

**Lahore College for Women University, Lahore**

**Department: Software Engineering**

**Submitted By: Iram Ahmad**

**Roll no: 2225165111**

**Submitted to: Sir Mohsin Nazir**

**Semester: VII**

**Session: 2022-2026**

**Assignment no 7**

**Course: Applied Data Science with AI**

**Semester:** BSSE 7th  
**Week #:** 7  
**Student Name:** Iram Ahmad  
**Roll Number:** 2225165111  
**Project Title:** House Price Prediction  
**GitHub Link:** https://github.com/Iram-Ahmad/Data-Science-AI-Course

**1. Reading Summary (½–1 page)**

**Reading Material for this Week:**

* *Evaluation Metrics* (Scikit-Learn documentation)
* *ROC & AUC Concepts* (Kaggle and Analytics Vidhya guides)

**Key Learnings:**

1. Learned how to measure model performance using **Precision, Recall, F1-score, and Accuracy**.
2. Understood the purpose of the **Confusion Matrix**, which shows correct and incorrect predictions.
3. Explored the concepts of **ROC curve (Receiver Operating Characteristic)** and **AUC (Area Under Curve)** for comparing model performance.

**Reflection:**  
These readings helped me understand that accuracy alone isn’t always the best metric, especially when the dataset is imbalanced. For my *House Price Prediction* classification task (High vs. Low price), **F1-score** is more useful as it balances both precision and recall — ensuring that the model performs well on both positive and negative predictions.

**2. Classroom Task Documentation**

**Task Performed in Class:**

* Computed and analyzed the **Confusion Matrix** for Random Forest and Logistic Regression models.
* Plotted **ROC Curves** and calculated **AUC scores** to evaluate which model performed better in classification.

**Screenshots / Code Snippets:**

from sklearn.metrics import confusion\_matrix, classification\_report, roc\_curve, auc

y\_pred = rf\_model.predict(X\_test)

print(confusion\_matrix(y\_test, y\_pred))

print(classification\_report(y\_test, y\_pred))

fpr, tpr, \_ = roc\_curve(y\_test, rf\_model.predict\_proba(X\_test)[:,1])

roc\_auc = auc(fpr, tpr)

print("AUC:", roc\_auc)

**3. Weekly Assignment Submission**

**Assignment Title:**  
*Model Evaluation using Precision, Recall, F1-score, and ROC-AUC*

**Steps Taken:**

1. Used predictions from Week 6 models (Logistic Regression and Random Forest).
2. Calculated **Precision, Recall, F1-score, and Accuracy** using scikit-learn metrics.
3. Visualized results using a **bar chart** for easy comparison.
4. Plotted **ROC Curve** and calculated **AUC value** for both models.
5. Wrote a reflection on the most suitable evaluation metric for the project.

**Output:**

* Logistic Regression F1-score ≈ 0.84
* Random Forest F1-score ≈ 0.88
* Random Forest AUC ≈ 0.91

**Challenges Faced:**

* Understanding the trade-off between Precision and Recall.
* Needed to adjust the threshold for ROC curve to improve sensitivity.

**GitHub Link:**  
https://github.com/Iram-Ahmad/Data-Science-AI-Course

**4. Project Progress Milestone**

**This Week’s Milestone:**

* Evaluated classification models using **Precision, Recall, F1, ROC, and AUC**.
* Chose **F1-score** as the final evaluation metric for the *House Price Prediction* project.

**Next Week’s Goal:**

* Perform **hyperparameter tuning** to improve model performance based on the chosen F1-score metric.

**5. ✅ Self-Evaluation**

☑️ I completed all tasks on time.  
⬜ I partially completed the tasks.  
⬜ I struggled with this week’s tasks and need help.

**6. Questions for Instructor (Optional)**

* How can we handle cases where increasing Recall significantly lowers Precision?
* Should we always use F1-score when the dataset is slightly imbalanced?