Evaluation Report for Customer Retention Prediction

1. **Performance metrics table:**

Model	Accuracy	Precision	Recall	F1-score
RandomForest	0.485	0.48	0.48	0.48
XGBClassifier	0.500	0.50	0.50	0.50
LGBMClassifier	0.540	0.54	0.53	0.53
LogisticRegression	0.495	0.48	0.48	0.47
TabNet	0.485	0.43	0.47	0.39
Simple ANN	0.525	0.28	0.53	0.36
TabTransformer	0.565	0.57	0.57	0.56

2. **Confusion Matrix:**

- **RandomForestClassifier:**

	Pred 0	Pred 1
Actual 0	41	54
Actual 1	49	56

- **XGBClassifier:**

	Pred 0	Pred 1
Actual 0	45	50
Actual 1	50	55

- **LGBMClassifier:**

	Pred 0	Pred 1
Actual 0	41	54
Actual 1	38	67

- **LogisticRegression:**

	Pred 0	Pred 1
Actual 0	27	68
Actual 1	33	72

Evaluation Report for Customer Retention Prediction

- **TabNet:**

	Pred 0	**Pred 1**
Actual 0	9	86
Actual 1	17	88

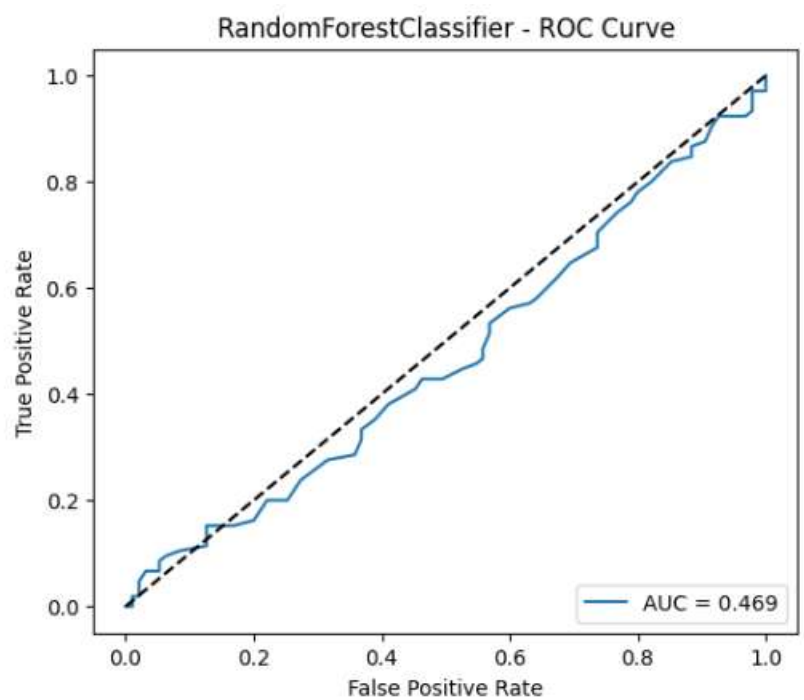
- **Simple ANN:**

	Pred 0	**Pred 1**
Actual 0	0	95
Actual 1	0	105

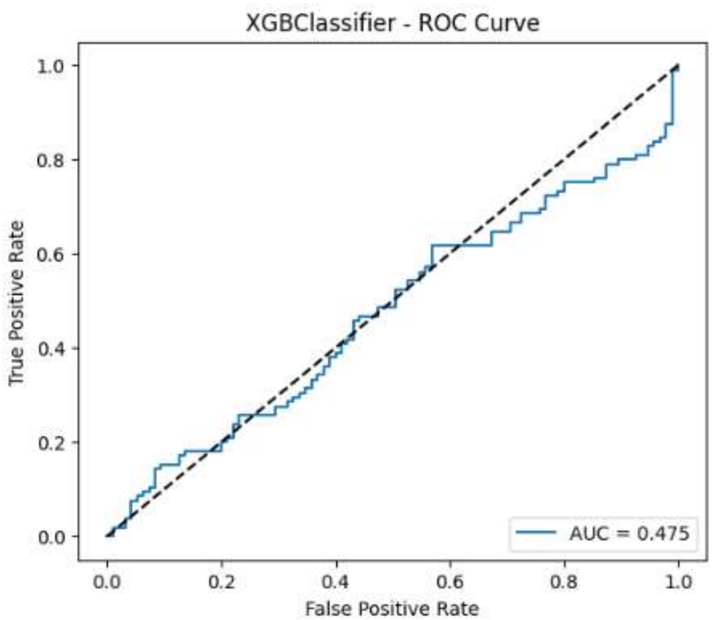
- **TabTransformer:**

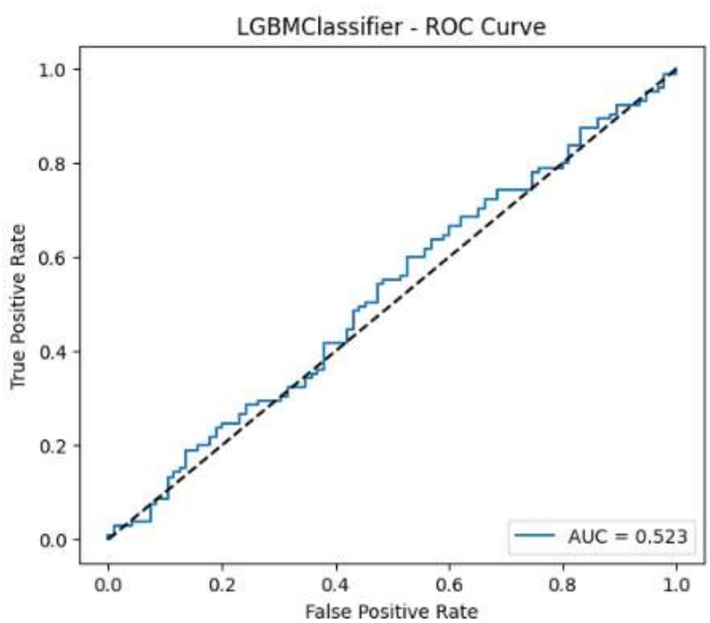
	Pred 0	**Pred 1**
Actual 0	56	38
Actual 1	49	57

3. ROC Curve:

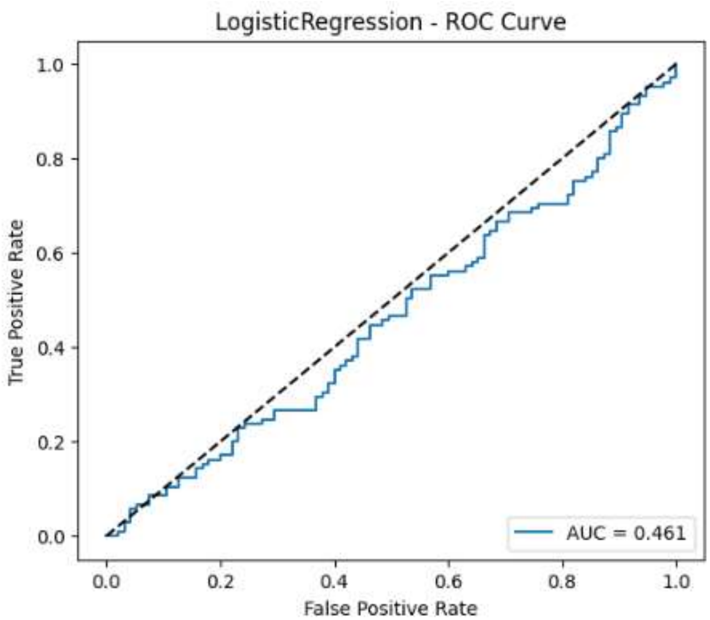


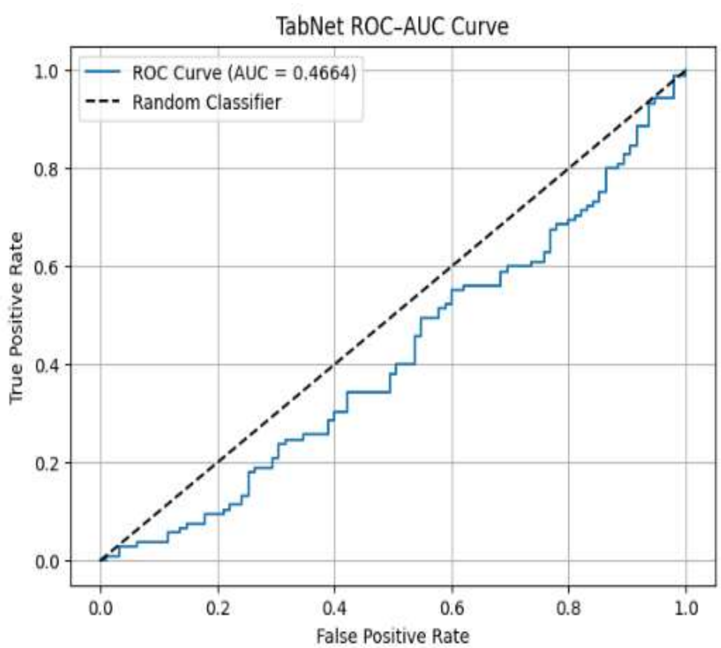
Evaluation Report for Customer Retention Prediction



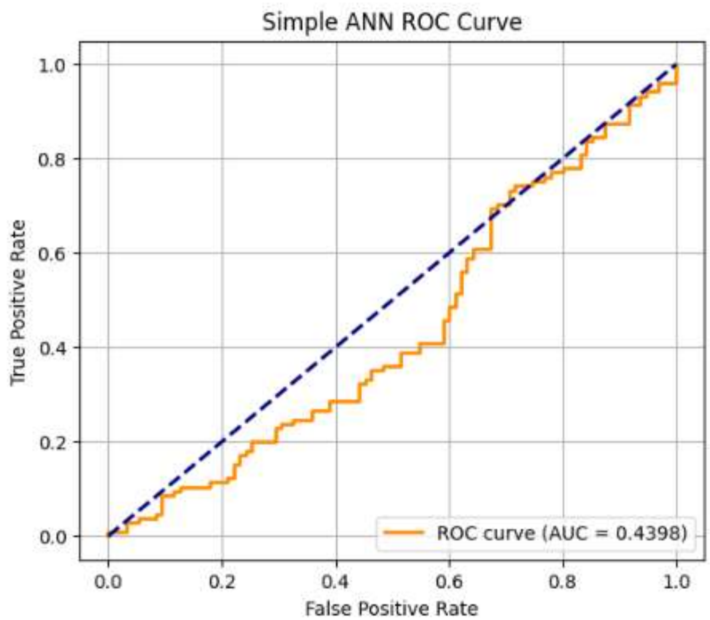


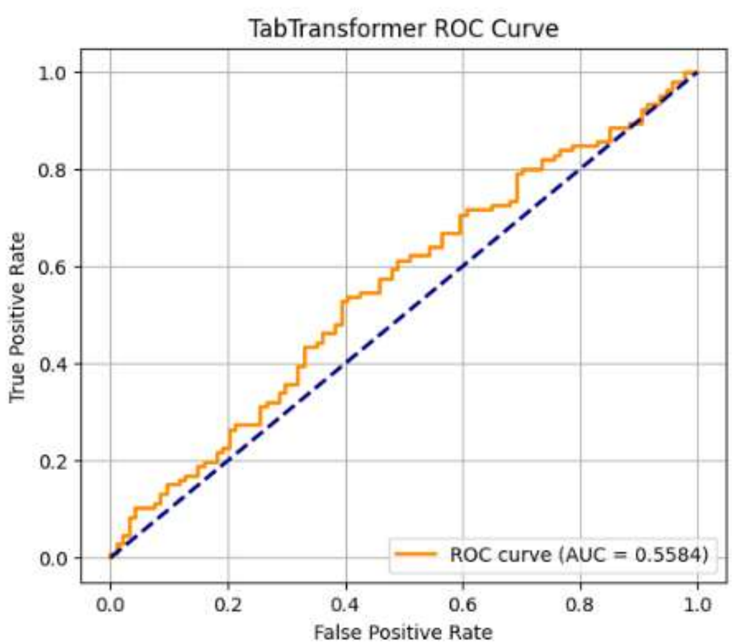
Evaluation Report for Customer Retention Prediction





Evaluation Report for Customer Retention Prediction





Evaluation Report for Customer Retention Prediction

4. **Summary of Key Findings**

1. Dataset Quality Confirmed

- No missing values were found in any of the features or target columns.
- No duplicated rows exist in the dataset.
- No significant outliers were detected in numerical features.
- Dataset is effectively clean, ensuring that poor model performance is not caused by data quality issues.

---

2. Balanced Dataset Confirmed

- Test set distribution: approximately 95 class 0 vs 105 class 1 across all models.
- Dataset is nearly perfectly balanced (~50/50 split).
- No class imbalance issues affecting model performance.

---

3. Poor Overall Model Performance

- All models demonstrate near-random predictive performance, with accuracy ranging from 48.5% to 56.5%.
- Random guessing in a balanced dataset would achieve ~50% accuracy.
- The best-performing model, TabTransformer, only provides a 6.5% improvement over random chance.
- ROC-AUC scores (0.46–0.56) indicate weak predictive power across all models.

---

4. Feature Predictive Power is Limited

- Weak correlations between features and churn suggest:
  - Individual features have low predictive power.
  - No clear linear or non-linear patterns exist for models to learn.
  - Churn behavior may be random or driven by unmeasured factors.

---

5. Model-Specific Observations

- Tree-based models: Random Forest, XGBoost, and LightGBM achieved similar performance (48.5%–54%), with LightGBM slightly better.
- Linear model: Logistic Regression (49.5%) shows limited capability, confirming non-linear relationships in the data.
- Neural networks: Mixed results:
  - TabTransformer performed best (56.5%).
  - Simple ANN completely failed (predicted all samples as class 1).
  - TabNet showed poor class balance handling.

---

6. Core Problem Lies in the Dataset

- Algorithm choice is not the primary issue.
- Available features (demographics, transaction history, support contacts) are insufficient to predict churn reliably.
- Key drivers of churn are likely missing, temporal, behavioral, or external factors not captured in this dataset.

---

7. Business Implications

- Current dataset provides limited churn prediction capability.
- Recommended actions:
- Feature Engineering: Create composite metrics, behavioral patterns, and time-series features.
- External Data Integration: Include customer feedback, product usage, and web analytics.
- Alternative Approaches: Focus on customer segmentation rather than direct churn prediction.
- Business Process Improvement: Collect more relevant data to capture actual churn drivers.

---

8. Root Cause Analysis

Churn may be:

- Essentially random within the measured feature space.
- Driven by unobserved or temporal factors**.**
- Influenced by data quality issues not visible in this analysis**,** such as inconsistent recording or measurement error.

Evaluation Report for Customer Retention Prediction

9. Conclusion:

- With the current dataset, reliable churn prediction is not feasible.
- Future improvements require better data collection, feature engineering, and external data integration to capture true churn drivers.
- Algorithm selection is secondary; the main limitation is the dataset itself.