Model Analysis Report

Comprehensive Evaluation and Interpretation

Generated on: February 26, 2025

Executive Summary

This report presents a comprehensive analysis of the predictive model performance, including data quality assessment, feature relevance, class balance, model performance metrics, and financial impact analysis. The insights provided in this report aim to support data-driven decision making.

Note: Some complex visualization images may not be included in this PDF due to formatting constraints. These images can be viewed directly in the project's models/plots directory.

Data Quality Analysis

Missing Values Summary

Feature	Missing Count	Missing Percentage
Feature_ae_0	0	0.00%
Feature_dn_1	0	0.00%
Feature_cn_2	0	0.00%
Feature_ps_3	0	0.00%
Feature_ps_4	0	0.00%
Feature_ee_5	0	0.00%
Feature_cx_6	0	0.00%
Feature_cx_7	0	0.00%
Feature_em_8	0	0.00%
Feature_nd_9	0	0.00%
Feature_jd_10	0	0.00%
Feature_md_11	0	0.00%
Feature_ed_12	0	0.00%
Feature_dd_13	0	0.00%
Feature_hd_14	0	0.00%
Feature_ld_15	0	0.00%
Feature_cd_16	0	0.00%
Feature_md_17	0	0.00%
Feature_dd_18	0	0.00%
Feature_pd_19	0	0.00%
Response	0	0.00%

Feature Relevance Analysis

Mutual Information with Target

Feature	Mutual Information Score
Feature_dn_1	0.0770
Feature_em_8	0.0726
Feature_cx_7	0.0694
Feature_cx_6	0.0668
Feature_nd_9	0.0653
Feature_ee_5	0.0540
Feature_ps_3	0.0389
Feature_pd_19	0.0326
Feature_md_17	0.0281
Feature_ps_4	0.0204
Feature_cd_16	0.0155
Feature_ae_0	0.0134
Feature_jd_10	0.0091
Feature_dd_13	0.0075
Feature_cn_2	0.0067
Feature_ed_12	0.0051
Feature_hd_14	0.0048
Feature_md_11	0.0044
Feature_dd_18	0.0017
Feature_ld_15	0.0011

Mutual Information measures the amount of information obtained about the target variable when observing each feature. Higher values indicate stronger relevance to the prediction task.

Highly Correlated Features

Feature 1	Feature 2	Correlation
Feature_em_8	Feature_ee_5	0.9723
Feature_nd_9	Feature_ee_5	0.9073

Feature_nd_9	Feature_em_8	0.9453

Highly correlated features provide similar information and may introduce redundancy in the model. Correlation values close to 1 or -1 indicate strong linear relationships between features.

Feature Correlation Matrix

The correlation matrix visualizes the pairwise correlation between numerical features. Darker red indicates strong positive correlation, darker blue indicates strong negative correlation, and light colors indicate weak correlation.

[Image 'correlation_matrix.png' is available in the project directory but not shown in the PDF due to formatting constraints.]

Class Balance Analysis

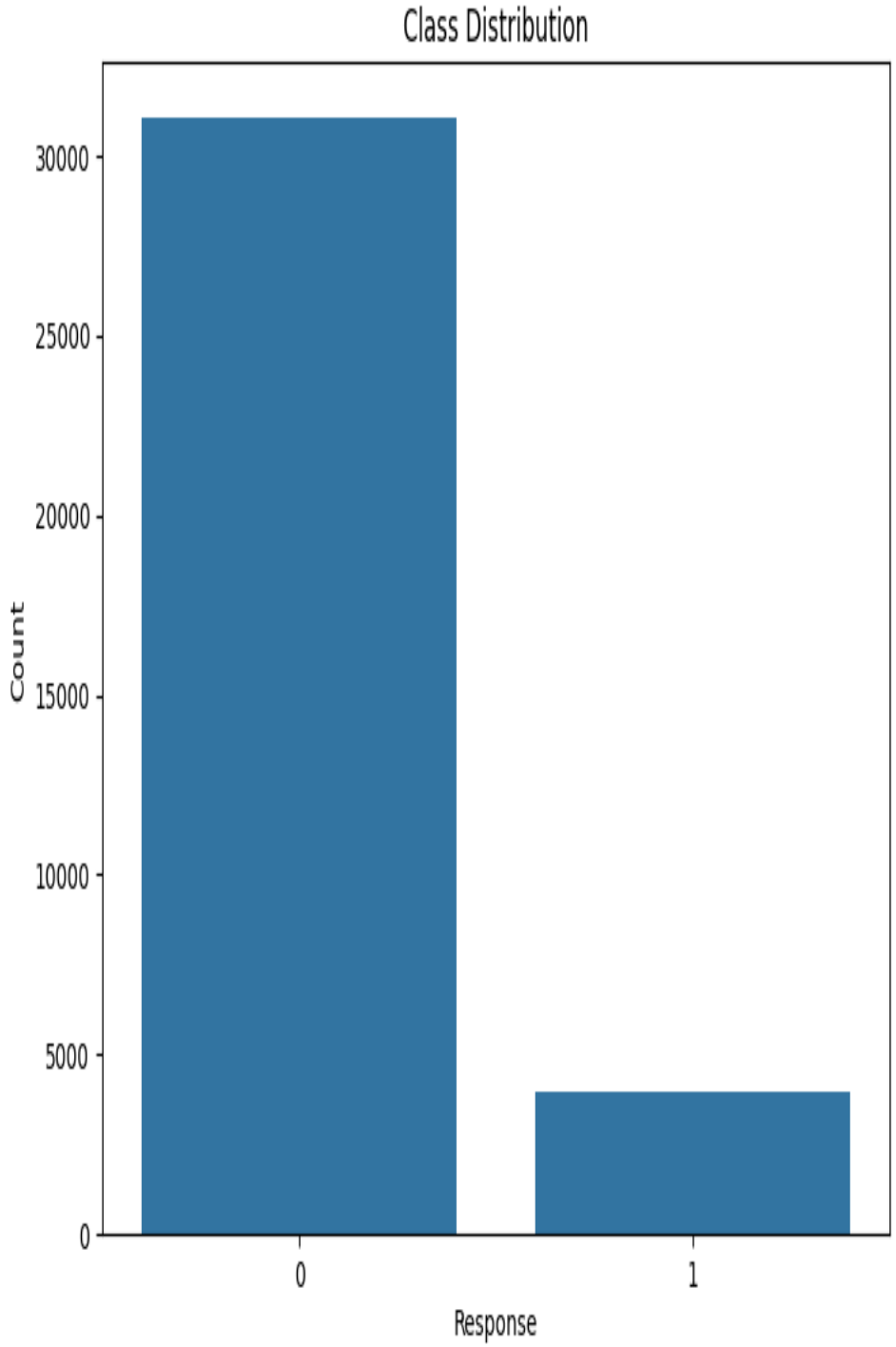
Class Counts

Class	Count
0	31070
1	3930

Class Proportions

Class	Proportion
0	88.77%
1	11.23%

Class imbalance can significantly impact model performance. The distribution shown above indicates the relative frequency of each class in the dataset.

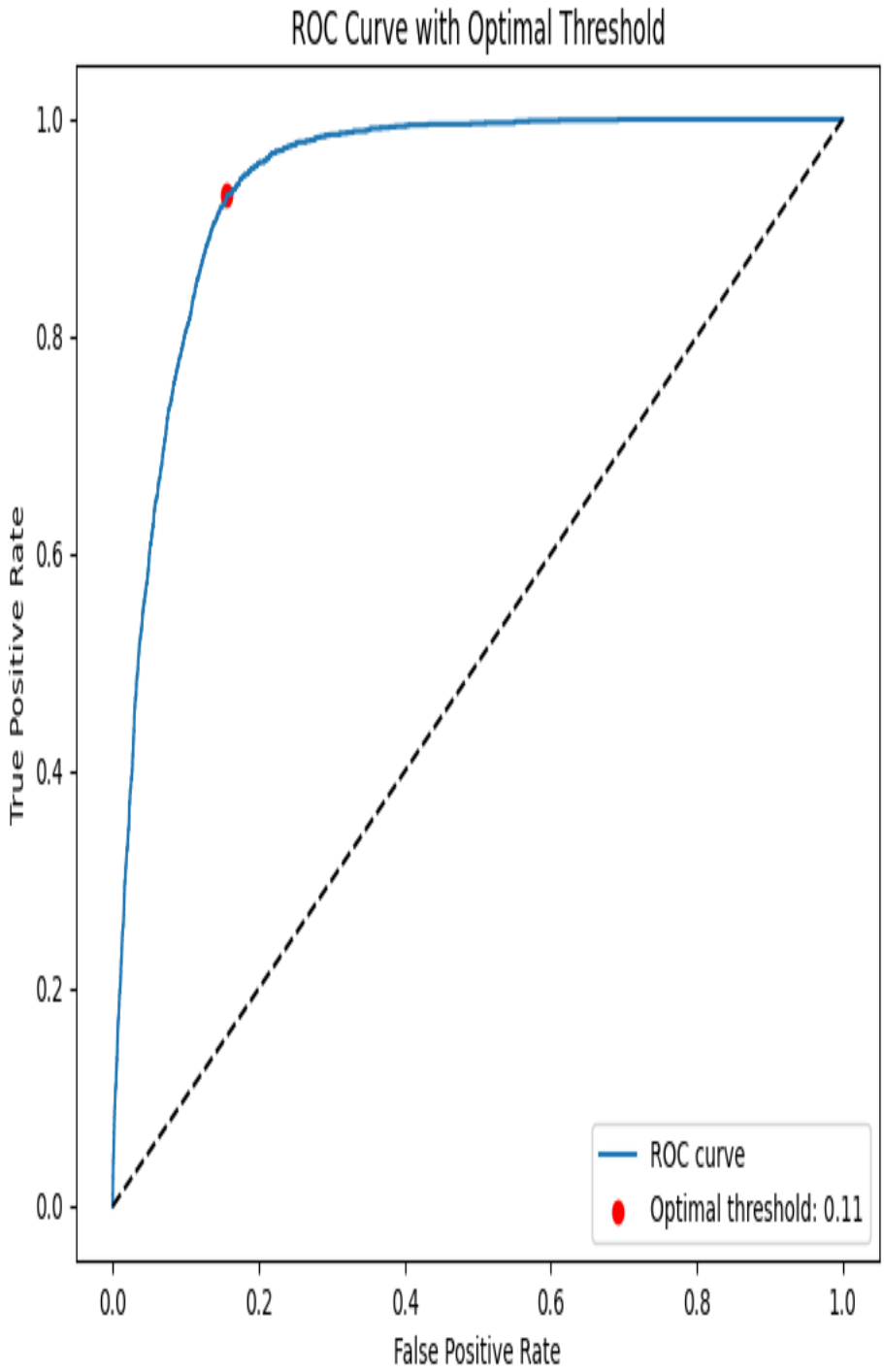

Class Distribution
30000
25000
20000
15000
10000
5000
0
Count
0
1
Response

Model Performance Analysis

Optimal Threshold

Metric	Value
Optimal Threshold	0.1074
Tpr	0.9300
Fpr	0.1563

The optimal threshold is the probability cutoff that maximizes model performance by balancing the trade-off between true positive rate and false positive rate. This threshold can be adjusted based on business requirements.


ROC Curve with Optimal Threshold
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve
Optimal threshold: 0.11

The plot above highlights the optimal threshold point on the ROC curve, which represents the best balance between true positive rate and false positive rate.

Financial Impact Analysis

Campaign Overview

Metric	Value
Campaign Size	10,000

Financial Metrics

Metric	Value
Total Profit	$-1356915.00
Opportunity Loss	$74945.00

Profit by Risk Band

Risk Band	Profit
High	$-211590.00
Medium	$-417975.00
Low	$-727350.00

Scaled Confusion Matrix

	Predicted Negative	Predicted Positive	Total
Actual Negative	812	8,065	8,877
Actual Positive	59	1,063	1,122
Total	871	9,128	9,999

The scaled confusion matrix shows the predicted distribution of customers in a campaign of 10000 customers. True positives (1,063) represent correctly targeted customers, while false positives (8,065) represent customers incorrectly targeted. False negatives (59) represent missed opportunities.

[Image 'financial_impact_detailed.png' is available in the project directory but not shown in the PDF due to formatting constraints.]