nearest Neighbor which gave us an accuracy of 91.81%

All our models were subjected to hyper parameterization through GridSearchCV or RandomSearchCV for optimal parameters.

The Gradient Boost model gave us the highest accuracy out of the 4 models.

Review Process

1. Define the question, the metric for success, the context, experimental design taken.
2. Read and explore the given dataset.
3. Define the appropriateness of the available data to answer the given question.
4. Find and deal with outliers, anomalies, and missing data within the dataset.
5. Perform univariate, bivariate recording your observations.
6. Performing oversampling
7. Performing *Random forest, Gradient Boost and SVM*
8. Challenging the Solution using *KNN*
9. Conclusion
10. Recommendations

Conclusions

Most customers who churned:

1. were male
2. used debit cards as their preferred mode of payment
3. had made complaints
4. used mobile phones as their preferred login device
5. lived a distance of 14 from the warehouse
6. were single
7. had 4 devices registered

Recommendations

1. Preemptively engaging at risk customers by offering them something special; a promo, discount, loyalty program, etc. This small effort can go a long way when it comes to showing your customers how much you value their business.

2. Establishing more warehouses to ensure customers have easy access to the warehouse, this would also reduce time taken in our doorstep delivery.

3. Aiding in partial coverage of the transaction cost incurred by customers in the different modes of payments

4. Develop a complaints handling policy. It should include reassuring customers that you value their feedback and you are committed to resolving their issues in a fair, timely and efficient manner.

5. Following up on customer complaints ,incase of defective product claims the involved seller should be barred from making sales in our websites/mobile applications.

6. Optimize your E-Commerce Website For Mobile Devices by Speeding up your website for mobile devices and ensure that your content is mobile-optimized and also ensuring there are no pop ups on your website