DataRobot vs. Open Source Forecasting Tools

Zachary Deane-Mayer

14 July, 2017

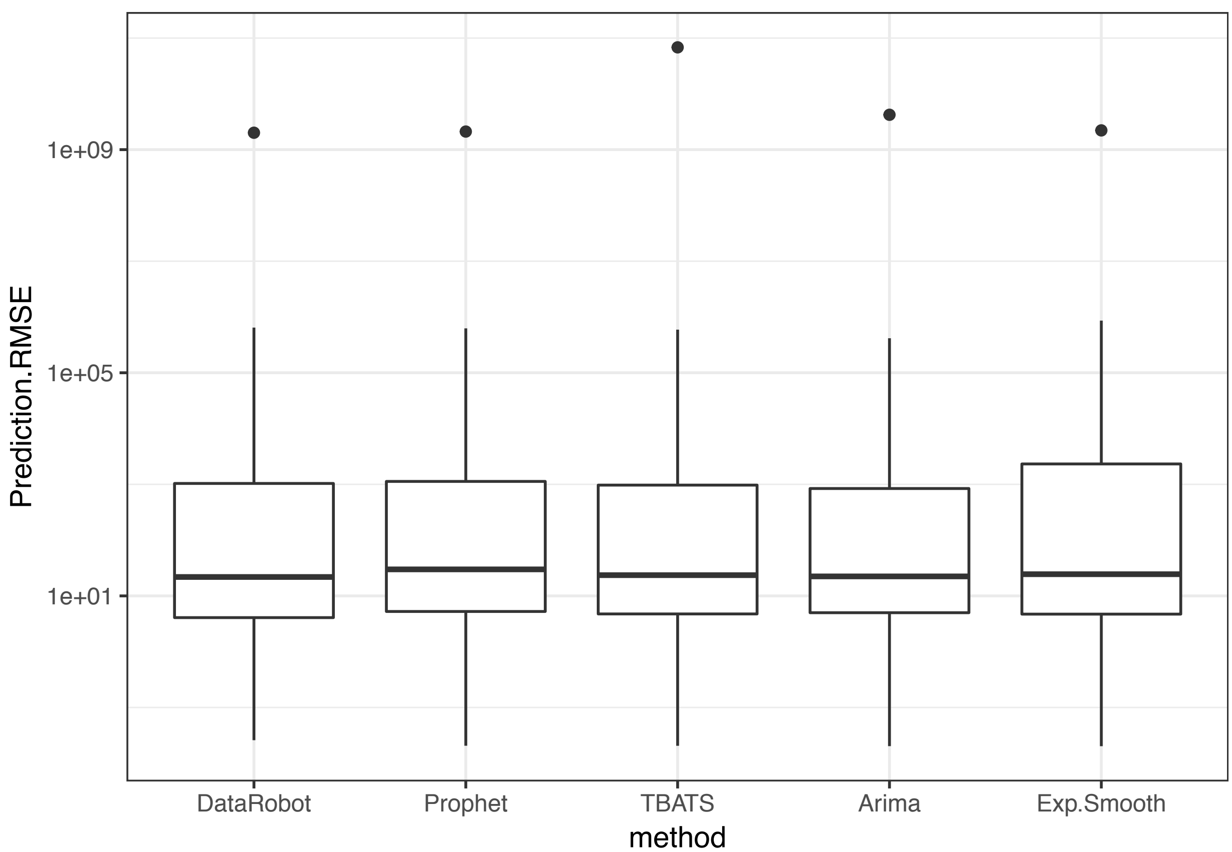
Methodology

I compared DataRobot's Out-of-Time Validation (OTV) models to R's Forecast package and Facebook's Prophet package on 70 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 rmse Score" or "rmse 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 rmse 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 RMSE, so to keep the comparison fair, I used RMSE as the metric for all DataRobot projects. For each dataset, the DataRobot model with the best rmse 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 rmse of 21.83 vs 29.9). TBATS is the best model from the Forecast package, but is typically worse that DataRobot or Prophet (median rmse of 23.57).

	method	min_rmse	median_rmse	mean_rmse	sd_rmse	max_rmse
1	DataRobot	0.03	21.83	28698484.82	239938277.65	2007487860.33
2	Prophet	0.02	29.90	30119580.90	251858397.93	2107215173.25
3	TBATS	0.02	23.57	973899307.57	8148078566.71	68171733940.64
4	Arima	0.02	22.36	60244214.16	503919915.14	4216110723.92
5	Exp.Smooth	0.02	24.57	31630703.99	264460777.40	2212659013.57

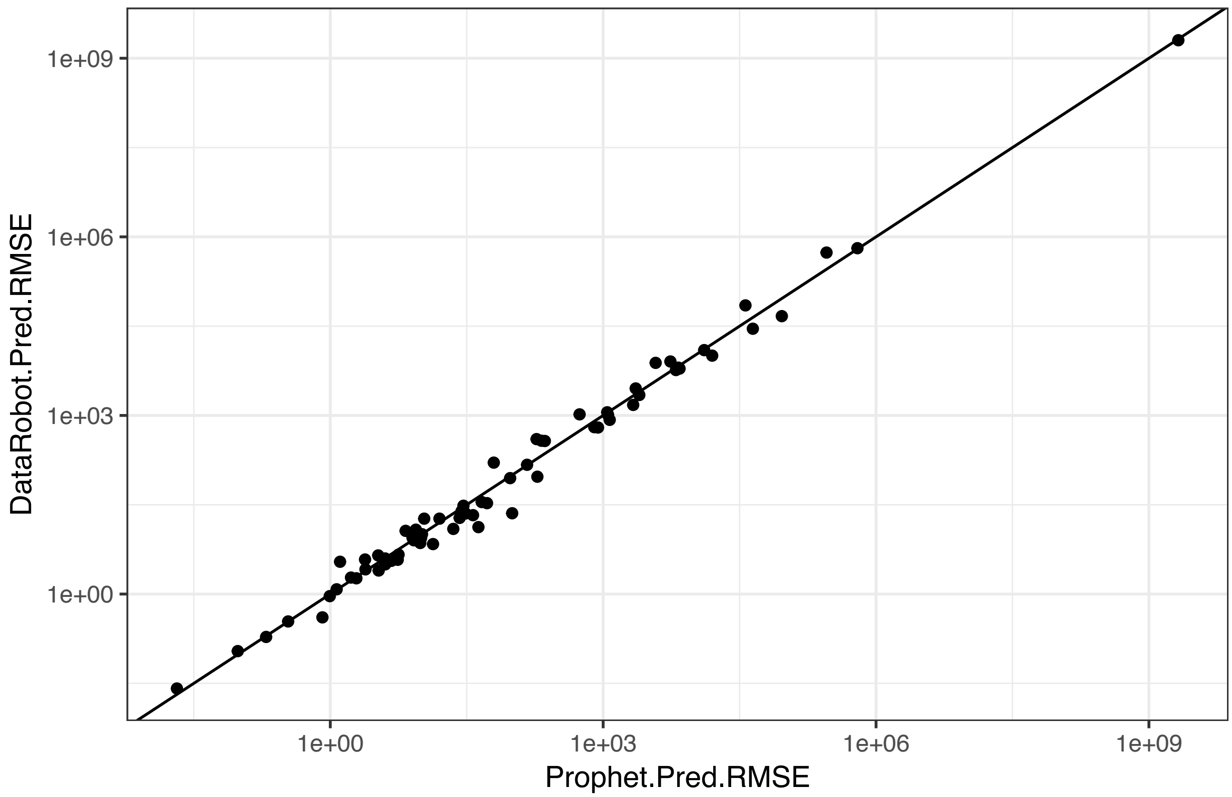
Table 1: Summary Results

Conclusion

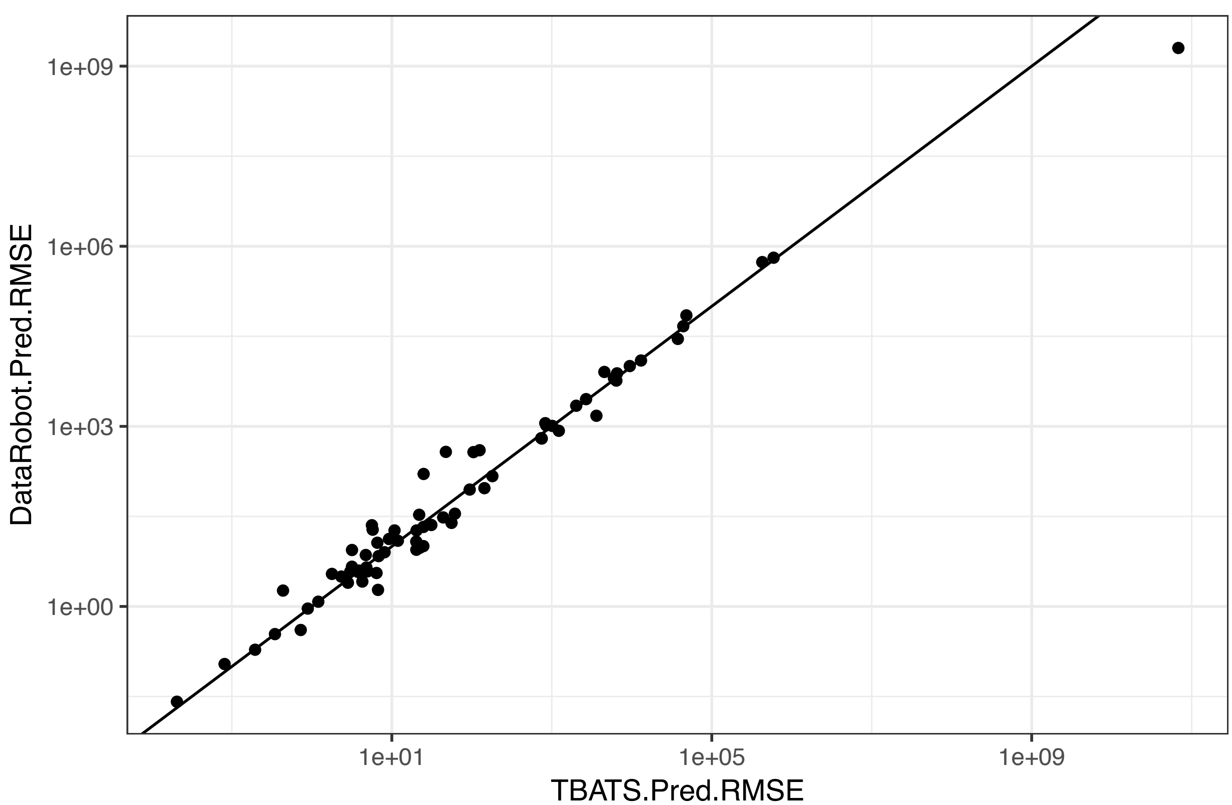
DataRobot's OTV models already perform well on a variety of time series problems, and are on average more accurate than all 4 of the open source forecasting models tried. DataRobot does especially well on datasets with complex seasonal patterns, e.g. electricity load, where the hour-of-day pattern may 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.

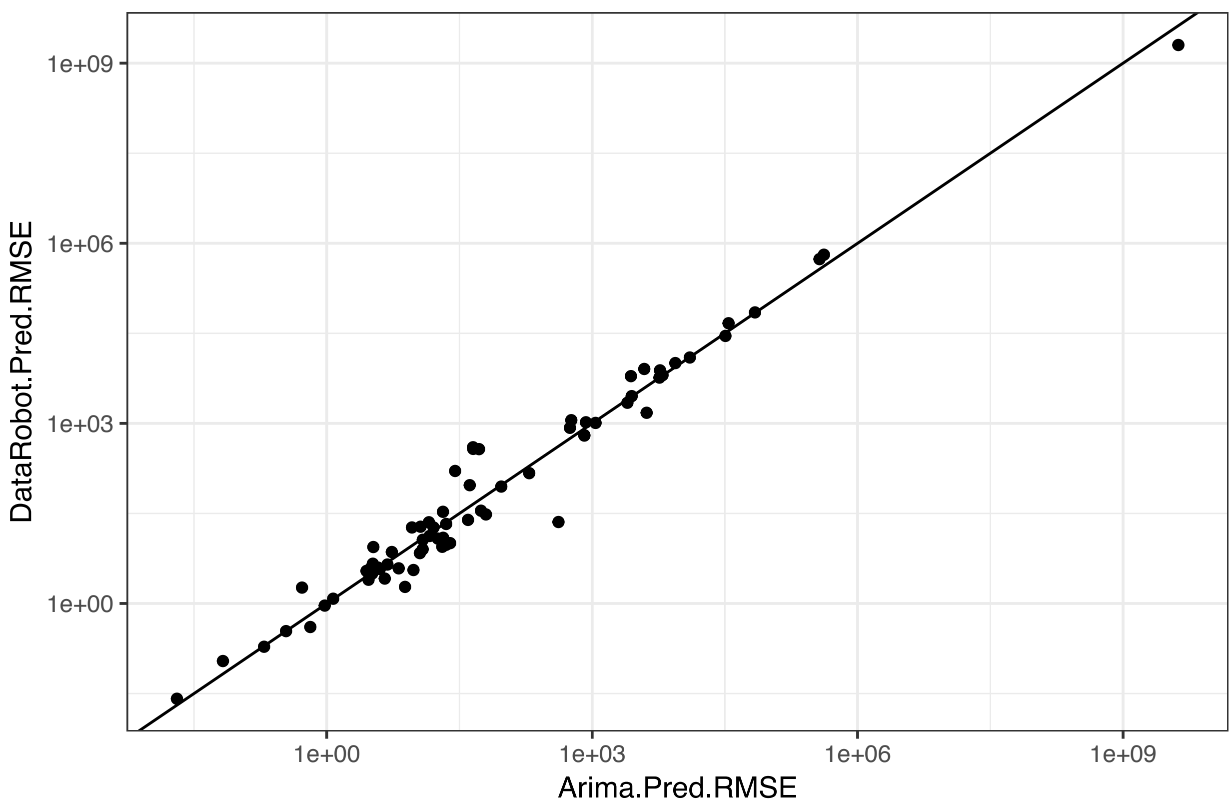
DataRobot vs Prophet out−of−sample performance



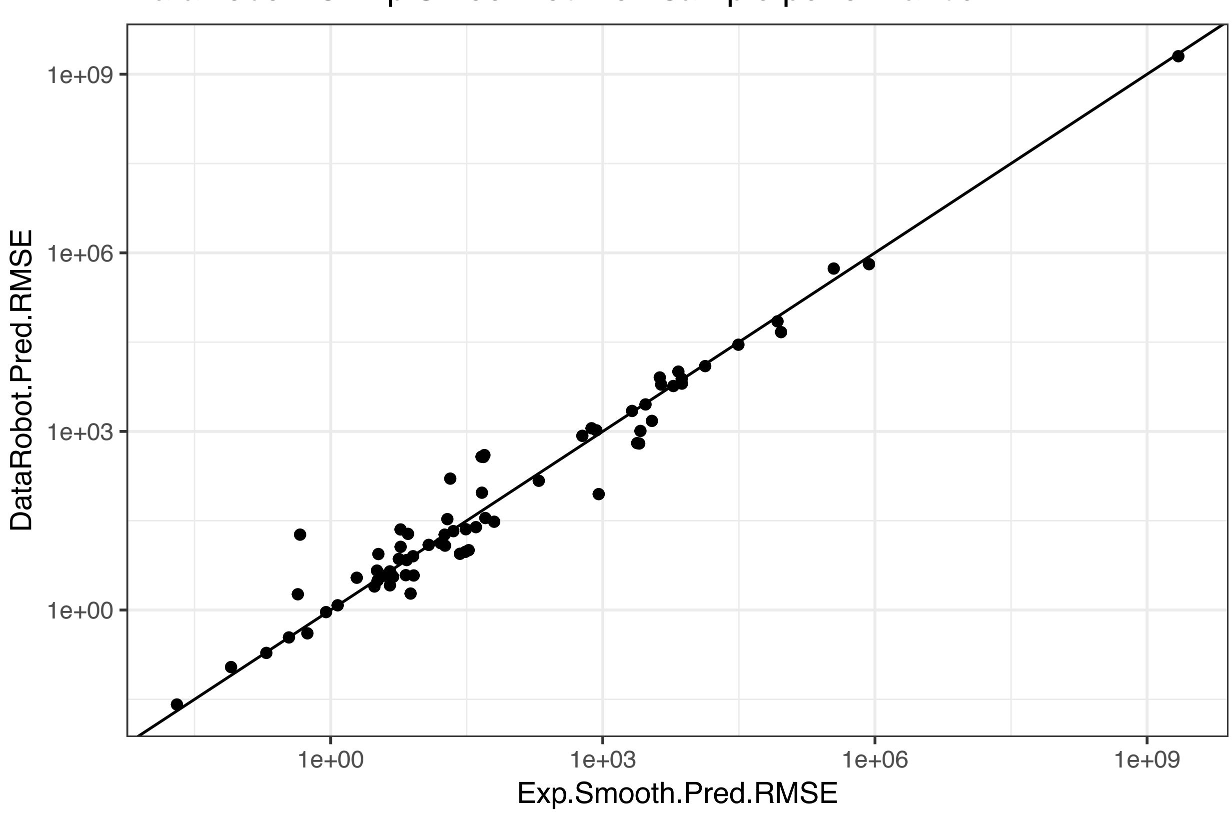
DataRobot vs TBATS out−of−sample performance



DataRobot vs Arima out−of−sample performance



DataRobot vs Exp.Smooth out−of−sample performance



	dataset	DataRobot.Pred.RMSE	Prophet.Pred.RMSE	TBATS.Pred.RMSE	Arima.Pred.RMSE	Exp.Smooth.Pred.RMSE	DR.Min.Pct.Improve
1	iso_ne_hourly_load_train.csv	1501.76	2139.93	3613.52	4127.05	3481.56	0.30
2	fpp_aus_air_traffic_first_train.csv	0.41	0.82	0.73	0.65	0.55	0.27
3	SWilliams_train.csv	22.76	99.94	31.20	416.11	30.94	0.26
4	dutch_politics_train.csv	35.11	46.23	61.49	55.38	50.76	0.24
5	SP500_SPG_train.csv	3.61	4.69	6.44	9.56	4.89	0.23
6	fpp_aus_air_traffic_economy_train.csv	3.85	5.17	4.97	6.51	6.74	0.23
7	facebook_wp_with_outliers1_train.csv	626.86	878.10	749.09	815.99	2521.25	0.16
8	facebook_wp_no_outliers_train.csv	635.83	799.48	738.32	814.16	2394.74	0.14
9	fpp_antidiabetic_drugs_train.csv	2.48	3.40	2.82	2.97	3.04	0.12
10	twitter_volume_over_time_train.csv	24.66	27.77	55.78	39.52	39.99	0.11
11	fpp_aus_departures_visshort_train.csv	28626.81	44217.69	37632.51	32073.34	31249.74	0.08
12	energydata_complete_train.csv	88.57	94.99	94.13	93.89	901.90	0.06
13	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_MALES _train.csv	21.12	37.09	24.84	22.38	22.52	0.06
14	internet_time_series_train.csv	2007487860.33	2107215173.25	68171733940.64	4216110723.92	2212659013.57	0.05
15	ozone_data_train.csv	0.19	0.20	0.20	0.20	0.20	0.03
16	facebook_peyton_manning_train.csv	12511.50	12908.31	13043.58	12690.84	13455.03	0.01
17	2017-05_Percent_Change_of_Sales_of_Existing_Detached_Homes_LA_AREA_train.csv	8.00	8.29	8.11	12.14	8.10	0.01
18	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_JUVENILES _train.csv	3.76	5.52	3.79	3.98	3.84	0.01
19	wunderground_Chicago_actual_max_temp_train.csv	10.12	10.20	24.80	24.80	33.24	0.01
20	Facebook-TotalClicks_train.csv	5787.78	6311.37	6413.69	5759.45	6002.42	-0.00
21	wunderground_Chicago_actual_precipitation_train.csv	0.35	0.34	0.35	0.35	0.35	-0.01
22	facebook_wp_with_outliers2_train.csv	1018.58	1133.07	1011.87	1092.37	2599.63	-0.01
23	facebook_data_train.csv	6.90	13.48	6.84	11.29	6.88	-0.01
24	hyndman_usa_gasoline_train.csv	148.24	145.79	180.79	194.28	197.01	-0.02
25	fpp_us_personal_income_change_train.csv	1.20	1.18	1.20	1.18	1.20	-0.02
26	fpp_us_personal_consumption_change_train.csv	0.92	0.99	0.89	0.95	0.89	-0.04
27	fpp_us_electricity_net_generation_train.csv	12.43	22.53	11.90	20.69	12.06	-0.04
28	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HOPE_VILLAGE _train.csv	3.96	4.01	3.89	3.79	4.65	-0.04
29	dr-commit-activity_train.csv	30.42	29.11	43.90	62.96	63.48	-0.04
30	Google-TotalSpend_train.csv	6354.14	6699.02	5979.69	6236.68	7452.23	-0.06
31	SP500_MSI_train.csv	2.60	2.44	4.28	4.53	4.50	-0.07
32	fpp_aus_departures_reslong_train.csv	2207.01	2497.82	2020.18	2501.77	2104.24	-0.09
33	wunderground_Chicago_actual_min_temp_train.csv	8.80	8.03	20.25	20.22	26.63	-0.10
34	weekly_earth_co2_train.csv	1.89	1.69	6.72	7.67	7.61	-0.12
35	wunderground_Chicago_actual_mean_temp_train.csv	9.49	8.42	22.33	22.34	30.55	-0.13
36	fpp_autralian_beer_train.csv	18.47	15.86	20.38	16.10	18.07	-0.16
37	hyndman_turkish_electricity_demand_train.csv	2842.54	2280.46	2684.24	2795.09	2959.85	-0.25
38	google_trends_debt_searches_predict_market_returns_train.csv	0.03	0.02	0.02	0.02	0.02	-0.28
39	SP500_URI_train.csv	4.46	3.37	4.81	4.85	4.51	-0.32
40	fpp_aus_departures_resshort_train.csv	46574.97	92087.13	44064.02	34789.91	92332.46	-0.34
41	SP500_VRSK_train.csv	3.15	4.02	2.33	3.26	3.30	-0.35
42	gdelt_wti_daily_train.csv	12.06	8.73	20.20	17.99	18.24	-0.38
43	HPI_PO_monthly_hist_PacRegion_train.csv	13.31	42.55	9.22	14.56	16.32	-0.44
44	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ FAIRVIEW _train.csv	4.59	5.63	3.17	3.32	3.24	-0.45
45	facebook_retail_train.csv	10106.93	15808.96	9441.04	8705.79	6809.95	-0.48
46	fpp_aus_departures_permanent_train.csv	843.26	1180.82	1224.17	560.70	595.26	-0.50
47	SP500_DLTR_train.csv	7.20	9.76	4.73	5.44	5.65	-0.52
48	Google-TotalImpressions_train.csv	645160.54	624725.12	593002.20	415881.48	861168.42	-0.55
49	methane-input-into-gas-furnace-c_train.csv	3.80	2.41	3.02	3.11	8.23	-0.58
50	fpp_aus_cortecosteroid_sales_train.csv	0.11	0.10	0.08	0.07	0.08	-0.64
51	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF _train.csv	33.65	52.86	21.99	20.58	19.36	-0.74
52	goog_trends_all_terms_predict_DJIA_raw_price_train.csv	1043.29	551.01	849.97	849.96	849.99	-0.89
53	Facebook-TotalImpressions_train.csv	544663.97	285606.64	427773.91	370916.37	351336.43	-0.91
54	BCHAIN-NTRAN_train.csv	70571.15	36685.60	48153.50	69216.67	84335.05	-0.92
55	fpp_iceland_debit_card_usage_train.csv	1129.71	1108.49	829.15	580.85	752.31	-0.94
56	fpp_eu_electric_equipment_train.csv	11.52	6.74	6.59	12.18	5.91	-0.95
57	Facebook-TotalSpend_train.csv	7645.03	3776.02	6542.10	5844.17	7413.37	-1.02
58	Google-TotalClicks_train.csv	8061.30	5490.38	4540.29	3878.20	4249.97	-1.08
59	fpp_aus_departures_vislong_train.csv	6122.36	6912.73	6089.28	2742.39	4405.99	-1.23
60	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_MALES _train.csv	93.46	189.19	143.82	41.33	46.49	-1.26
61	monthly_earth_co2_train.csv	3.49	1.28	1.78	2.81	1.94	-1.72
62	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HWH _train.csv	8.72	9.97	3.18	3.36	3.36	-1.75
63	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_FEMALES _train.csv	19.01	26.44	5.76	11.49	7.13	-2.30
64	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_FEMALES _train.csv	22.54	30.70	5.63	14.28	5.90	-3.00
65	fpp_aus_air_traffic_business_train.csv	1.84	1.94	0.44	0.53	0.43	-3.23
66	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC _train.csv	372.52	228.56	104.91	52.60	48.34	-6.71
67	Federer_train.csv	161.18	62.73	24.98	28.25	20.84	-6.73
68	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_MALES _train.csv	376.10	210.14	47.36	45.21	46.10	-7.32
69	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF _train.csv	400.21	185.33	125.07	45.04	49.58	-7.89
70	RecessionData_train.csv	18.46	10.78	10.81	9.19	0.46	-39.09

Table 2: Full Results