Model Validation Report

Generated on: {{validation\_date}}

Validated by: {{validated\_by}}

Model File: {{model\_path}}

Task: {{task\_type}}

Table of Contents

[1. Model Metadata 3](#_Toc210534342)

[2. Performance Metrics 3](#_Toc210534343)

[a. Classification Metrics 3](#_Toc210534344)

[b. Classification Visuals 3](#_Toc210534345)

[b.1. ROC Curve 3](#_Toc210534346)

[b.2. Precision–Recall Curve 3](#_Toc210534347)

[b.3. Cumulative Gain / Lift 4](#_Toc210534348)

[b.4. Calibration (Reliability) Curve 4](#_Toc210534349)

[b.5. Confusion Matrix (image) 4](#_Toc210534350)

[b.6. Confusion Matrix (Classification) 4](#_Toc210534351)

[b.7. Decile Lift Table 4](#_Toc210534352)

[b.8. KS Curve (Kolmogorov–Smirnov) 4](#_Toc210534353)

[3. Data Quality (Cleaned Data) 4](#_Toc210534354)

[4. Raw Data Validation 5](#_Toc210534355)

[5. Correlation Analysis 5](#_Toc210534356)

[a. Top High-Correlation Feature Pairs (|r| ≥ Threshold) 5](#_Toc210534357)

[b. Correlation Heatmap 5](#_Toc210534358)

[6. Variance Inflation Factor (VIF) Check 5](#_Toc210534359)

[7. Exploratory Data Analysis (EDA) 5](#_Toc210534360)

[8. Benchmark (Logistic Regression Baseline) 6](#_Toc210534361)

[a. Logistic Regression Results 6](#_Toc210534362)

[b. Logistic Regression Coefficients 6](#_Toc210534363)

[c. Baseline Metrics 6](#_Toc210534364)

[9. Stress Testing Results 6](#_Toc210534365)

[10. Input Cluster Coverage Check 6](#_Toc210534366)

[a. Cluster Summary Table 6](#_Toc210534367)

[b. Cluster Distribution Plot 6](#_Toc210534368)

[11. Rule Engine Check 6](#_Toc210534369)

[12. Model Contents Summary 7](#_Toc210534370)

[a. Model Class: {{ModelMetaCheck.model\_class}} 7](#_Toc210534371)

[b. Module: {{ModelMetaCheck.module}} 7](#_Toc210534372)

[c. Hyperparameters: 7](#_Toc210534373)

[d. Attributes: 7](#_Toc210534374)

[13. Explainability 7](#_Toc210534375)

[a. SHAP Summary Plot 7](#_Toc210534376)

[b. Top Features by Mean |SHAP| (Bar) 7](#_Toc210534377)

[c. Top SHAP Features 7](#_Toc210534378)

## Model Metadata

Model Type: {{ModelMetaCheck.model\_type}}

Feature Count: {{ModelMetaCheck.n\_features}}

Feature Names: {{ModelMetaCheck.feature\_names}}

Training Rows: {{ModelMetaCheck.n\_train\_rows}}

Target Balance: {{ModelMetaCheck.target\_balance}}

## Performance Metrics

### Classification Metrics

AUC-ROC: {{classification\_summary.AUC2}}

KS: {{classification\_summary.KS2}}

F1: {{classification\_summary.F12}}

PR AUC: {{classification\_summary.PR\_AUC2}}

Gini: {{classification\_summary.GINI2}}

Precision: {{classification\_summary.Precision2}}

Recall: {{classification\_summary.Recall2}}

Accuracy: {{classification\_summary.Accuracy2}}

Brier Score: {{classification\_summary.Brier2}}

### Classification Visuals

### ROC Curve

Shows the trade-off between True Positive Rate (sensitivity) and False Positive Rate (1 – specificity) across thresholds. The closer the curve is to the top-left corner, the better the model’s discrimination ability.

[[IMG:roc]]

### Precision–Recall Curve

Plots precision vs recall at different thresholds. Particularly useful when the positive class is rare, since it focuses on the model’s ability to find positives without producing too many false alarms.

[[IMG:pr]]

### Cumulative Gain / Lift

Compares model targeting to random selection. A model with good lift quickly captures a high proportion of true positives in the top deciles, showing value for prioritization strategies.

[[IMG:lift]]

### Calibration (Reliability) Curve

Plots predicted probability vs observed probability. A perfectly calibrated model will follow the diagonal; deviations indicate over- or under-confidence in predicted probabilities.

[[IMG:calibration]]

### Confusion Matrix (image)

Visual heatmap of predicted vs actual classes. Helps identify where the model is making the most errors.

[[IMG:confusion]]

### Confusion Matrix (Classification)

### Decile Lift Table

### KS Curve (Kolmogorov–Smirnov)

Measures maximum separation between the cumulative distribution of positives and negatives. Higher KS indicates better class separation.

[[IMG:ks]]

## Data Quality (Cleaned Data)

Missing (Train): {{DataQualityCheck.train\_avg\_missing}}

Missing (Test): {{DataQualityCheck.test\_avg\_missing}}

Columns With Missing (Train): {{DataQualityCheck.train\_cols\_missing}}

Columns With Missing (Test): {{DataQualityCheck.test\_cols\_missing}}

Constant Columns: {{DataQualityCheck.constant\_columns\_str}}

## Raw Data Validation

Total Rows: {{RawDataCheck.total\_rows}}

Total Columns: {{RawDataCheck.total\_columns}}

Average Missing: {{RawDataCheck.avg\_missing\_pct}}

Columns With Missing: {{RawDataCheck.columns\_with\_missing\_str}}

Duplicate Rows: {{RawDataCheck.duplicate\_rows}}

Constant Columns: {{RawDataCheck.constant\_columns\_str}}

## Correlation Analysis

### Top High-Correlation Feature Pairs (|r| ≥ Threshold)

**Note:** Displayed only if high-correlation pairs exist; otherwise, none are generated.

### Correlation Heatmap

[[IMG:correlation\_heatmap]]

Method: {{CorrelationCheck.method2}}, Threshold: {{CorrelationCheck.threshold2}},

Features plotted: {{CorrelationCheck.plotted\_features2}} of {{CorrelationCheck.n\_numeric\_features2}},

High-corr pairs ≥ threshold: {{CorrelationCheck.n\_pairs\_flagged\_ge\_threshold2}} (of {{CorrelationCheck.n\_pairs\_total2}})

Full CSV: {{CorrelationCheck.top\_pairs\_csv\_path}}

Pearson Correlation CSV: {{correlation\_pearson\_path}}

Spearman Correlation CSV: {{correlation\_spearman\_path}}

{{CorrelationCheck.notes\_text}}

## Variance Inflation Factor (VIF) Check

{{VIFCheck.note\_text}}

## Exploratory Data Analysis (EDA)

Summary Stats CSV: {{eda\_summary\_path}}

Missing Values CSV: {{eda\_missing\_path}}

**Distribution Plots**

{{eda\_count\_note}}

## Benchmark (Logistic Regression Baseline)

### Logistic Regression Results

{{LogitStats.summary\_text}}

### Logistic Regression Coefficients

### Baseline Metrics

\* Calculated on the same dataset that was used to train and evaluate the primary model.

## Stress Testing Results

{{StressTest.note\_text}}

## Input Cluster Coverage Check

A K-Means clustering algorithm was applied to the cleaned dataset to evaluate input space coverage. This analysis highlights the distribution of samples across different clusters, assisting in the identification of any imbalance or concentration that could affect model behavior.

Input cluster distribution (saved to CSV): {{InputClusterCheck.cluster\_csv}}

### Cluster Summary Table

{{InputClusterCheck.note\_text}}

### Cluster Distribution Plot

[[IMG:cluster\_plot]]

## Rule Engine Check

Overall Pass: {{RuleEngineCheck.overall\_pass\_str}}

Rule Details: {{RuleEngineCheck.rules\_text}}

## Model Contents Summary

### Model Class: {{ModelMetaCheck.model\_class}}

### Module: {{ModelMetaCheck.module}}

### Hyperparameters:

{{ModelMetaCheck.hyperparams\_text}}

### Attributes:

{{ModelMetaCheck.attributes\_text}}

## Explainability

### SHAP Summary Plot

SHAP (SHapley Additive exPlanations) is used to explain the impact of each input feature on the model's predictions. The SHAP summary plot below shows the magnitude and direction of influence of the top features across a sample of predictions. Features at the top have the highest average impact. Red indicates higher feature values, and blue indicates lower values. This plot helps identify which features the model relies on most and whether they align with domain knowledge. Only a subset of the test data is used for SHAP due to performance considerations.

{{SHAP.note\_text}}

[[IMG:shap\_beeswarm]]

### Top Features by Mean |SHAP| (Bar)

[[IMG:shap\_bar]]

### Top SHAP Features

{{SHAP.top\_features\_note}}