

## Summary

### Lead Score – Case Study

The analysis is carried out for X Education in an effort to attract more business professionals to their courses. We learned a lot from the fundamental data on how potential customers use the site, how long they stay there, how they got there, and the conversion rate.

These are the procedures used:

- **Understanding & cleaning of data set:** The majority of the data was clean, save for a few null values, and the option pick had to be changed to a null value because it provided little useful information. To avoid losing too much data, a few of the null values were changed to "not provided." Nevertheless, they were later taken out while manufacturing dummies. The elements were altered to "India," "Outside India," and "not provided" because there were many people from India and few from elsewhere.
- **EDA:** A brief EDA was performed to assess the state of our data. It was discovered that several of the categorical variables' components were unnecessary. The numerical figures are accurate, and no anomalies were discovered.  
3. Dummy Variables: After the dummy variables were generated, those that had the phrase "not provided" were later eliminated. We utilised the MinMaxScaler to scale numerical numbers. The split for train and test data was done at 70% and 30%, respectively.
- **Model Building:** RFE was first used to identify the top 15 pertinent variables. Later, based on the VIF values and pvalue, the remaining variables were manually deleted (the variables with VIF 5 and p-value 0.05 were retained).
- **Model Evaluation:** A confusion matrix was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 80% each. Using an optimal cutoff of 0.35 and a prediction accuracy, sensitivity, and specificity of 80%, the test data frame was used. Precision - Recall method was also utilized to retest, and on the test data frame, a cut off of 0.41 was discovered with Precision around 73% and recall around 75%.

The factors that affected potential purchasers the most were discovered to be: The overall amount of time spent on the Website. The overall volume of visits.

If the lead came from one of the following: Google, Direct traffic, Organic search & The Welink website.

The most recent activity was when:

SMS exchanges and conversations on Olark If the format of the lead add is the lead origin. If they are a working professional at the time.

With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.