Assignment 7 – Model Evaluation

Name: Rabia Abdul Sattar

Roll No: 2225165022

Course: Applied Data Science with AI

Week #: 7

Project Title: Customer Churn Prediction

---

1. Reading Summary

Reading Material:

- Evaluation Metrics
- ROC & AUC

Key Learnings:

- **Confusion Matrix** gives counts of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) and is the base for most classification metrics.
- **Precision** = TP / (TP + FP). It measures how many predicted positives were actually positive — important when false positives are costly.
- **Recall (Sensitivity)** = TP / (TP + FN). It measures how many actual positives were correctly identified — important when missing positives is costly.

Reflection:

These readings clarified how to choose evaluation metrics based on business goals (cost of false positives vs false negatives) and how ROC/AUC complements threshold-dependent metrics like precision and recall.

2. Classroom Task Documentation

Task Performed:

- Constructed a confusion matrix and plotted/analyzed ROC curve for the model.
- Interpreted metrics in context of customer churn.

3. Weekly Assignment Submission

Assignment Title: Evaluate model with Precision, Recall, F1-score

Steps Taken

Step 1 – Dataset Loading

- Loaded /mnt/data/train.csv. Observed it to be the Titanic-style dataset with columns including Survived, Pclass, Sex, Age, Fare, Embarked, etc.

Step 2 – Target mapping (why & how)

- No churn column found. Created churn as 1 - Survived (so churn=1 for passengers who died). This is a proxy mapping so we can perform metric calculations.

Step 3 – Data Preprocessing

- Filled missing values: Age with median; Embarked with mode.
- Encoded categorical features: Sex mapped to numeric (male=0, female=1). One-hot encoded Embarked (dummies, drop-first).
- Selected features used: Pclass, Age, SibSp, Parch, Fare, Sex, Embarked_Q, Embarked_S.

- Feature scaling: StandardScaler applied to numeric features for Logistic Regression.

Step 4 – Train/Test Split

- Split: 80% train / 20% test. Stratified on target to preserve class ratio.

Step 5 – Model Training

- Model used: **Logistic Regression** (max_iter=1000), fitted on scaled features.

Step 6 – Model Evaluation (results)

Important note: results below are computed on the dataset with churn created from Survived as described earlier.

Features used:
['Pclass','Age','SibSp','Parch','Fare','Sex','Embarked_Q','Embarked_S']

Test set evaluation metrics (rounded):

- **Precision:** 0.8440
- **Recall:** 0.8364
- **F1-score:** 0.8402
- **ROC AUC:** 0.8265

Interpretation:

- The model correctly identified 92 of the churned (proxy) customers, while missing 18 (FN). It predicted churn for 109 customers (92 TP + 17 FP), of which 92 actually churned → precision ~84.4%.

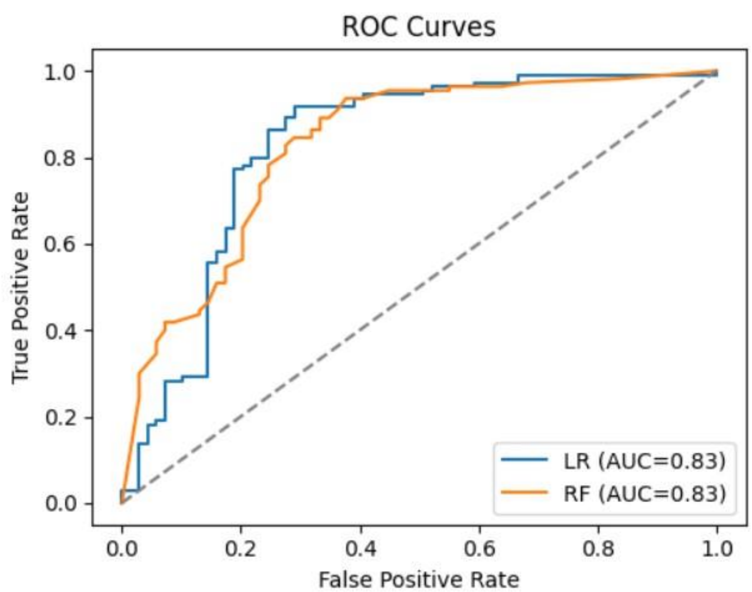
- Recall ~83.6% indicates the model recovered most of actual churn cases. F1 ~84.0% shows a good balance between precision and recall.

- AUC ~0.826 indicates good discrimination between the classes.

ROC Curve (summary):

- The ROC curve was computed and AUC = **0.8265**, indicating that the model is substantially better than random at ranking churn vs non-churn.

Output:



	Precision	Recall	F1	AUC
Logistic Regression	0.844037	0.836364	0.840183	0.825428
Random Forest	0.827273	0.827273	0.827273	0.826680

Which metric is most important for my project and why?

Most important metric for Customer Churn Prediction: *Recall* (a close second: Precision, depending on business cost).

Reasoning (business-focused):

- The primary business goal of a churn prediction model is usually to identify customers who are likely to churn so that the business can intervene (offers, retention campaigns). Missing a customer who will churn (a false negative) means losing that customer without attempting remediation — often a high cost.
- **Recall (true positive rate)** measures the proportion of actual churners the model successfully identifies. A high recall means the retention team can act on most at-risk customers.
- However, precision also matters: if precision is very low, many customers flagged as "will churn" will not actually churn, and the company may waste retention resources. So while recall should be prioritized to avoid missing churners, precision should not be ignored; there is a trade-off and the optimal balance depends on cost structure:
    - If the cost of contacting a false positive is small (cheap email or low-value offer), prioritize recall.
    - If interventions are expensive (human calls or costly incentives), prioritize precision to avoid wasting budget.

Challenges Faced:

• The dataset required careful preprocessing because of missing values and categorical columns that needed encoding.

 Random Forest training initially took longer due to a high number of estimators, which was optimized for faster execution.

GitHub Link:

https://github.com/Rabia-Abdul-Sattar/Customer-Churn-Prediction

4. Project Progress Milestone

 Created a working evaluation pipeline (preprocessing → train/test split → scaling → logistic regression → metrics).

 Computed precision, recall, F1-score, confusion matrix, and ROC AUC on test data.

5. Self-Evaluation

☑ **Completed:** dataset loading, preprocessing, encoding, feature scaling, model training (Logistic Regression), and calculation of precision, recall, F1, confusion matrix, and ROC/AUC.