

## **Summary**

Analysis was done for X Education to identify and select most promising leads that can be converted to paying customers. And to implement process of lead generating that will be efficient in conversions.

The data which was provided us gave us insight regarding the behavior of the potential leads, such as how they visit the site, how they enquire about the courses, how much time they spent, how the company is reaching out to them, how are they responding to the company initiated communication and also the conversion status of these leads.

The steps that are used are:

- Data cleaning and preparation.
  - Unique value columns are dropped
  - We tried to minimize the dropping of features instead we replaced level 'Select' in the categorical features with np.nan
  - We replaced all management related specialization to a single value as to get most of the insight from these levels
  - We did the similar treatment with Lead\_Source, Tags, last\_Activity.
  - We dropped the features which have very imbalanced information
- Exploratory Data Analysis.
  - EDA was performed to check the correlation and also the check the outlier presence. The outlier treatment also have been done
- Dummy Variable Creation
  - Dummies of the categorical variables have been created.

- Test Train Split
  - The data was split to train data set and test data set.
- Scaling of Numeric Values
  - Standard Scaler was used to scale the data
- Model Building
  - First the coarse tuning was done using RFE
  - Then fine tuning was done manually by looking at the p-value and the VIF.
  - The features with p-value < 0.05 and VIF < 4 was kept
- Model Evaluation
  - By using confusion matrix and ROC curve the optimum cut-off value was found and the accuracy, sensitivity and specificity came around 90%
- Prediction
  - The prediction was done and the accuracy was 90%, Sensitivity was 89% and Specificity was 90%

It was also found out that the variables which mattered the most in identifying revenue generating leads are:

1. Lead Origin\_Landing Page Submission
2. Tags\_Will revert after reading the email
3. Lead Origin\_Lead Add Form
4. Last Activity\_SMS Sent
5. Last Notable Activity\_Modified
6. Tags\_Other\_Tags
7. Lead Source\_Welingak Website
8. Tags\_Ringing

9. Total Time Spent on Website
10. Tags\_Interested in other courses
11. Tags\_Busy
12. Tags\_Lost to EINS
13. Last Notable Activity\_Olark Chat Conversation