

Model Performance Comparison Report

1. Overview

This report presents a comparative evaluation of four machine learning models - **LightGBM**, **XGBoost**, **Random Forest**, and an **Ensemble Voting Classifier** - based on the analysis performed in the "Milling Machine 3.ipynb" notebook.

The evaluation focuses on the minority class (Class 1, representing "Machine Failure"), which is the critical target for predictive maintenance.

2. Summary of Key Metrics (Class 1 Focus)

The table below summarizes the performance of each model on the test set as recorded in the notebook.

Model	F1-Score (Class 1)	Precision (Class 1)	Recall (Class 1)	Accuracy
LightGBM (Optimized)	0.9242	0.95	0.90	0.99
Ensemble Voting	0.9200	0.95	0.90	0.99
XGBoost	0.9023	0.92	0.88	0.99
Random Forest	0.7581	0.84	0.69	0.98

3. Analysis of Results

- **LightGBM (Top Performer):** LightGBM achieved the highest F1-score (0.9242) for the failure class. It demonstrates an excellent balance, correctly identifying 90% of failures (Recall) while maintaining a high trust level in its alerts (95% Precision).
- **Ensemble Voting:** The Voting Classifier performed nearly identically to the best single model (LightGBM), achieving the same high precision (0.95) and recall (0.90). While it didn't significantly outperform LightGBM in this specific test run, it offers robustness by combining predictions from multiple algorithms.

- **XGBoost:** A strong runner-up with an F1-score of 0.9023. It is slightly less sensitive than LightGBM, missing a few more failures (Recall 0.88 vs 0.90), but remains a highly effective model.
- **Random Forest:** This model showed the weakest performance on the minority class ($F1=0.7581$). Its lower recall (0.69) means it missed approximately 31% of actual failures, making it less suitable for a critical safety system compared to the gradient boosting methods.

4. Visual Performance Summary

Metric	Best Model	Runner-Up	Least Effective
F1-Score (Class 1)	LightGBM (0.92)	Ensemble Voting (0.92)	Random Forest (0.76)
Recall (Class 1)	LightGBM / Voting (0.90)	XGBoost (0.88)	Random Forest (0.69)
Precision (Class 1)	LightGBM / Voting (0.95)	XGBoost (0.92)	Random Forest (0.84)

5. Conclusion

Based on the notebook results, **LightGBM** is the recommended model for deployment. It achieved the highest individual performance metrics for detecting machine failures. While the **Ensemble Voting Classifier** matched this performance, the added complexity of maintaining three models (RF, XGB, LGBM) inside an ensemble may not be justified when a single LightGBM model performs just as well.