

## LEAD SCORING CASE STUDY SUMMARY

### **Problem Statement**

1. X Education company sells online courses to industry professionals. X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%
2. Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to Target Potential Leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted. Achieve lead conversion rate to be around more than 80%.
3. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.
4. X Education has appointed you to help them select the most promising leads, i.e., the leads that are most likely to convert into paying customers.
5. Finding the Top three variables in your Model which contribute most towards the Probability of a Lead Getting Converted.

**The following technical steps are used for the analysis.**

1. **Data cleaning, Manipulation (Missing value treatment) & outlier analysis**
2. **Exploratory Data Analysis**
3. **Data Preparation for Modelling**
4. **Model Prediction**
5. **Model Evaluation**
6. **Lead score calculation**
7. **Feature Importance Determination**
8. **Business Insights**
9. **Results & Recommendation**

## CONCLUSION & RECOMMENDATIONS

**We tried a number of models before settling on model number 5, which had the following features:**

- All factors have p-values below 0.05, indicating that they all significantly influence lead conversion.
- Since all of the characteristics have extremely low VIF values, there is virtually any multicollinearity between them. The heat map demonstrates this.
- On the test dataset, the overall accuracy of almost 80% at a probability threshold of 0.34 is likewise extremely respectable.
- **For Train Dataset**
  - Accuracy : 80.81%
  - Sensitivity: 82.23%
  - Specificity : 79.91%
- **For Test Dataset**
  - Accuracy: 79.64%
  - Sensitivity: 81.07%
  - Specificity : 78.76%
- **Based on sensitivity, specificity, and accuracy** trade-offs, the model's ideal threshold was found to be 0.36. This threshold can be adjusted to boost or reduce a particular measure in accordance with company demands.
- High specificity ensures sure that most of the leads who are not likely to convert are accurately forecasted, while high sensitivity makes sure that most of the leads who are likely to convert are correctly anticipated.