Lead Scoring Case Study

Submitted by

- Sarika Pande
- Sathish Rangarajan
- Shahrukh Naved

Problem statement

- X Education sells online courses to industry professionals.
- Professionals visit the website through marketing efforts on various platforms.
- Visitors may browse courses, fill forms, or watch videos.
- Leads are generated from form submissions, referrals, and website interactions.
- Sales team initiates communication via calls and emails to convert leads.
- Average lead conversion rate is approximately 30%.

Business Goal:

- X Education seeks assistance to prioritize high-conversion probability leads.
- A lead scoring model is required to assign scores based on conversion likelihood.
- Higher lead scores correspond to increased conversion potential.
- CEO's target lead conversion rate is approximately 80%.

Approach

- Develop a logistic regression model which assigns score to leads based on different variables
- At first, we identify the relevant variables which are useful in modelling using EDA
- Prepare and clean the data like, removing nulls, outliers, variables with no variance, etc.
- Prepare the data for modelling: creating dummy variables, binary mapping, scaling etc.
- Build and train the model using train dataset and identifying top features using RFE and manual feature elimination
- Identifying the best model based on p value and VIF
- Evaluating the model based on multiple metrics like accuracy, sensitivity, specificity, etc.
- Predicting lead conversion using the test data
- Evaluate the model using test data based on similar metrics
- Present the model and findings to relevant stakeholders