Boost Trials

1. Ss

Validation Metrics: AUC=0.9637, Accuracy=0.9856, Precision=0.9599, Recall=0.6152, F1=0.7499, Confusion Matrix=[[91096, 85], [1272, 2034]]

Test Metrics: AUC=0.9654, Accuracy=0.9855, Precision=0.9532, Recall=0.6160, F1=0.7484, Confusion Matrix=[[113850, 125], [1587, 2546]]

Random Forest Trials

**Version 1 of Feature Engineering:**

**Validation Metrics:**

* **AUC**: 0.9339 (Strong model discrimination)
* **Accuracy**: 97.95% (High overall correctness)
* **Precision**: 94.49% (Good at identifying class 1 without many false positives)
* **Recall**: 44.10% (Moderate detection of class 1)
* **F1 Score**: 60.14% (Balance between precision and recall)
* **Log Loss**: 0.0829 (Well-calibrated probabilities)
* **Brier Score**: 0.0171 (Reliable predictions)
* **Average Precision**: 71.56% (Good for imbalanced classes)

**Validation Confusion Matrix:**

* **True Negatives**: 91,096
* **False Positives**: 85
* **False Negatives**: 1,848
* **True Positives**: 1,458

**Test Metrics:**

* **AUC**: 0.9220 (Strong model discrimination)
* **Accuracy**: 97.89% (High overall correctness)
* **Precision**: 94.13% (Good at identifying class 1 without many false positives)
* **Recall**: 42.32% (Moderate detection of class 1)
* **F1 Score**: 58.39% (Balance between precision and recall)
* **Log Loss**: 0.0978 (Well-calibrated probabilities)
* **Brier Score**: 0.0189 (Reliable predictions)
* **Average Precision**: 70.76% (Good for imbalanced classes)

**Test Confusion Matrix:**

* **True Negatives**: 113,866
* **False Positives**: 109
* **False Negatives**: 2,384
* **True Positives**: 1,749

**Comparison with Version 2 of Feature Engineering (Provided):**

* **AUC**: Version 1 (0.9339) performs slightly better than Version 2 (0.9324).
* **Accuracy**: Version 1 is slightly higher at 97.95% compared to Version 2's 97.92%.
* **Precision**: Version 1 leads slightly with 94.49% compared to Version 2's 94.03%.
* **Recall**: Both versions have a similar recall, but Version 1 (44.10%) is slightly better than Version 2 (43.41%).
* **F1 Score**: Version 1 again has a slight edge (60.14% vs. 59.39%).
* **Log Loss**: Version 1 has a lower log loss (0.0829 vs. 0.0961), indicating better-calibrated probabilities.
* **Brier Score**: Both versions are similar, with Version 1 slightly better at 0.0171 vs. 0.0172.
* **Average Precision**: Both versions are close, with Version 1 at 71.56% and Version 2 at 71.15%.

A graph with a line

AI-generated content may be incorrect.

XGBoost Trials

**✅ 1) Feature Engineering - Small**

* **AUC:** 0.9689 **Accuracy:** 98.51% **Precision:** 96.18%
* **Recall:** 59.76% **F1 Score:** 0.7372 **Log Loss:** 0.0519
* **Brier Score:** 0.0123 **Avg. Precision:** 0.8321
* **Fraud Detection:** 2470 TP, 1663 FN

**✅ 2) Feature Engineering - Full**

* **AUC:** 0.9664 **Accuracy:** 98.35% **Precision:** 96.95%
* **Recall:** 54.61% **F1 Score:** 0.6987 **Log Loss:** 0.0563
* **Brier Score:** 0.0135 **Avg. Precision:** 0.8289
* **Fraud Detection:** 2286 TP, 1847 FN

**🔍 Comparison**

| **Metric** | **Small FE** | **Full FE** | **Better** |
| --- | --- | --- | --- |
| AUC | 0.9689 | 0.9664 | Small |
| Accuracy | 98.51% | 98.35% | Small |
| Precision | 96.18% | 96.95% | Full |
| Recall | 59.76% | 54.61% | Small |
| F1 Score | 0.7372 | 0.6987 | Small |
| Log Loss | 0.0519 | 0.0563 | Small |
| Brier Score | 0.0123 | 0.0135 | Small |
| Avg. Precision | 0.8321 | 0.8289 | Small |