Forecasting time-series stock prices (Supplementary)

Machine Learning Capstone Project

Ihab Sultan

Contents

Charging Bull - New York City by Sam valadi used under CC BY 2.0
Bankers by Tim Kwee used under CC BY-NC 2.0

1. appendix

1.1 Supported Visualizations

As we have to pick a handful of stocks samples in this section, we'll pick symbols from S&P 500 information technology sector.

The technology sector of S&P 500, at the time of this writing, constitutes of 68 symbols, where the top 10 constituents (by index weight) are [**?**]: AAPL, MSFT, FB, GOOGL, GOOG, V, INTC, CSCO, IBM, and ORCL. We'll pick from these symbols for our analysis moving forward, the dates we'll explore cover the last 10 years, and lie between January 1st 2013 and July 1st 2016.

1.1.1 Statistics Summary

	SPY	AAPL	MSFT	FB	GOOG
count	860.0	860.0	860.0	860.0	860.0
mean	182.63	91.21	40.19	69.81	557.5
std	20.87	22.89	8.38	27.0	103.06
min	135.7	52.11	23.99	22.9	351.08
25%	166.81	70.58	33.41	49.4	505.12
50%	188.64	94.43	40.95	74.69	546.78
75%	201.48	110.8	45.75	89.45	609.25
max	209.18	129.88	56.07	120.5	776.6
Sharpe ratio	1.03	0.47	1.01	1.33	0.97

Figure 1.1: *Stats summary exploration table*

Stats summary [Figure 1.1] gives all default statistics per stock for the given period, in addition to Sharpe ratio which is calculated assuming risk free rate of 0%.

1.1.2 Simple Moving Average and Moving Standard Deviation

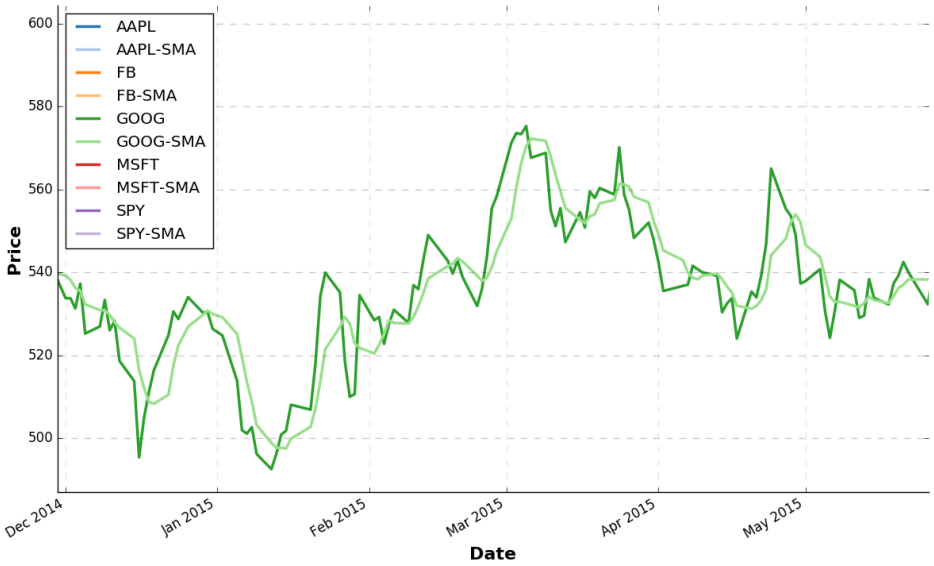

Figure 1.2: *Simple moving average*

Simple moving average (or rolling average) [Figure 1.2] is parametrized over averaging window, and clearly lags the actual price as it relies on accumulation of values instead of instantaneous measurements. Moving average can be utilized as an indicator in some predictive models.

Similarly, rolling standard deviation [Figure 1.3] is used as another indicator, and is commonly used as a measure of stock volatility.

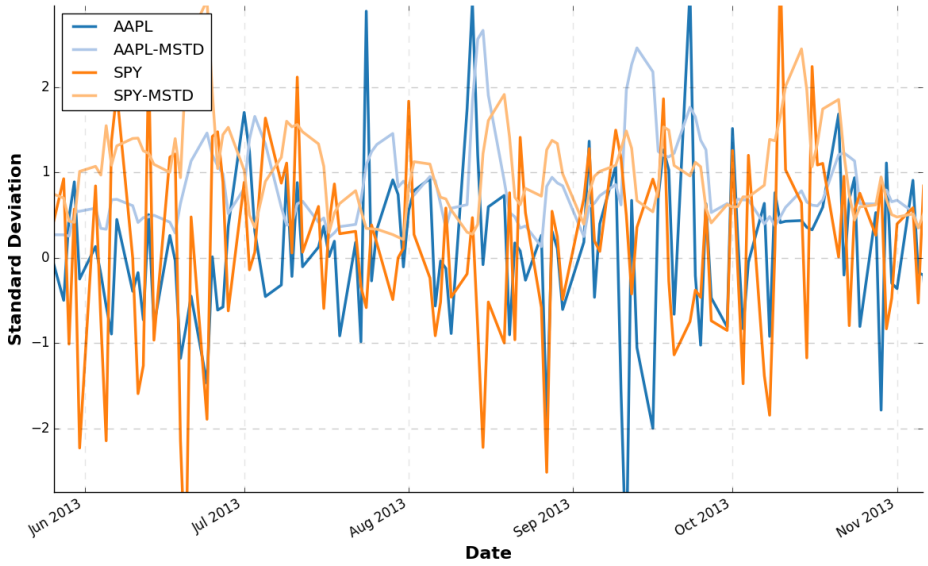

Figure 1.3: *Moving standard deviation*

1.1.3 Bollinger Bands

Figure 1.4: *Bollinger bands*

Bollinger bands [Figure 1.4] is another measure of volatility that is used in technical analysis of stocks. The upper band is one standard deviation above moving average, and lower is one standard deviation below.

1.1.4 Daily Percent Return

The daily percent returns is analyzed in three different forms: plot [Figure 1.5], statistics [Figure 1.6], and histogram [Figure 1.7].

Figure 1.5: *Daily percent return plot*

	SPY	AAPL	FB	GOOG	MSFT
count	860.0	860.0	860.0	860.0	860.0
mean	0.05	0.05	0.19	0.09	0.1
std	0.82	1.64	2.32	1.54	1.55
min	-4.21	-12.36	-6.93	-5.32	-11.4
25%	-0.35	-0.74	-1.02	-0.69	-0.68
50%	0.07	0.02	0.12	0.01	0.05
75%	0.52	1.01	1.3	0.86	0.84
max	3.84	8.2	29.61	16.05	10.45
Sharpe ratio	1.03	0.47	1.33	0.97	1.01

Figure 1.6: *Daily percent return statistics*

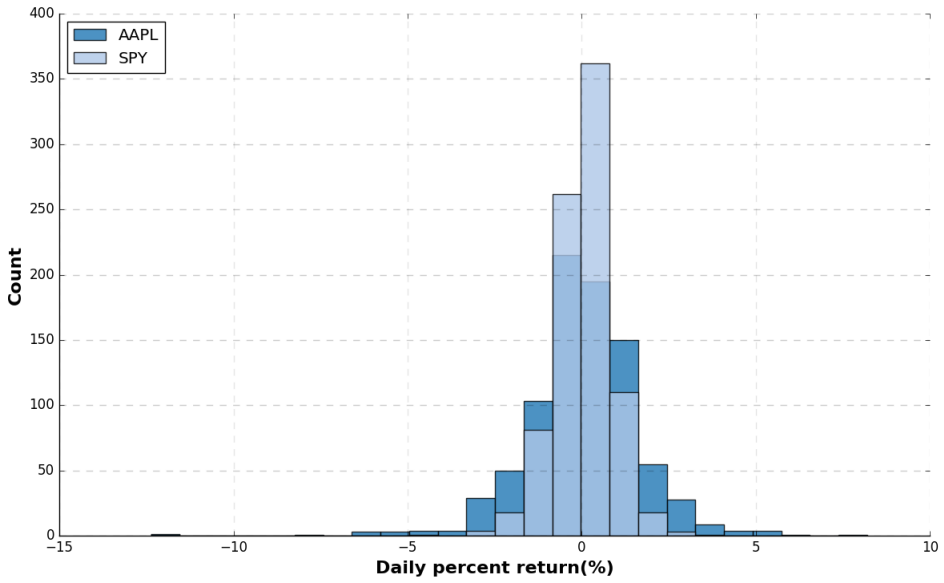

Figure 1.7: *Daily percent return histogram*

The three types of explorations give different views about daily percent returns, which is defined as:

$$\text{DPR}_t = \frac{y_t - y_{t-1}}{y_{t-1}} * 100\% = (\frac{y_t}{y_{t-1}} - 1) * 100\%$$

1.1.5 Cumulative Percent Returns

Cumulative returns [Figure 1.4] is similar to daily returns except that the reference is fixed in time, formally defined as:

$$\text{CPR}_t = \frac{y_t - y_r}{y_r} * 100\% = (\frac{y_t}{y_r} - 1) * 100\%$$

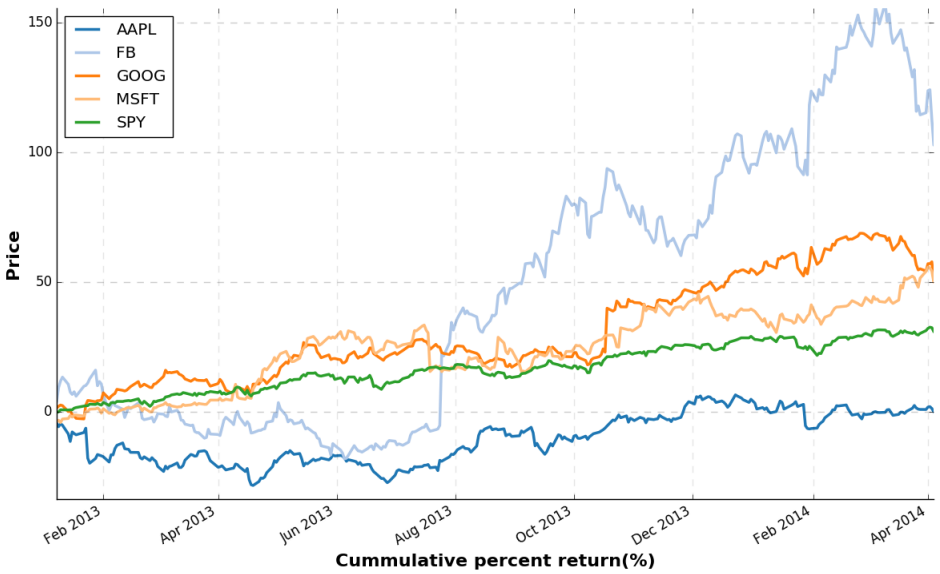

Figure 1.8: *Cumulative percent return plot*

	SPY	AAPL	GOOG	FB	MSFT
count	860.0	860.0	860.0	860.0	860.0
mean	34.12	25.23	54.32	149.31	60.45
std	15.32	31.42	28.53	96.43	33.48
min	-0.35	-28.45	-2.82	-18.21	-4.2
25%	22.5	-3.11	39.82	76.41	33.38
50%	38.53	29.64	51.35	166.73	63.49
75%	47.96	52.12	68.64	219.48	82.67
max	53.62	78.32	114.97	330.36	123.86

Figure 1.9: *Cumulative percent return summary*

1.1.6 Scatter Plot and Correlation Statistics

Scatter plot [Figure 1.10] displays the relation between two stocks. On top of the scatter plot, a linear fit is presented with calculated β (slope) and α (intercept).

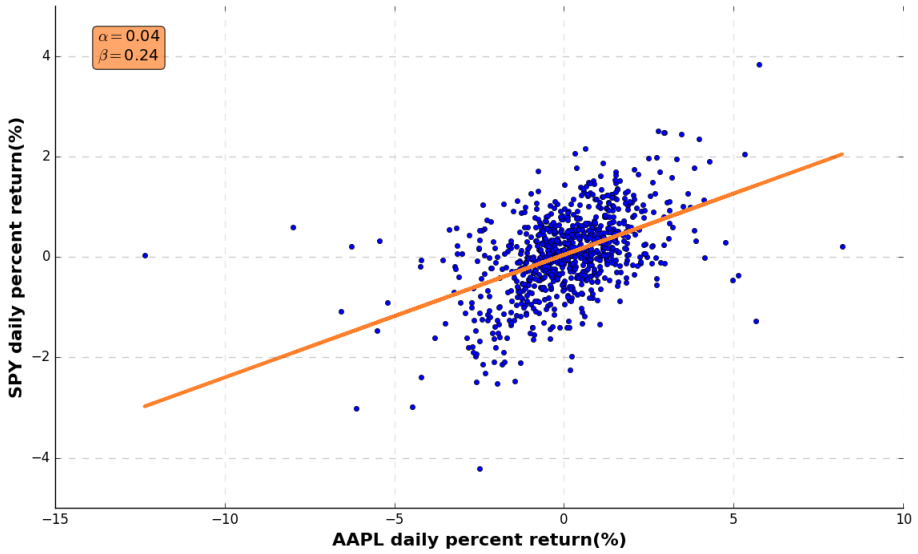

Figure 1.10: *Scatter plot*

Statistical correlation between chosen stocks is displayed in table form in [Figure 1.11].

	SPY	AAPL	INTC	FB	GOOG
SPY	1.0	0.93	0.91	0.9	0.77
AAPL	0.93	1.0	0.87	0.81	0.6
INTC	0.91	0.87	1.0	0.81	0.67
FB	0.9	0.81	0.81	1.0	0.91
GOOG	0.77	0.6	0.67	0.91	1.0

Figure 1.11: *Correlation statistics*

1.2 Results for GOOG

1.2.1 Training Time Results

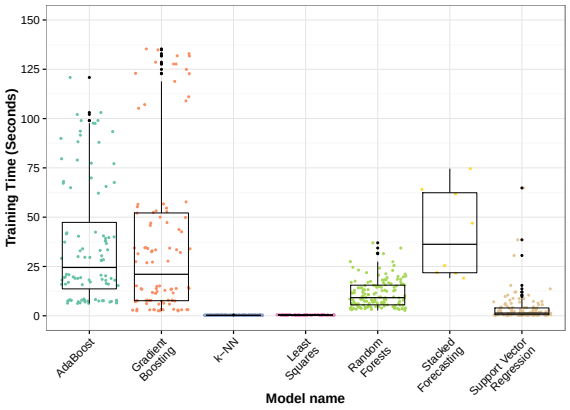

Figure 1.12: *Training time plot: displays the total training time per hyper-parameter permutation of each of the models for GOOG stock.*

1.2.2 Validation Time Results

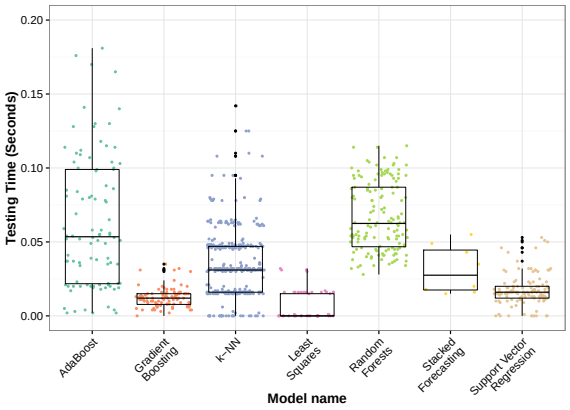

Figure 1.13: *Validation time plot: displays the total validation time per hyper-parameter permutation of each of the models for GOOG stock.*

1.2.3 Validation MAPE Results

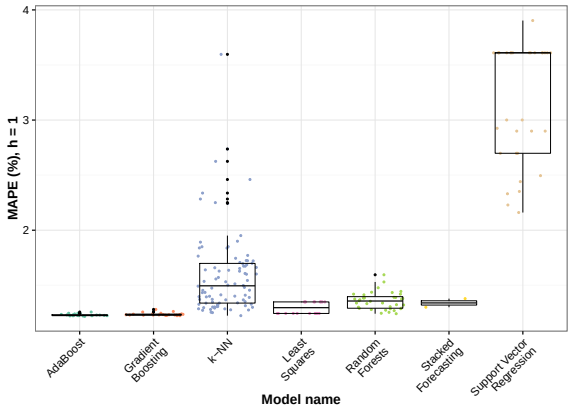

Figure 1.14: *MAPE plot: displays MAPE result for 1 day horizon per predictive model for GOOG stock. Support Vector Regression shows a slightly lacking performance compared to other regression models in this project.*

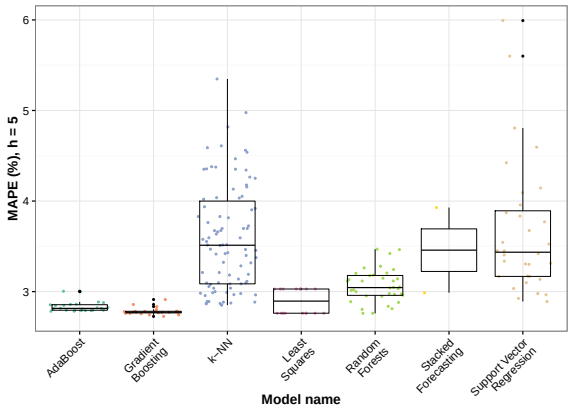

Figure 1.15: *MAPE plot: displays MAPE result for 1 week horizon per predictive model.*

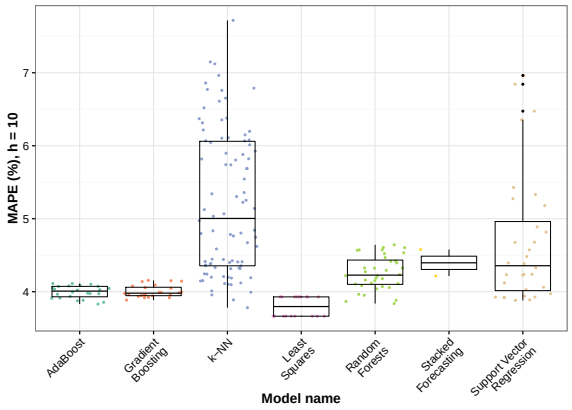

Figure 1.16: *MAPE plot: displays MAPE result for 2 weeks horizon per model.*

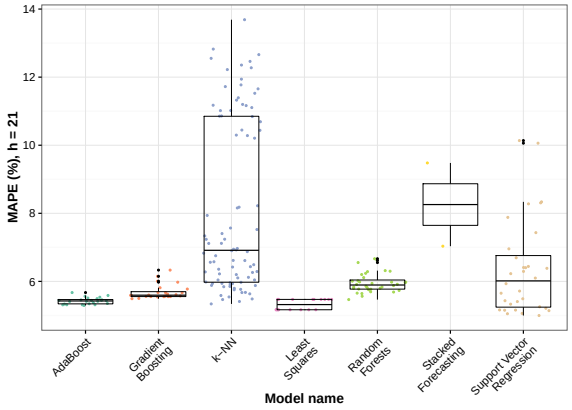

Figure 1.17: *MAPE plot: displays MAPE result for 1 month horizon per model.*

1.2.4 Validation Hit Rate

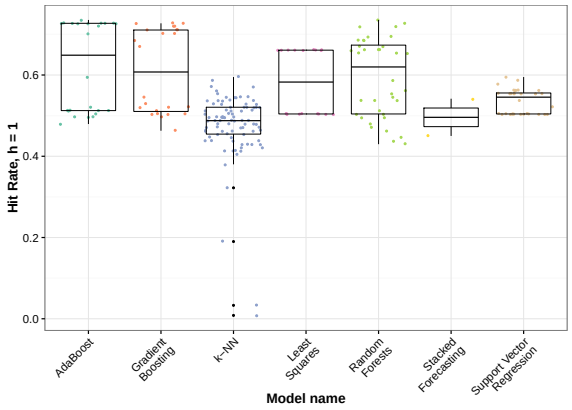

Figure 1.18: *Hit rate plot: displays hit rate result for 1 day horizon per model.*

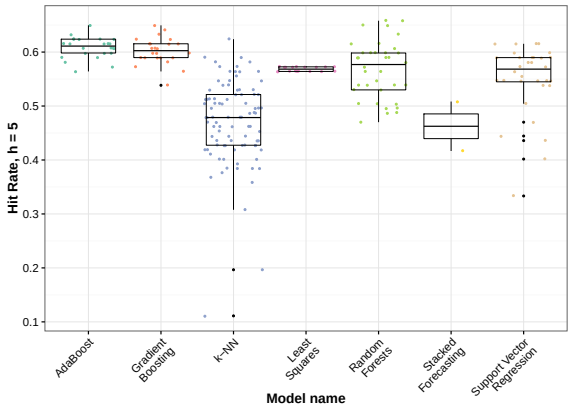

Figure 1.19: *Hit rate plot: displays hit rate result for 1 week horizon per model.*

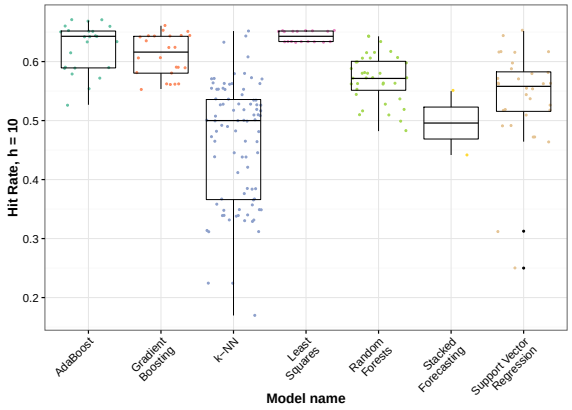

Figure 1.20: *Hit rate plot: displays hit result for 2 weeks horizon per model.*

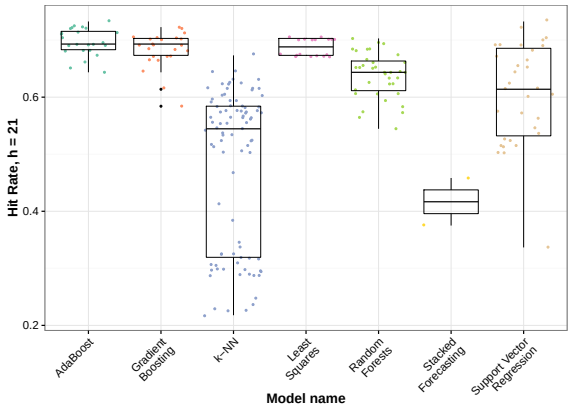

Figure 1.21: *Hit rate plot: displays hit result for 1 month horizon per model.*

1.2.5 Testing Results

Predictive Model	Horizon	MAE	MAPE	MSE	RMSE	Hit Rate
AdaBoost	1	7.78	1.09%	116.34	10.79	0.53
Gradient Boosting	1	7.78	1.09%	114.67	10.71	0.52
k-NN	1	7.78	1.09%	114.47	10.70	0.00
Least Squares	1	7.78	1.09%	115.34	10.74	0.52
Random Forests	1	8.63	1.21%	129.32	11.37	0.45
Stacked Forecasting	1	8.39	1.18%	121.92	11.04	0.58
Support Vector	1	16.14	2.26%	366.88	19.15	0.52

Figure 1.22: *Testing Metrics Table: displays all metrics results for testing with 1 day forecast horizon.*

Predictive Model	Horizon	MAE	MAPE	MSE	RMSE	Hit Rate
AdaBoost	5	18.78	2.65%	704.19	26.54	0.51
Gradient Boosting	5	18.52	2.60%	592.85	24.35	0.44
k-NN	5	19.95	2.80%	671.40	25.91	0.48
Least Squares	5	17.73	2.50%	573.45	23.95	0.56
Random Forests	5	19.55	2.75%	682.10	26.12	0.56
Stacked Forecasting	5	19.52	2.74%	604.65	24.59	0.65
Support Vector	5	22.39	3.16%	923.66	30.39	0.50

Figure 1.23: *Testing Metrics Table: displays all metrics results for testing with 1 week forecast horizon.*

Predictive Model	Horizon	MAE	MAPE	MSE	RMSE	Hit Rate
AdaBoost	10	24.03	3.39%	1020.52	31.95	0.52
Gradient Boosting	10	23.04	3.25%	979.64	31.30	0.55
k-NN	10	24.14	3.40%	1014.62	31.85	0.52
Least Squares	10	23.91	3.38%	997.54	31.58	0.52
Random Forests	10	25.49	3.60%	1131.79	33.64	0.52
Stacked Forecasting	10	24.38	3.43%	915.70	30.26	0.57
Support Vector	10	34.93	4.95%	1884.88	43.42	0.52

Figure 1.24: *Testing Metrics Table: displays all metrics results for testing with 2 weeks forecast horizon.*

Predictive Model	Horizon	MAE	MAPE	MSE	RMSE	Hit Rate
AdaBoost	21	35.58	5.04%	1853.47	43.05	0.48
Gradient Boosting	21	32.83	4.63%	1463.24	38.25	0.42
k-NN	21	48.75	6.87%	3379.54	58.13	0.32
Least Squares	21	36.15	5.11%	1878.40	43.34	0.43
Random Forests	21	34.54	4.87%	1758.79	41.94	0.52
Stacked Forecasting	21	33.54	4.64%	1686.53	41.07	0.53
Support Vector	21	37.54	5.33%	2090.61	45.72	0.44

Figure 1.25: *Testing Metrics Table: displays all metrics results for testing with 1 month forecast horizon.*