

Capstone Presentation

Social Media Tourism

Business Problem Understanding

Defining Problem

Aviation Company trying to make use of ***digital marketing platform*** to come up with a plan for targeted customer

To achieve this objective, Aviation company has collaborated with social networking platform ***to understand*** the digital and social behavior of the customers.

Provide the ***digital advertisement*** on the user page of the targeted customers who have a high propensity to take up the product.

Propensity of buying tickets is different for different login devices. Hence, you have to create 2 models separately for ***Laptop and Mobile***. [Anything which is not a laptop can be considered as mobile phone usage.]

The advertisements on the digital platform are a bit ***expensive***; hence, you need to be very accurate while creating the models.

Business Problem Understanding (Contd.)

Need of Study / Project

This project will help the aviation company get a ***clearer picture*** about their customers and help them according to their expectation.

Targeted advertisement not only helps the ***company's revenue grow*** but also bring ***satisfaction to the customers*** as their individual need is met.

Various tour packages can be ***framed*** and sent to the specific group of people

Understanding business/social opportunity

The business opportunity here is that it helps the company give ***attention*** to targeted audiences.

Eventually, this will lead to an ***increase in revenue***. Once the company revenue increases, the company size will also grow and the company will expand.

A company with good growth will have various ***employment opportunities***. This can be considered the social opportunity of this project.

Modelling Approach Used & Why

Understanding of attributes

There are **17 variables**, out of which 3 variables are of type float, 7 variables are of type int and 7 variables are of type object. The data given is for **11760** individuals.

Univariate Analysis

There is **skewness** in all continuous / numerical variables

From count plot, we understand that **preferred_device** for majority of the users is Mobile.

Bivariate and Multivariate Analysis

From Pairplot and Heat Maps, we can understand that there is **minimal corelation** between numerical features.

From boxplot, we understand that, the device **mobile and laptop** almost have the **same median** when it comes to users taking the product

Modelling Approach Used & Why (Contd.)

Metric considered for this project

Precision is the parameter we should evaluate for this approach since the cost of not targeting the appropriate client is higher than the cost of targeting the wrong customer, who may not buy the product. As a result, the organization will lose potential clients, hence I'm contemplating Precision as a desired criteria for this project

F1 and Recall were also considered as **recall** is computed as the ratio of Positive samples that were correctly categorized as Positive to the total number of Positive samples. The recall of the model assesses its ability to detect Positive samples. The more positive samples identified, the larger the recall.

Smote was used to balance the minority class in the target variable and it was used to build Random Forest model.

Data frame was split into **2 dataset** one for Mobile users and the other for Laptop users and models are built separately for each dataset.

Modelling Approach Used & Why (Contd.)

Models	Precision-Laptop	Precision-Mobile	Accuracy-Laptop	Accuracy-Mobile	Recall-Laptop	Recall-Mobile	F1-Laptop	F1-Mobile
Cart	0.94	0.90	0.94	0.97	0.85	0.89	0.90	0.90
Cart using pruning	0.75	0.68	0.81	0.88	0.48	0.47	0.59	0.56
Smote	0.75	0.68	0.81	0.88	0.48	0.47	0.59	0.56
Logistic Regression	0.65	0.00	0.75	0.85	0.25	0.00	0.36	0.00
Linear Discriminant Analysis	0.81	0.63	0.80	0.85	0.41	0.14	0.55	0.22
Random Forest	1.00	1.00	0.97	0.97	0.88	0.82	0.94	0.90
KNN	0.93	0.95	0.95	0.98	0.91	0.91	0.92	0.93
Smote – Logistic Regression	0.46	0.23	0.68	0.61	0.78	0.67	0.58	0.35
Smote - LDA	0.51	0.29	0.73	0.68	0.80	0.76	0.63	0.42
Smote – KNN	0.89	0.91	0.96	0.98	0.97	0.96	0.93	0.93
Smote - RF	1.00	0.98	0.98	0.98	0.94	0.87	0.97	0.92
Logistic Regression-GridSearch CV	0.80	0.62	0.80	0.85	0.41	0.12	0.54	0.19
LDA- GridSearch CV	0.81	0.63	0.80	0.85	0.41	0.14	0.55	0.23
Random Forest-GridSearch CV	1.00	1.00	0.98	0.99	0.94	0.91	0.97	0.95
Naïve Bayes	0.60	0.61	0.78	0.86	0.71	0.18	0.65	0.27
Bagging	0.96	0.97	0.98	0.97	0.97	0.84	0.96	0.90
Bagging-GridSearch CV	1.00	0.92	0.84	0.86	0.43	0.07	0.60	0.14
ADA Boosting	0.78	0.74	0.84	0.87	0.61	0.28	0.69	0.40
Gradient Boosting	0.97	0.89	0.94	0.89	0.81	0.36	0.89	0.51
XG Boost	0.99	0.99	0.99	0.98	0.97	0.89	0.98	0.94
XG Boost Hyperparameter Tune	0.95	0.95	0.95	0.91	0.87	0.45	0.91	0.61



Insights from Analysis

Insights from EDA

The customers with **3 or 4 family members** are more likely to go on trips. Therefore, we can focus on these customers to get a good response.

The customers opting for the location **“beach” and “financial”** contribute to the maximum count. We can come up with a separate tour package for these users.

The **total likes and comments** given by the users indicate they are active users in the media platform. The highly active customers should be given high priority as there is a strong chance that they will take the product.

From **following company page**, we can see that if following or not following users have still bought the product. This means this variable **doesn't carry too value** and can the company can concentrate on other important variables to design the right strategy.

From **working flag**, we can see that the not working users have bought the product more than the working class. This means the users are either young adults still studying in college or not working and the right strategy could be derived to target this class of users.

Insights from Analysis (Contd.)

Insights from Model

Based on the integrated results, we can conclude that, **Random Forest** is the best performing model for both mobile phone users and laptop users, with a 98% percent accuracy rate.

Precision, the required criterion for this problem, is also found to be much higher for Random Forest models, with 100 percent for Laptop and Mobile phone users.

Precision quantifies the number of positive class predictions that actually belong to the positive class based on this observation

Recall and F1 score are also higher for random forest models.

As a result, this should be taken as **100%** of total consumers who use **Laptops** and were expected to purchase the product actually bought the product.

Similarly, among all customers who use **mobile phones**, **100%** of all customers forecasted to purchase the product do so. Which is the highest level of precision that the model can achieve.

Insights from Analysis (Contd.)

Insights from Model

The prediction capabilities of a model are considerably improved by selecting the proper model. This increases the model's dependability for decision making. When we select the ***Random Forest model***, we train the model using the entire dataset as the train set, and the resulting model is ready to generate predictions with independent variables.

In our case, given the social media components of a person, such as the time spent on travel websites, the number of likes received and given, and so on, we will now be able to predict the likelihood of that customer purchasing the travel packages offered by the aviation company with a ***98 percent accuracy***.

This, in turn, gave a tailored way for the aviation company to reach their consumer base, lowering costs and maximizing the return on investment in digital marketing initiatives

Recommendations

The analysis on the dataset has given the company an understanding how the ***users behave*** on social media platform's and if the organization focuses on the 10 variables it can turn the revenue game in their favor.

As digital advertising is ***costly***, the company can take decision on which device it wants to deploy their best advertisement and, in this case, it should be hand held or mobile device which happens to be the favorite to browse the net amongst the users.

The company can make ***tailored packages*** for fun destinations like beach for instance and include a clause if taken in a group more than 3 individuals then the user will get a discount.

The company can include ***complimentary services*** for travelers travelling for the purpose of finance or business that could be included with their stay if they opt for the package offered to them.

Special discounts could be given to travelers travelling for medical purpose. By doing so, they can gain the trust of the customers and in return they can have a loyal customer following.

Recommendations (Contd.)

For the customers to stay on their page and follow the page the company could use and introduce ***interactive games*** which offers the customer discounts or complimentary services or a free stay with an affiliated hotel and so on if they participate and win the games.

They could ask the customers to ***write or comment*** on the company page and get points that can be redeemed in the future to get travel discounts.

Coupons could be provided to repeat customers for their family and friends for their next travel plan and in return the company can have a new customer base, thus increasing the footfall and revenue