

WEEKLY STOCK OUTPERFORMANCE SCREENING USING S&P 500 DATA

More Data from Yahoo Finance
News

BUSINESS PROBLEM & STAKEHOLDER

- Stakeholder: Retail trading platform
- Challenge: Identifying stocks likely to outperform the S&P 500 over the next week
- Key risk: Missing stocks that actually outperform
- Objective: Build a recall-oriented weekly screening model

PREDICTION TASK & TARGET DEFINITION

- **Modeling Objective**
- Binary classification task
- Target: Whether a stock outperforms the S&P 500 over the next 5 trading days
- Label construction:
 - Stock 5-day forward return vs index 5-day forward return

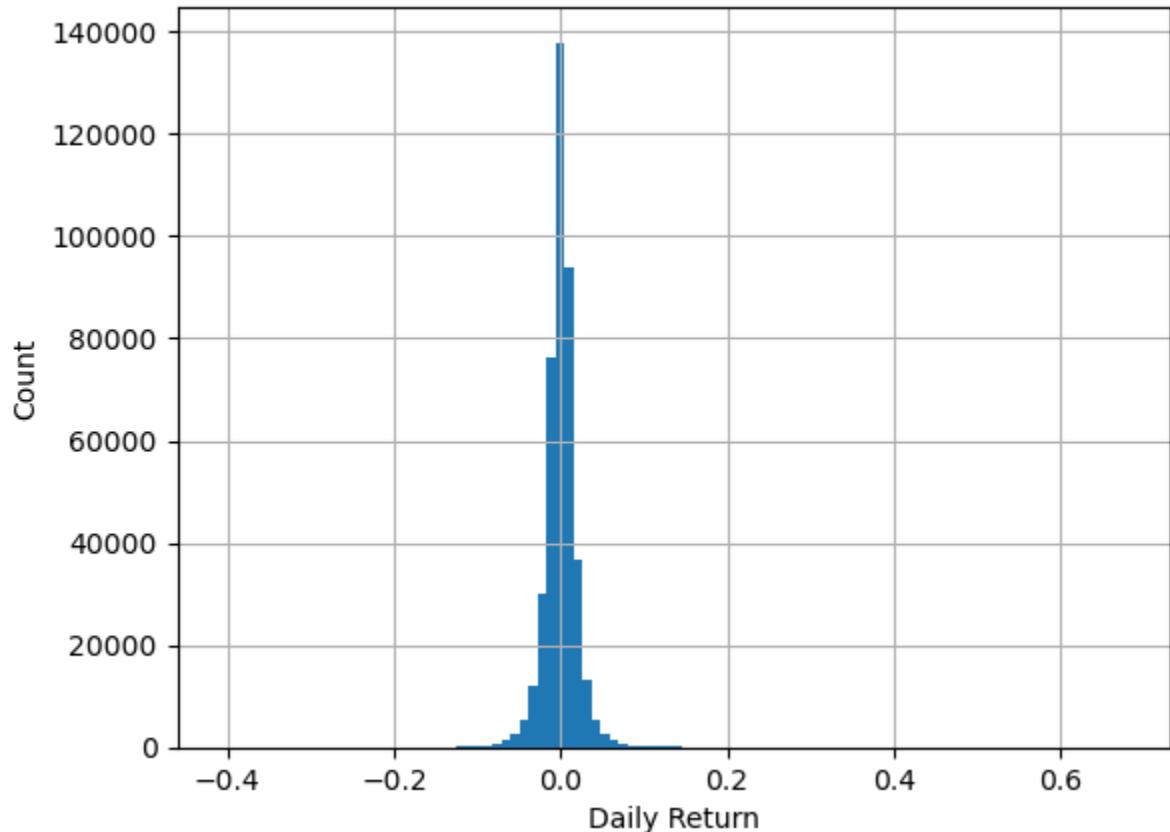
DATA OVERVIEW

- S&P 500 stock-level daily prices
- S&P 500 index prices
- Company metadata (sector)
- Coverage:
 - 172 stocks
 - ~420,000 observations
 - Weekly forward-looking labels

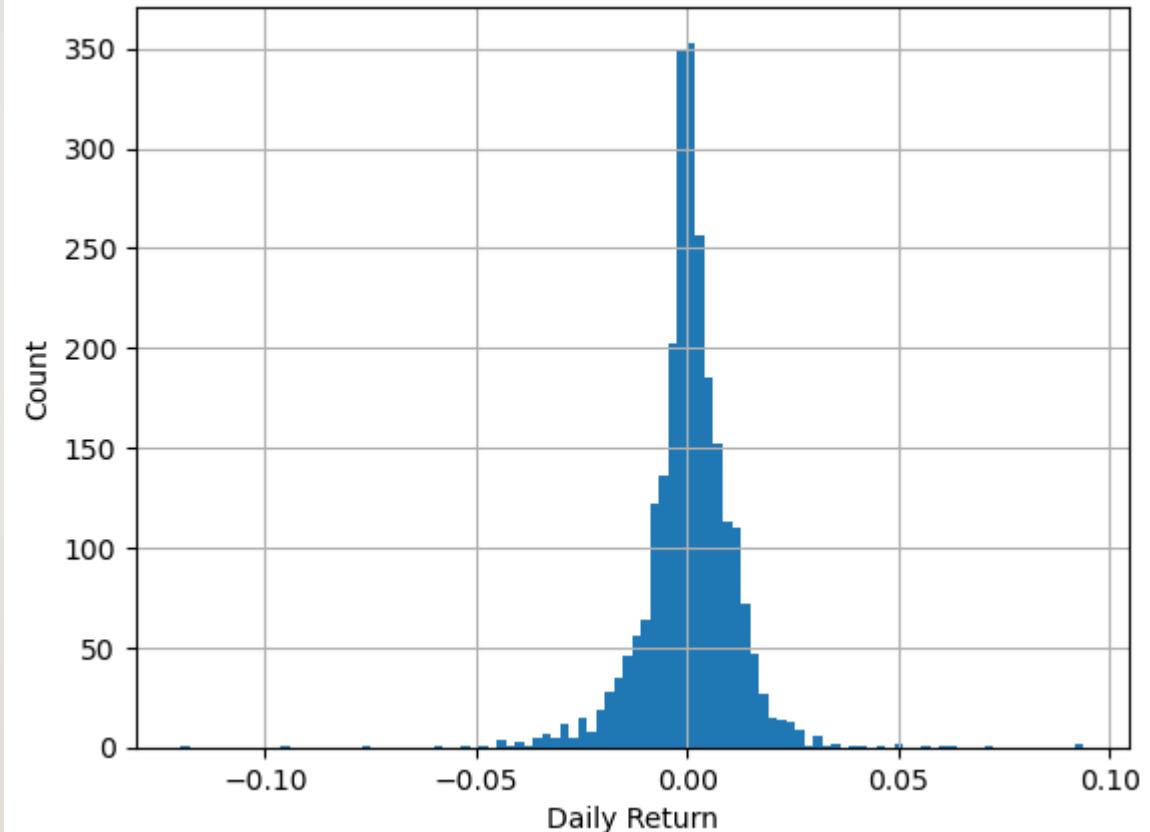
FEATURE ENGINEERING

- Price-based features:
 - Rolling returns (1d, 5d, 10d)
 - Rolling volatility
 - Moving averages
- Relative performance:
 - Excess returns vs index
- Volume features
- Sector encoding

Distribution of Daily Stock Returns (Adj Close)



Distribution of Daily S&P 500 Returns (S&P500)



EVALUATION STRATEGY

- Time-aware train/test split
- Primary metric: Recall (outperformers)
- Supporting metrics:
 - Precision
 - F1 score
 - ROC-AUC
- Why recall is prioritized for the business case

BASELINE MODEL: LOGISTIC REGRESSION

Why Logistic Regression

- Interpretable and stable
- Strong baseline for classification
- Suitable for noisy financial data

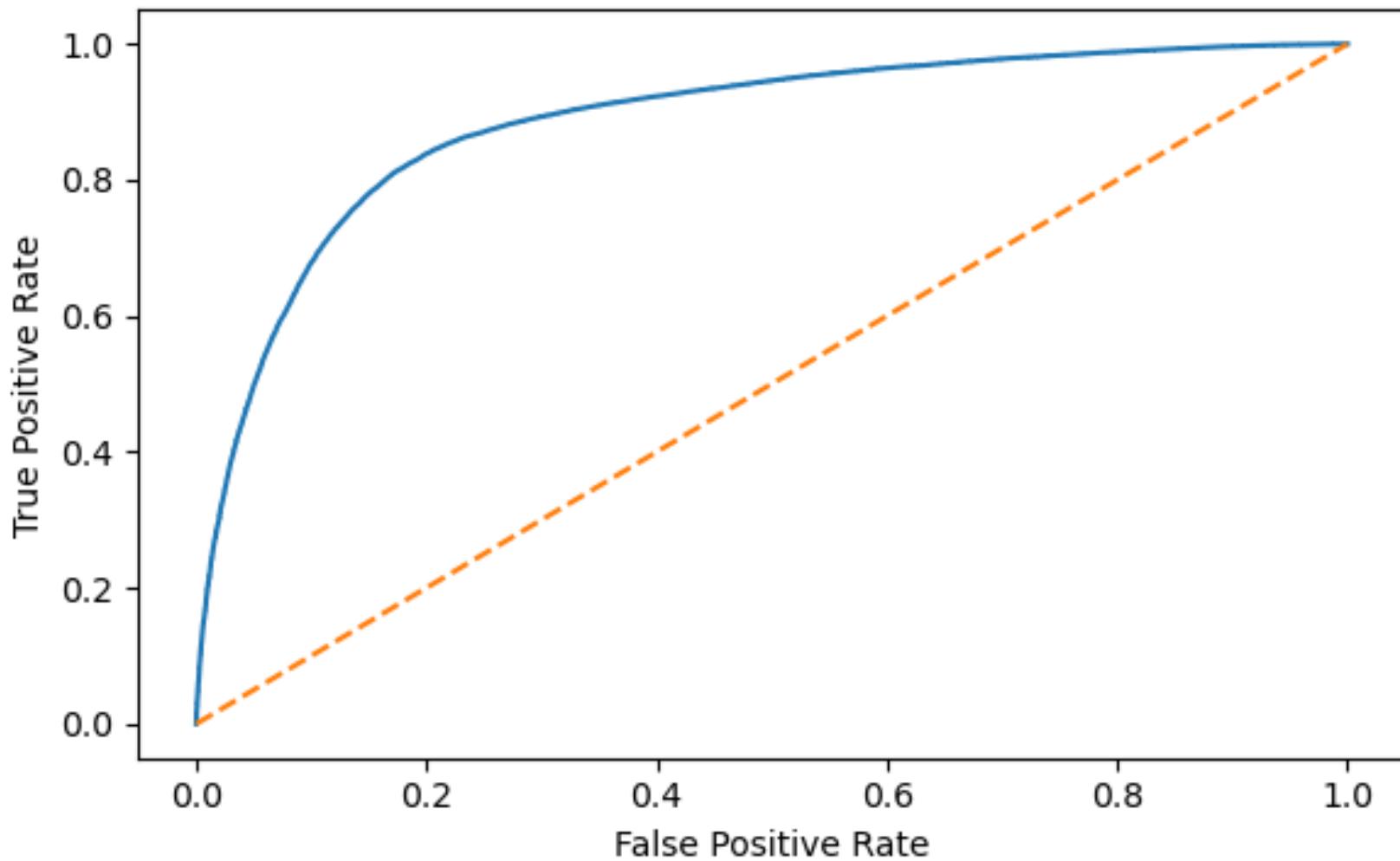
Key Results

- Recall (outperform): **87.0%**
- ROC-AUC: **0.8846**
- Accuracy: **80.8%**

NONLINEAR MODEL: DECISION TREE

- **Decision Tree Exploration**
- Default tree:
 - Overfits (100% train accuracy)
- Tuned tree:
 - High recall (98.6%)
 - Collapses in accuracy and ROC-AUC
- **Key Insight**
- Nonlinear trees do not generalize well for this problem

Logistic Regression ROC Curve



MODEL COMPARISON

| Model | Recall (I) | ROC-AUC | Accuracy |
|-------------------------|------------|---------|----------|
| Logistic Regression | 0.87 | 0.88 | 0.81 |
| Decision Tree (default) | 0.77 | 0.73 | 0.73 |
| Decision Tree (tuned) | 0.99 | 0.55 | 0.50 |

Final Choice: Logistic Regression

BUSINESS RECOMMENDATIONS

How the Platform Should Use the Model

- Deploy as a **weekly screener**, not a trading signal
- Rank stocks by predicted probability
- Surface top candidates (e.g., top 10–20%)
- Combine with human judgment or technical analysis

LIMITATIONS & CONCLUSION

Key Takeaways

- Model performs well but is not a guarantee
- Sensitive to market regimes
- Designed for decision support, not automation

Final Value

- Reduces missed short-term opportunities
- Transparent, interpretable, and deployable