

## **SUMMARY**

X Education deals with online courses for industry professionals. The company's current conversion rate is very low at 30%. X Education wants to build a model aimed at assigning lead scores, with the objective of prioritizing leads having higher probability of conversion. The CEO wants to achieve 80% lead conversion rate.

The following steps were involved

### **Data Cleaning**

- Checked and handled duplicate data
- Dropped columns with over 40% null values.
- Imputed values, wherever necessary. Replaced 'Select' with NaN as mentioned in Problem Statement and replaced Null values with 'others' or mode of the column
- Removed unwanted columns

### **EDA**

- Checked data imbalance and calculate Converted variable.
- Performed univariate and bivariate analysis for categorical and numerical variables.
- Removed highly skewed and unique values
- Compared variables with Target variable
- Handled outliers and combined low frequency values into one in a column.

### **Data Preparation**

- Created dummy features for categorical variables.
- Split the data into train and test sets, with 70:30 ratio
- Dropped highly correlated columns.
- Reduced the number of variables using recursive feature elimination (RFE) and manual feature reduction.

### **Model Building**

- Used Logistic Regression approach for model building
- Built four models before arriving at the final model.
- Ensured stability with p-values < 0.05 and no multicollinearity with VIF < 5.

### **Final Model**

- The final model, comprised 17 variables.
- Used ROC curve to see the area covered. It was 0.89
- Evaluated the model by constructing a confusion matrix.
- Improved Specificity by plotting curve between Accuracy, Sensitivity and Specificity. Found the optimal probability as 0.4

## Test Model

- Predicted the values on the test data and evaluated.

Parameters	Train Model	Test Model
Accuracy	81.3	80.3
Specificity	82.2	80.1
Sensitivity	79.9	80.8

## Hot leads

- Assigned the lead score to leads

Leads with more than 80 lead score were suggested as Hot Leads.

- The company should connect with them as they have high probability of converting

## Recommendations:

### Do's

- Company should make call to leads coming from "Lead Source\_welingak website" and 'reference', they are most likely to be converted.
- The company should contact 'working professionals' as they are more likely to be converted.
- Leads who have 'Last Activity' as 'had a phone conversation' are more likely to be converted.
- The company should consider 'total time spent on the website' as important feature and connect with leads spending more time on website. The company can make the website more engaging so that leads can spend more time on the website.

### Don'ts

- The company should not make calls to leads whose 'Last Activity' was 'Olark Chat Conversation' and 'email bounced'. The company should not call to leads who have clicked 'Do Not Email'.
- The company should not contact leads whose 'last Notable Activity' is 'page visited on website', 'modified', 'email link clicked'. They have very low chances of converting.

Based on the analysis, it's recommended to focus more on Welingak Website and references that convert to leads, and target working professionals.

To maximize lead conversion, prioritize hot leads identified by the model and implement personalized outreach.