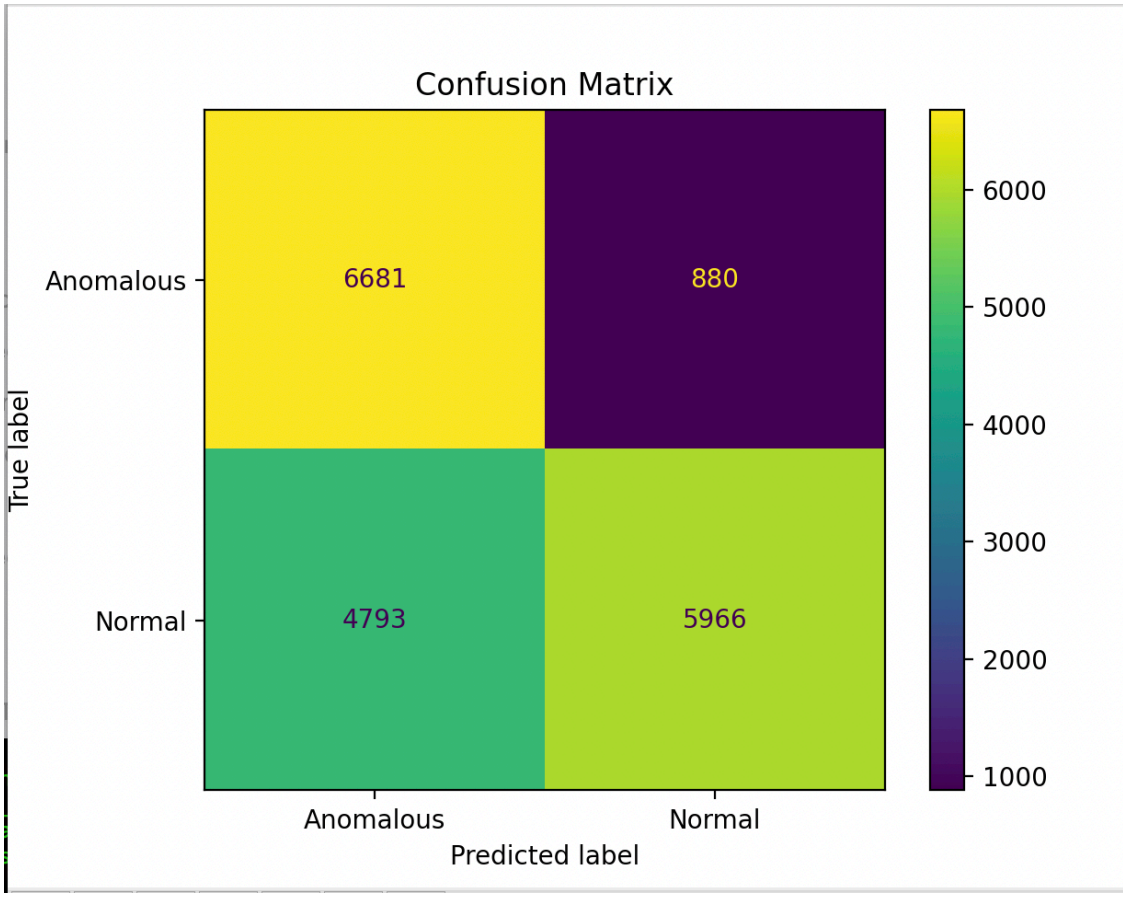
1. Performance Metrics

The following metrics were evaluated across three models: **Random Forest**, **Logistic Regression**, and **Gradient Boost**. Key metrics include accuracy, precision, recall, and F1-score.

Model	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
Random Forest	69%	0.73	0.72	0.69
Logistic Regression	65%	0.64	0.64	0.64
Gradient Boost	68%	0.72	0.71	0.68

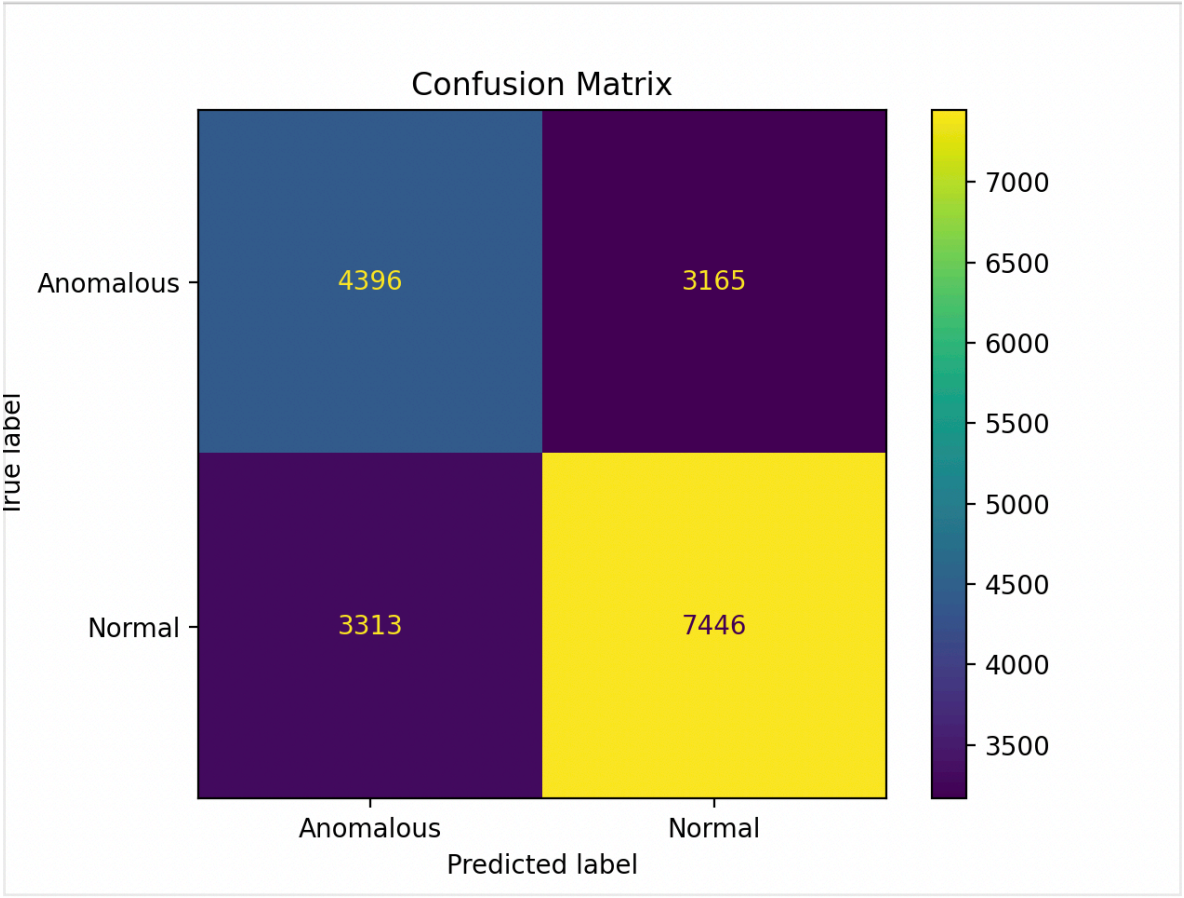
2. Confusion Matrix Analysis

- **Random Forest**

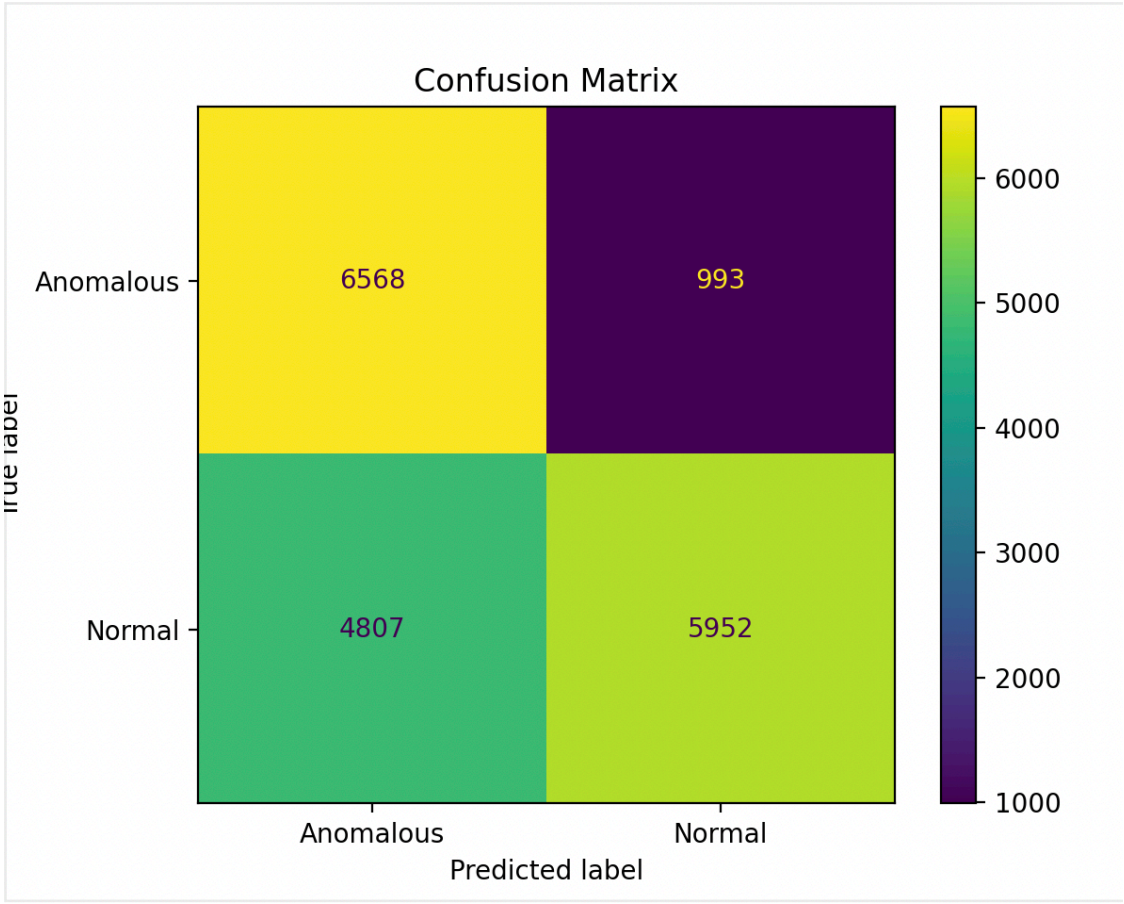


  - **True Positive (Normal):** 5966
  - **False Negative (Normal as Anomalous):** 4793
  - **True Negative (Anomalous):** 6681
  - **False Positive (Anomalous as Normal):** 880

- **Logistic Regression**



  - Achieved lower precision and recall compared to the other models.
  - **True Positive (Normal):** 7446
  - **False Negative (Normal as Anomalous):** 3313
  - **True Negative (Anomalous):** 4396
  - **False Positive (Anomalous as Normal):** 3165

- **Gradient Boost**
  - 

  - **True Positive (Normal):** 7446
  - **False Negative:** 3313

3. Comparison of Models

- **Best Performing Models: Random Forest and Logistic Regression**
  Both models achieved an **accuracy of 69%** and had balanced macro-averaged precision and recall scores.
  - **Random Forest:** Performed better at identifying "Anomalous" samples, evidenced by its recall score.
  - **Logistic Regression:** Balanced overall, with slightly better precision on normal samples.
- **Gradient Boost:**
  Gradient Boost underperformed with an accuracy of **65%**, showing lower recall and precision values.

4. Feature Importance

Random Forest Top Features:

1. **has_imagenes** - 0.2483
2. **Method** - 0.1434
3. **has_miembros** - 0.0780

4. **has_index_jsp** - 0.0590
5. **has_B1** - 0.0590

Gradient Boost Top Features:

1. **has_imagenes** - 0.3614
2. **has_index_jsp** - 0.0916
3. **Method** - 0.0863

5. Conclusion

Both **Random Forest** and **Logistic Regression** demonstrated similar performance with **68-69% accuracy**, balancing precision and recall effectively. Random Forest demonstrated an advantage by not falsely identifying normal samples as much as others .Gradient Boost fell short, achieving only **65% accuracy** and lower recall values. Feature importance analysis indicates that has_imagenes is consistently a influential feature across Models.