

# Summary

The Lead Scoring Case Study is done for X Education find the potential buyers who are more likely to convert.

Approach:

1. Data Preparation:

Data was cleaned by removing columns having more null values, imputing null values and replacing certain categories from multiple columns like Country, City, etc.

2. EDA:

Outliers were checked using boxplot and treatment was done.

Univariate and Bivariate analysis was done.

3. Feature Engineering:

Dummy encoding was done to create new variable from categorical columns.

4. Data Scaling:

Standard Scaler was used for scaling data.

5. Model Building:

70:30 data split for model building. Final model had 16 variables with VIF less than 5.

6. Model Evaluation:

Model was evaluated using Confusion Matrix, ROC Curve was built and had 0.97 AUC, metrics such as accuracy, sensitivity of 0.86 and specificity of 0.96.

## 7. Prediction:

Prediction was done on test data. And performance metrics used were Accuracy score = 0.93, Sensitivity = 0.895 and Specificity = 0.955

The Conclusion from the model was:

- People spending more time on the website are more likely to be converted.
- People headed through Google, Direct traffic and organic search are more likely to be converted.
- People with current occupation as working professionals are most likely to be converted among others.
- People with Last Activity as SMS and Chat conversation are more likely to be converted.
- Keeping in mind the above hot leads, X Education can focus on those potential buyers and convert them.