Summary

X Education needs help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. A model is required to be built wherein a lead score is assigned to each of the leads such that the customers with higher lead score have higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Data Cleaning:

- Columns with >40% nulls were dropped.
- Categorical null values are imputed with mode value.
- Numerical columns with null values are imputed with median
- Other steps performed: checking for unique values, checking the value counts in each column for the data distribution and data imbalance.

EDA:

● Checked the co-relation between the numeric columns using Heatmap and Pairplot.
● Found outliers and capped them.
● Visualized Categorical analysis using countplot.

Data Preparation:

● Created dummy features for categorical variables.
● Splitting Train & Test Sets with 70:30 ratio.
● Feature Scaling using Standardization

Model Building:

● Used RFE to reduce variables from 94 to 20 This will make dataframe more manageable.
● Manual Feature Reduction process was used to build models by dropping variables with p – value > 0.05.
● Total 7 models were built before reaching final Model 8 which was stable with (p-values < 0.05). No sign of multicollinearity with VIF < 5.
● Model 8 was selected as final model with 11 variables, we used it for making prediction on train and test set.

Model Evaluation:

● Confusion matrix was made and cut off point of 0.42 was selected based on accuracy, sensitivity, specificity plot and precision recall plot this cut off gave accuracy, specificity and precision all above 85%.
● Accuracy 89.6%,Sensitivity 85% ,specificity 92%, precision 86% ,recall:86%precision_score:87%recall_score:85% is given by our model

Making Predictions on Test Data:

● Making Predictions on Test: Scaling and predicting using final model.
● Evaluation metrics for train & test are very close to around 89.6%.
● Lead score was assigned.
    ● Top 3 features are:
        Total Time Spent on Website
        Tags_Will revert after reading the email
        What is your current occupation_Working Professional

Recommendations:

- The company should make calls to the leads coming from the ‘Tags_Closed by Horizzon’ , ‘Tags_Lost to EINS’ ‘Last Notable Activity_Had a Phone Conversation’, ‘Lead Origin_Lead Add Form ‘, ‘Last Notable Activity_Email Bounced ‘ and ‘Last Notable Activity_SMS Sent’ as these are more likely to get converted.
- The company should make calls to the leads who are the ‘working professionals’ as they are more likely to get converted.
- The company should make calls to the leads coming from ‘Google’ , ‘Lead Source_Welingak Website’ and who spent ‘more time on the websites’ as these are more likely to get converted.