Presentation: Optimizing Lead Conversion Strategies at X Education

Objective: Enhance lead conversion rates through data-driven strategies tailored for different operational phases.

Data Analysis and Model Overview

Data Preprocessing:

- Loaded and cleaned dataset to handle missing values.
- Encoded categorical variables using one-hot encoding.
- Dropped irrelevant columns (e.g., Prospect ID, Lead Number).

Logistic Regression Model:

- Trained a logistic regression model to predict lead conversion.
- Evaluated model performance using metrics like accuracy, ROC-AUC.

Key Model Insights

Top Three Variables Contributing to Conversion Probability:

Identified through coefficients:

1. Last Activity_Approached upfront
2. Lead Source_NC_EDM
3. Last Notable Activity_Had a Phone Coversation

Top Three Categorical/Dummy Variables negatively Impacting Conversion:

Leveraging categorical features:

1. Lead Source_Olark Chat
2. Lead Source_Google
3. Lead Source_Direct Traffic

Technical Analysis (1/2)

Step 1: Data Preprocessing

1. **Load the Data**: Read the dataset into a pandas DataFrame.
2. **Checking the Dataset**: Checking type of data in data set.
3. **Handle Missing Values**: Deal with missing values by either imputing them or removing the rows/columns with excessive missing data. In some cases, missing values have been replaced with 'Not Provided' to avoid loss of data as those columns look like important for current analysis.
4. **Drop Irrelevant Columns**: Remove columns that are not useful for the prediction task, such as Prospect ID, Country, Tags etc.

Step 2: Data Visualization and Exploratory Data Analysis

1. **Univariate Analysis: To understand each variable better.**
2. **Bi-variate Analysis: To understand relationship between two variables.**
3. **Introduction of Dummies : Dummies introduced for all categorical variables and removed categorical variables as those are already covered under dummies.**

Technical Analysis (2/2)

Step 3: Model Training

1. **Split the Data**: Split the data into training and testing sets. We have used 70: 30 ratio for splitting the data.
2. **Re-scaling:** Re-scaling the features of train data set by using Min-Max method.
3. **Data division:** Dividing data into X & Y sets for model building.
4. **Train the Model**: Use logistic regression to train the model on the training set. After multiple iteration, achieved a model where p value is lower than 0.05 and VIF is lower than 5.
5. **Evaluate the Model**: Evaluate the model on the testing set using metrics like Confusion metrics, Accuracy, ROC.

Step 4: Assign Lead Scores

1. **Predict Probabilities**: Use the trained model to predict the probability of conversion for each lead.
2. **Scale to 0-100**: Scale these probabilities to a range of 0-100 to assign lead scores.

Note: The model shows 70% accuracy which is close to study's objective

Business Implications

Strategy for Maximizing Lead Conversions During Internship Period

- **Objective**: Convert almost all high probability leads (predicted as 1).
- **Recommendation**: Implement an aggressive calling strategy focusing on top probability leads:
  - Prioritize high probability leads for manual calling.
  - Use automated systems for efficient outbound calls.
  - Provide personalized scripts based on model insights.

Strategy to Minimize Useless Phone Calls Post Early Target Achievement

- **Objective**: Focus on new work and minimize unnecessary phone calls.
- **Recommendation**: Shift to automated email campaigns and strategic engagement:
  - Refine lead scoring to focus only on high probability leads.
  - Automate personalized email sequences based on lead behavior.
  - Explore new market opportunities with freed resources.

Strategic Recommendations

Peak Hiring Period Strategy: Maximizing Lead Conversion

Approach:

- Focus on high probability leads predicted by the model (scored as 1).
- Implement automated calling campaigns for efficient outreach.
- Provide interns with personalized scripts and continuous feedback.

Benefits:

- Maximizes conversion during peak periods.
- Utilizes interns effectively with structured calling campaigns.

Early Target Achievement Strategy: Minimizing Useless Calls

Approach:

- Refine lead scoring to focus only on high probability leads.
- Shift to automated email campaigns for engagement.
- Explore new market opportunities and product enhancements.

Benefits:

- Reduces operational costs associated with unnecessary calls.
- Aligns sales efforts with strategic business development goals.

Implementation and Next Steps

- **Implementation Steps:**
- Deploy refined strategies based on model insights.
- Integrate automated systems for efficient lead management.
- Monitor performance metrics for continuous optimization.
- **Next Steps:**
- Conduct A/B testing on communication strategies
- Enhance CRM capabilities for real-time lead prioritization.

Conclusion

- **Impact:** Achieve higher conversion rates through targeted strategies.

- **Long-term Vision:** Align data-driven insights with business growth objectives.

- **Continuous Improvement:** Iterative refinement based on real-time performance feedback.

Our logistic regression model provides actionable insights for optimizing lead conversion strategies. By leveraging predictive analytics, we can target high-value leads effectively while minimizing resource wastage during non-optimal periods.

Thank You