Marketing Promotion Campaign Uplift Modelling

(Customer Retention data for Churn Prediction or Uplift Modelling)

Aim:

Predict customer's conversion rate
Increase the IRR (Incremental Response Rate)
Given the details of a customer, predict if he falls into the persuadable category or not

Theory:

Uplift model helps us to predict the lift for each customer and then empowers us to target only the high-lift customers.Based on the actions that a customer takes, they can be put into 4 groups. The four groups are shown in the figure. We are interested in finding out customers who fall in the first group: Persuadables. And if we target only these customers, then the campaign will give high lift as they wouldn't have taken any action without any campaign (Low response rates in Control group and high response rates in Treatment group). The second group: The Sure Things is the one which would have anyway activated and hence should be left out of the campaign. And obviously, group three and four should also not be targeted. Thus, to summarize target Group 1 and leave out all the other groups.

Uplift model also uses previous campaign's data to predict the lift of each customer in the future campaigns.

We have used the following key metric to track the performance of our model:

Incremental Response Rate (IRR):

IRR measures how many more customers purchased the product with the promotion, as compared to if they didn't receive the promotion.

Mathematically, it's the ratio of the number of purchasers in the promotion group to the total number of customers in the purchasers group (*treatment*) minus the ratio of the number of purchasers in the non-promotional group to the total number of customers in the non-promotional group (*control*).

$$IRR = \frac{purch_{treat}}{cust_{treat}} - \frac{purch_{ctrl}}{cust_{ctrl}}$$

Technical aspects:

If we take a look at the "conversion" column -

Number of customers who bought the product (No. of 1s in the conversion column) = 8455
Number of customers who did not buy the product (No. of 0s in the conversion column) = 49145

This shows that the dataset is highly imbalanced and we need a technique to handle such an imbalance. Failing to do so would result in our Machine Learning model to predict everyone to have not bought the product.

We will be using **SMOTE (Synthetic Minority Oversampling Technique)** for handling such an imbalance i.e. We will be up-sampling the minority classes in the training data with SMOTE. Up-sampling would result in an equal number of 0s and 1s (number of 1s getting increased to match the number of 0s) in the conversion column while generating dummy values for the rest of the columns which mimic the real data that we have. Note that we'll be performing up-sampling only on our training data.

```
# up sample only the train dataset with SMOTE
sm = SMOTE(random_state=42, ratio = 1.0)
X_train_upsamp, Y_train_upsamp = sm.fit_sample(X_train, Y_train)

X_train_upsamp = pd.DataFrame(X_train_upsamp, columns=features)

Y_train_upsamp = pd.Series(Y_train_upsamp)
```

Y_train_upsamp now has 49145 0s and an equal number of 1s.

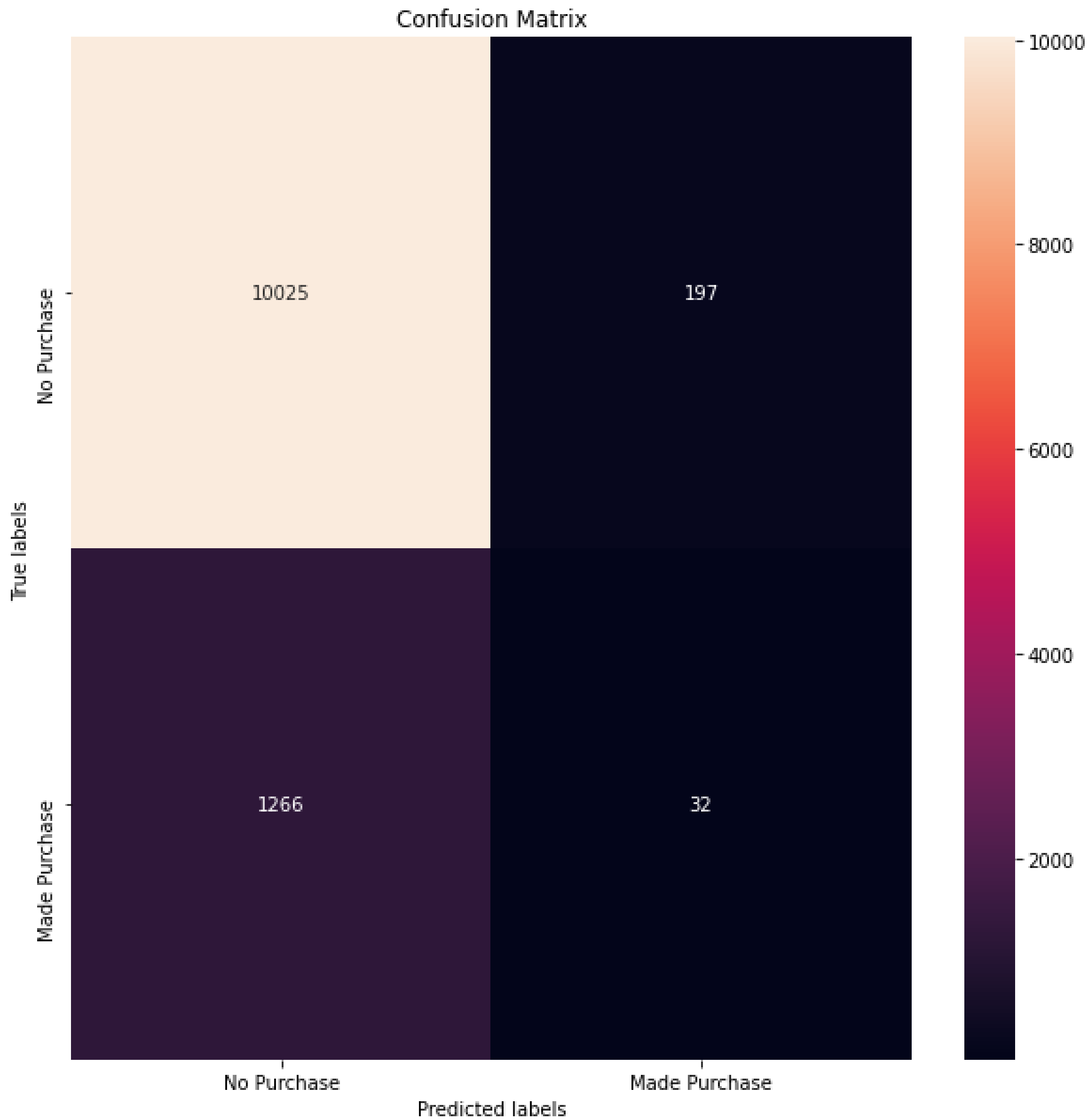
We have used the XGBoost algorithm to train our Machine Learning model. XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. Presented in a research paper in 2016, this algorithm is considered the best in class for small-to-medium structured/tabular data.

```
# Train an xgboost model
eval_set = [(X_train_upsamp, Y_train_upsamp), (X_valid, Y_valid)]
model = xgb.XGBClassifier(learning_rate = 0.1,\
                          max_depth = 7,\
                          min_child_weight = 5,\
                          objective = 'binary:logistic',\
                          seed = 42,\
                          gamma = 0.1,\
                          silent = True)
model.fit(X_train_upsamp, Y_train_upsamp, eval_set=eval_set,\
          eval_metric="auc", verbose=True, early_stopping_rounds=30)
```

Hyperparameter tuning was done to come up with the parameters used in this model.

Now our model is ready to make predictions. If the model predicts a label of 1 for that individual, then it is likely that the individual will respond favorably to the promotion campaign and we should send him or her the promotion. Otherwise, we should not send a promotion.

We obtained the following results:



This model achieved an **IRR of 10.66%** on the validation set which is 4.50% more than what our baseline model (sending promotions to everyone) produced which is pretty remarkable.

Procedure:

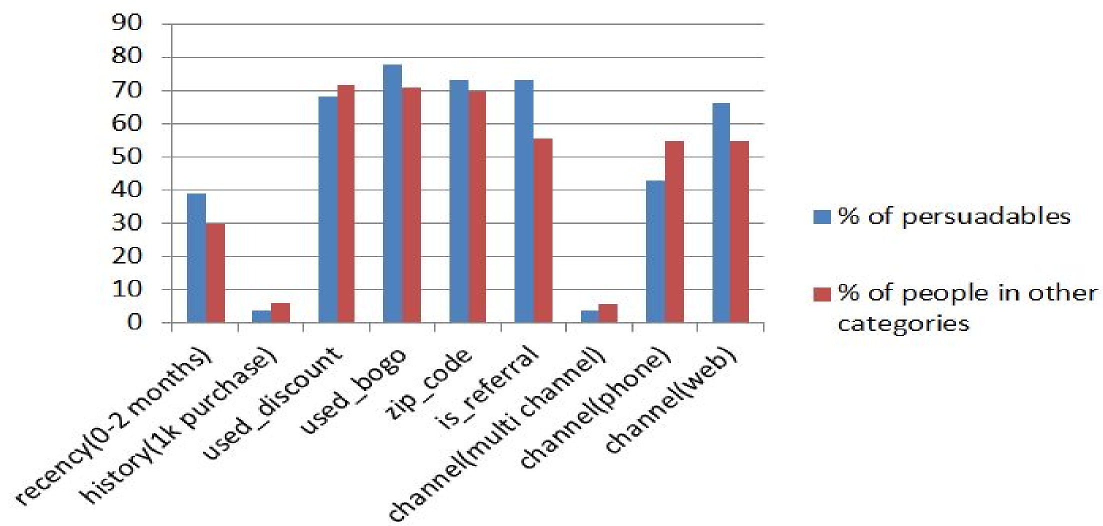
Total count of dataset values= 57,520
Total count of dataset values in training set= 46,000
Total count of dataset values in testing/valid set= 11,520
Total no.of persuadables=5138

Total no.of people in remaining 3 categories(excluding persuadables)= 46,000-5136= 40,942

Average no. of people in remaining 3 categories(excluding persuadables)= 40,942/3=13,647

Note: In the given data, there were 3 categories of zip_code- urban, suburban and rural. Out of all the 3 categories, sub_urban showed the highest data values, and hence only sub_urban values
have been taken into consideration while plotting the graph in the feature zip_code.

Persuadables Vs. Unpersuadables



After plotting the training dataset on the graph,the following observations are made:

- Persuadables have bought more products in the last 0-2 months as compared to people in other categories by 9%.
- Persuadables have less purchase history of Rs.1000 as compared to people in other categories by a factor of 2.4%.
- Persuadables have used less discounts as compared to people in other categories by a factor of 3.7%.
- Persuadables have used more buy one get one as compared to people in other categories by a factor of 6.73%.
- Persuadables have used referral more one as compared to people in other categories by a factor of 3.4%.
- Persuadables belong to suburban areas more as compared to people in other categories by a factor of 17.4%.
- Persuadables are targeted via multi channels modes less as compared to people in other categories by a factor of 2%.
- Persuadables are targeted via phone less as compared to people in other categories by a factor of 12.1%.
- Persuadables are targeted via the web more as compared to people in other categories by a factor of 11.15%.

Result (Graph Analysis) :

- Recency- It can be deduced from the graph that persuadables are inclined towards buying more products when they're targeted in a shorter time span of their last buying as compared to people from other categories.
- History- It can be deduced from the graph that persuadables have less buying history than people from other categories because they are more influenced by the marketing and promotional activities.
- Used_discount- It can be deduced from the graph that persuadables are less interested in buying discounted products as compared to people from other categories.
- Used_bogo- It can be deduced from the graph that persuadables are more interested in buying products using buy one get one offer as compared to people from other categories.
- This implies that persuadables are more inclined towards buying a product using buy one get one offer as compared discounts. This pointer can be taken into consideration while designing future marketing campaigns by various companies, to target the right set of people using the right strategy.
- Is_referral- It can be deduced from the graph that persuadables are more inclined towards using referral codes while buying products as compared to people from other categories. Referral codes play a major role to efficiently connect the referral to the customer who sent them to the business.
- zip_code- It can be deduced from the graph that persuadables belong to suburban cities more as compared to people from other categories. This is because suburban cities majorly consist of middle class households, where equal importance is given to both the necessary product as well as to the discounts/offers associated with it.
- channel(multichannel/phone/web)- It can be deduced from the graph that persuadables are inclined towards buying more products when they're targeted through the web,for example through Google ads,social media platforms, SEO prioritization as compared to people from other categories. The order in which persuadables respond to the various marketing campaigns (in descending order) are through web, phone and multi channels.
- IRR - The baseline model had an IRR of 6.16% on the training set. The trained model was able to improve it upto 10.66% on the validation set which is an increase of 4.50% from the baseline model.

Conclusion:

We have successfully conducted our project of predicting the customer's conversion rate.We have applied the XGBoost algorithm on our dataset. After doing so, we concluded that the best features that are required to determine the design of a marketing campaign are recency of the latest bought products, history of purchased items, history of previously used buy one get one offer, referral code and online marketing. We found out that the persuadables have an inclination towards buy one get one offer more, as compared to discounted products. Our final

Incremental Response Rate(IRR) value is 10.66% on the validation set.This means that 10.66% more customers purchased the product with the promotion, as compared to if they didn’t receive the promotion.