Train Data:

Accuracy : 65.3 %

Sensitivity : 64.7 %

Specificity : 86.5 %

Test Data:

Accuracy : 63.3 %

Sensitivity : 85.1 %

Specificity : 50.8 %

The model seems to predict the Conversion Rate okay and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 60-80%.

- The company **should not make calls** to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.
- The company **should not make calls** to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.
- The company **should make calls** to the leads whose lead origin is " Lead Add Form" as they are not likely to get converted.
- The company **should make calls** to the leads coming from the lead sources "Welingak Website" as these are more likely to get converted.
- The company **should not make calls** to the leads coming from the last activity "Email Bounced" as these are more likely to get converted.
- The company **should not make calls** to the leads coming from the last activity "had a phone conversation" as these are more likely to get converted.
- The company **should not make calls** to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.
- The company **should not make calls** to the leads whose specialization was "Others" as they are not likely to get converted.
- The company **should make calls** to the leads whose current occupation is "Working Professionals" as they are not likely to get converted.

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.

The following are the steps used:

1. Cleaning data: 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.

2. EDA: A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good and no outliers were found.

3. Dummy Variables: The dummy variables were created and later on the dummies with not provided' elements were removed. For numeric values we used the MinMaxScaler.

4. Train-test Split- The split was done at 70%

5. Model Building: Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with VIF<5 and p-value <0.05 were kept).