Exercise 3C: Tree-Based Models

BUSI 722: Data-Driven Finance II

Using the same ranked features, ranked target, and train/test split from Exercise 3B, fit tree-based models and compare to the linear models from Exercise 3B.

Submission

Submit a **Jupyter notebook** (`.ipynb`) containing all code, output, and charts. Use markdown cells for any written discussion.

Using the same train/test split:

1. Fit a **Random Forest** (100 trees, max depth 4, `random_state=42`). Predict on the test set and compute the monthly Spearman rank correlation.
2. Fit a **LightGBM** model (100 trees, learning rate 0.05, max depth 6, `random_state=42`). Predict on the test set and compute the monthly Spearman rank correlation.
3. For LightGBM, display the **feature importances** (by split gain) as a bar chart. In a markdown cell, discuss which features matter most.
4. Compare mean Spearman correlations across all four models (Ridge, Lasso, Random Forest, LightGBM). In a markdown cell, state which model produces the best predictions and why.