

Summary

This analysis was conducted for **X Education** to identify strategies for attracting more industry professionals to enroll in their courses. The provided dataset offered valuable insights into customer behavior, including how potential leads visit the website, the time they spend on it, their sources of arrival, and the overall conversion rate.

Steps Followed

1. Data Cleaning:

- The dataset was mostly clean, except for a few missing values.
- The “option select” entries were replaced with null values due to their lack of significance.

2. Exploratory Data Analysis (EDA):

- A quick EDA was performed to assess the dataset’s condition.
- Many categorical variable elements were found to be irrelevant and were removed.
- Numeric variables appeared well-distributed, with no significant outliers detected.

3. Dummy Variable Creation:

- Dummy variables were generated.
- Numeric variables were standardized using **MinMaxScaler**.

4. Train-Test Split:

- The dataset was split into **70% training** and **30% testing** data.

5. Model Building:

- **Recursive Feature Elimination (RFE)** was used to select the **top 15 most relevant variables**.
- Additional variables were manually removed based on **Variance Inflation Factor (VIF < 5)** and **p-value (< 0.05)** to ensure a robust model.

6. Model Evaluation:

- A **confusion matrix** was created to analyze model performance.
- The **optimal cutoff value**, determined using the **ROC curve**, provided an **accuracy, sensitivity, and specificity of approximately 76%**.

7. Prediction:

- Predictions were made on the test dataset using an **optimal cutoff of 0.44**, achieving **72% of accuracy, 73% of sensitivity, and 66% of specificity**.

8. Precision-Recall Analysis:

- The **Precision-Recall** method was used for further validation.
- An optimal cutoff of **0.44** was identified, yielding **72% precision** and **73% recall** on the test data.

Key Factors Influencing Potential Buyers (Ranked in Descending Order):

1. **Total Time Spent on the Website** – A strong indicator of conversion likelihood.
2. **Total Number of Website Visits** – More visits correlate with higher engagement and conversion.
3. **Lead Source:**
 - **Google**
 - **Direct Traffic**
 - **Organic Search**
 - **Welingak Website**
4. **Last Activity:**
 - **SMS Sent**
 - **Olark Chat Conversation**
5. **Lead Origin: Lead Add Form Submission** plays a crucial role in lead conversion.
6. **Current Occupation: Working professionals** show the highest conversion rates.

Conclusion

By leveraging these insights, **X Education** can enhance its marketing strategies to **effectively target high-potential leads**, thereby **maximizing conversion rates and driving growth**. Given the strong conversion indicators, strategic interventions focused on high-performing channels and engagement methods will significantly improve the likelihood of course enrollments.