X EDUCATION LEAD SCORING

CASE STUDY ANALYSIS

Presented By –
Ankit Uikey

Introduction:

X Education, an education company 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.

We need analysis of lead conversion using the Logistic Regression model and propose actionable strategies for maximizing efficiency during both high and low conversion periods.

Problem Statement:

X Education *gets a lot of leads, but its* **lead conversion rate is very poor, typically lead conversion rate is around 30%.**

For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as **'Hot Leads'**.

Goal of our Analysis:

Our goal is to improve lead conversion efficiency, especially during peak periods when sales interns are available, and minimize unnecessary phone calls when conversion targets are met.

Analysis Approach:

Data Understanding and Preparation:

❑ For data understanding, I have used Python Jupyter Notebook, to load the required dataset as a DataFrame and then checking various stats like Info and Descriptive Statistical Summary.

❑ Next, I proceeded with Data Cleaning and Pre-processing by Removing the Case Mismatch, and then checking the Data for Duplicates Values and Null Values.

Feature Engineering & Handling Missing Values:

❑ After checking the Null Values, I have dropped Variables (Features) that are having more than 30% of missing values.

❑ For the rest of the Missing Values, I have used appropriate methods like Imputation using Mode() or Median() or Mean() and by creating new features like new categories

❑ Also, I have created new features by Grouping Data in the categorical columns having many categories to avoid over fitting of the data.

Analysis Approach:

Exploratory Data Analysis of all the Features:

❑ I have treated Outliers in the numerical variables using The Percentile Method, by Capping Outliers at 99 Percentile.

❑ For categorical variables I have visualize the distribution using Countplot and dropping Highly biased variables.

❑ And then creating the dummies of all the categorical variables to make them compatible with the logistic regression model.

Model Building using Logistic Regression:

❑ Firstly, I have scaled the dataset using Min-Max Normalization to get all the Features at a common scale between 0 and 1.

❑ Then, I used Logistic Regression to model the probability of lead conversion. Logistic regression is well-suited to this classification problem because it predicts the likelihood of binary outcomes (conversion or no conversion).

❑ After cleaning and preparing the data, I selected 15 Features, including both numerical variables and categorical dummy variables using Recursive Feature Elimination (RFE).

Key Model Insights:

81.6%	74%	79.6%	69.2%
Accuracy Score	**F1 Score**	**Precision**	**Recall**

Model Performance

clasification report:

	precision	recall	f1-score	support
0.0	0.83	0.89	0.86	1436
1.0	0.80	0.69	0.74	874
accuracy			0.82	2310
macro avg	0.81	0.79	0.80	2310
weighted avg	0.81	0.82	0.81	2310

confussion matrix:
[[1281   155]
 [269   605]]

Top Factors Influencing Conversion:

Top 5 Feature Variables:	Coefficients
Total Time Spent on Website	3.6572
Lead Origin - Lead Add Form	3.4154
What is your current occupation_WORKING PROFESSIONAL	2.3091
Grouped_Last_Activity_Messaging Activity	1.6541
Grouped_Last_Activity_Phone & Event Interaction	1.2344

Conclusion:

Model-Driven Decision Making:

Logistic regression models, when built correctly, can provide actionable insights into which factors drive business outcomes. In this case, it helped identify key variables (like Total Time Spent on Website) that had the most influence on lead conversion.

Optimizing Sales Strategy:

The ability to dynamically adjust conversion strategies based on model predictions can lead to significant efficiency gains. By raising thresholds and adjusting focus, the company can avoid unnecessary resource expenditures (like excess phone calls) when it's not critical.

Balancing Business and Technical Needs:

By presenting the results in the form of presentation to the Stakeholders required to translate complex technical insights into clear, actionable business strategies. It highlighted the importance of aligning data science outcomes with business goals to ensure maximum impact.

Recommendations:

✓ **Aggressive Conversion:**

Leverage interns during peak periods to target high-probability leads with focused phone calls.

✓ **Target Met Efficiency:**

Reduce phone calls when the target is reached by focusing only on high-probability leads and using automated follow-ups for the rest.

✓ **Continuous Monitoring:**

Adjust probability thresholds dynamically based on real-time sales trends.

THANK YOU!

Ankit Uikey

ankit00741@gmail.com

7000-126-157