

A/B HYPOTHESIS TESTING

AD CAMPAIGN PERFORMANCE

INTRODUCTION

- Brand awareness is key in every businesses. It plays a big role in driving consumers to a company. Smart Ad is an advertising company that seeks to test the brand awareness by assessing the users engagement with an ad on online platforms.
- The goal of this research is to use A/B testing to determine if the Ads contribute to a significant lift in brand awareness by splitting the users into two experimental groups, control group and exposed group.

OBJECTIVES

- The main objective of this project is to test if an ads that an advertising company runs resulted in a significant lift in brand awareness.

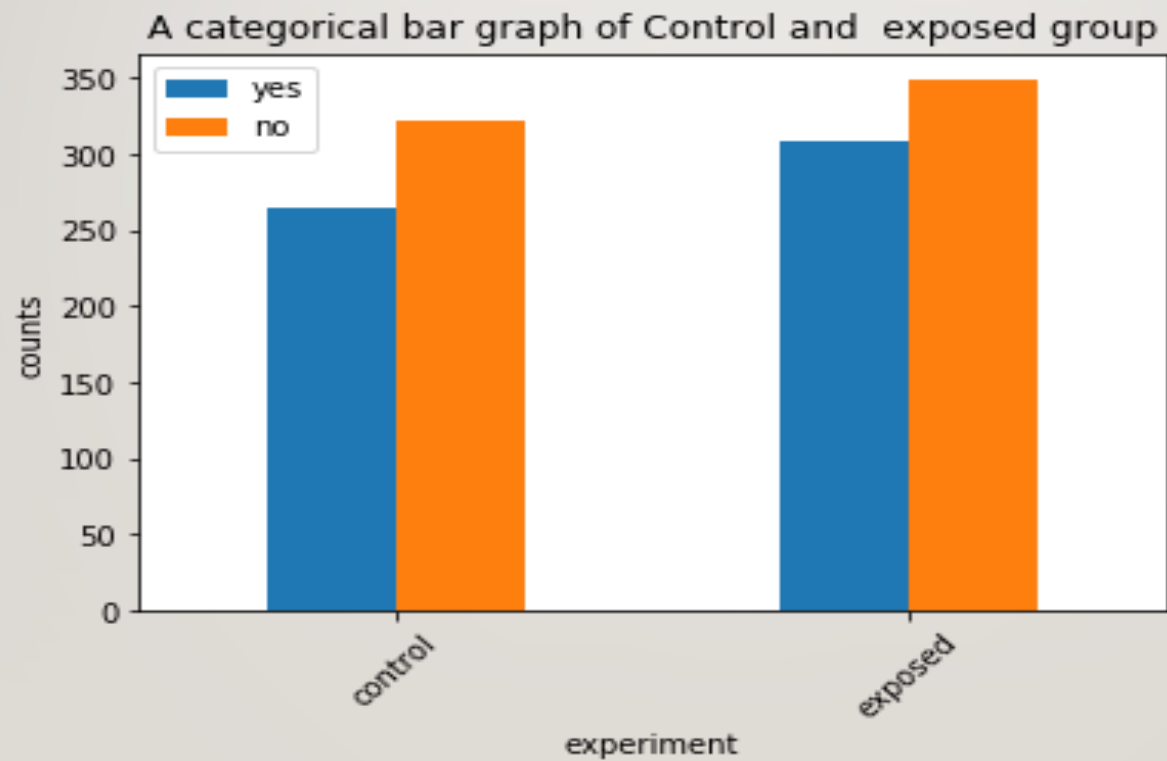
METHODS

- There are three different methods that are used in performing A/B testing:
 - i. Classical/ frequentist A/B testing
 - ii. Sequential A/B testing
 - iii. Using Machine learning
- Sequential A/B testing, in comparison to classical optimizes necessary traffic volumes while reducing the likelihood of mistakes at the same time.
- Machine learning requires formatted data that can be used for modelling. This will involve doing some feature engineering, dropping unnecessary values, handling missing data and rearranging the columns.
- In this project, we only used Classical A/B testing and Machine learning.

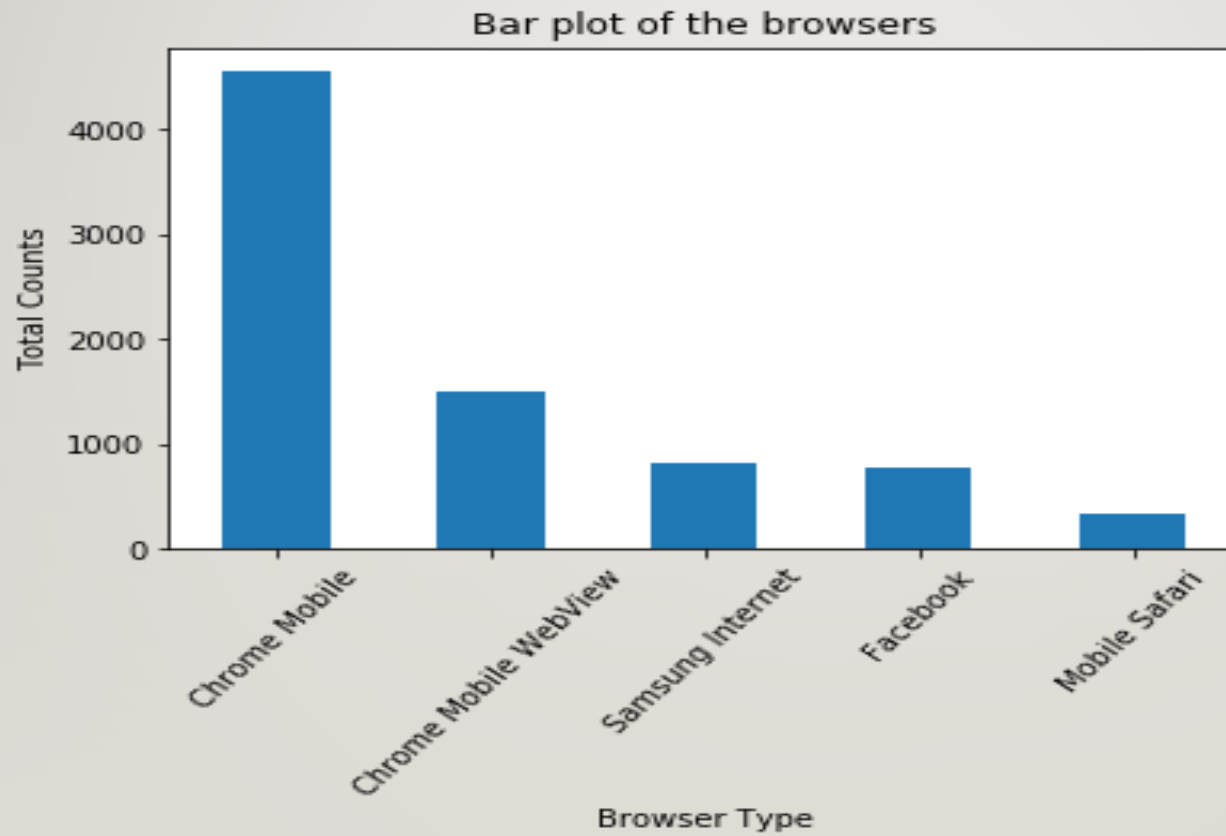
DATA

- Data is obtained using Brand Input Optimizers (BIO) questionnaire. The total population of the users in the website is split randomly into control and exposed groups by being shown a dummy ad and a creative, an online interactive ad, with the Smart Ad brand respectively. The users belonging to the control group are the users who have been shown a dummy ad while the ones belonging to the exposed group are the users who have been shown a creative, an online interactive ad, with the Smart Ad brand. Each user can be tracked by users' session identifier which is stored by the user's cookies.
- The following slides shows the total counts of different categorical variables.

DATA EXPLORATION



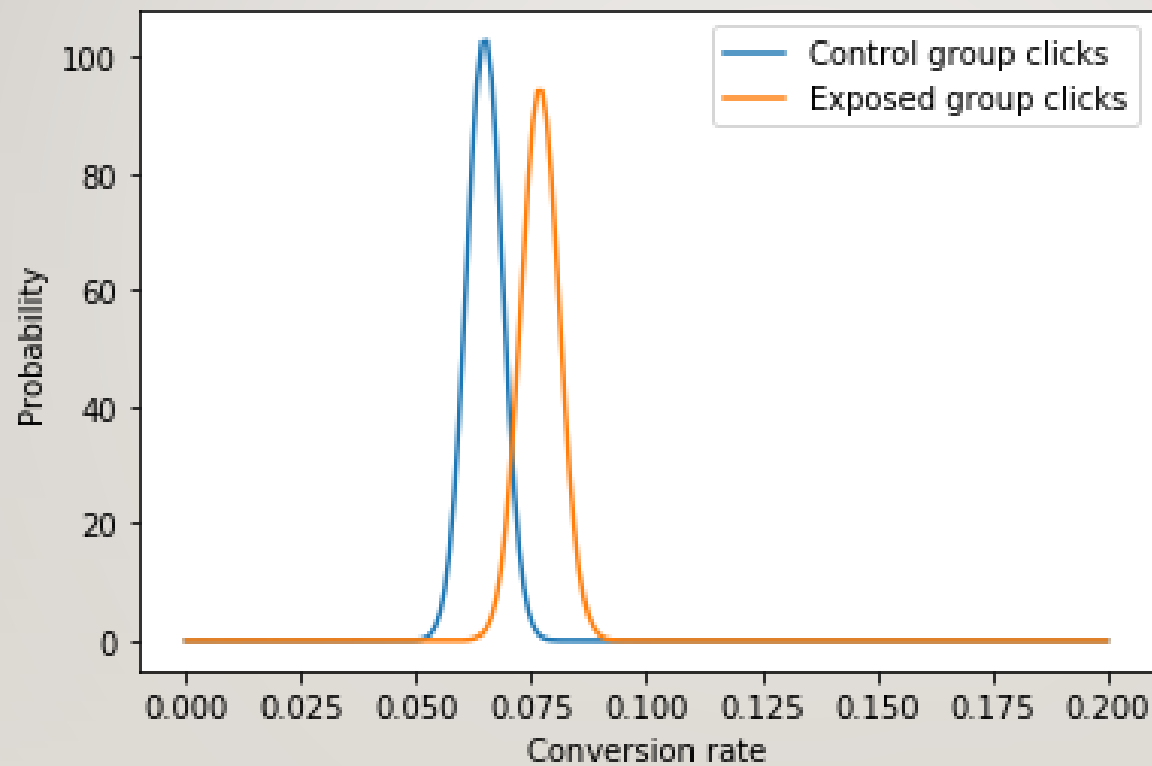
DATA EXPLORATION



RESULTS

- We performed an A/B hypothesis test under the following hypothesis.
 - I. H_0 : There is no significant difference between the two groups
 - II. H_1 : There is a significant difference between the two groups
- We then reject or fail to reject the null hypothesis depending on the calculated p-value comparison with the determined significant level, for this case we use 5%
- We first determine the conversion rate by using click rates, the probability conversion rate is shown in the following slide.

CONVERSION RATE PER GROUP

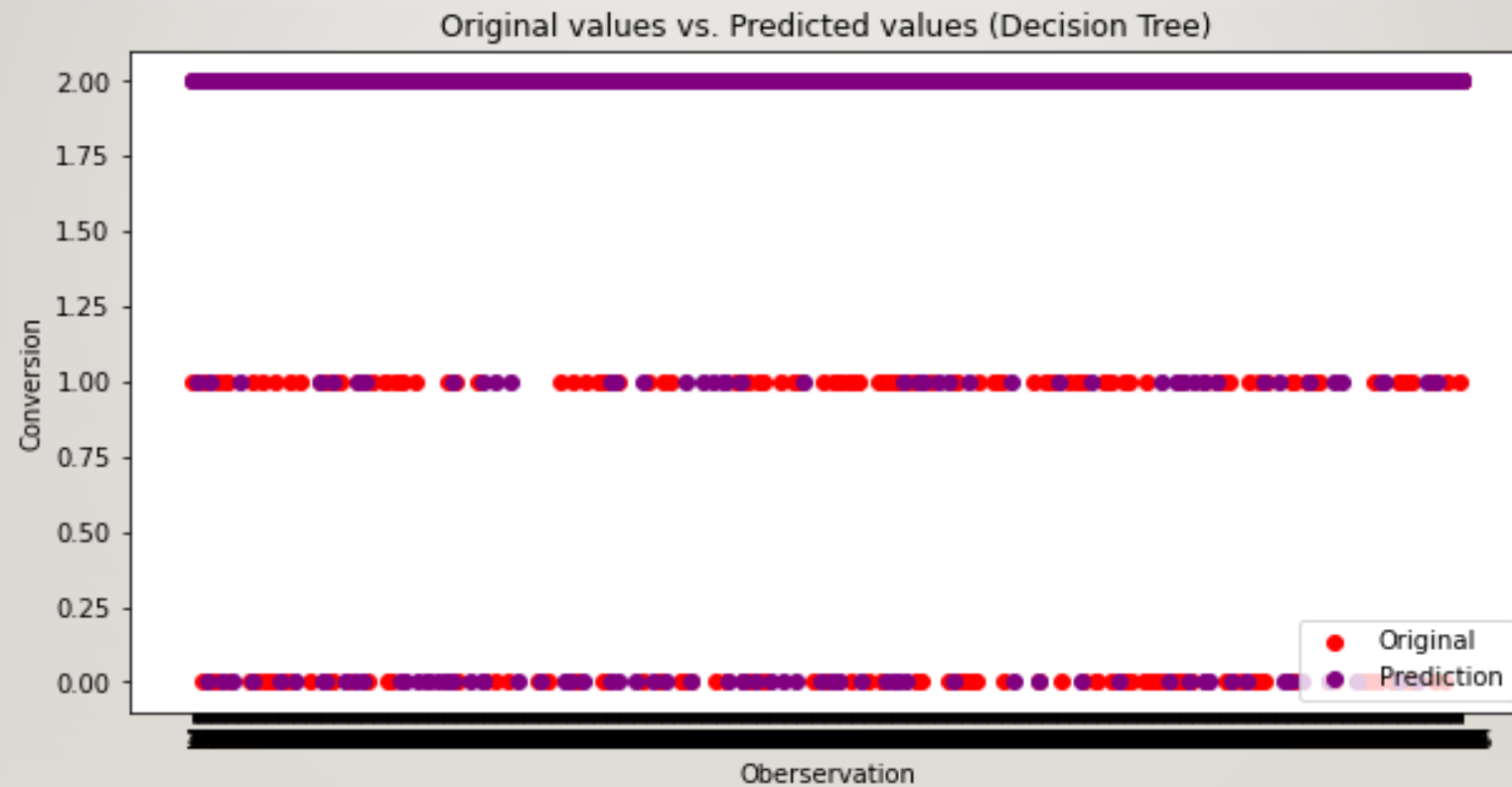


-
- From the plot above, the control group has a higher conversion probability than the exposed group based on click rates.
 - Using click rates only may not be conclusive enough since click may have happen by random chance hence the need to determine the p-value from the z-score to ascertain which group has a higher conversion probability.
 - We found the p-value to be 0.018 which is less than the significant level hence we rejected the null hypothesis. Therefore there is significant lift in brand awareness in the exposed group.

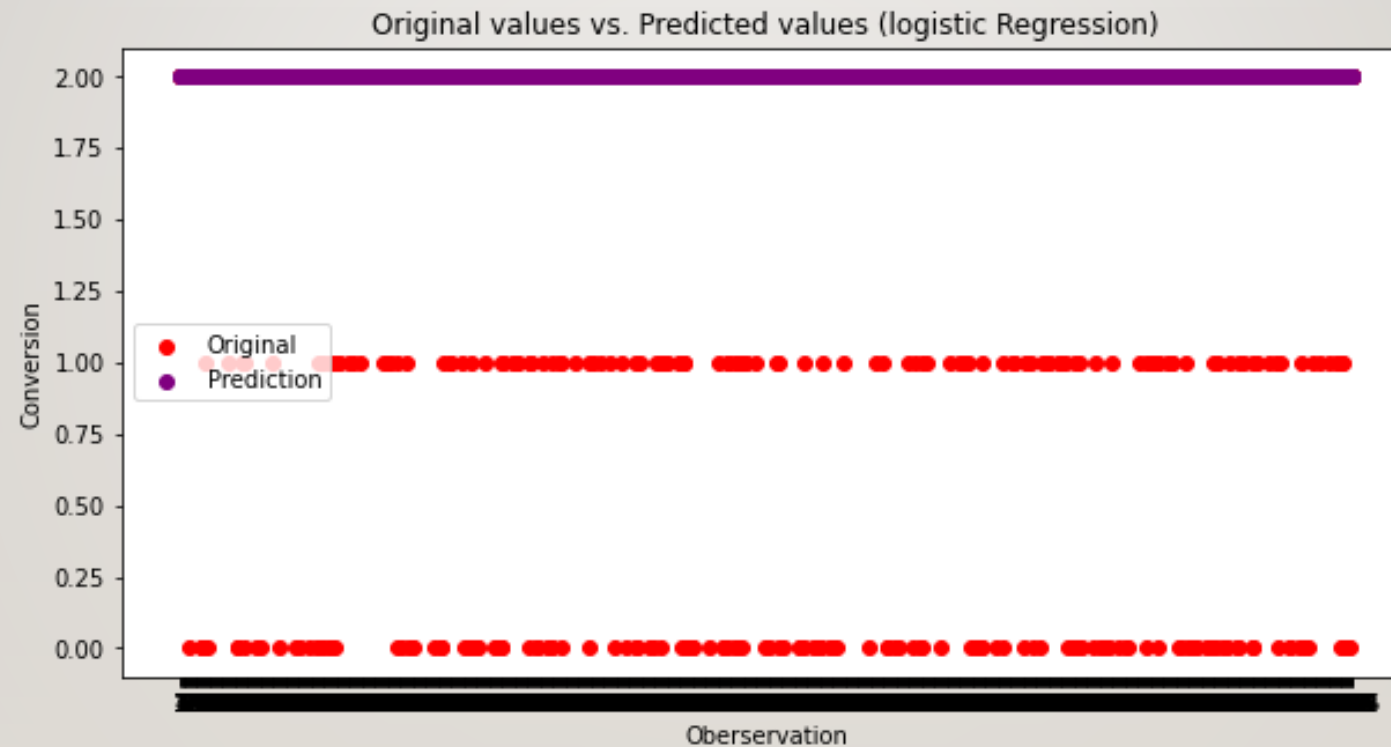
MACHINE LEARNING

- Using different models in machine learning we classify the training features into different clusters to predict the target variable.

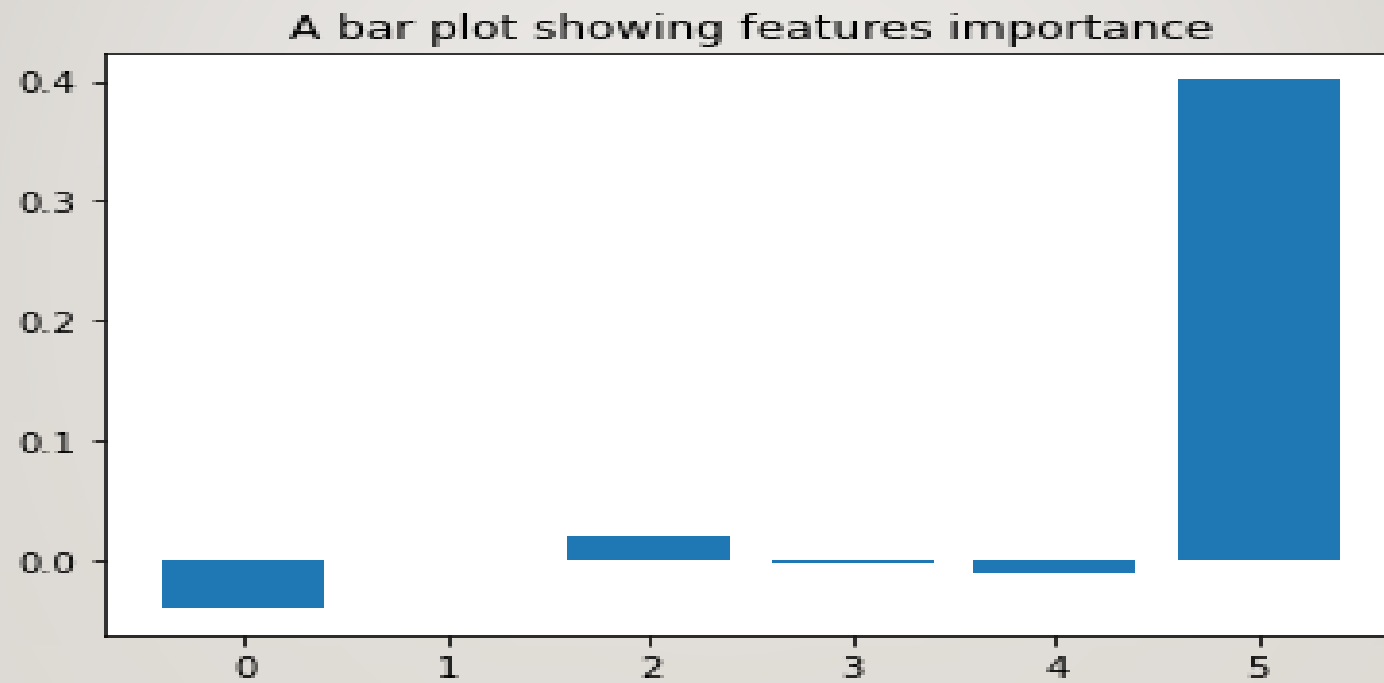
A COMPARISON OF PREDICTED VALUES AGAINST TEST VALUES USING DECISION TREE



A COMPARISON OF PREDICTED VALUES AGAINST TEST VALUES USING LOGISTIC REGRESSION

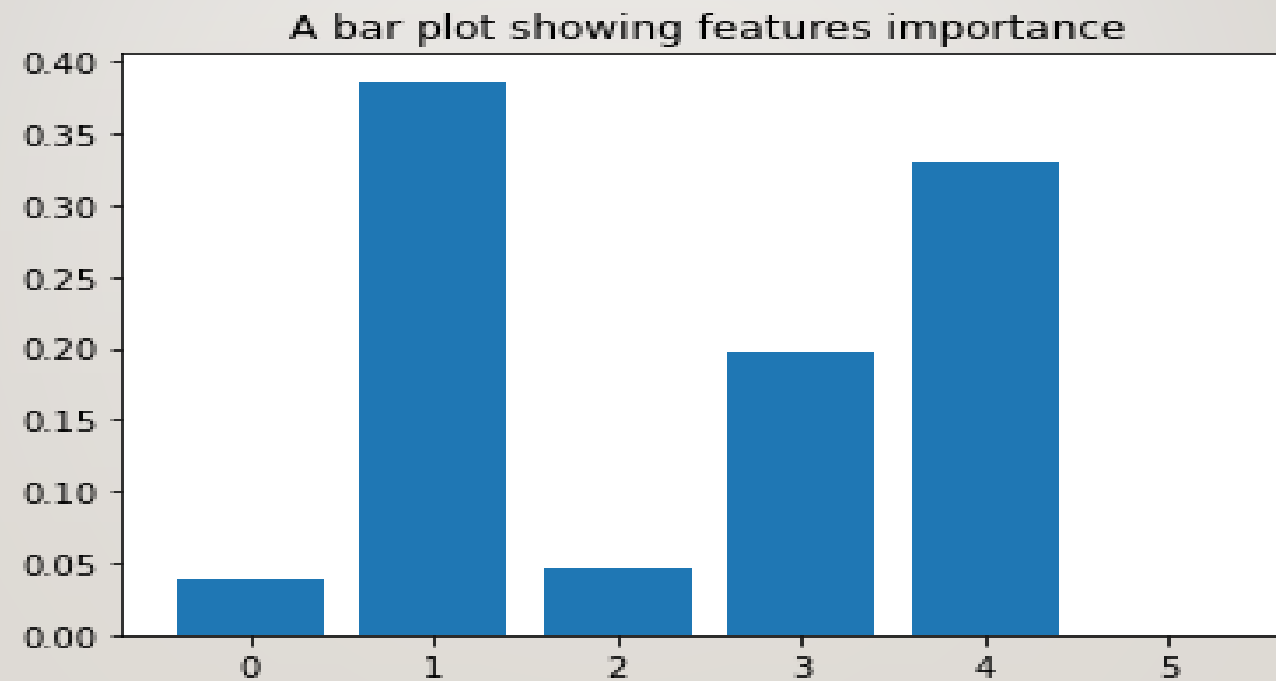


FEATURE IMPORTANCE FOR LOGISTIC REGRESSION



-
- The feature importance plot produced both negative and positive values since the target variable is categorical.
 - The most positive contributed highly to selecting a YES.
 - The most negative contributed highly in selecting a NO.

FEATURE IMPORTANCE FOR DECISION TREE



FEATURE IMPORTANCE FOR DECISION TREE

- Three features seems to have contributed highly in decision tree clustering.
- Using k-fold cross validation improved the model accuracy scores comparing to just train and test split.

RECOMMENDATION

- SmartAd advertisements are effective, hence they can reach to more clients or increase charges.

LIMITATIONS

- Majority of the users didn't engage with the BIO and dropping them would highly affect the modelling process.

REFERENCES

- <http://www.qubit.com/wp-content/uploads/2017/12/qubit-research-ab-test-results-are-illusory.pdf>
- <https://www.austinrochford.com/posts/2014-01-01-intro-to-sequential-testing.html>
- <https://medium.com/analytics-vidhya/a-b-testing-clearly-explained-56488430156>