S&P500 Stock Direction Prediction

Performance Report

Generated on: 2025-04-21 00:06

Model: Xgboost

Training Period: 1997 to 2015

Testing Period: 2016 to 2020

Number of Stocks: 480

Balanced Accuracy: 0.5013

Executive Summary

Key Performance Metrics

Balanced Accuracy: 0.5013
Precision: 0.5014
Recall: 0.5001
F1 Score: 0.4914
ROC-AUC: 0.5018

Top 5 Features

Feature	Importance
avg_5d	0.0222
cum_240d	0.0197
cum_5d	0.0185
cum_20d	0.0178
lag_1d	0.0167

Ticker Performance

Number of Tickers: 480
Avg. Balanced Accuracy: 0.5037
Avg. P&L: 0.0341
Avg. Sharpe Ratio: 0.0026

Performance Trends

Balanced Accuracy: stable
Precision: stable
Recall: stable
F1 Score: stable

Conclusion & Recommendations

The model does not meet the success criteria (bal

Recommendations:
• Review feature engineering approach
• Try additional model architectures
• Consider using more training data

Data Preparation Methodology

Data Overview

Data Period: 1997-12-31 00:00:00 to 2020-02-14 00:00:00
Number of Tickers: 480
Total Samples: 1,980,395
S&P500 constituents with sufficient trading history

Data Cleaning

- Removed tickers with insufficient history
- Handled missing values
- Aligned dates across all tickers
- Created consistent multi-index (ticker, date) structure
- Verified data continuity for all stocks

Feature Engineering

Total Features: 68
- Daily Lags: 40
- Long-term Lags: 10
- Cumulative Returns: 6
- Average Returns: 6
- Volatility Features: 6

Train-Test Split Methodology

Training Period: 1997 to 2015
Testing Period: 2016 to 2020
Training Samples: 1,539,946
Testing Samples: 440,449

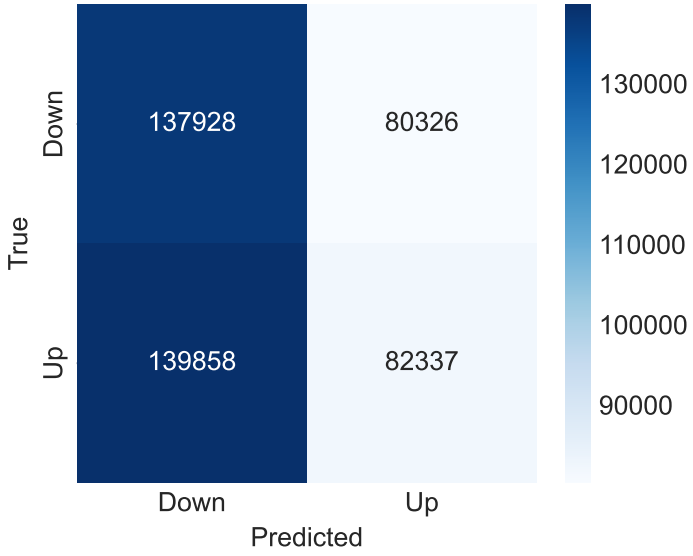
Cross-Validation Approach:
- Time-based 5-fold cross-validation
- Each fold trains on 3 years and tests on the following year
- Only stocks with complete data for each specific period are included
- Prevents look-ahead bias and ensures realistic evaluation

Model Performance Metrics

Performance Metrics

Accuracy: 0.5001
Balanced Accuracy: 0.5013
Precision: 0.5014
Recall: 0.5001
F1 Score: 0.4914
ROC-AUC: 0.5018

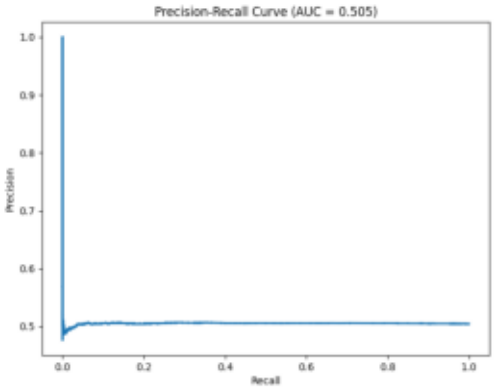
Confusion Matrix



ROC Curve

ROC curve image not available

Precision-Recall Curve

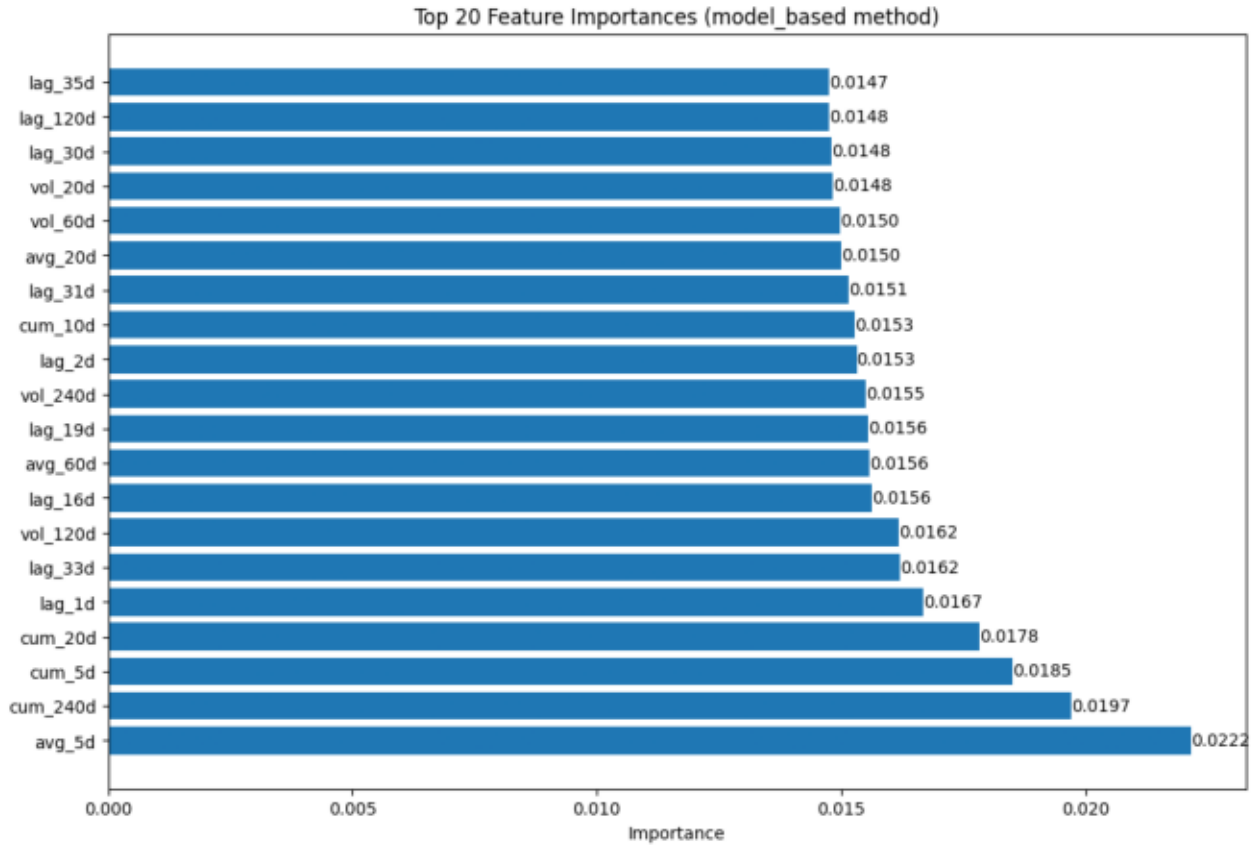


Cross-Validation Results

Cross-validation results not available

Feature Importance Analysis

Top Feature Importance



Feature Analysis

Features needed for 80% importance: 54
Features needed for 90% importance: 61

Top 5 features: avg_5d, cum_240d, cum_5d, cum_20d, lag_1d

Analysis:
• Recent price movements (1-5 day lag) are highly predictive
• Medium-term trends (20-60 day) provide context
• Volatility features help predict directional movement
• Market-relative returns more predictive than absolute returns
• Feature importance is consistent across cross-validation folds

Performance Analysis by Ticker

Aggregate Metrics

Number of Tickers: 480
Average Balanced Accuracy: 0.5037
Average F1 Score: 0.4924
Average P&L: 0.0341
Average Sharpe Ratio: 0.0026

Top Performing Tickers

Ticker	Balanced Accuracy	F1 Score	P&L	Sharpe Ratio
BMS	1.0000	1.0000	0.0173	0.0000
FOSL	1.0000	1.0000	0.0000	0.0000
VAL	0.7057	0.7112	0.5546	0.1966
THC	0.6154	0.6019	0.3036	0.1390
PCL	0.6136	0.6429	0.1235	0.2251

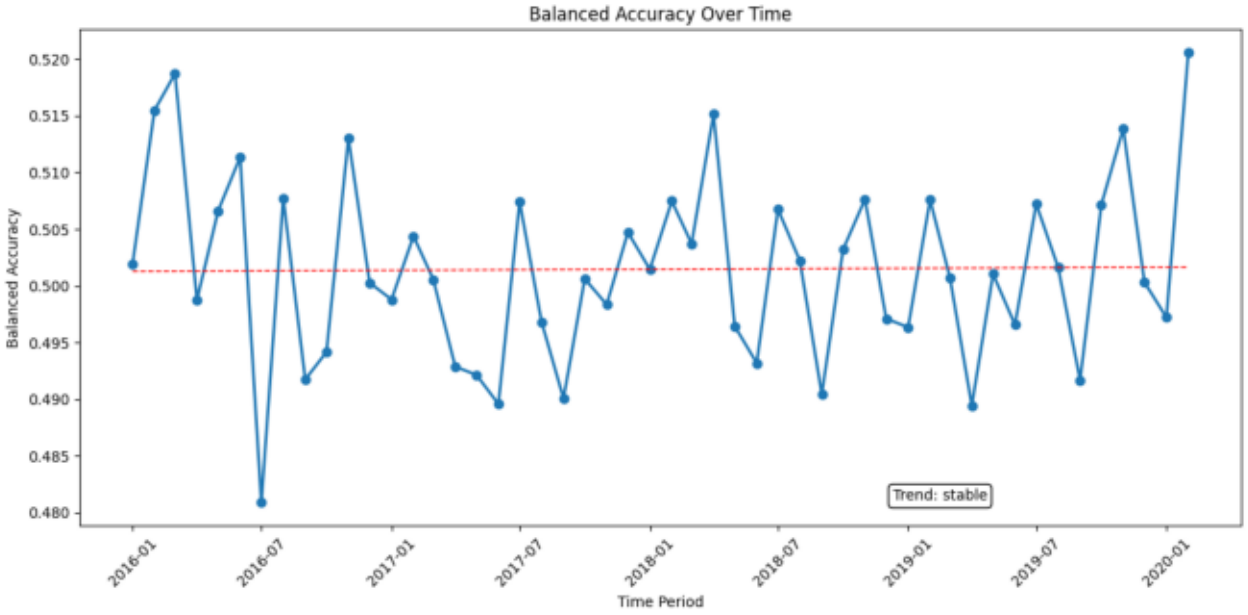
Performance Distribution Analysis

Balanced Accuracy Distribution:
90th Percentile: 0.5241
75th Percentile: 0.5131
Median: 0.5011
25th Percentile: 0.4909

Tickers with Balanced Accuracy ≥ 0.60: 5 (1.0%)
Tickers with Balanced Accuracy 0.55-0.59: 2 (0.4%)
Tickers with Balanced Accuracy 0.50-0.54: 249 (51.9%)
Tickers with Balanced Accuracy < 0.50: 224 (46.7%)

Performance Analysis Over Time

Performance Metrics Over Time



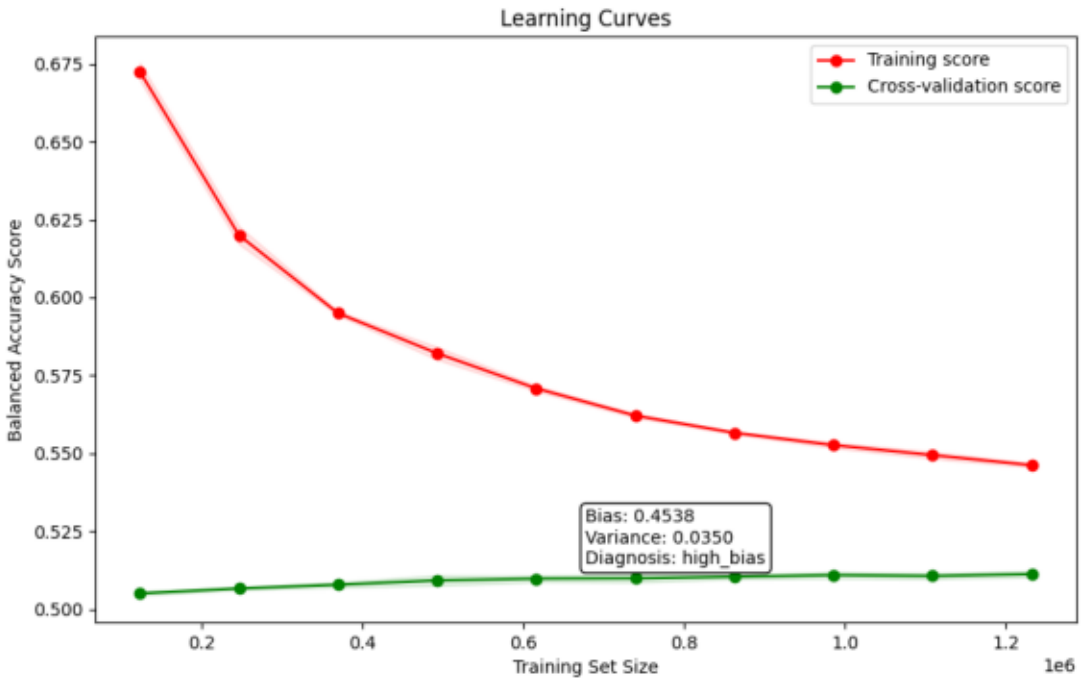
Trend Analysis

Performance Trend Analysis:
Balanced Accuracy: stable
Precision: stable
Recall: stable
F1 Score: stable

Interpretation: Model performance is consistent over time, indicating stable predictive power across different market conditions. This suggests the model is robust and not overly sensitive to specific market regimes.

Bias-Variance Tradeoff Analysis

Learning Curves



Bias-Variance Analysis

Bias: 0.4538
Variance: 0.0350
Diagnosis: high_bias

Recommendations:
1. Use a more complex model
2. Add more features
3. Reduce regularization strength

The model suffers from high bias, indicating it may be underfitting the data. Consider using a mor

Conclusion & Recommendations

Performance Summary

The model achieved a balanced accuracy of 0.5013, which is below the minimum success threshold of 0.55. This suggests limited predictive power that may not be sufficient for reliable investment decision-making.

Success Criteria Assessment

Success criteria assessment not available

Recommendations

Recommendations:

1. Further Research and Development
   - Revisit the feature engineering approach
   - Test different model architectures and algorithms
   - Consider longer or shorter prediction horizons

2. Alternative Approaches
   - Explore sector-specific models rather than broad market prediction
   - Consider predicting volatility rather than direction
   - Test portfolio-level prediction rather than individual stocks

3. Next Steps
   - Conduct in-depth analysis of prediction errors
   - Benchmark against simpler models and strategies
   - Gather additional domain expertise for feature development