

Summary -

We created a Logistic Regression Model to predict the lead which are most likely to be converted. We did the exploratory data analysis by imputing the missing values and dropping unnecessary columns. After that, we converted all the categorical variables to numerical variables. We used logistic regression model to classify the “converted” value. We find the optimal cutoff value for predicting the test data. We used the intersection point of specificity, accuracy and sensitivity.

We found that these are the important variables -

tags (Busy, Closed by Horizon, Will revert after reading the email, Lost to EINS)

lead_quality (High in Relevance, Low in Relevance)

lead_source (Welingak Website)

lead_profile (Other Leads)

last_activity (Others, SMS Sent)

occupation (Working Professional)

last_notable_activity (Modified)

aai (03.LOW)

to be considered for increasing the conversion rate.