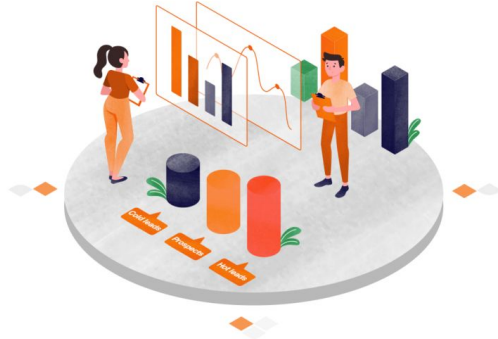
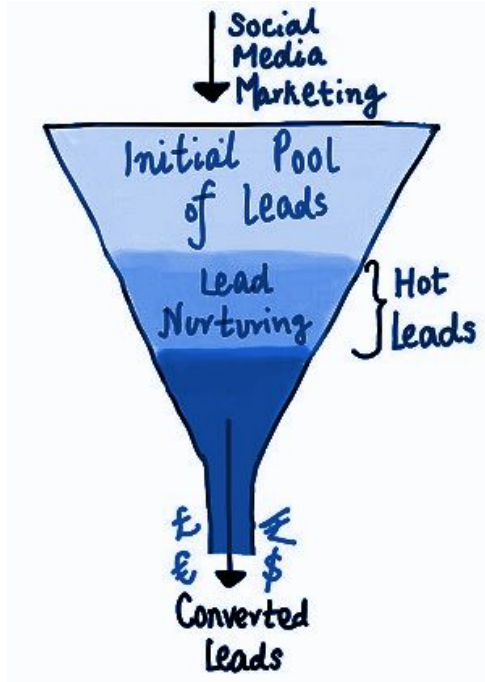


Lead Scoring Assignment

Prediction Model: Approach Explanation



PROBLEM STATEMENT



An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. Now, although X Education gets a lot of leads, its lead conversion rate is very poor.

X Education wants to select the most promising leads, i.e. the leads that are most likely to convert into paying customers through a model wherein each of the leads is assigned a lead score such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

SOLUTIONS | PART A: EDA

1, Data Cleaning

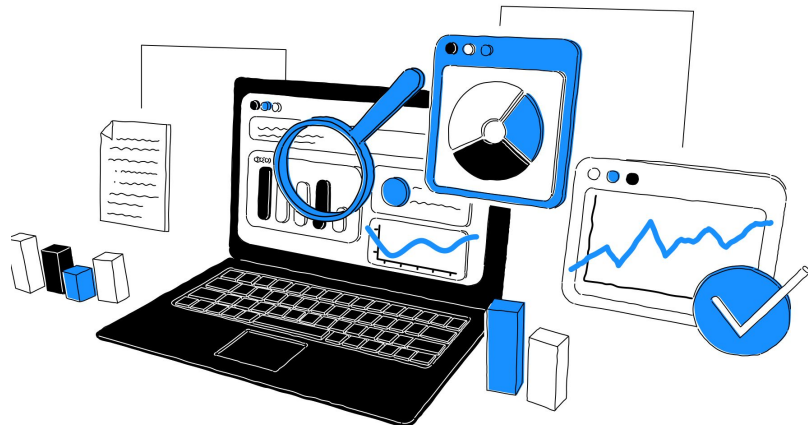
- Check dataset's dimensions
- Drop columns with null percentage > 45%

2, Categorical Variables Analysis

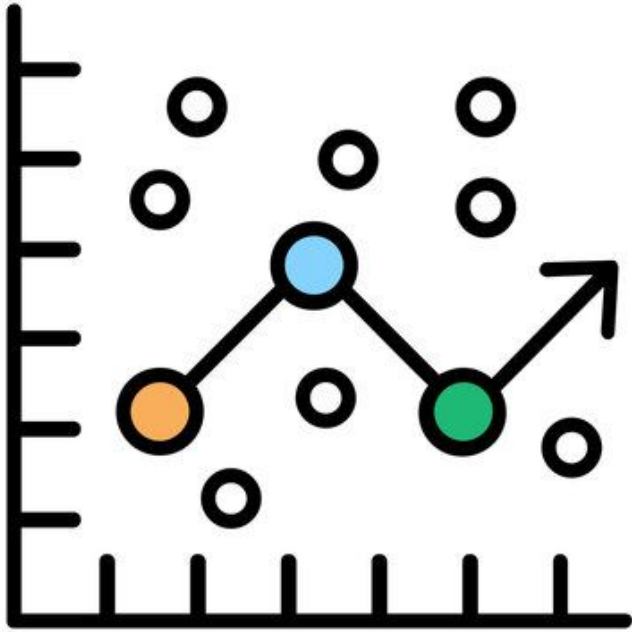
- Impute null values for columns with null percentage > 25%
- Visualize values frequency using countplot
- Group low frequency values into 'Other'

3, Numerical Variables Analysis

- Visualize numerical values using boxplot
- Handle outliers



SOLUTIONS | PART B: REGRESSION MODEL BUILDING



1, Data Preparation

- Perform Train-Test Split
- Impute Yes/No to non-binary values & Create dummy variables for categorical variables
- Normalize numerical values using StandardScaler

2, Model Training

- Select features for model training using RFE (starting with 15 features and drop one by one)
- Train model using GLM and select optimal cutoff point by looking at ROC curve

3, Model Evaluation: Calculate Accuracy and Sensitivity, Specificity via Confusion Matrix

4, Prediction on Test set: Calculate Accuracy and Sensitivity, Specificity via Confusion Matrix



CONCLUSION

The selected model seems to predict the Conversion Rate very well by looking at the following values:

Train Data:

- Accuracy: 89.01%
- Sensitivity: 90.92%
- Specificity: 86.82%

Test Data:

- Accuracy : 88.53%
- Sensitivity : 89.26%
- Specificity : 88.11%



BUSINESS RECOMMENDATIONS

Strategy X Education should employ during periods with several interns hired:

- Prioritize leads from high potential sources (Welingak Website, Reference) and profiles with 'Working Professional' as occupation
- Follow up frequently with customers willing to receive information through email (Tags_Will revert after reading the email) and SMS recipients (Last Notable Activity_SMS Sent)
- Contact as many highly-scored leads as possible during the 2-month period

Strategy X Education should employ during periods when only extremely necessary leads are to be contacted:

- Focus on only high potential channels and occupation profile (as in previous questions)
- Automate communications with potential but not top priority channels eg. Olark Chat and/or profiles eg. Unemployed
- Deprioritize ineligible profiles and customers with low willingness to engage: Tags_Interested in other courses, Tags_Ringing, Last Activity_Email Bounced, Do Not Email