Summary Report

Lead Conversion Model for X Education

Introduction

X Education, an online course provider, faces a significant challenge in converting leads into paying customers. With a lead conversion rate of around 30%, the company seeks to improve efficiency by identifying 'Hot Leads'—those most likely to convert. The objective of this project was to build a logistic regression model to score leads based on conversion probability, thereby optimizing sales efforts and increasing the conversion rate to approximately 80%.

Data Analysis and Preprocessing

The dataset provided contained around 9000 records with various attributes influencing lead conversion. Key preprocessing steps included:

- Handling missing values by imputing or removing records.
- Encoding categorical variables using dummy variables.
- Treating the 'Select' category in certain columns as null values.
- Scaling numerical variables for improved model performance.

Model Development

A logistic regression model was chosen due to its interpretability and suitability for binary classification. The steps involved in model development included:

- Splitting the data into training and test sets (70:30 ratio).
- Performing feature selection based on correlation and statistical significance.
- Using recursive feature elimination (RFE) to identify the most relevant variables.
- Evaluating model performance using metrics such as accuracy, precision, recall, and F1-score.

Key Findings

1. **Top Influencing Variables**:
    - Tags_Lost to EINS
    - Tags_Closed by Horizzon
    - Tags_Will revert after reading the email
2. **Key Categorical Variables**:
    - Lead Source (where leads originate)
    - Last Activity (most recent engagement)

    - Specialization (industry of the lead)

3. **Model Performance**:
    - Achieved an accuracy of ~80% on the test dataset.
    - Precision and recall balanced to ensure minimal false negatives.

Business Recommendations

To maximize lead conversion efficiency, the following strategies were recommended:

- **During Peak Hiring Periods (Intern Hiring Phase)**
    - Lower the classification threshold (e.g., 0.4 instead of 0.5) to capture more potential leads.
    - Increase frequency of follow-ups with leads who showed interest but haven't responded.
    - Prioritize engagement with leads tagged as "Will revert after reading the email."
- **When Quarterly Targets Are Achieved Early**
    - Raise the classification threshold (e.g., 0.7 or 0.8) to focus only on the most promising leads.
    - Prioritize highly engaged leads with strong intent signals (e.g., requested course details, spent long durations on the website).
    - Reduce phone calls for low-engagement leads and rely on email/SMS outreach instead.

Conclusion

The logistic regression model successfully identified high-potential leads, allowing X Education to allocate resources more efficiently. By implementing a data-driven lead-scoring system, the company can increase conversion rates, improve sales efficiency, and enhance overall profitability. The strategic adjustments recommended for different business scenarios will further optimize sales team efforts, ensuring a sustainable growth model for X Education.