# DataRobot vs. Open Source Forecasting Tools

*Zachary Deane-Mayer*

*14 July, 2017*

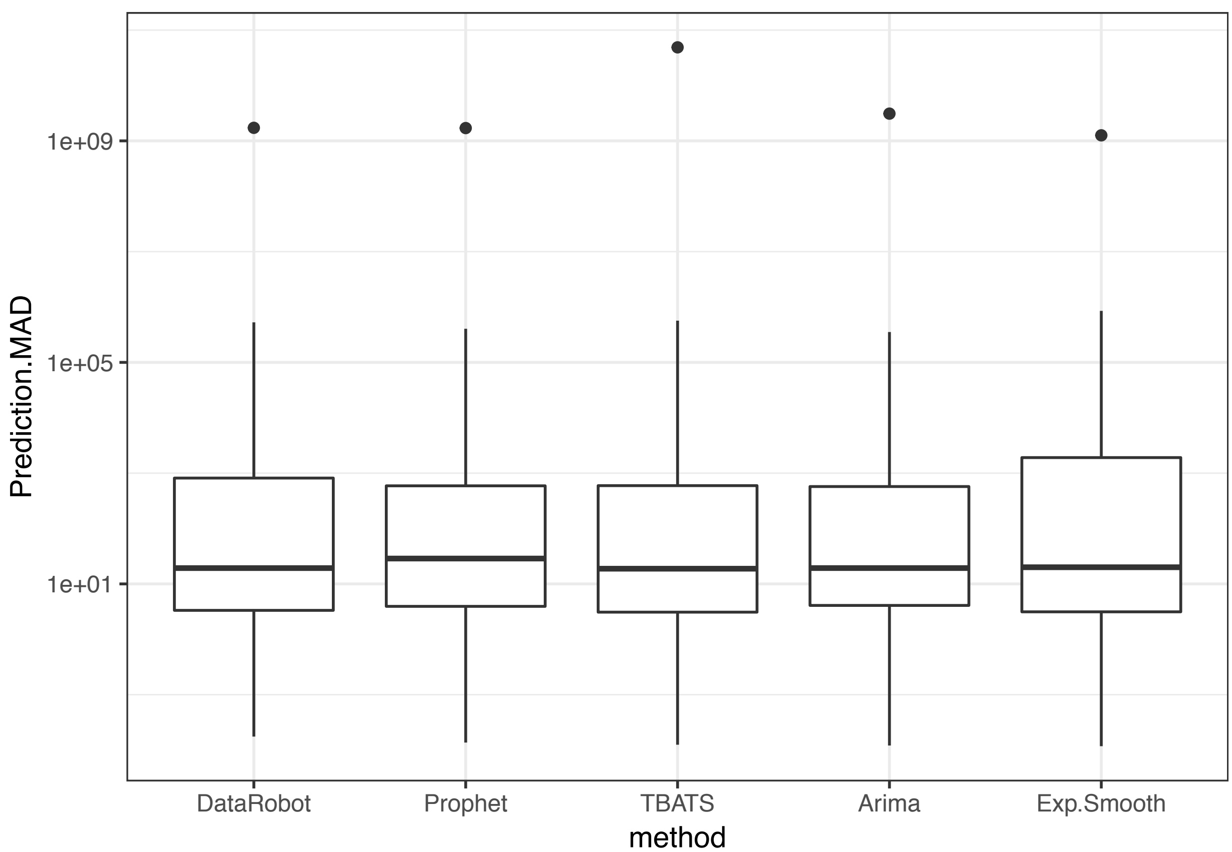
## Methodology

I compared DataRobot's Out-of-Time Validation (OTV) models to R's Forecast package and Facebook's Prophet package on 69 datasets. I split these datasets into training sets and test sets, and ran each algorithm on the training set and calculated accuracy on the unseen test set. For DataRobot, I used the test set as a prediction set, so that it was not used by the autopilot for model training or selection. I used the metric "Normalized mad Score" or "mad Norm" to compare algorithms, which ranges from -1 (perfectly anti-predictive) to 0 (equivalent to random guessing) to 1 (perfectly predictive).

With time series problems, we see negative mad scores more often than with traditional machine learning problems, as models can be fooled by randomness in the time series and end up extrapolating "trends" that do not exist.

I compared DataRobot to a total of 4 open source models: Prophet (from Facebook), auto.arima (from Forecast), ets (automated exponential smoothing from Forecast) and TBATS (a trigonometric function based model from Forecast). Note that all 4 of the open-source forecasting models minimize MAD, so to keep the comparison fair, I used MAD as the metric for all DataRobot projects. For each dataset, the DataRobot model with the best mad Norm on the holdout set was used to make forecasts on the prediction set.

## Results



On average, DataRobot is slightly more accurate than Facebook's Prophet package (median mad of 19.31 vs 28.73). TBATS is the best model from the Forecast package, but is typically worse that DataRobot or Prophet (median mad of 18.81).

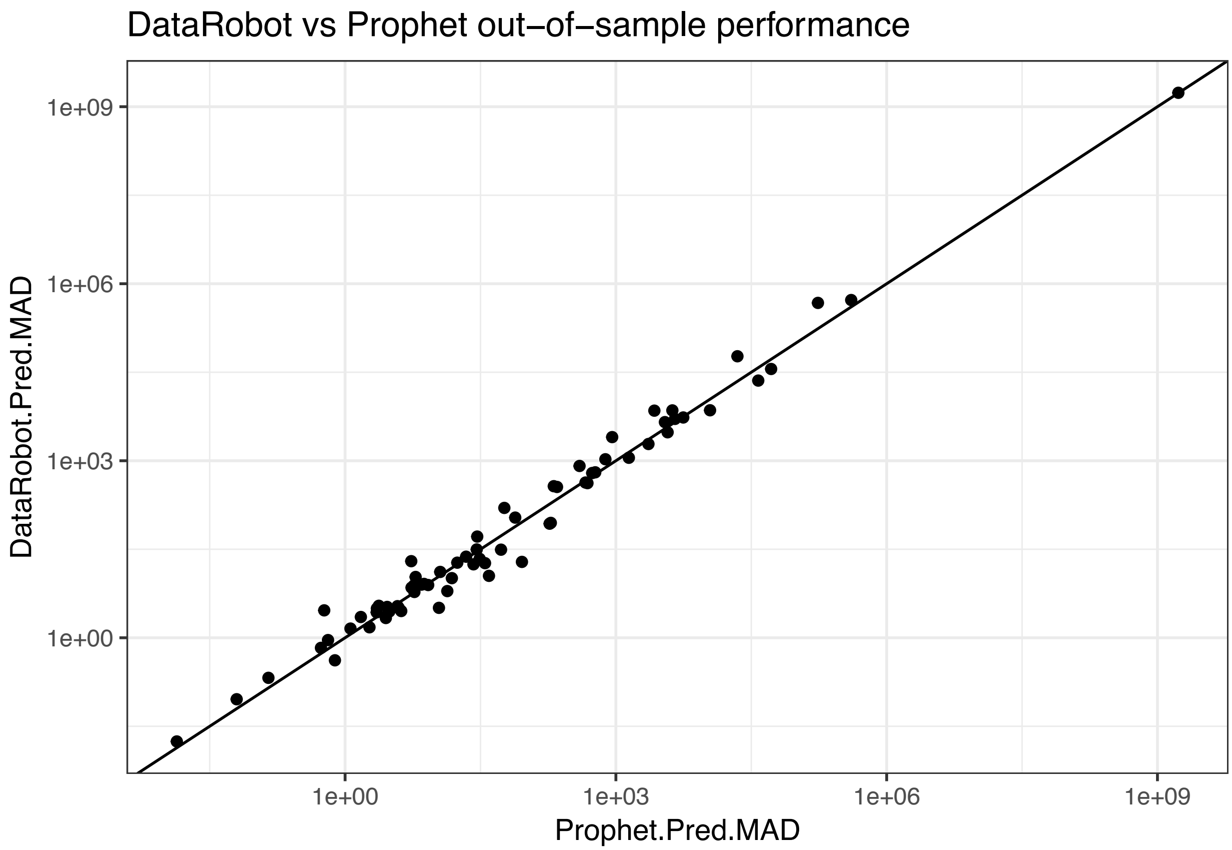
| | method | min_mad | median_mad | mean_mad | sd_mad | max_mad |
|---|---|---|---|---|---|---|
| 1 | DataRobot | 0.02 | 19.31 | 24955365.11 | 207151902.42 | 1720750000.45 |
| 2 | Prophet | 0.01 | 28.73 | 24697403.86 | 205062305.79 | 1703386189.96 |
| 3 | TBATS | 0.01 | 18.81 | 707727098.72 | 5878698777.44 | 48832154474.78 |
| 4 | Arima | 0.01 | 19.31 | 44995568.87 | 373669711.86 | 3103944737.31 |
| 5 | Exp.Smooth | 0.01 | 20.00 | 18273883.81 | 151636494.98 | 1259606018.87 |

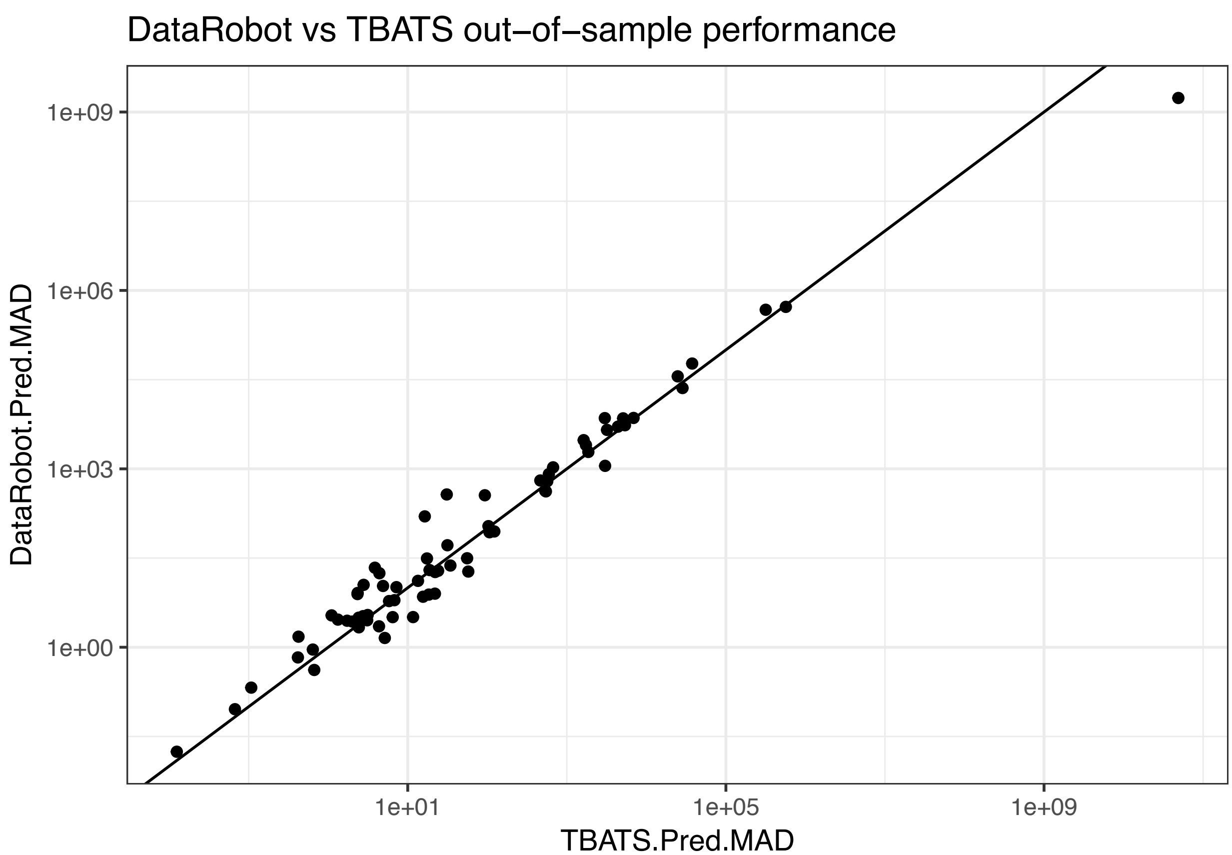
Table 1: Summary Results

## Conclusion

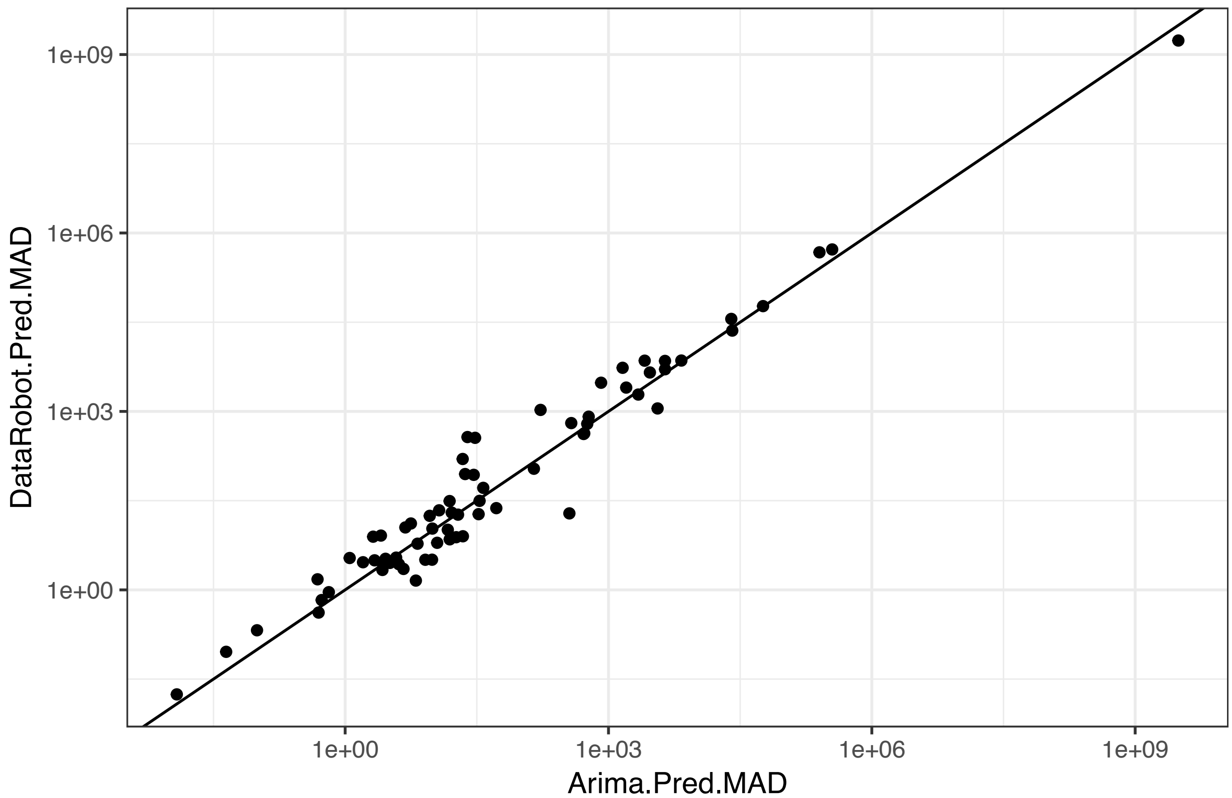
DataRobot's OTV models already perform well on a variety of time series problems, and are on average more accurate than all 4 of the open source forecasting models tried. DataRobot does especially well on datasets with complex seasonal patterns, e.g. electricity load, where the hour-of-day pattern may differ in a dramatic (and predictable) way between the winter and summer months. This sort of dynamic seasonality can be extremely difficult for traditional time series models to capture, but is modeled beautifully by DataRobot's "seasonal dummies + XGBoost" approach. DataRobot is also able to make use of covariates, while arima is the only open-source model with this capability.

However, there remains some room for improvement in DataRobot, which can be over-confident based on the holdout set when picking the model to use for forecasting. It might be beneficial to add some heuristics to DataRobot to cause it to prefer simpler models during model selection for time series models. This could help prevent DataRobot from picking complicated models that happen to get lucky in the holdout set but end up extrapolating the wrong trend. On random walk datasets like the S&P 500 stock market data, or the incarceration data, the correct forecast is usually a flat line from the last point (also known as a naive forecast), and the open source models tend to correctly predict this, while DataRobot tends to be overconfident in extrapolating trends.

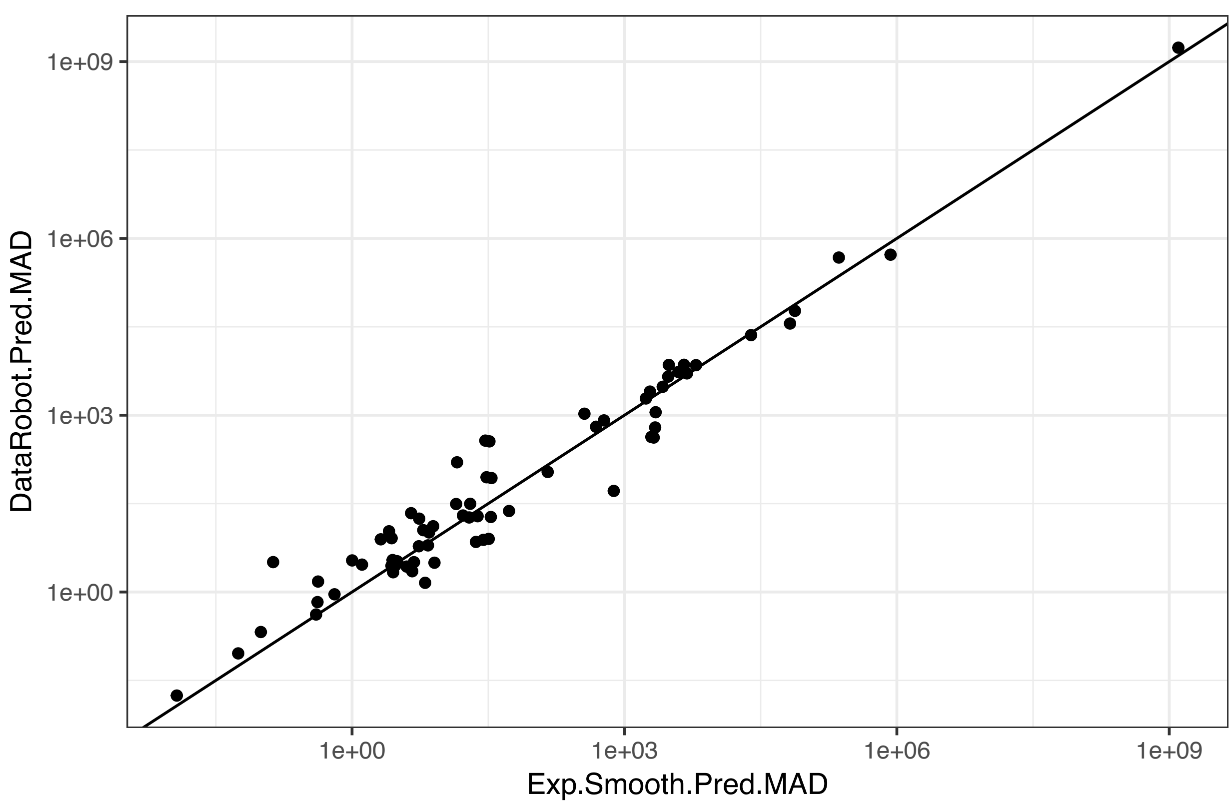

DataRobot vs Prophet out−of−sample performance
1e+09
1e+06
1e+03
1e+00
DataRobot.Pred.MAD
1e+00
1e+03
1e+06
1e+09
Prophet.Pred.MAD


DataRobot vs TBATS out−of−sample performance
1e+09
1e+06
1e+03
1e+00
DataRobot.Pred.MAD
1e+01
1e+05
1e+09
TBATS.Pred.MAD

DataRobot vs Arima out−of−sample performance




DataRobot vs Exp.Smooth out−of−sample performance



| | dataset | DataRobot.Pred.MAD | Prophet.Pred.MAD | TBATS.Pred.MAD | Arima.Pred.MAD | Exp.Smooth.Pred.MAD | DR.Min.Pct.Improve |
|---|---|---|---|---|---|---|---|
| 1 | SWilliams_train.csv | 19.31 | 91.00 | 24.08 | 358.51 | 24.13 | 0.20 |
| 2 | iso_ne_hourly_load_train.csv | 1122.79 | 1392.45 | 3029.61 | 3624.61 | 2205.24 | 0.19 |
| 3 | SP500_SPG_train.csv | 3.21 | 3.93 | 6.47 | 8.19 | 4.80 | 0.18 |
| 4 | facebook_wp_with_outliers1_train.csv | 417.65 | 484.28 | 545.87 | 520.80 | 2097.78 | 0.14 |
| 5 | fpp_antidiabetic_drugs_train.csv | 2.16 | 2.83 | 2.43 | 2.66 | 2.83 | 0.11 |
| 6 | facebook_data_train.csv | 6.19 | 13.56 | 6.79 | 11.18 | 6.85 | 0.09 |
| 7 | fpp_aus_departures_visshort_train.csv | 22843.47 | 37774.24 | 28499.47 | 25701.34 | 24803.70 | 0.08 |
| 8 | facebook_wp_no_outliers_train.csv | 427.45 | 460.95 | 534.80 | 529.40 | 1977.75 | 0.07 |
| 9 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_JUVENILES _train.csv | 2.84 | 4.19 | 3.09 | 3.23 | 3.00 | 0.05 |
| 10 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_MALES _train.csv | 18.39 | 35.48 | 22.12 | 19.31 | 19.48 | 0.05 |
| 11 | fpp_aus_air_traffic_first_train.csv | 0.42 | 0.77 | 0.67 | 0.50 | 0.40 | -0.03 |
| 12 | twitter_volume_over_time_train.csv | 18.74 | 17.50 | 57.71 | 33.15 | 33.74 | -0.07 |
| 13 | dr-commit-activity_train.csv | 23.67 | 21.91 | 34.38 | 52.72 | 53.55 | -0.08 |
| 14 | 2017-05_Percent_Change_of_Sales_of_Existing_Detached_Homes_LA_AREA_train.csv | 5.97 | 5.86 | 5.82 | 6.68 | 5.44 | -0.10 |
| 15 | fpp_aus_departures_reslong_train.csv | 1922.87 | 2294.82 | 1864.47 | 2185.38 | 1722.89 | -0.12 |
| 16 | facebook_wp_with_outliers2_train.csv | 618.91 | 546.34 | 567.55 | 572.60 | 2177.80 | -0.13 |
| 17 | wunderground_Chicago_actual_max_temp_train.csv | 7.96 | 7.03 | 21.96 | 21.94 | 31.98 | -0.13 |
| 18 | Google-TotalSpend_train.csv | 5118.15 | 4515.23 | 4396.71 | 4409.00 | 4876.00 | -0.16 |
| 19 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HOPE_VILLAGE _train.csv | 3.33 | 2.92 | 2.74 | 2.88 | 3.13 | -0.22 |
| 20 | weekly_earth_co2_train.csv | 1.43 | 1.15 | 5.15 | 6.39 | 6.38 | -0.25 |
| 21 | wunderground_Chicago_actual_min_temp_train.csv | 7.07 | 5.45 | 15.58 | 15.54 | 23.12 | -0.30 |
| 22 | wunderground_Chicago_actual_mean_temp_train.csv | 7.68 | 5.78 | 18.40 | 18.41 | 28.00 | -0.33 |
| 23 | fpp_aus_air_traffic_economy_train.csv | 2.70 | 2.24 | 2.01 | 4.09 | 3.99 | -0.34 |
| 24 | internet_time_series_train.csv | 1720750000.45 | 1703386189.96 | 48832154474.78 | 3103944737.31 | 1259606018.87 | -0.37 |
| 25 | hyndman_usa_gasoline_train.csv | 108.75 | 76.66 | 103.47 | 141.29 | 142.80 | -0.42 |
| 26 | fpp_us_personal_income_change_train.csv | 0.91 | 0.65 | 0.64 | 0.65 | 0.64 | -0.43 |
| 27 | fpp_aus_departures_resshort_train.csv | 35908.85 | 52470.53 | 24801.96 | 25067.93 | 66601.82 | -0.45 |
| 28 | methane-input-into-gas-furnace-c_train.csv | 3.14 | 2.25 | 2.42 | 2.16 | 8.08 | -0.45 |
| 29 | fpp_us_electricity_net_generation_train.csv | 10.25 | 15.23 | 7.21 | 14.76 | 7.05 | -0.45 |
| 30 | SP500_URI_train.csv | 3.48 | 2.36 | 3.13 | 3.81 | 2.79 | -0.47 |
| 31 | google_trends_debt_searches_predict_market_returns_train.csv | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | -0.49 |
| 32 | Google-TotalImpressions_train.csv | 528780.69 | 404501.46 | 567186.03 | 353430.13 | 852257.07 | -0.50 |
| 33 | SP500_MSI_train.csv | 2.25 | 1.49 | 4.35 | 4.62 | 4.60 | -0.50 |
| 34 | Facebook-TotalClicks_train.csv | 4516.70 | 3490.82 | 3201.66 | 2962.34 | 3029.14 | -0.52 |
| 35 | dutch_politics_train.csv | 31.37 | 28.73 | 55.80 | 33.99 | 20.00 | -0.57 |
| 36 | facebook_retail_train.csv | 7170.17 | 11049.36 | 6904.73 | 6761.74 | 4526.81 | -0.58 |
| 37 | SP500_VRSK_train.csv | 2.79 | 3.09 | 1.73 | 2.67 | 2.72 | -0.61 |
| 38 | fpp_us_personal_consumption_change_train.csv | 0.67 | 0.54 | 0.42 | 0.54 | 0.42 | -0.62 |
| 39 | fpp_aus_departures_permanent_train.csv | 636.91 | 590.45 | 464.52 | 376.78 | 486.67 | -0.69 |
| 40 | energydata_complete_train.csv | 51.90 | 29.09 | 31.53 | 37.44 | 762.07 | -0.78 |
| 41 | fpp_aus_cortecosteroid_sales_train.csv | 0.09 | 0.06 | 0.07 | 0.04 | 0.06 | -1.06 |
| 42 | goog_trends_all_terms_predict_DJIA_raw_price_train.csv | 815.63 | 395.67 | 594.74 | 594.74 | 594.74 | -1.06 |
| 43 | wunderground_Chicago_actual_precipitation_train.csv | 0.21 | 0.14 | 0.11 | 0.10 | 0.10 | -1.12 |
| 44 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF _train.csv | 31.12 | 53.39 | 17.45 | 15.49 | 14.00 | -1.22 |
| 45 | fpp_autralian_beer_train.csv | 13.06 | 11.32 | 13.45 | 5.61 | 7.82 | -1.33 |
| 46 | Facebook-TotalSpend_train.csv | 7062.36 | 2668.13 | 5103.91 | 4402.27 | 6142.13 | -1.65 |
| 47 | BCHAIN-NTRAN_train.csv | 58965.88 | 22206.64 | 37657.69 | 57626.12 | 75431.28 | -1.66 |
| 48 | Facebook-TotalImpressions_train.csv | 473146.33 | 173171.57 | 316659.80 | 253105.80 | 229707.11 | -1.73 |
| 49 | Google-TotalClicks_train.csv | 7131.86 | 4223.05 | 3002.81 | 2584.44 | 3080.50 | -1.76 |
| 50 | hyndman_turkish_electricity_demand_train.csv | 2515.93 | 909.31 | 1737.96 | 1591.75 | 1909.25 | -1.77 |
| 51 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF _train.csv | 85.83 | 184.50 | 106.67 | 29.16 | 34.30 | -1.94 |
| 52 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ FAIRVIEW _train.csv | 3.43 | 3.81 | 1.10 | 1.13 | 1.00 | -2.42 |
| 53 | SP500_DLTR_train.csv | 8.18 | 7.49 | 2.34 | 2.55 | 2.73 | -2.49 |
| 54 | fpp_aus_air_traffic_business_train.csv | 1.51 | 1.86 | 0.42 | 0.48 | 0.42 | -2.57 |
| 55 | gdelt_wti_daily_train.csv | 19.86 | 5.40 | 18.81 | 16.38 | 16.65 | -2.68 |
| 56 | facebook_peyton_manning_train.csv | 3035.60 | 3741.50 | 1629.21 | 823.60 | 2639.11 | -2.69 |
| 57 | fpp_aus_departures_vislong_train.csv | 5411.35 | 5576.32 | 5385.52 | 1448.26 | 3927.16 | -2.74 |
| 58 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HWH _train.csv | 7.83 | 8.32 | 2.34 | 2.08 | 2.07 | -2.78 |
| 59 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_MALES _train.csv | 88.55 | 190.06 | 122.57 | 23.27 | 30.27 | -2.81 |
| 60 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_FEMALES _train.csv | 17.57 | 26.45 | 4.40 | 9.19 | 5.49 | -2.99 |
| 61 | HPI_PO_monthly_hist_PacRegion_train.csv | 11.21 | 39.25 | 2.79 | 4.85 | 6.08 | -3.02 |
| 62 | fpp_eu_electric_equipment_train.csv | 10.72 | 6.04 | 4.89 | 9.85 | 2.55 | -3.20 |
| 63 | monthly_earth_co2_train.csv | 2.91 | 0.59 | 1.32 | 1.60 | 1.29 | -3.97 |
| 64 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_FEMALES _train.csv | 21.75 | 30.95 | 3.86 | 11.76 | 4.47 | -4.63 |
| 65 | fpp_iceland_debit_card_usage_train.csv | 1060.14 | 765.27 | 672.86 | 168.29 | 362.38 | -5.30 |
| 66 | Federer_train.csv | 158.72 | 58.09 | 16.38 | 21.82 | 14.33 | -10.08 |
| 67 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC _train.csv | 359.67 | 222.65 | 93.23 | 30.21 | 32.57 | -10.90 |
| 68 | incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_MALES _train.csv | 370.66 | 205.13 | 30.93 | 24.76 | 29.34 | -13.97 |
| 69 | RecessionData_train.csv | 3.22 | 10.96 | 11.68 | 9.77 | 0.14 | -22.80 |

Table 2: Full Results