

# Summary

This analysis is done for X Education, so that more industry professionals to join their courses. The Lead Conversion rate is only 30%. The company requires us to build a model which can calculate Lead Score for every potential lead and to achieve lead conversion rate of 80%.

## **Analysis Approach:**

### **Data Cleaning:**

- Firstly, we checked for Column for missing or null value and the columns with missing values greater than 40% were dropped. The columns where distribution of the data is very heavily skewed were also dropped from dataframe.
- Columns have only one unique value were dropped.
- Outlier Treatment was done for Numerical Columns, binary variables mapping was done and all the low frequency categories were grouped to a Single Category.

### **EDA:**

- Performed univariate and bivariate analysis for categorical and numerical variables.
- Only 37.9% leads were Converted while checking Data imbalance.

### **Data Preparation:**

- Created dummy variable for categorical variables.
- Data was split into Training and Test Sets.
- Feature Scaling was done using StandardScaler.
- Using Heatmap Correlation was checked and, Highly Correlated columns were removed.

### **Model Building:**

- top 15 relevant variables were selecting using RFE.
- After creating model, we manually removed the variable with high P- value  $< 0.05$ .
- In the end checking the model for multicollinearity using VIF. All value was under control i.e.  $(VIF < 5)$ .

### **Model Evaluation:**

- Confusion matrix was made. at cut off point of 0.346 calculated using ROC curve. This cut off gave accuracy, specificity and precision all around 80%.

### **Precision – Recall:**

- This method was also used to recheck and a cut off of 0.41 was found with Precision and recall around 75% on the test data frame.

### **Making Predictions on Test Data:**

- Scaling and predicting using final model was done.
- Evaluation metrics for train & test are very close to around 80%.
- Lead score feature was added to test dataframe.
- These were the Top 3 features:
  - Lead Source\_Welingak Website
  - Lead Source\_Reference
  - Last Activity\_SMS Sent

### **Conclusion:**

The analysis tells us that more money should be spend on website engagement, some kind of bonus should be provided for Reference Lead. Automated mails and SMS alerts should be sent periodically to Leads. Working Professional should be targeted because they are financial freedom and take up the course to enhance their skill.