Summary

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate. CEO target for lead conversion is 80%.

1. DATA CLEANING:

- 17 columns have missing values.
- Columns with > 25% null values were dropped.
- Some columns have only one type of data value, so dropped.
  Ex: Do not call, Newspaper, Magazine etc.
- Other activities like outlier's treatment, fixing invalid data, grouping low frequency values, mapping binary categorical values were carried out.

2. EDA:

- Performed univariate for categorical and numerical variables. 'Lead Origin', 'Last activity', 'Lead Source', etc. provide valuable insight on effect on target variable.
- Time spend on website shows positive impact on lead conversion.

3.DATA PREPARATION:

- Created dummy features (one-hot encoded) for categorical variables.
- Splitting Train & Test Sets: 70:30 ratio
- Feature Scaling using Standardization

- Dropped few columns, they were highly correlated with each other.

4.MODEL BUILDING:

- Used RFE to reduce variables from 48 to 15. This will make data frame more manageable.
- Manual Feature Reduction process was used to build models by dropping variables with p – value > 0.05.
- Total 6 models were built before reaching final Model 4 which was stable with (p-values < 0.05). No sign of multicollinearity with VIF < 5.

5.MODEL EVALUATION:

- Confusion matrix was made and cut off point of 0.37 was selected based on accuracy, sensitivity and specificity plot. This cut off gave accuracy, specificity and precision all around 80%. Whereas precision recall view gave less performance metrics around 79%
- As to solve business problem CEO asked to boost conversion rate to 80%, but metrics dropped when we took precision-recall view. So, we will choose sensitivity-specificity view for our optimal cut-off for final predictions
- Lead score was assigned to train data using 0.37 as cut off

RECOMMENDATIONS:

To improve the potential lead conversion rate X-Education will have to mainly focus on TOP 2 features responsible for good conversion rate are: -

1. Lead Origin_Lead Add Form - Leads who have engaged through 'Lead Add Form' having higher conversion rate so company can focus on it to get a greater number of leads cause have a higher chance of getting converted.

2. Last Notable Activity_SMS Sent - Lead whose last activity is sms sent can be potential lead for company.