

Summary Report on Lead Scoring Model

Introduction: The purpose of this project was to help X Education improve their lead conversion rate by building a logistic regression model that identifies potential 'Hot Leads'-those most likely to convert into paying customers. The company currently converts only 30% of leads, and the CEO's target is to raise this to 80%. The provided dataset contained 9,000 leads, each with multiple features such as Lead Source, Total Visits, Last Activity, and more.

Data Preprocessing: The dataset required significant cleaning before model building. Several columns had missing values, with some columns missing over 45% of the data. Columns with excessive missing values, such as 'Lead Quality' and 'Asymmetrique Activity Index', were dropped. For the remaining columns with missing values (such as 'Specialization' and 'City'), appropriate imputation methods were applied-either filling with 'Unknown' or using mean values for numerical features. Categorical features were one-hot encoded to handle multiple categories effectively.

Model Building: A logistic regression model was chosen for its interpretability and ability to output probabilities that can be used to assign a lead score to each customer. The dataset was split into training and testing sets (80% train, 20% test), and the model was trained on the clean data. Several key variables contributed significantly to the prediction of lead conversion, including 'Total Time Spent on Website', 'Lead Source', and 'Last Activity'.

Model Performance: The logistic regression model performed well, with the following evaluation metrics: - Accuracy: 93.77% - Precision: 93.35% - Recall: 90.96% - F1-Score: 92.14% - ROC-AUC Score: 97.76% These results show that the model is capable of accurately identifying leads that are likely to convert, with a strong balance between precision and recall.

Key Insights and Recommendations:

1. 'Total Time Spent on Website' was found to be the most important feature. Leads who spend more time browsing the courses are more likely to convert, indicating that X Education should encourage website engagement.

2. Leads sourced from Google and Direct Traffic had higher conversion rates, suggesting that these channels should be a focus for marketing efforts.

3. 'Last Activity' such as 'Email Opened' was another strong predictor of conversion. Leads actively engaging with emails are more likely to convert, so the sales team should prioritize follow-up with such leads.

Future Recommendations:

- The company should use this model to focus efforts on the top-scoring leads, ensuring that the sales team spends their time on the leads most likely to convert.

- In periods of aggressive lead conversion (e.g., during the intern phase), the focus should be on high-probability leads, with more frequent calls and follow-ups.

- In times of reduced lead conversion effort (e.g., after hitting quarterly targets), the model can help minimize unnecessary calls by focusing only on the highest scores.

Conclusion: By applying this model, X Education can significantly improve its lead conversion rate by targeting the most promising leads, thereby increasing efficiency and achieving the CEO's target conversion rate.