TELCO INNOVATIONS

Telco Customer Churn Prediction

Presented by HamidReza Eslami and Ali Kashipazha

The Churn Challenge

Understanding Revenue Risks

- **Definition**: Customer churn represents the phenomenon where customers discontinue a service.
- **Business Impact**: In the highly competitive telecommunications market, churn directly threatens revenue and long-term sustainability.
- **The Goal**: The project aim is to design an intelligent system to identify at-risk customers before they leave, enabling proactive retention.
- **Economic Strategy**: Retaining an existing customer is significantly more cost-effective than the high cost of acquiring a new one.

Dataset

Data Characteristics and Structure

- **Data Composition**: The Telco Customer Churn dataset includes a mix of numerical and categorical variables describing customer behavior.
- **Variable Types**: Initial exploration identified 33 attributes, including integers, floating-point values, and categorical strings.
- **Target Variable**: The core focus is the binary "Churn Value," indicating whether a customer has discontinued the service.

Dataset Summary:

- **Total Samples**: 7,043 customer records.
- **Original Features**: 33 attributes covering demographics, service usage, and billing.
- **Distribution**: 73.46% Non-Churn vs. 26.54% Churn.

Data Integrity Measures

Ensuring Quality and Accuracy

- **Leakage Prevention:** Removed variables containing post-hoc information—such as Churn Label, Churn Score, CLTV, and Churn Reason—to prevent the model from using target-derived data.
- **Noise Reduction**: Eliminated non-informative features like CustomerID, City, and Zip Code that encourage memorization over generalization.
- **Feature Cleaning**: Converted Total Charges to a numerical type and logically filled missing values for new customers with zero tenure.
- **Final Goal**: Ensured the pipeline extracts true behavioral signals rather than accidental correlations.

Feature Engineering Techniques

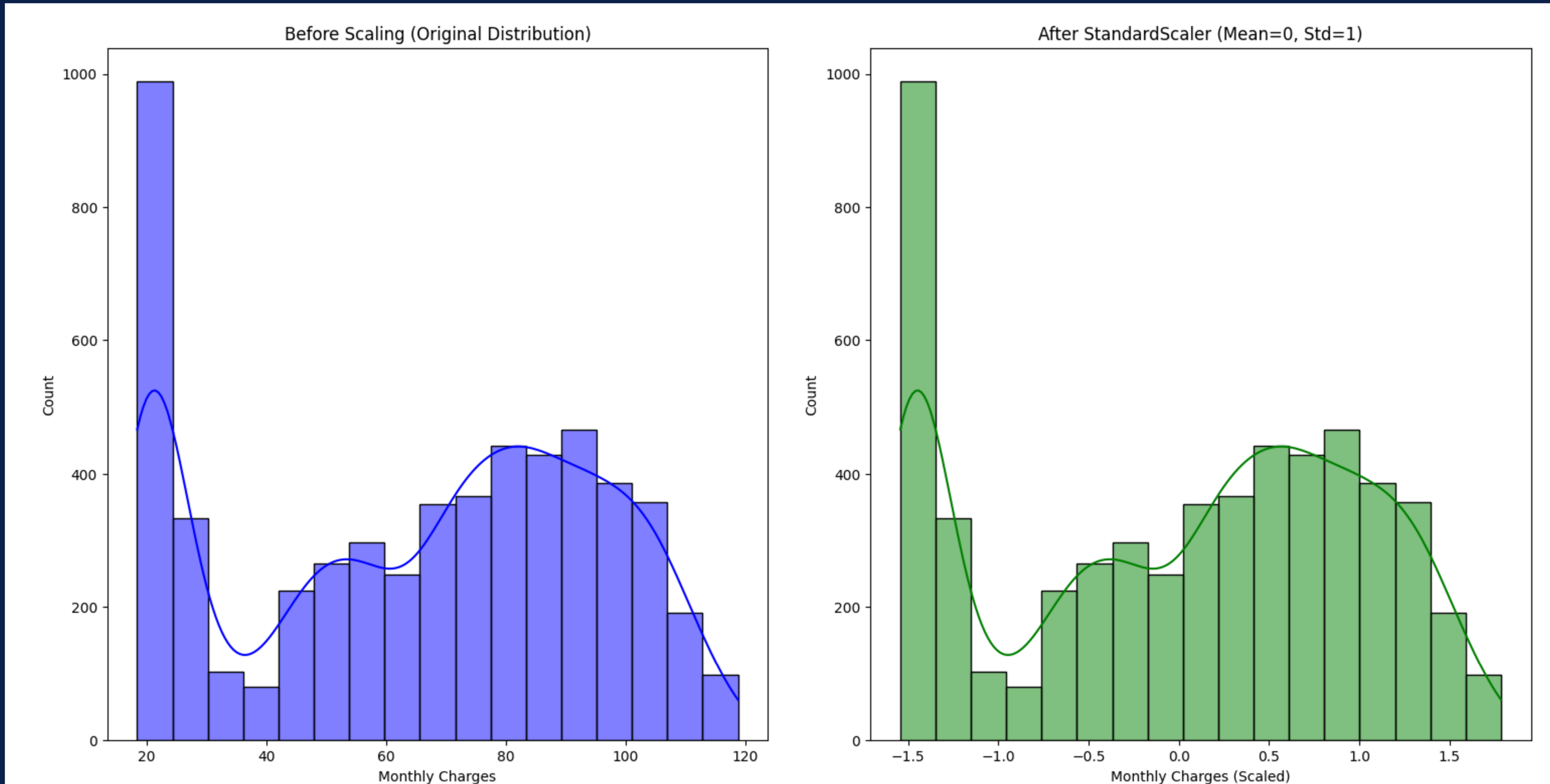
Crafting Key Predictive Features

- Average Monthly Charges: Computed to establish a long-term historical baseline of customer spending habits.
- Charge Difference Ratio: Defined to identify recent cost increases.
- Dissatisfaction Indicators: A positive ratio signals that a customer’s costs have risen relative to their historical norm, serving as a strong driver of churn.
- Value Addition: These engineered features transform static billing data into dynamic indicators of customer behavior over time.

Data Scaling

Normalization Techniques

Z-score normalization significantly impacts model performance by standardizing feature distributions, enhancing the Gradient Boosting model's ability to predict customer churn effectively.



Model Selection

Comparing Approaches for Predicting Customer Churn

Random Forest

Random Forest leverages bagging techniques to reduce variance, making it robust against overfitting while maintaining high predictive accuracy.
HyperParams :

- MaxDepth = 10

Gradient Boosting

Gradient Boosting employs boosting methods to minimize bias, enhancing predictive performance by sequentially correcting errors from previous models.
HyperParams :

- MaxDepth = 3
- Learning rate = 0.1

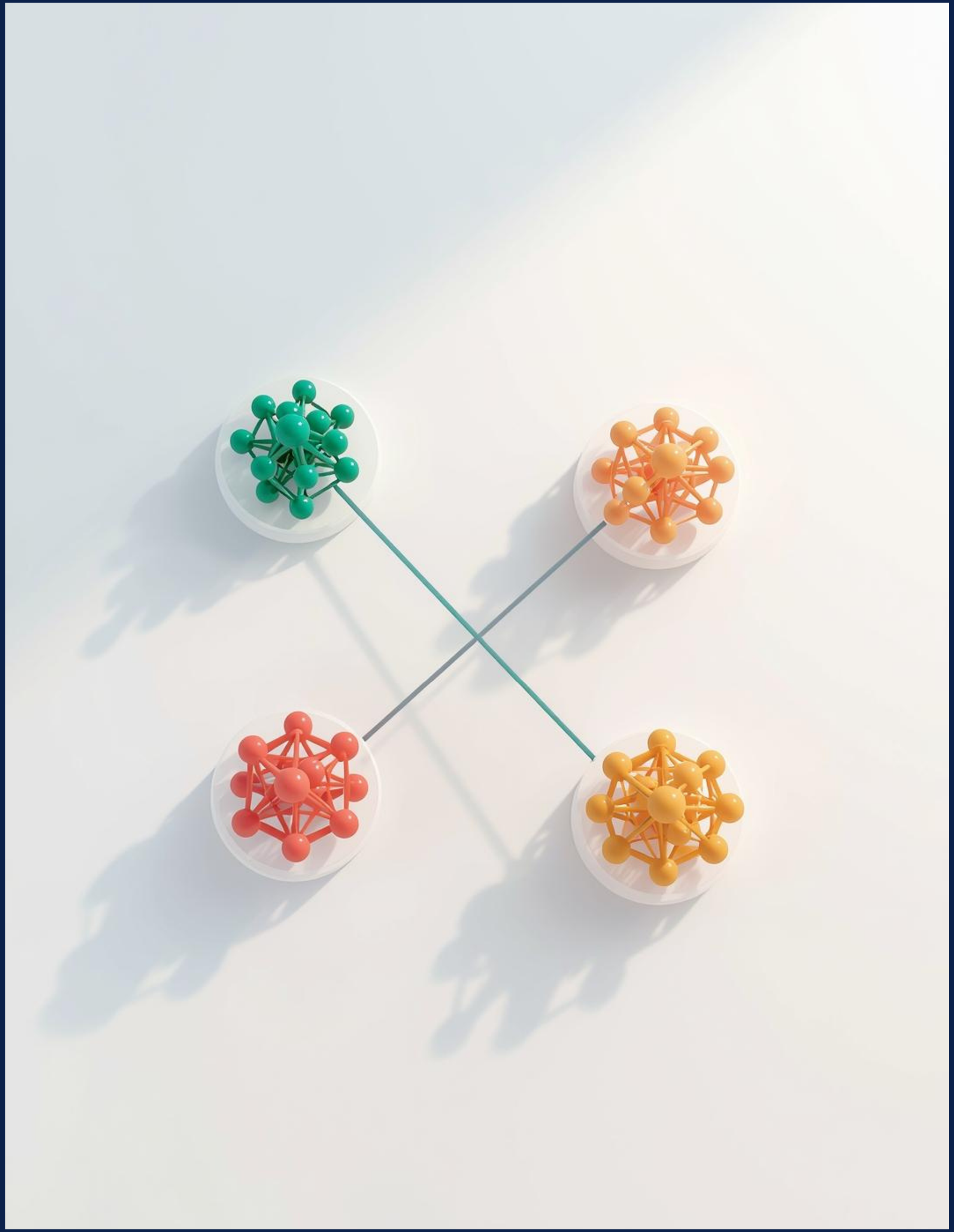
Strengths Comparison

Both models offer unique strengths; Random Forest excels in handling noise, while Gradient Boosting is more sensitive to underlying patterns.

Smart Penalty Ensemble

Hybrid Model Overview

- **Beyond Standard Voting**: Instead of relying on simple majority rules, a custom hybrid ensemble was engineered to integrate model-specific strengths.
- **Dual-Role Logic:**
  - Random Forest as the "Hunter": Acts as a high-sensitivity component designed to catch as many potential churners as possible.
  - Gradient Boosting as the "Precision Filter": Serves as a secondary check to validate the "Hunter's" findings.
- **Dynamic Penalty Mechanism:** If Random Forest predicts churn but Gradient Boosting assigns a very low probability, a penalty is applied to suppress the false alarm.
- **Operational Goal:** This architecture allows the system to retain the high recall required for business safety while effectively controlling the false-positive rate.



Risk-Aware Evaluation

0.3 rule

- Sub-optimal defaults in churn prediction can lead to ineffective outcomes due to equal error treatment.
- In telecommunications, losing a customer is costlier than providing a minor retention incentive.
- Lowering the classification threshold enhances model sensitivity, helping marketing identify most at-risk individuals.

Lowering the threshold enhances recall, vital for minimizing churn losses.

Random Forest Results

Performance Metrics Overview

Model Paradigm	Precision	Recall	F1-Score	Accuracy
Random Forest (Bagging)	0.47	0.89	0.61	0.71
Gradient Boosting (Boosting)	0.53	0.77	0.63	0.76

Table 2: Comparison of Base Models at Threshold 0.3

- **Key Metrics (at 0.3 Threshold):**
- Recall (Churn): 0.89 – Successfully identified 89% of all actual churners.
- Precision (Churn): 0.47 – Reflects a “bold” prediction strategy where sensitivity is prioritized over error suppression.
- Overall Accuracy: 0.71.
- Technical Outcome: By leveraging parallel trees and variance reduction, the model captured a wide variety of churn signals but produced 372 False Positives.

Business Value: Provides the necessary coverage for an aggressive early-warning system, ensuring very few at-risk customers are missed.

Results of Gradient Boosting

Performance Metrics Overview

Model Paradigm	Precision	Recall	F1-Score	Accuracy
Random Forest (Bagging)	0.47	0.89	0.61	0.71
Gradient Boosting (Boosting)	0.53	0.77	0.63	0.76

Table 2: Comparison of Base Models at Threshold 0.3

Performance Metrics (0.3 Threshold):

- **Recall**: 0.77 – Maintained high sensitivity.
- **Precision**: 0.53 – Superior ability to minimize false alarms.
- **Accuracy**: 0.76 – Highest individual accuracy.
- **Technical Outcome**: Effectively reduced bias, making the model more selective about flagging at-risk customers.
- **Business Value**: Ensures retention resources are targeted toward customers with a high statistical probability of leaving.

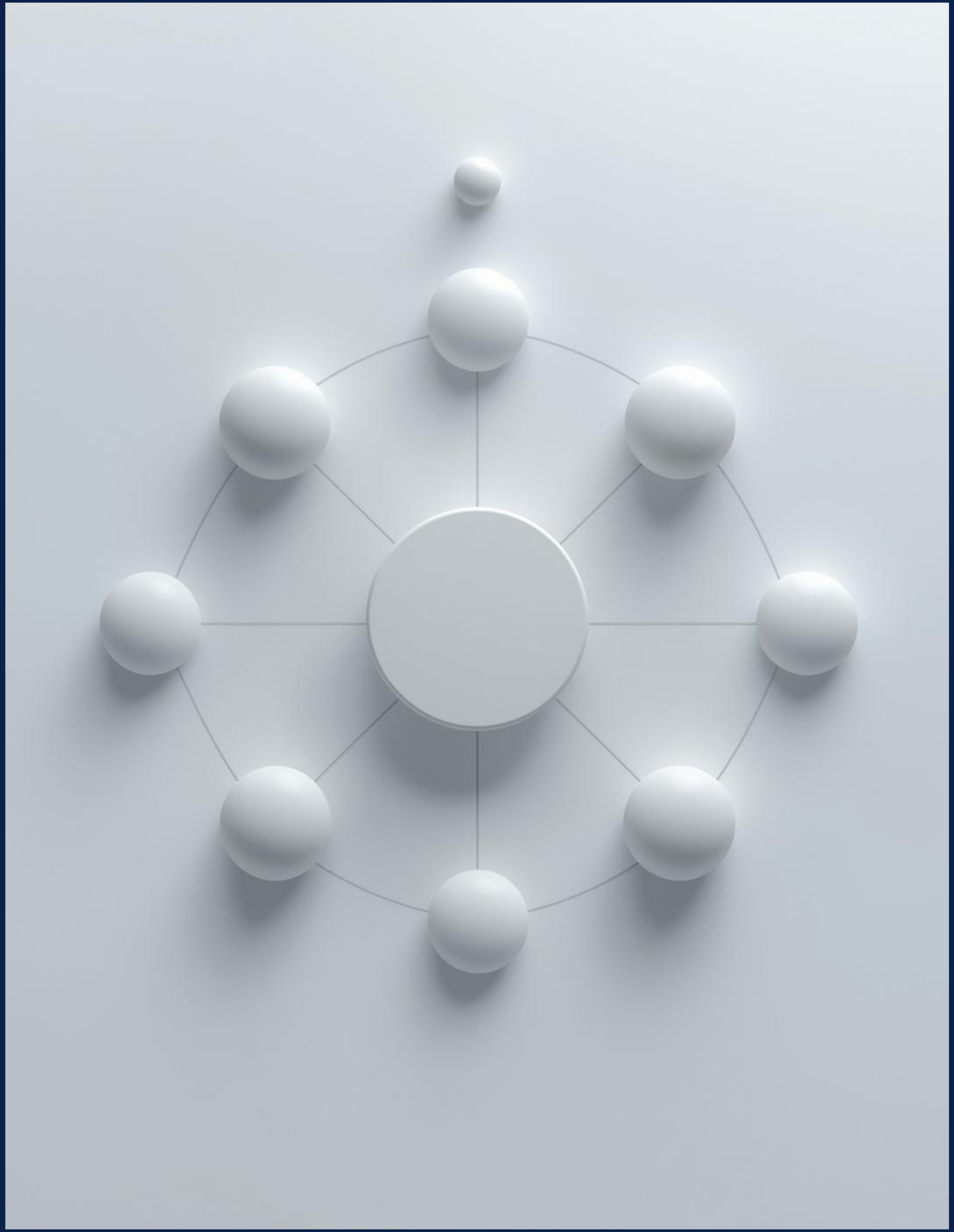
Results – Voting Ensemble

Combined Model Performance

The voting ensemble demonstrates improved **balance in metrics**, with a balanced recall of 0.84 and accuracy of 0.74, significantly enhancing prediction capabilities for customer churn.

Class	Precision	Recall	F1-Score	Support
Stayed (0)	0.92	0.70	0.80	1035
Churn (1)	0.51	0.84	0.63	374
Accuracy	—		**0.74**	1409

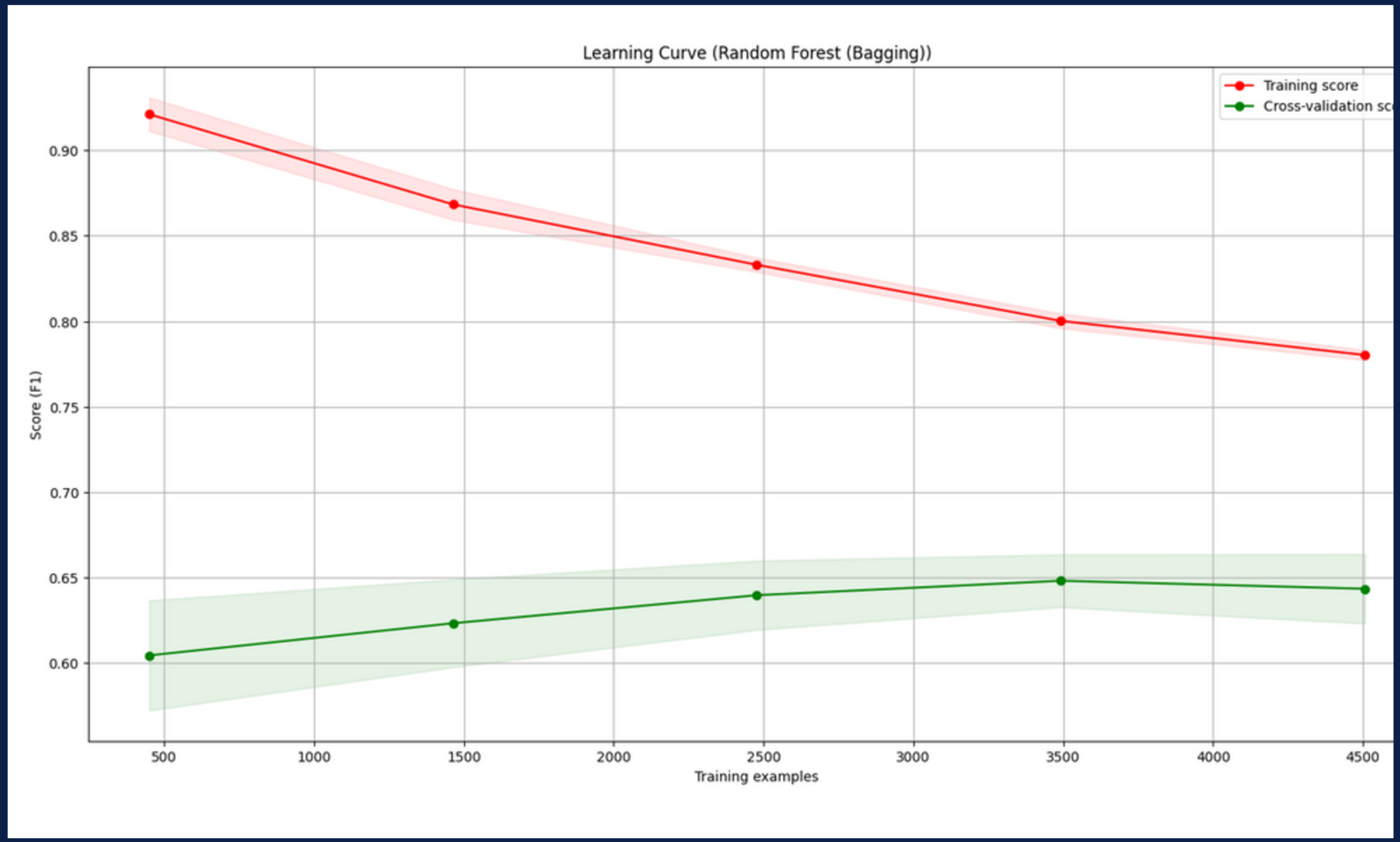
Table 3: Detailed Classification Report for the Final Ensemble

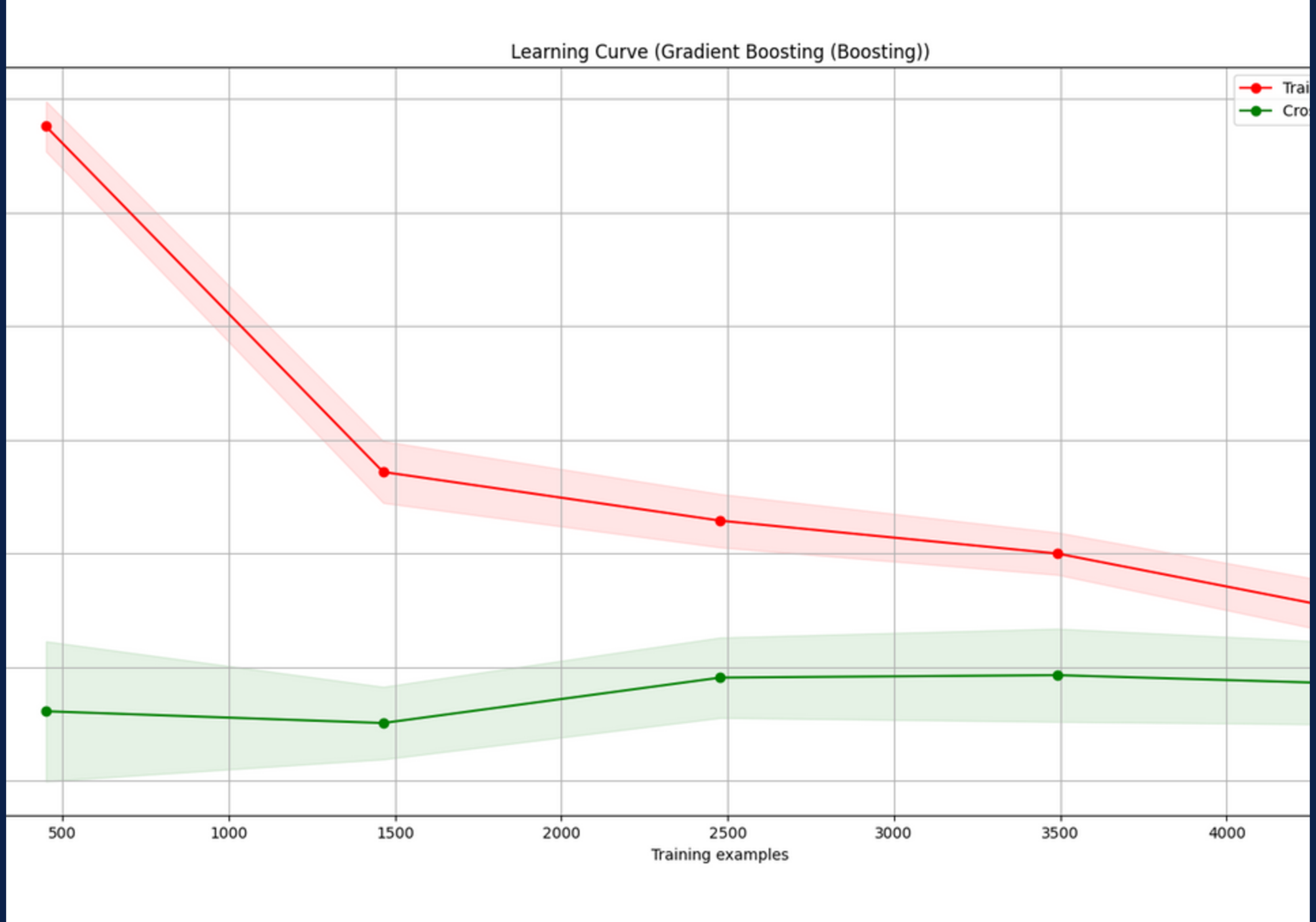


Model Stability Analysis

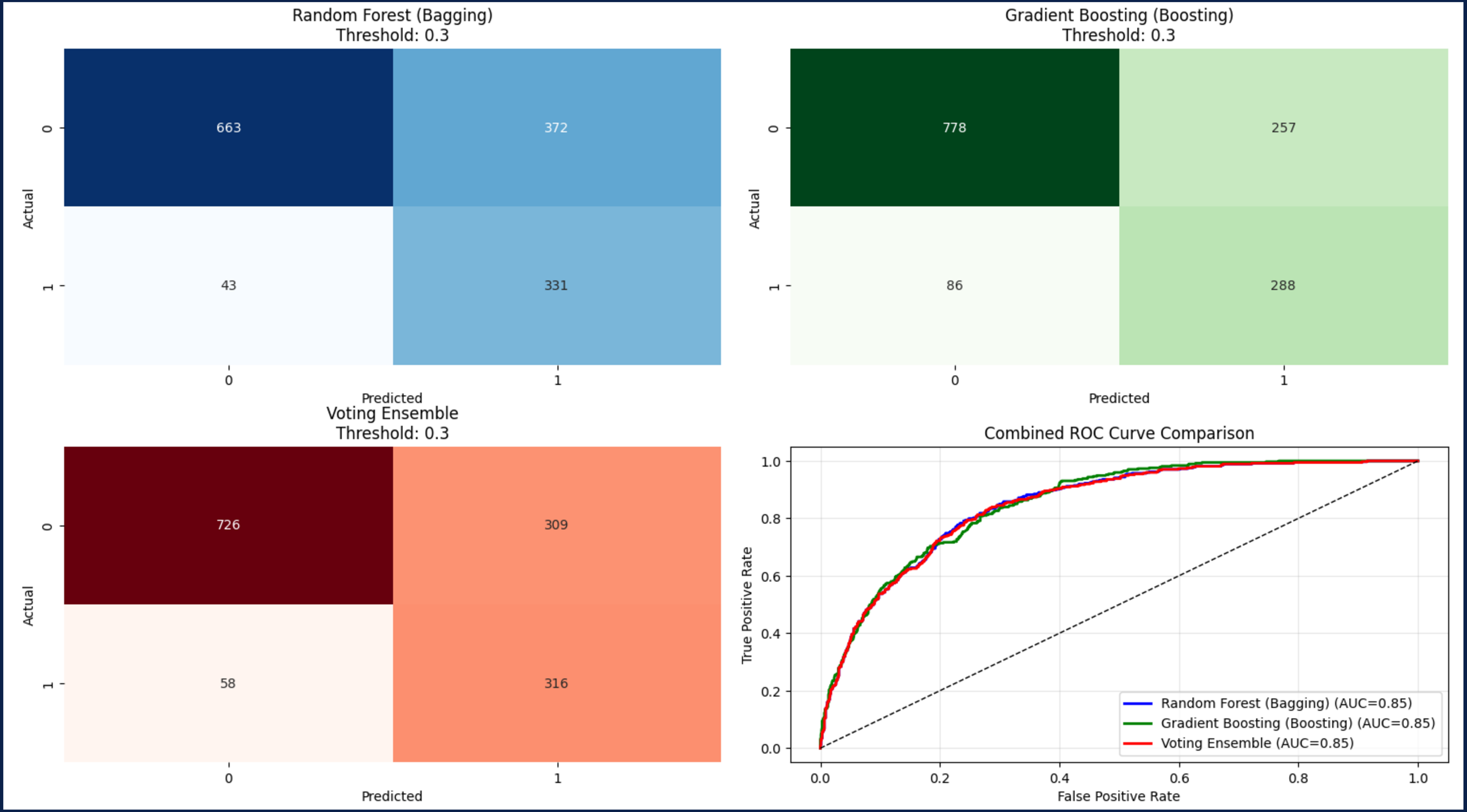
Evaluating Learning Curves

- **Convergence Success**: The learning curves for both models show training and validation scores converging as the number of training examples increases.
- **No Overfitting**: The small gap between training and validation performance confirms that the models effectively generalize to new data rather than memorizing the training set.
- **Data Utilization**: The plateau in the curves indicates that the models have successfully utilized the available dataset to reach their optimal predictive capacity.
- **Reliability:** This stability ensures that the "Hunter" and "Filter" logic is based on robust behavioral patterns, like spending spikes, rather than statistical noise.





Confusion Matrixes and Roc



Any questions for us?

Thank you for your attention and engagement today