

Summary Lead Score Case Study

- The dataset was loaded into the python notebook and explored statistically and visually get an idea of distribution of data, feature redundancies, and so on Correlations between the variables were also identified using heat map. The redundant variables were removed. Some of the columns had string values "Select" which meant that the user did not select a particular value in the form. This means it is as good as null values. Most of the customer was identified in landing page submission and converted as well. The maximum source of lead conversion in Google and is also high than others sources ex: Social medias , blogs which means the people are maximum spent their time using in Internet so, it would be good opportunity to Lead team to focus on strategic behind Digital marketing to the internet default browsers users who using google as browsers. which are highly chances to converting the leads.
- In Mumbai city have highly conversion leads rate than others Ex: Thane & Outskirts also we see that highly conversion rate for the people who will revert after reading the emails. For the occupation we see that the unemployed people are maximum conversion rate , So lead team have to focused on the Professional people who are upgrading their skill that would be chance to higher conversion rate. For Thane & Out skirt Region Peoples using maximum time in the website according to EDA analysis After finishing the EDA process , Based on the analysis that many columns are not adding any information to the model, hence we can drop them and move to the dummy variable for Categorical columns, and there are some columns are highly correlated each other so we are going to dropped this columns to avoiding multicollinearity Issue. After that we move to model Building so we take the target variable is (Converted) rest of the values are depending. So using the LogisticRegression technique, Rfe support for feature selection, Stats Model for checking p- Values,VIF These are steps are follow for the Model Building, Than we calculate the scores for prediction (TEST and TRAIN Data set)
- For (Train data accuracy score) is 91%
- sensitivity: 86%
- specificity: .95%
- false +ve rate: 40%
- +ve predictive value: 91%
- -ve predictive value: 91%
- For (Test data accuracy score) is 91%
- sensitivity: 86%
- specificity: 94%
- false +ve rate: 50%
- +ve predictive value: 91%
- -ve predictive value: 91%
- For comparison for both data set we see the Model the Test model is performing in good Accuracy.