Social Media Tourism

Rohit Agrawal

Business Problem Understanding

Business Problem

An aviation company that provides domestic as well as international trips to the customers now wants to apply a targeted approach instead of reaching out to each of the customers. This time they want to do it digitally instead of tele calling. Hence, they have collaborated with a social networking platform, so they can learn the digital and social behaviour of the customers and 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 2models 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

Need of Study

- Digital media has penetrated all aspects of tourism and have led to fundamental changes in the way tourism experiences are planned, consumed, evaluated and marketed.
- In the tourism industry, websites and social media provide a wealth of information with regards to experiences and review of the destination, views, likes, comments, travel check ins.
- Social media marketing generates more business exposure, increased traffic and improved search, generating leads and improved sales at lower cost.
- To understand what kind of information consumers seek online and how they actually use information acquired online from other consumers to make their travel and hospitality decisions.

Understanding Business/Social Opportunity

- The leading trends towards the Social Networking has drawn high public attention from past ‘two’ decades. For both small businesses and large corporations, social media is playing a key role in brand building and customer communication.
- The understanding of the customer’s behaviours on a social media platform will result in targeting advertisements according to the needs and wants of the specific set of customers that can result in high propensity to take up the product.
- Apart from this company can understand the problems associated with the customers that have posted bad reviews. Then, instead of calling each and every customer company can utilize its resources to improve revenue

Modelling Approach Used & Why

Steps used before Model Building

- Treating bad data
- Removal of unwanted variables
- In missing value treatment replaced missing value in numerical column using median and object column using mode.
- In outlier treatment replaced those outliers with the upper limit and lower limit of the particular columns.
- Variable Transformation / Addition of New Variables
- We have created separate data for Mobile and Laptop for modelling
- We have data in different scales, the variables with larger scale will dominate. After scaling there is variance look similar across all data.
- We have split dataset into Train Test for 70:30 Ratio.

Different types of models used

1. **Logistic Regression Model -** Logistic regression is a process of modelling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.
2. **KNN Model -** KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).
3. **Naïve Bayes Model -** Naïve Bayes is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
4. **Bagging -** Bagging is designed to improve the performance of existing ML algorithms used in statistical classification or regression. It is most used with tree-based algorithms. It is a parallel method.
5. **ADA Boosting -** This model is used to increase the efficiency of binary classifiers, but now used to improve multiclass classifiers as well. ADA boosting can be applied on top of any classifier method to learn from its issues and bring about a more accurate model and this it is called “best out of the box classifier”

6. **Gradient Boosting -** This model is just like the ADA Boosting works by sequentially adding the misidentified predictors and under-fitted predictions to the ensemble, ensuring the errors identified previously are corrected. The major difference lies in what it does with the mis-identified values of the previous weak learner.

Three types of approach used

We have build all models which are mentioned in previous slide for the approaches give below:

1. **Base Models –** In this all models build without using any type of hyper tuning parameters.

2. **Model Tuning -** Tuning is process of maximizing a model's performance without overfitting or creating too high of a variance. In ML, this is accomplished by selecting appropriate "hyper-parameters.

3. **SMOTE Technique -** SMOTE (synthetic minority oversampling technique) is one of the most commonly used oversampling methods to solve the imbalance problem. The data is highly imbalanced as out of 11760 customers there are 9864 customers that are not interested in purchasing our product which constitutes around 83.9%. This can be treated with the help of SMOTE

Results

Base Models comparison for Laptop

Basic Model		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.84	0.83	0.85	0.83	0.95	0.97	0.9	0.89	0.82	0.87
	Yes Taken Product			0.74	0.87	0.46	0.49	0.57	0.63		
KNN	No Taken Product	0.96	0.88	0.96	0.88	0.99	0.96	0.98	0.92	0.99	0.94
	Yes Taken Product			0.96	0.88	0.87	0.68	0.91	0.77		
Naïve Bayes	No Taken Product	0.83	0.84	0.88	0.87	0.91	0.91	0.89	0.89	0.81	0.85
	Yes Taken Product			0.66	0.75	0.59	0.66	0.62	0.7		
Bagging	No Taken Product	1	0.94	1	0.93	1	1	1	0.96	1	0.99
	Yes Taken Product			1	1	0.99	0.8	1	0.89		
Ada Boosting	No Taken Product	0.95	0.87	0.94	0.89	0.99	0.94	0.97	0.91	0.99	0.93
	Yes Taken Product			0.97	0.82	0.79	0.69	0.87	0.75		
Gradient Boosting	No Taken Product	0.99	0.96	0.98	0.94	1	1	0.99	0.97	0.99	0.99
	Yes Taken Product			1	1	0.94	0.85	0.97	0.92		

Results

Model Tuning comparison for Laptop

Grid Search Model Tuning		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.84	0.83	0.85	0.83	0.95	0.97	0.9	0.89	0.82	0.87
	Yes Taken Product			0.74	0.87	0.46	0.49	0.57	0.63		
KNN	No Taken Product	0.87	0.8	0.86	0.79	0.99	0.97	0.92	0.87	0.96	0.92
	Yes Taken Product			0.95	0.85	0.45	0.35	0.61	0.49		
Naïve Bayes	No Taken Product	0.83	0.84	0.88	0.87	0.91	0.91	0.89	0.89	0.81	0.85
	Yes Taken Product			0.66	0.75	0.59	0.66	0.62	0.7		
Bagging	No Taken Product	1	0.89	1	0.87	1	1	1	0.93	1	0.99
	Yes Taken Product			1	1	1	0.62	1	0.77		
Ada Boosting	No Taken Product	0.95	0.88	0.95	0.89	0.99	0.95	0.97	0.92	0.99	0.95
	Yes Taken Product			0.96	0.84	0.84	0.72	0.9	0.77		
Gradient Boosting	No Taken Product	1	0.99	1	0.98	1	1	1	0.99	1	1
	Yes Taken Product			1	1	1	0.96	1	0.98		

Results

SMOTE Technique comparison for Laptop

SMOTE		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.74	0.73	0.74	0.9	0.75	0.7	0.74	0.79	0.82	0.86
	Yes Taken Product			0.74	0.52	0.74	0.8	0.74	0.63		
KNN	No Taken Product	0.99	0.91	1	0.96	0.97	0.91	0.99	0.94	1	0.96
	Yes Taken Product			0.98	0.81	1	0.92	0.99	0.86		
Naïve Bayes	No Taken Product	0.71	0.65	0.78	0.89	0.6	0.59	0.68	0.71	0.81	0.84
	Yes Taken Product			0.67	0.44	0.83	0.82	0.74	0.58		
Bagging	No Taken Product	1	0.98	1	0.99	1	0.98	1	0.99	1	0.99
	Yes Taken Product			1	0.96	1	0.97	1	0.96		
Ada Boosting	No Taken Product	0.92	0.83	0.93	0.93	0.92	0.83	0.92	0.88	0.98	0.93
	Yes Taken Product			0.92	0.66	0.93	0.83	0.92	0.74		
Gradient Boosting	No Taken Product	0.99	0.95	1	0.99	0.98	0.95	0.99	0.97	0.99	0.99
	Yes Taken Product			0.99	0.88	1	0.97	0.99	0.92		

Results

Base Models comparison for Mobile

Basic Model		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.87	0.87	0.87	0.88	0.99	0.98	0.93	0.93	0.79	0.8
	Yes Taken Product			0.71	0.74	0.2	0.23	0.31	0.35		
KNN	No Taken Product	0.99	0.97	0.99	0.98	1	0.99	0.99	0.98	0.99	0.99
	Yes Taken Product			0.98	0.95	0.93	0.86	0.95	0.9		
Naïve Bayes	No Taken Product	0.86	0.85	0.88	0.88	0.96	0.95	0.92	0.92	0.77	0.77
	Yes Taken Product			0.56	0.53	0.28	0.31	0.37	0.39		
Bagging	No Taken Product	1	0.96	1	0.96	1	1	1	0.98	1	0.99
	Yes Taken Product			1	1	1	0.77	1	0.87		
Ada Boosting	No Taken Product	0.88	0.88	0.89	0.9	0.98	0.97	0.93	0.93	0.88	0.86
	Yes Taken Product			0.73	0.71	0.33	0.39	0.46	0.5		
Gradient Boosting	No Taken Product	0.91	0.9	0.91	0.9	0.99	0.99	0.95	0.94	0.94	0.92
	Yes Taken Product			0.91	0.88	0.48	0.43	0.63	0.57		

Results

Model Tuning comparison for Mobile

Grid Search Model Tuning		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.87	0.87	0.87	0.87	0.99	0.99	0.93	0.93	0.79	0.8
	Yes Taken Product			0.72	0.74	0.2	0.23	0.31	0.35		
KNN	No Taken Product	0.95	0.92	0.94	0.92	1	0.99	0.97	0.95	0.99	0.97
	Yes Taken Product			0.98	0.94	0.67	0.5	0.79	0.66		
Naïve Bayes	No Taken Product	0.86	0.85	0.88	0.88	0.96	0.95	0.92	0.92	0.77	0.77
	Yes Taken Product			0.56	0.53	0.28	0.31	0.37	0.39		
Bagging	No Taken Product	1	0.9	1	0.89	1	1	1	0.94	1	0.99
	Yes Taken Product			1	1	0.98	0.34	0.99	0.51		
Ada Boosting	No Taken Product	0.89	0.87	0.89	0.9	0.98	0.98	0.94	0.93	0.88	0.87
	Yes Taken Product			0.78	0.74	0.34	0.37	0.48	0.5		
Gradient Boosting	No Taken Product	1	0.97	1	0.97	1	1	1	0.98	1	0.99
	Yes Taken Product			1	0.98	0.88	0.82	0.99	0.89		

Results

SMOTE Technique comparison for Mobile

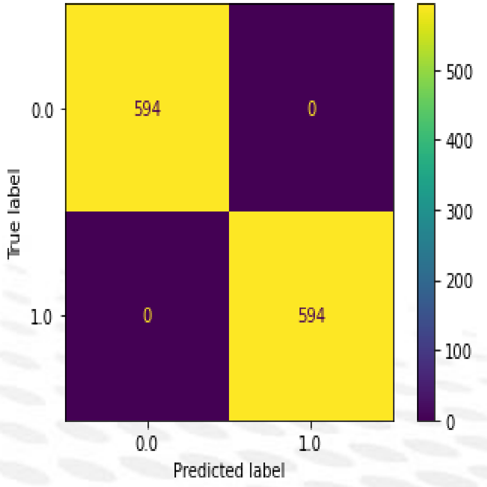
SMOTE		Accuracy		Precision		Recall		F1 Score		AUC	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Logistic Regression	No Taken Product	0.73	0.71	0.73	0.94	0.72	0.7	0.72	0.81	0.79	0.79
	Yes Taken Product			0.72	0.32	0.73	0.76	0.73	0.45		
KNN	No Taken Product	0.99	0.97	1	1	0.98	0.97	0.99	0.98	1	0.99
	Yes Taken Product			0.99	0.85	1	0.98	0.99	0.91		
Naïve Bayes	No Taken Product	0.68	0.66	0.7	0.93	0.65	0.64	0.67	0.76	0.77	0.77
	Yes Taken Product			0.67	0.28	0.72	0.75	0.7	0.4		
Bagging	No Taken Product	1	0.98	1	0.98	1	0.99	1	0.99	1	0.99
	Yes Taken Product			1	0.96	1	0.92	1	0.94		
Ada Boosting	No Taken Product	0.84	0.93	0.83	0.93	0.85	0.85	0.84	0.89	0.93	0.85
	Yes Taken Product			0.85	0.44	0.83	0.64	0.84	0.52		
Gradient Boosting	No Taken Product	0.9	0.87	0.89	0.94	0.92	0.91	0.9	0.92	0.97	0.9
	Yes Taken Product			0.91	0.58	0.89	0.69	0.9	0.63		

Interpretation of the most optimum model

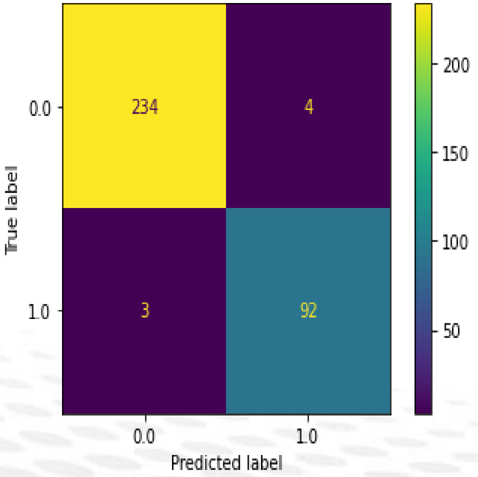
- Based on our model evaluation, performing visual inspection, stacking and bagging models. We finally are able to combine all the results and can clearly infer that after Using SMOTE Technique Bagging using base estimator as Random Forest is the best performing model for both Mobile phone users and Laptop users, with the highest accuracy of 98%.
- Bagging is performing well in terms of Recall, Precision and F1-Score for both Laptop and Mobile users.
- The desired metric for this problem which is Precision, is also observed to be significantly the highest for Random Forest models with 96% for Laptop users and 96% for Mobile phone users.

For Laptop

Train Data

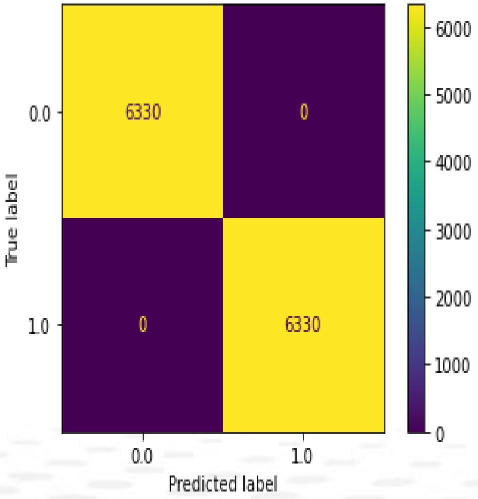


Test Data

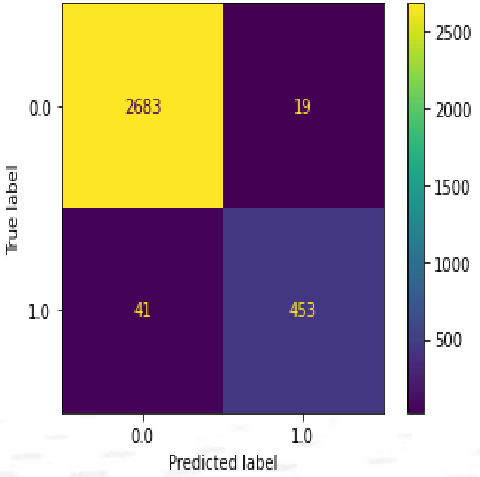


For Mobile

Train Data



Test Data



Interpretation of the most optimum model

- From this observation Precision quantifies the number of positive class predictions that actually belong to the positive class. Hence this should be interpreted as 96% of total customers who use Laptops who were predicted to purchase the product actually purchases the product.
- Similarly, among the total customers who use Mobile phones 96% of all the customers predicted to purchase the product actually buys the product.
- We the help of this model company can identify which customers can purchase the product in the new future. Also, this helps in better reach and target the audience accordingly.

Insights from Analysis

- Customers that visits once in a year does not like to purchase the ticket from the company.
- As the family size of the customer increases the chances of purchasing the product decreases.
- The people that travel for beaches and financial purposes which are two major reasons for travel are less likely to buy the company’s products.
- Customers following the company’s page have high chances of purchasing the products.
- The working category have high chances of purchasing the product as compared to the people that are not working.
- It is observed that despite any preferred devices there is very high attrition rate amongst the customers as they are not interested in purchasing the product at all.
- company should come up with discount offer the user who travels for medical related travels as this will have good customer experience in these unprecedented times and it will increase brand value.
- The people who don’t follow company page have high average view on company page and people who follow company page has less view.

Recommendations

- Target the customers who have not checked in in the last few weeks as outstation checkin is the most important feature.
- Plan a campaign for people who spend a lot of time on the page and target the customers more on the mobile device.
- Adults play a critical role in the buying decision, thus they should be targeted more wisely.
- The aviation company should invest their budget in acquiring the dataset from the networking platform to learn about their behavior and target these customers.
- Using optimum model i.e. SMOTE Technique Bagging using base estimator as Random Forest can also increase the traffic on the company’s site resulting in better minimizing the click per cost expense for the company.
- As, the higher the number of hits on website increases, more chances of purchasing the product also increases bringing in the surge in revenues for the company.
- This in turn provided a targeted approach for the aviation company to approach their customer base, thereby reducing cost and making the most returns out of the expenditure they put in digital marketing campaigns.

Thank You