SUMMARY

LEAD SCORING CASE STUDY

The Logistic Regression Model building and prediction is done for X Education Company and to find out how potential customers can be generated. Problem statement is clearly understood and Logistic Regression Machine Learning Algorithm model has been built. Found 80% accuracy model, which is indeed a good model to predict.

The various steps followed in the model building are as follows.

i. Reading and understanding the data:

   The dataset is loaded as 'lead_data' and basic understanding of the data is done.

ii. Data cleaning:

   - Basic data cleanings are done.
   - Dropping variables having unique values.
   - Columns with 'Select' values are changed into null values.
   - Drop columns with null values greater than 45%.
   - Imbalanced and redundant variables are removed.
   - Outlier treatments are done.

iii. Data preparation

   - Binary variables are changed into '0' and '1'.
   - Dummy variables are created for categorical variables.

iv. Train-Test split and Feature Scaling

   - Dataset is split into train and test sets.
   - StandardScaler is used to scale the numerical variables.

v. Model Building

   - Using Recursive Feature Elimination, 18 variables are selected.
   - Model is built and using the P-values and VIFs insignificant values are dropped.
   - Repeating the process, finally made a model with 8 most significant variables.
   - Accuracy, Sensitivity, Specificity of the final model is built.
   - ROC curve is plotted.
   - Optimal cut-off point of the final model is found.
   - Precision and Recall is calculated.
   - Accuracy obtained for train data is 79%.

- The trade cut-off obtained is 0.36.
- Test dataset is analysed and its accuracy is obtained as 80%.
- Finally, the lead score is calculated.

CONCLUSION

- ✓ The top variables contributing to conversion are:
    - Lead Origin
    - Lead Source

  Among these, Lead Origin_Lead Add Form and Lead Source Welingak Website has high value and they give good results.
- ✓ The customers who fill the form are potential customers.
- ✓ It is better to focus on customers who spent significant time on websites.
- ✓ Also, it is better to focus on working professionals.