

SUMMARY OF LEAD SCORE CASE STUDY

X Education gets a lot of leads, its lead conversion rate is very poor at around 30%. The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have higher conversion chance. CEO's target for lead conversion rate is around 80%.

Data Cleaning:

- Columns with >40% nulls were dropped. Value counts within categorical columns were checked to decide appropriate action: if imputation causes skew, then column was dropped, created new category (others), impute high frequency value, drop columns that don't add any value.
- Numerical categorical data were imputed with mode and columns with only one unique response from customer were dropped.
- Other activities like outliers' treatment, fixing invalid data, grouping low frequency values, mapping binary categorical values were carried out.

EDA:

- Data imbalance checked- only 38.5% leads converted.
- Performed univariate and bivariate analysis 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.

Data Preparation:

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

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 3 models were built before reaching final Model 4 which was stable with (p-values < 0.05). No sign of multicollinearity with VIF < 5.

Model Evaluation:

- ROC Curve has a value of 0.96 was selected based on accuracy, sensitivity and specificity plot. We have the following values for the Train Data: Accuracy : 89.80% Sensitivity : 88.81% Specificity : 90.40%
- 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

Making Predictions on Test Data:

- Making Predictions on Test: Scaling and predicting using final model.
- After running the model on the Test Data these are the outcomes we obtain:

Accuracy : 89.33% Sensitivity : 88.14% Specificity : 90.09%

- Evaluation metrics for train & test are very close to around 80%.
- Lead score was assigned.
- Top 3 features are: o Lead Source_Welingak Website o Lead Source_Reference o Current_occupation_Working

Professional Recommendations:

- More budget/spend can be done on Welingak Website in terms of advertising, etc.
- Incentives/discounts for providing reference that convert to lead, encourage to provide more references.
- Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.