XGBoost Model Comparison Using Python and R

Introduction

This report presents a comparative analysis of XGBoost models built using Python and R on datasets of different sizes. Python implementation uses scikit-learn with 5-fold cross-validation, while R implementation uses direct xgboost() and caret packages. The goal is to compare predictive performance and computational time for each method.

Methodology : Tools Used:

Python (scikit-learn), R (xgboost(), caret)

5-fold Cross-Validation

Metrics Used:

Accuracy

Time taken for model fitting (in seconds)

Python XGBoost Results

Dataset Size	Accuracy	Time (s)
100	0.9400	3.10
1,000	0.9520	0.42
10,000	0.9755	1.44
100,000	0.9868	4.46
1,000,000	0.9917	51.62
10,000,000	0.9931	427.75

R XGBoost Results

Method	Dataset Size	Accuracy	Time (s)
R xgboost() direct CV	100	0.8784	0.97
R xgboost() direct CV	1,000	0.9290	1.77
R xgboost() direct CV	10,000	0.9578	2.44
R xgboost() direct CV	100,000	0.9710	8.12
R xgboost() direct CV	1,000,000	0.9750	143.72

R xgboost() direct CV	10,000,000	0.9757	547.55
R caret xgboost() 5-fold CV	100	0.9204	1.88
R caret xgboost() 5-fold CV	1,000	0.9480	1.94
R caret xgboost() 5-fold CV	10,000	0.9736	4.29
R caret xgboost() 5-fold CV	100,000	0.9839	21.43
R caret xgboost() 5-fold CV	1,000,000	0.9885	219.30
R caret xgboost() 5-fold CV	10,000,000	0.9896	1348.75

Analysis & Recommendation

All implementations demonstrate accuracy improvements with larger datasets. However, Python's XGBoost offers better performance at scale, achieving high accuracy (0.9931) on large datasets with significantly reduced computation time compared to R caret (0.9896 in 1348 seconds). Therefore, Python XGBoost is recommended for large datasets. R caret, with structured CV, is ideal for smaller projects emphasizing interpretability.

Conclusion

XGBoost remains a reliable model across platforms. Python stands out for speed and scalability while R (especially caret) offers better interpretability and structured validation. Choose the method based on project scale and objectives.