



---

# LEAD SCORING

---

Edtech Case Study

Ashutosh Singh | Avinash Nayak | Aslam Nadeem

## **DS C58 BATCH 2023**

Analysis conducted for X Education to attract more industry professionals to their courses. Data analysis included information on customer site visits, time spent, site entry methods, and conversion rate.

Steps taken during the analysis:

1. Data Cleaning:

- Partially cleaned data, handled null values.
- Replaced 'select' option with null value.
- Dropping the columns which have NA >30 %

After Cleaning we have left with 70 % of Rows good enough for analysis.

2. Exploratory Data Analysis (EDA):

- Checked data quality.
- Removed irrelevant elements in categorical variables.
- No outliers found in numeric values.

3. Dummy Variables:

- Created dummy variables
- Applied MinMaxScaler to numeric values.

4. Train-Test Split:

- Split data into 70% for training and 30% for testing.

5. Model Building:

- Conducted Recursive Feature Elimination (RFE) to select top 15 relevant variables.
- Manually removed other variables based on VIF values and p-values ( $VIF < 5$  and  $p\text{-value} < 0.05$ ).

6. Model Evaluation:

- Created a confusion matrix.
- Determined optimum cutoff value using ROC curve, achieving around 80% accuracy, sensitivity, and specificity.

7. Prediction:

- Made predictions on the test data with an optimum cutoff of 0.42, resulting in 75-80% accuracy, sensitivity, and specificity.

#### 8. Precision-Recall:

- Utilized Precision-Recall method, found a cutoff of 0.41 with Precision around 73% and Recall around 75% on the test data.

- Identified key variables influencing potential buyers in descending order

1. Current occupation as Unemployed

2. Total time spent on the website.

3. Total number of visits.

4. Lead source:

- Google

- Direct traffic

- Organic search

5. Last activity:

- SMS

- Olark chat conversation

6. Lead origin: Lead add format.

Suggested that X Education can enhance its success by targeting potential buyers based on these critical variables, increasing the likelihood of them purchasing their courses.

X---X---X---X---X---X---X---X