Logistic Regression Analysis for Lead Conversion

DATE: [19-10-2024]

Objective

Goal: Predict lead conversion likelihood for targeted marketing campaigns.

Business Impact: Improve conversion rates by focusing on high-probability leads.

Data Overview

Dataset: Customer data including web visits, lead source, specialization, activity, etc.

Target Variable: Lead conversion (binary classification: 0 for not converted, 1 for converted).

Feature Selection

Method: RFE (Recursive Feature Elimination) used to select top features.

Key Features: Total Time Spent on Website, Lead Source, Last Activity, Specialization

Business Relevance: Time spent and lead source indicate lead interest and engagement.

Logistic Regression Model

Algorithm: Logistic Regression

Why Logistic Regression? Suitable for binary classification and interpretable coefficients.

Technical Summary: Train/Test Split and Scaling applied to numerical variables.

Model Performance (Train Set)

Accuracy: 78.6% (threshold: 0.5)

Confusion Matrix: TP: 1627, FP: 455, TN: 2595, FN: 693

Sensitivity (Recall): 70.1% | Specificity: 85.1%

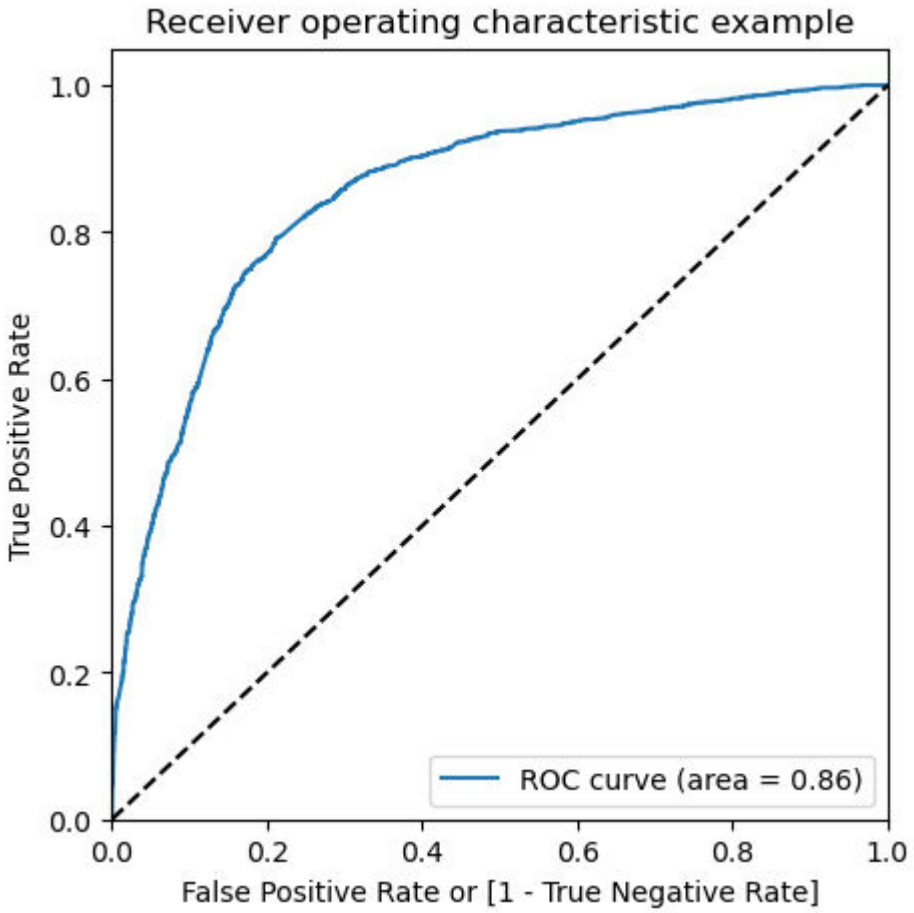
ROC and AUC

ROC Curve: Visualized for performance evaluation.

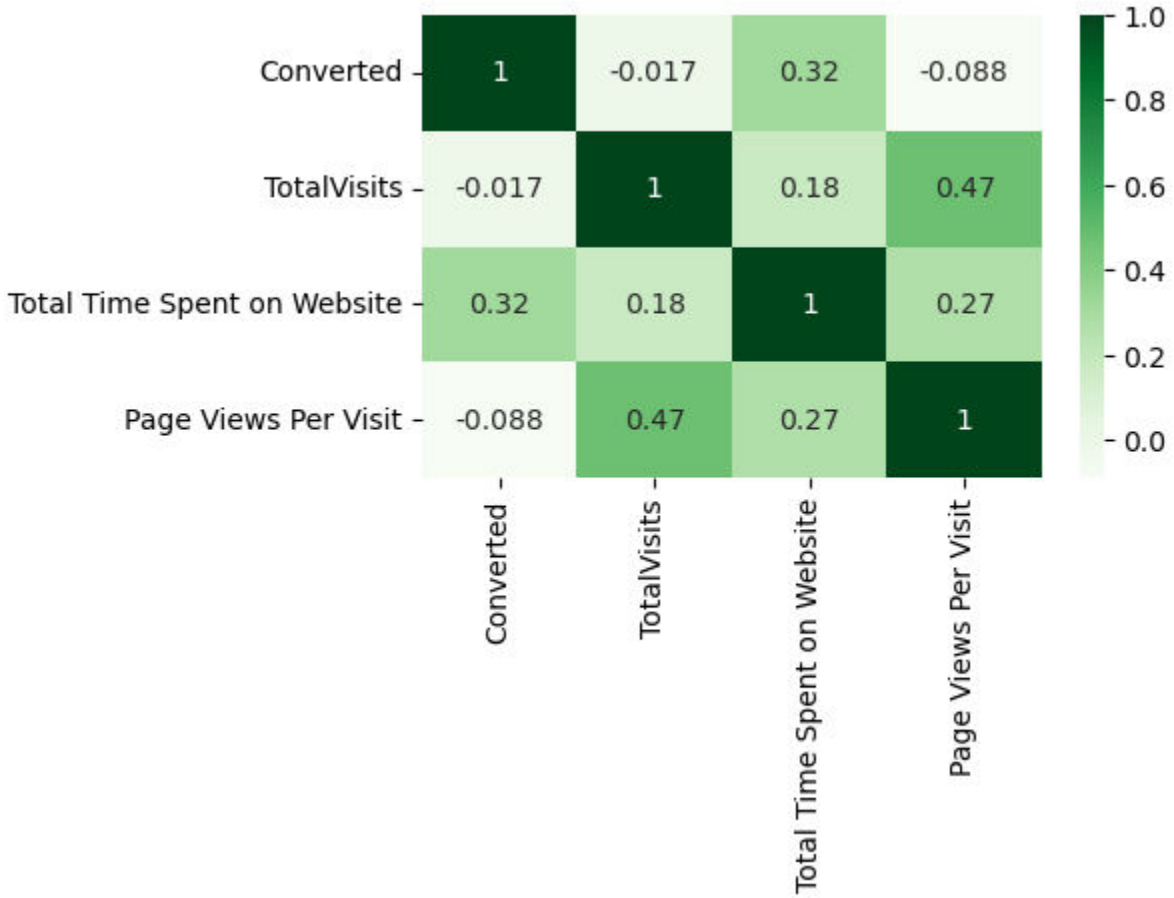
AUC Score: 0.85

Business Insight: AUC indicates the model's ability to distinguish between converted and non-converted leads.

ROC Curve


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.86)

Heatmap to find the correlation



Optimal Probability Threshold

Analysis: Tested thresholds from 0.1 to 0.9 for best performance.

Optimal Cut-off: 0.4 (Based on accuracy, sensitivity, and specificity balance).

Business Implication: Higher sensitivity ensures fewer missed opportunities in lead conversion.

Model Performance (Test Set)

Accuracy: 78.4%

Confusion Matrix: TP: 765, FP: 272, TN: 1040, FN: 225

Precision: 73.8% | Recall: 77.3%

Business Insight: Consistent performance across training and test data indicates reliable predictions.

Business Recommendations

- Target high-probability leads: Focus marketing efforts on leads with >40% conversion probability.
- Optimize content: Engage leads spending more time on the website.
- Monitor lead sources: Prioritize high-conversion lead sources like 'Olark Chat'.

Conclusion

Model Summary: Logistic regression is effective for lead conversion prediction with a 78% accuracy.

Next Steps: Implement model in marketing campaigns, track real-world performance.