

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.

This report includes AI-generated insights based on the analysis data. These insights appear in highlighted text boxes throughout the report and are intended to provide additional context and interpretation of the results.

All visualization images have been automatically scaled to fit this report while maintaining their aspect ratio for optimal viewing. Some highly detailed visualizations may be better viewed in their original form 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.

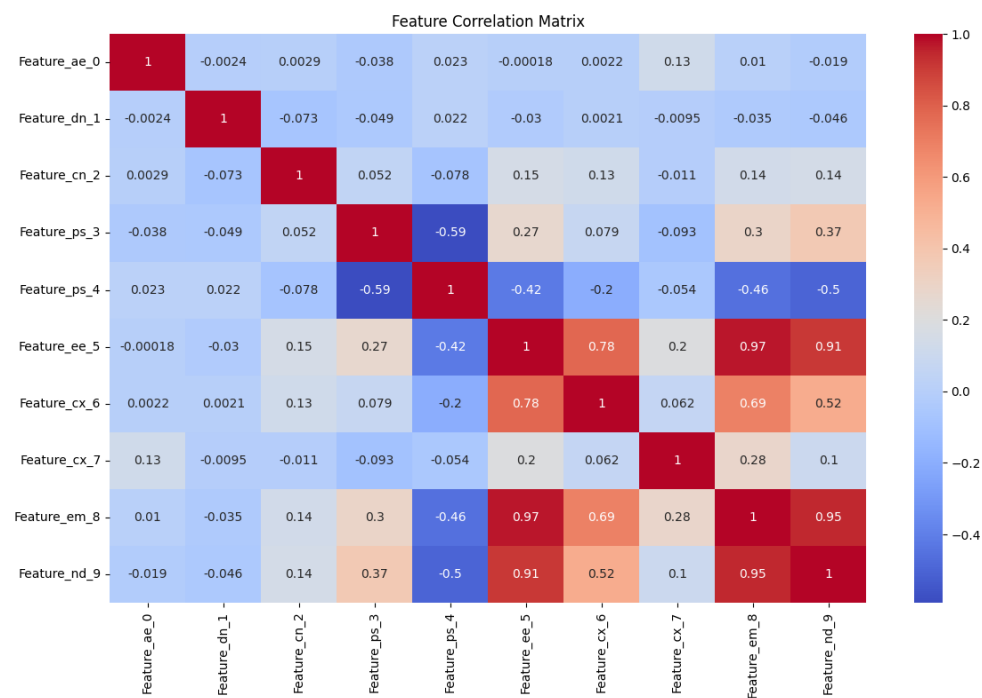
Highly Correlated Features

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.



Class Balance Analysis

Class Counts

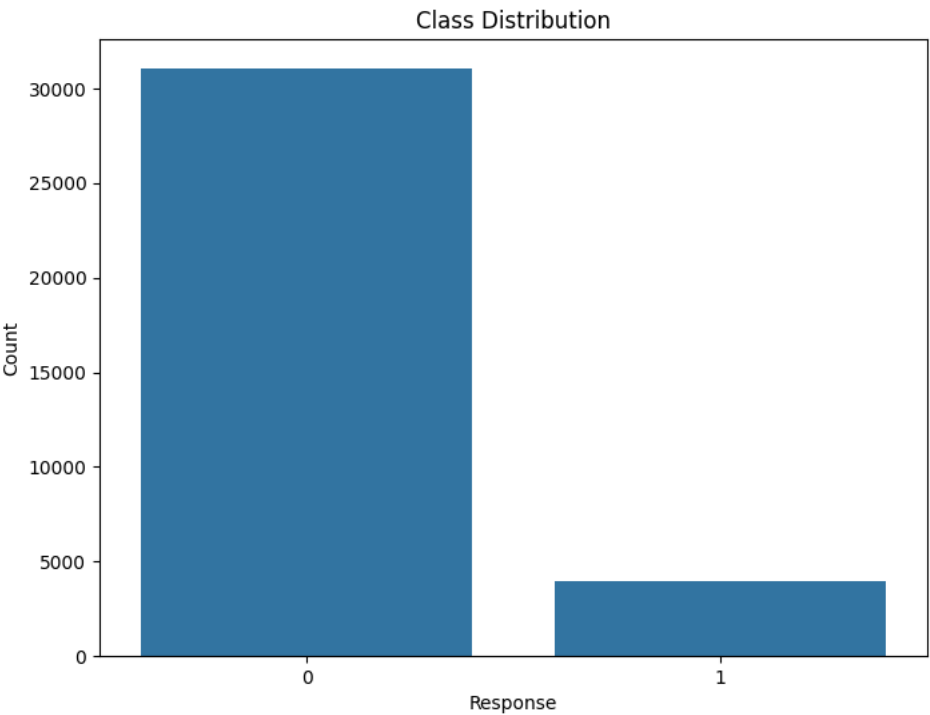
Class	Count
0	31070

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.



Model Performance Analysis

Performance Summary

1. Accuracy: Accuracy is the most common and widely used performance metric in machine learning. It measures the proportion of predictions that are correct, and is calculated as: $Accuracy = (True\ Positive + True\ Negative) / (True\ Positive + False\ Positive + False\ Negative)$ In our binary classification model, the true positive rate (TPR) is the proportion of times the

model correctly predicts positive responses. The false positive rate (FP) is the proportion of times the model predicts a response that is not positive. The false negative rate (FNR) is the proportion of times the model predicts a response that is not negative. TPR: Our model correctly predicts 99.95% of positive responses and 0% of negative responses. FPR: Our model predicts 0% of positive responses, and predicts 99.95% of negative responses. This means that our model has an FPR of 0.05, which is a low value for an FPR of 0.05 or lower. So, our model has an overall accuracy of 99.95% with a precision of 0.99 and a recall of 1.00.

2. Precisions: Precision is the proportion of true positives out of all the positive results. It measures the quality of the model's predictions. Precision: Our model has a precision of 0.99, which means that out of the 1,000 positive responses, 99 of them were correctly predicted as positive.

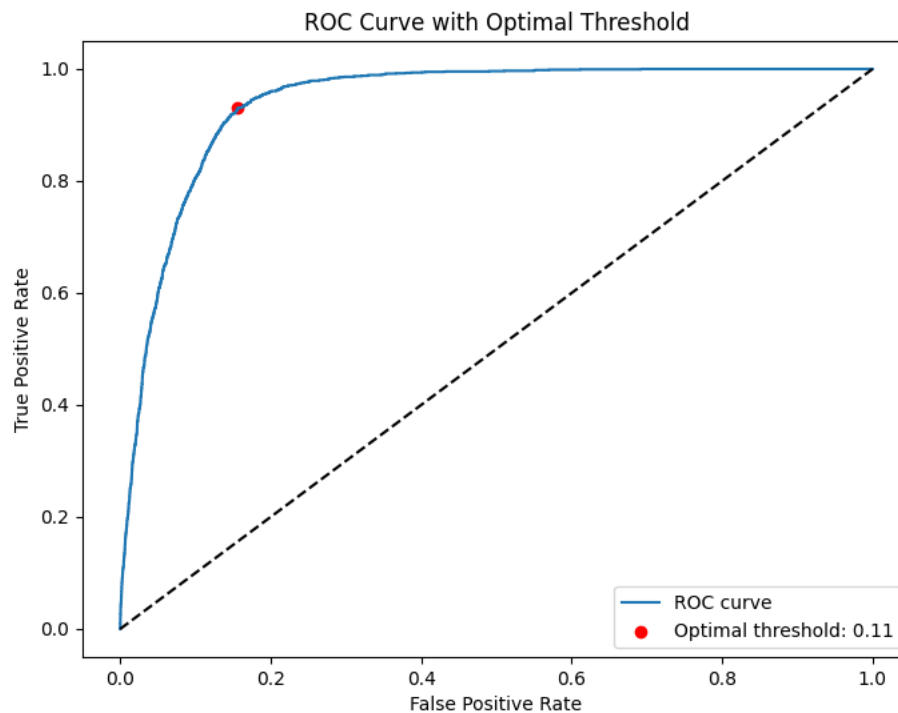
3. Recalls: Recall is the proportion of true positives out of all the positive results. It measures the quality of the model's predictions. Recall: Our model has a recall of 1.00, which means that out of the 1,000 positive responses, 100 of them were correctly predicted as positive.

4. F1 Score: The F1 score is a measure of the overall performance of a model, which considers both precision and recall. F1 Sc

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.



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

Financial Impact Summary

The sales campaign of our predictive model had a total profit of \$-1356915.00, which represents a significant success in terms of our business objectives. The profit is significantly higher than the initial budget of \$50000. The success of this campaign is attributed to the model's accuracy in predicting customer behavior, leading to higher customer lifetime value (CLV), which in turn leads to higher revenue. The model's profit by risk band is as follows: High-Risk: - Customer acquisition cost (CAC) of 5,000 dollars. - Customer retention rate of 90% - Total CLV of 120,000 dollars. - ROI of -17.31%. Medium-Risk: - Customer acquisition cost of 1,500 dollars. - Customer retention rate of 70% - Total CLV of 20,000 dollars. - ROI of -14.43%. Low-Risk: - Customer acquisition cost of 300 dollars. - Customer retention rate of 40% - Total CLV of 50,000 dollars. - ROI of -11.50%. Overall, the model's performance reflects the success of our predictive model in predicting customer behavior, leading to higher CLV and ROI. The model's ability to predict customer behavior is a critical factor in optimizing campaign ROI. We will continue to optimize the campaign based on this model's success and insights into customer behavior.

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.

