

Lead Scoring Summary Report

Introduction

The objective of this analysis is to predict the likelihood of lead conversion using a logistic regression model and assign a lead score between 0 and 100 to each lead. This analysis aims to identify key factors influencing lead conversion and provide actionable recommendations to improve conversion rates. The insights gained from this analysis will help the sales team prioritize leads and allocate resources more effectively.

Data Preprocessing

The dataset used for this analysis was created by merging three CSV files into a single dataframe. The following preprocessing steps were performed:

- **Dropped Irrelevant Columns:** Columns such as 'Prospect ID', 'Lead Number', 'Lead Origin', 'Lead Source', 'Do Not Email', 'Do Not Call', 'Last Activity', 'Country', 'City', and 'Last Notable Activity' were removed to focus on relevant features.
- **Handled Missing Values:** Missing values were filled with a placeholder value 'Unknown' to ensure the model could handle missing data without errors.
- **Converted Categorical Variables:** Categorical variables were converted into dummy/indicator variables to enable the logistic regression model to process categorical data effectively.

Model Building

The logistic regression model was built using the following steps:

- **Feature and Target Variables:** The feature variables (X) and target variable (y) were defined. The target variable 'Converted' indicates whether a lead was converted (1) or not (0).
- **Data Splitting:** The dataset was split into training (80%) and testing (20%) sets to evaluate the model's performance on unseen data.
- **Model Training:** A logistic regression model was trained on the training set. Logistic regression was chosen for its simplicity and effectiveness in binary classification tasks.

Model Evaluation

The model's performance was evaluated using the following metrics:

- **Accuracy:** The model achieved an accuracy of approximately 94% on the test set, indicating good predictive performance.
- **Classification Report:** The classification report provided detailed metrics such as precision, recall, and F1-score, helping to understand the model's performance in terms of correctly identifying converted and non-converted leads.

Lead Scoring

Lead scores were calculated using the predicted probabilities from the logistic regression model:

- Predicted Probabilities: Probabilities for each lead in the dataset were predicted, indicating the likelihood of each lead converting.
- Lead Scores: Lead scores were calculated by multiplying the predicted probabilities by 100 and added to the original dataframe.
- Saved Dataframe: The dataframe with lead scores was saved to a new CSV file, allowing the sales team to access and use the lead scores for prioritizing leads.

Key Variables Influencing Conversion

The analysis identified the following key variables influencing lead conversion:

- Top Three Variables:
 1. Total Time Spent on Website: Leads who spend more time on the website are more likely to convert as they show higher engagement and interest.
 2. Page Views Per Visit: Leads who view more pages per visit are more engaged and have a higher probability of conversion.
 3. TotalVisits: Leads with a higher number of visits are more likely to convert as they show consistent interest.
- Top Three Categorical/Dummy Variables:
 1. Lead Source_Reference: Leads that come through references have a higher probability of conversion. Focusing on referral programs can help increase conversions.
 2. What is your current occupation_Working Professional: Leads who are working professionals are more likely to convert. Tailoring marketing strategies to target working professionals can improve conversion rates.
 3. Specialization_Human Resource Management: Leads with a specialization in Human Resource Management have a higher probability of conversion. Creating targeted content and offers for this specialization can enhance conversions.

Recommendations

Based on the analysis, the following recommendations are provided:

For Aggressive Lead Conversion:

1. Prioritize High-Scoring Leads: Use the lead scores to prioritize leads with the highest scores. These leads have the highest probability of conversion and should be contacted first.
2. Allocate Interns to High-Scoring Leads: Assign interns to focus on calling and following up with high-scoring leads to ensure immediate attention.
3. Personalized Communication: Train interns to use personalized communication strategies when contacting leads. Addressing leads by their name and referring to their specific interests can increase conversion chances.
4. Follow-Up Schedule: Implement a structured follow-up schedule to ensure leads are contacted multiple times if necessary. Persistence can lead to higher conversion rates.

For Minimizing Useless Phone Calls:

1. Focus on Low-Scoring Leads: Use the lead scores to identify leads with lower probabilities of conversion. These leads can be deprioritized for phone calls during this period.
2. Automated Communication: Implement automated email or SMS campaigns to engage with low-scoring leads, allowing the sales team to maintain contact without making phone calls.
3. Content Marketing: Develop and share valuable content (e.g., blog posts, webinars, e-books) to nurture low-scoring leads over time, keeping them engaged without direct phone calls.
4. Data Analysis: Use this period to analyze data and identify trends or patterns in lead behavior, providing insights for future marketing strategies and improving overall lead conversion processes.

Conclusion

This analysis provides valuable insights into the factors influencing lead conversion and offers actionable recommendations to improve conversion rates. By leveraging the lead scores and focusing on key variables, the sales team can prioritize leads more effectively and allocate resources to maximize conversions. The recommendations provided will help the sales team adopt a more strategic approach to lead conversion, ultimately driving better business outcomes.