

Report on Customer Offer Response Analysis

Introduction

Customer behavior analysis is essential for businesses to enhance their customer base and maintain customer loyalty. It helps them to understand their customers' likes, dislikes, and habits that can be used to provide a personalized experience to each customer. In this report, we present an analysis of customer offer response, using data provided by Starbucks. The dataset is provided in three files, portfolio.json, profile.json, and transcript.json. The dataset has been analyzed using Python and the relevant libraries.

Data Overview

The data provided contains information on customer demographics, their transaction details, and the different types of offers available to customers. The portfolio dataset contains information on the various offers available, including the offer type, duration, and reward. The profile dataset contains information on customer demographics such as age, income, and gender, as well as the date on which they became a member. The transcript dataset contains the transaction details, including the time of the transaction, the type of event, and the offer details.

Data Preprocessing

The first step in data analysis is data preprocessing, which involves cleaning and transforming the data to make it ready for analysis. The preprocessing steps for each dataset are as follows.

Portfolio Dataset

The portfolio dataset contains information on the different types of offers available. The first step is to extract the different channels available for each offer and create a new column for each channel. The value in each channel column is 1 if the channel is available for the offer, and 0 otherwise. We then create a new column that sums the values of all the channels for each offer. We then plot a pie chart to visualize the distribution of offer types.

Profile Dataset

The profile dataset contains information on customer demographics. The first step is to convert the date format of the `became_member_on` column into a datetime format. We then extract the year, month, and day from the date and create separate columns for each. We then remove all rows with missing gender values and convert the gender column into a binary column where 1 represents male and 0 represents female. We then sort the dataset based on the `became_member_on` column in descending order and rename it to `became_member_on`.

Transcript Dataset

The transcript dataset contains transaction details, including the time of the transaction, the type of event, and the offer details. We first plot a bar chart to visualize the distribution of event types. We then extract the `offer_id`, amount, and reward values from the value column and create separate columns for each. We then convert the time column from hours to days. We then use one-hot encoding to create binary columns for each event type, group the data by `user_id`, and sum the values for each column.

Modeling

The next step is modeling, which involves using the preprocessed data to build a predictive model. We merge the three datasets based on the `user_id` and `offer_id` columns to create the final dataset for modeling. We then select the relevant columns for modeling and split the data into training and testing sets. We use a Decision Tree Classifier, Linear Regression, and Logistic Regression models to predict the offer completed column. We then evaluate the performance of each model using the R2 score, Mean Squared Error (MSE), Accuracy Score, and Confusion Matrix.

Normalization

The final step is normalization, which involves scaling the numerical data to a common range. We use `MinMaxScaler` to scale the time, amount, reward, age, and income columns to a range between 0 and 1.

Results

From the accuracy scores we find that the Decision tree classifier has the best accuracy. It seems that the classification models work better with the dataset but it only show if the user will accept an offer or not, but won't show the number of offers accepted per user it depends on what do stakeholders need.

Conclusion

The goal of this project is to predict how the customer will interact with the offers that Starbuck will present. Firstly, I took the provided data that requires some cleaning activity, then I did the needed changes to analyze that data that required to be applied before starting the data exploring activity. Finally, I used three models first LinearRegression model I got nearly 84%, Second, DecisionTreeClassifier I got 100%, third, LogisticRegression I got 97% accuracy. It seems that the classification models work better with the dataset but it only shows if the user will accept an offer or not but won't show the number of offers accepted per user it depends on what do stakeholders need. Moreover, Starbuck can use these models to enhance their offers periodically after each offer to know the real benefits to aim their offer to the correct audience