Summary Report

Education company named X Education sells online courses to industry professionals.On any given day, many professionals who are interested in the courses land on their website and browse for courses.When these people fill up a form providing their email address or phone number, they are classified to be a lead. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.

So we have to find the leads that are most likely to convert into paying customers. The company requires to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

For this Logistic regression is performed. The target variable, in this case, is the column 'Converted' which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn't converted. For modelling We have leads dataset from the past with around 9000 data points.

As a first step data cleaning is performed on the data set.This include null value treatment,dropping of unimportant column,outlier treatment,data type check etc.Dummy variables are created for the categoric variable,After that modelling is performed.During 20 variables are selected using feature selection. Variable having p value >0.05 and VIF >4 are dropped.we have got an overall accuracy of **89.84** percentage for the train data and **88.69** percentage for test data.

Tags - Closed by Horizzon,Tags_Lost to EINS,Total Time Spent on Website are the Top three variables in the model which contribute most towards the probability . Tags_Will revert after reading the email,Lead Origin_Lead Add Form,Tags_Busy are Top 3 variables that need improvement to convert a lead

While calling the lead Use the predictive model to identify potential leads that have been predicted as 1, indicating a higher probability of conversion. Prioritize these leads as they are more likely to convert. Encourage the interns to start calling leads from the highest-potential segment first. Once they have exhausted the calls in that segment, they can move on to the next segment. This approach ensures that the most valuable leads are contacted first and increases the chances of conversions.