DataRobot vs. Open Source Forecasting Tools

Zachary Deane-Mayer

31 July, 2017

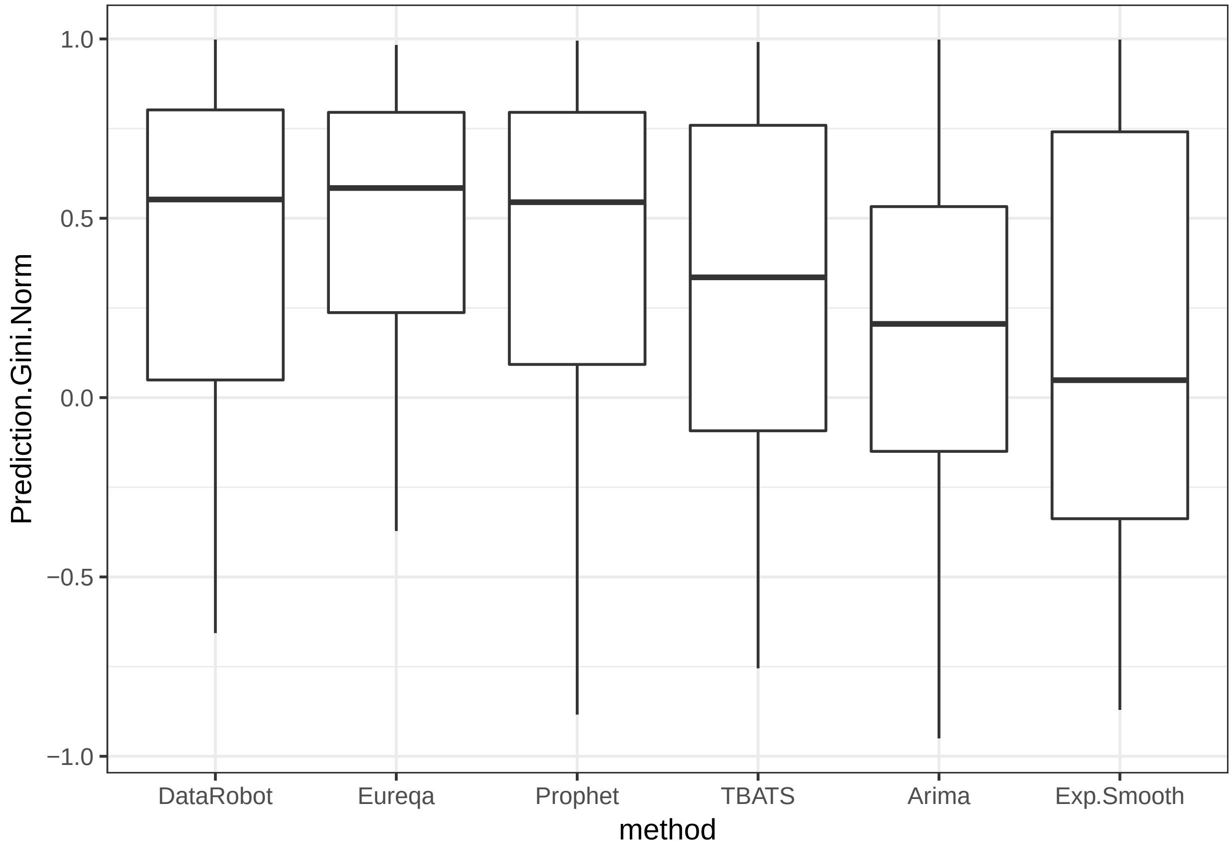
Methodology

We compared DataRobot's Out-of-Time Validation (OTV) models to Eureqa, R's Forecast package and Facebook's Prophet package on 70 datasets. We 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, we 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 Gini Score" or "Gini 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 Gini 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.

We compared DataRobot and Eureqa 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 Gini Norm on the holdout set was used to make forecasts on the prediction set.

Results



On average, Eureqa is slightly more accurate than Datarobot (median Gini of 0.58 vs 0.55). DataRobot is slightly more accurate than Facebook's Prophet package (median Gini of 0.55 vs 0.54). TBATS is the best model from the Forecast package, but is typically worse that DataRobot or Prophet (median Gini of 0.34).

	method	min_gini	pct_25_gini	median_gini	mean_gini	sd_gini	pct_75_gini	max_gini
1	DataRobot	-0.66	0.05	0.55	0.41	0.46	0.80	1.00
2	Eureqa	-0.37	0.24	0.58	0.51	0.35	0.80	0.98
3	Prophet	-0.88	0.09	0.54	0.39	0.51	0.80	1.00
4	TBATS	-0.75	-0.09	0.34	0.28	0.51	0.76	0.99
5	Arima	-0.95	-0.15	0.21	0.19	0.52	0.53	1.00
6	Exp.Smooth	-0.87	-0.34	0.05	0.12	0.60	0.74	1.00

Table 1: Summary Results

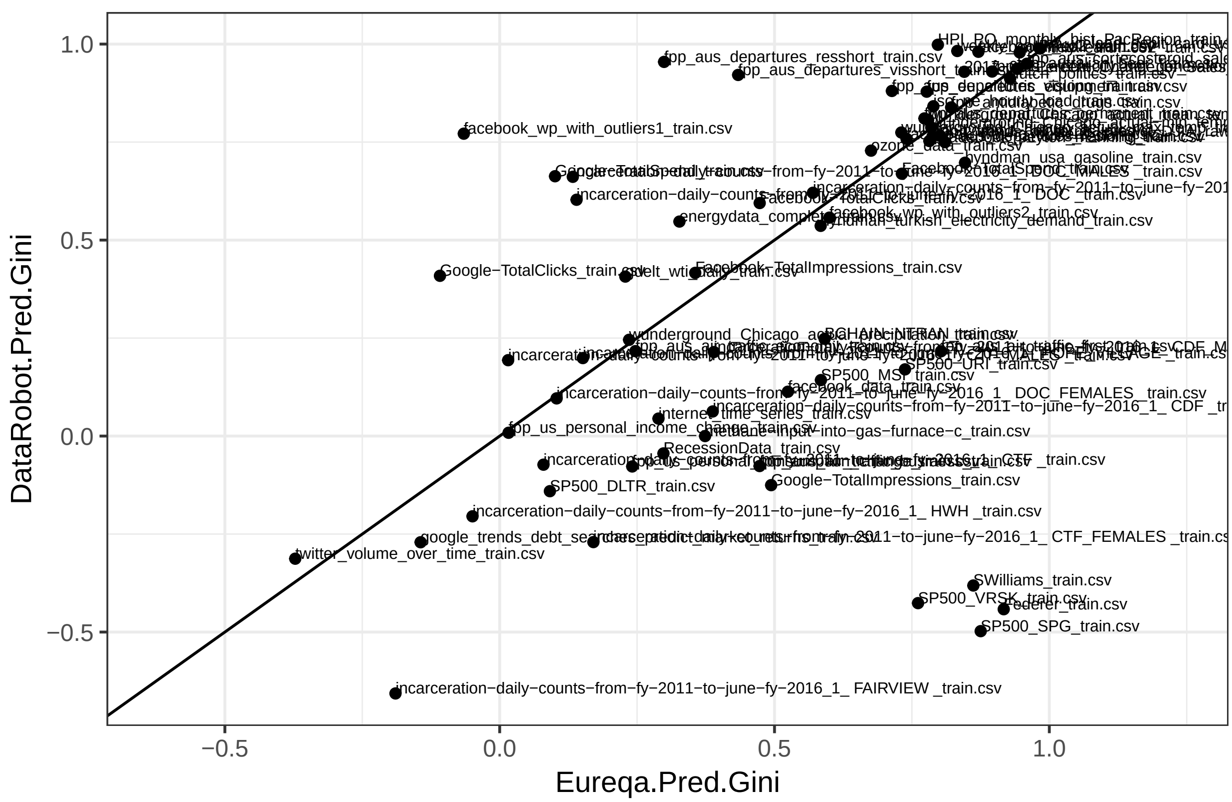
Conclusion

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.

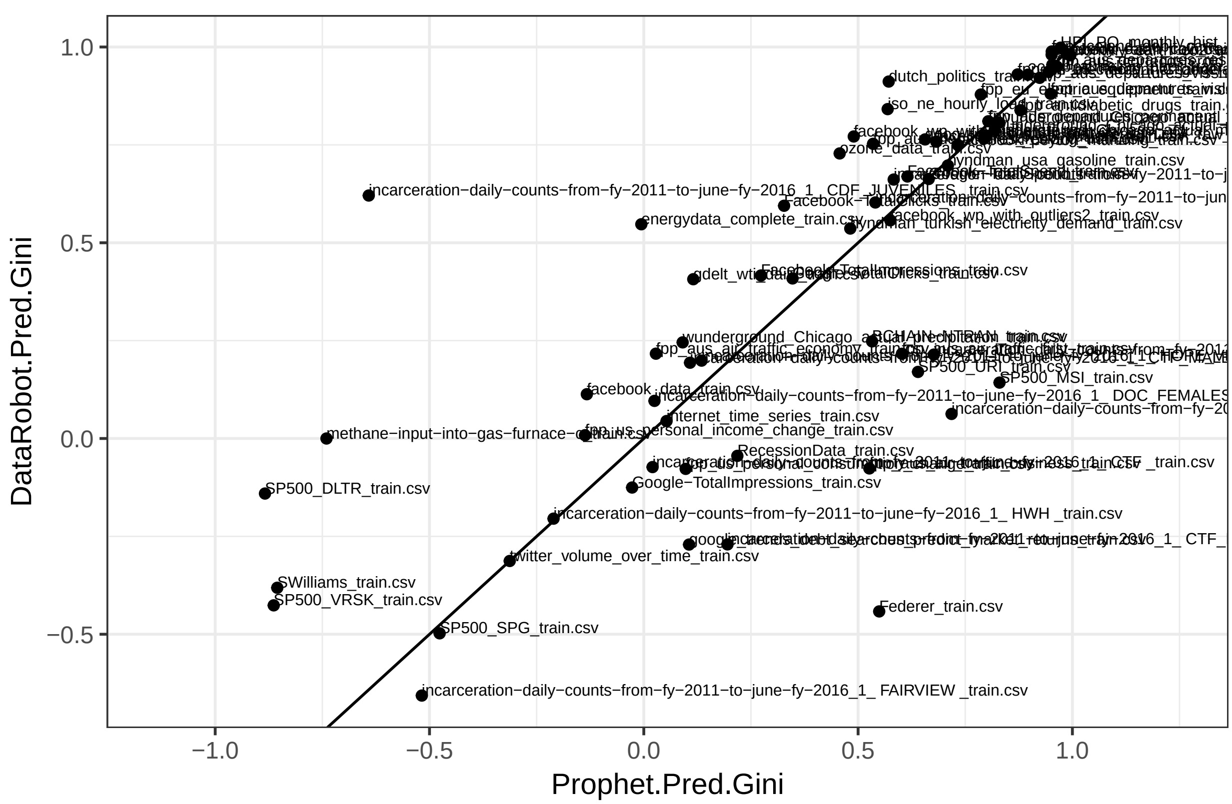
However, there remains some room for improvement in DataRobot, which can be over-confident based on the holdout set when picking the model to use for forecasting. It might be beneficial to add some heuristics to DataRobot to cause it to prefer simpler models during model selection for time series models. 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.

We could gain additional accuracy in DataRobot, especially on more difficult problems, by adding Eurea and Arima blueprints to the autopilot.

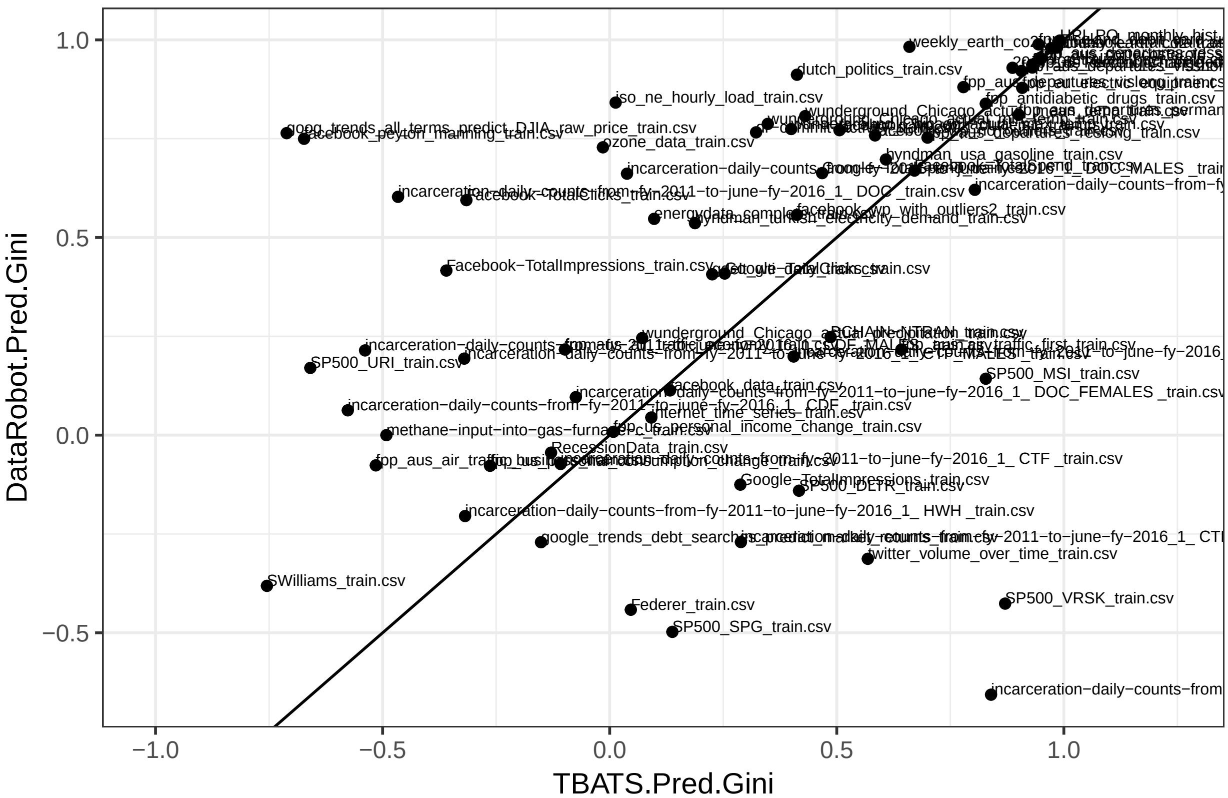
DataRobot vs Eureqa out−of−sample performance



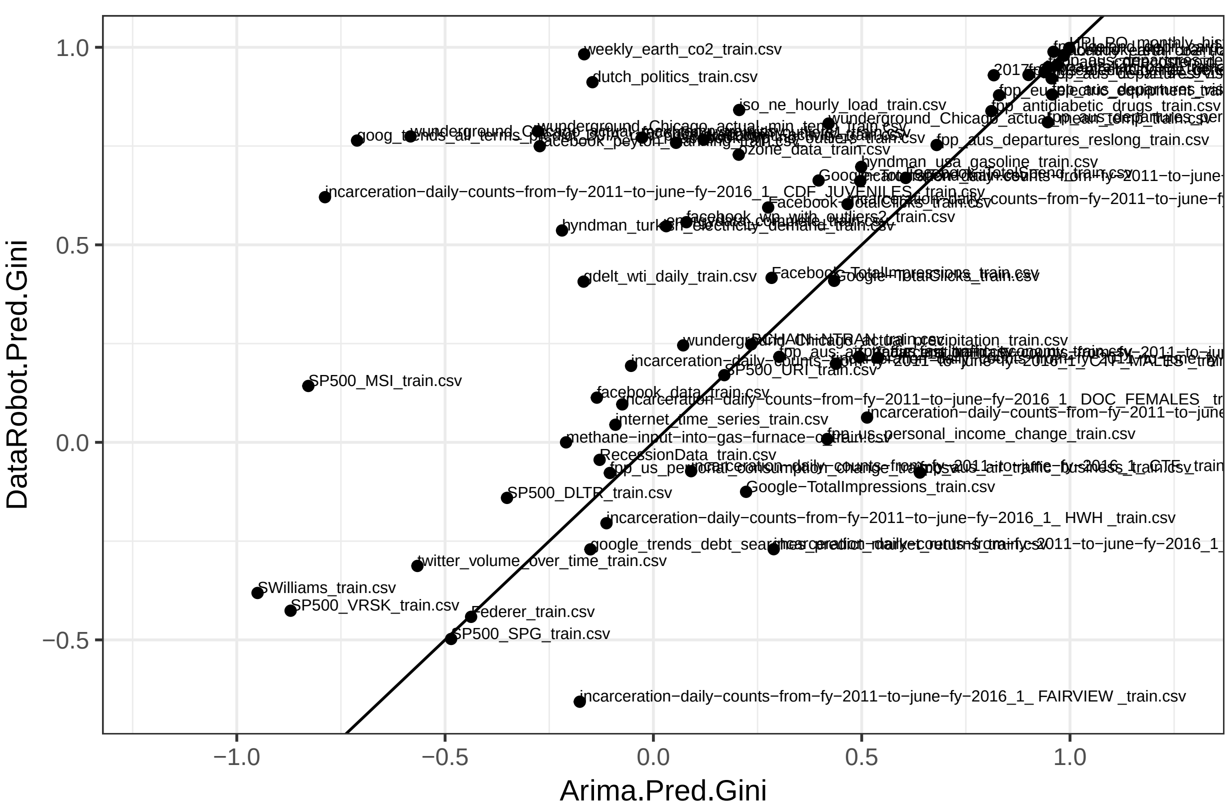
DataRobot vs Prophet out−of−sample performance

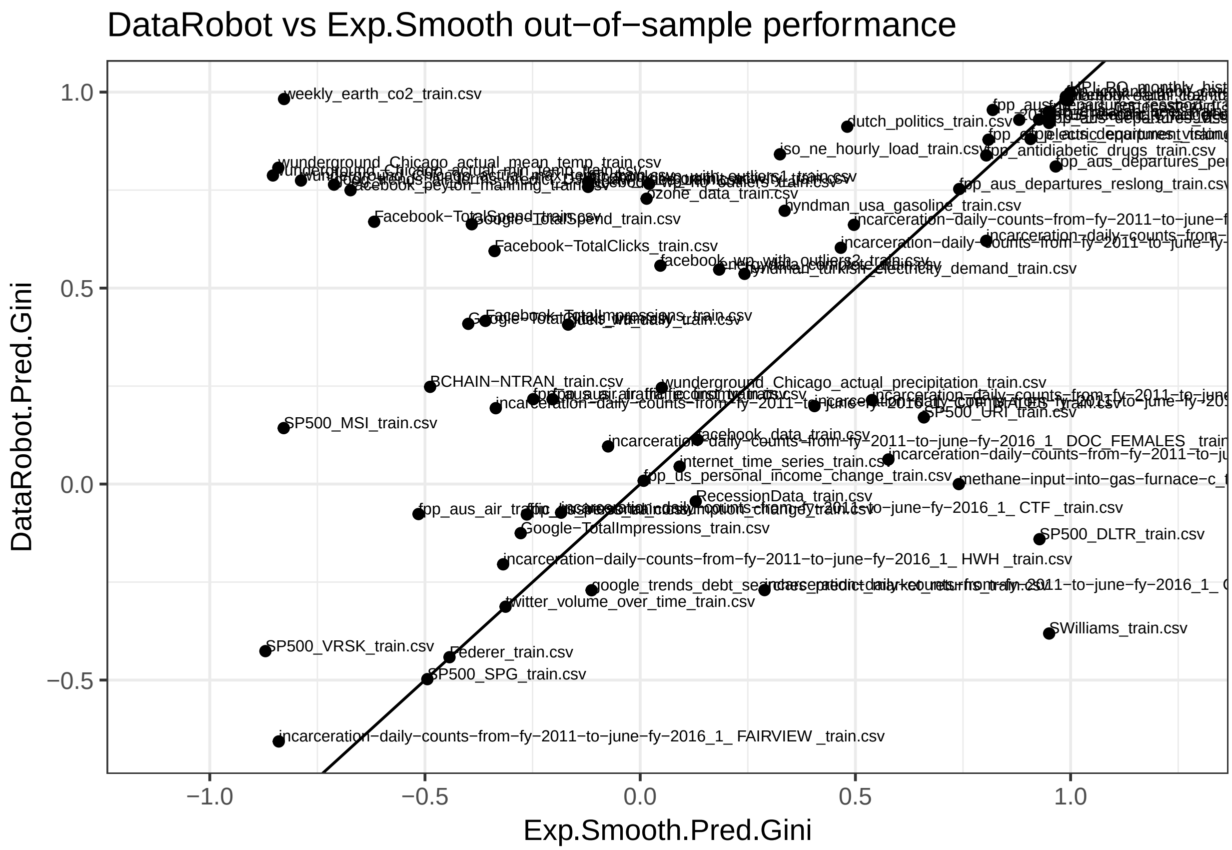


DataRobot vs TBATS out−of−sample performance



DataRobot vs Arima out−of−sample performance




DataRobot vs Exp.Smooth out−of−sample performance
DataRobot.Pred.Gini
Exp.Smooth.Pred.Gini
1.0
0.5
0.0
−0.5
−1.0
−0.5
0.0
0.5
1.0
weekly_earth_co2_train.csv
dutch_politics_train.csv
iso_ne_hourly_load_train.csv
fpp_antidiabetic_drugs_train.csv
fpp_aus_departures_reslong_train.csv
wunderground_Chicago_actual_mean_temp_train.csv
ozone_data_train.csv
hyndman_usa_gasoline_train.csv
Facebook−TotalClicks_train.csv
hyndman_turkish_electricity_demand_train.csv
BCHAIN−NTRAN_train.csv
wunderground_Chicago_actual_precipitation_train.csv
SP500_MSI_train.csv
SP500_URI_train.csv
internet_time_series_train.csv
fpp_us_personal_income_change_train.csv
RecessionData_train.csv
Google−TotalImpressions_train.csv
SP500_DLTR_train.csv
incarceration−daily−counts−from−fy−2011−to−june−fy−2016_1_ HWH _train.csv
twitter_volume_over_time_train.csv
SWilliams_train.csv
SP500_VRSK_train.csv
Pederer_train.csv
SP500_SPG_train.csv
incarceration−daily−counts−from−fy−2011−to−june−fy−2016_1_ FAIRVIEW _train.csv

	dataset	DataRobot.Pred.Gini	Eureqa.Pred.Gini	Prophet.Pred.Gini	TBATS.Pred.Gini	Arima.Pred.Gini	Exp.Smooth.Pred.Gini	DR.Min.Improve
1	facebook_wp_with_outliers1_train.csv	0.77	-0.07	0.49	0.51	-0.03	-0.12	0.27
2	energydata_complete_train.csv	0.55	0.33	-0.01	0.10	0.03	0.18	0.22
3	gdelt_wti_daily_train.csv	0.41	0.23	0.12	0.23	-0.17	-0.17	0.18
4	Facebook-TotalClicks_train.csv	0.59	0.47	0.33	-0.32	0.28	-0.34	0.12
5	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_MALES _train.csv	0.19	0.02	0.11	-0.32	-0.05	-0.34	0.09
6	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_MALES _train.csv	0.66	0.13	0.58	0.04	0.50	0.50	0.08
7	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC _train.csv	0.60	0.14	0.54	-0.47	0.47	0.47	0.06
8	Facebook-TotalImpressions_train.csv	0.42	0.36	0.27	-0.36	0.28	-0.36	0.06
9	ozone_data_train.csv	0.73	0.68	0.46	-0.01	0.20	0.01	0.05
10	iso_ne_hourly_load_train.csv	0.84	0.79	0.57	0.01	0.21	0.32	0.05
11	2017-05_Percent_Change_of_Sales_of_Existing_Detached_Homes_LA_AREA_train.csv	0.93	0.85	0.90	0.89	0.82	0.88	0.03
12	facebook_wp_no_outliers_train.csv	0.76	0.74	0.68	0.58	0.05	-0.12	0.02
13	wunderground_Chicago_actual_precipitation_train.csv	0.25	0.24	0.09	0.07	0.07	0.05	0.01
14	HPI_PO_monthly_hist_PacRegion_train.csv	1.00	0.80	0.97	0.99	1.00	1.00	0.00
15	fpp_iceland_debit_card_usage_train.csv	0.99	0.98	0.95	0.94	0.96	0.99	-0.00
16	fpp_us_electricity_net_generation_train.csv	0.93	0.90	0.87	0.93	0.90	0.93	-0.00
17	Google-TotalSpend_train.csv	0.66	0.10	0.66	0.47	0.40	-0.39	-0.00
18	weekly_earth_co2_train.csv	0.98	0.83	0.99	0.66	-0.17	-0.83	-0.00
19	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ DOC_FEMALES _train.csv	0.10	0.10	0.02	-0.07	-0.07	-0.07	-0.01
20	facebook_retail_train.csv	0.98	0.87	0.95	0.98	0.98	0.99	-0.01
21	fpp_autralian_beer_train.csv	0.94	0.94	0.94	0.93	0.94	0.95	-0.01
22	fpp_aus_cortecosteroid_sales_train.csv	0.95	0.96	0.96	0.95	0.95	0.95	-0.01
23	fpp_aus_departures_resshort_train.csv	0.95	0.30	0.95	0.95	0.97	0.82	-0.01
24	monthly_earth_co2_train.csv	0.98	0.95	1.00	0.97	0.99	0.99	-0.02
25	dutch_politics_train.csv	0.91	0.93	0.57	0.41	-0.15	0.48	-0.02
26	wunderground_Chicago_actual_max_temp_train.csv	0.77	0.73	0.80	0.40	-0.58	-0.79	-0.02
27	wunderground_Chicago_actual_mean_temp_train.csv	0.81	0.78	0.83	0.43	0.42	-0.84	-0.02
28	goog_trends_all_terms_predict_DJIA_raw_price_train.csv	0.76	0.79	0.66	-0.71	-0.71	-0.71	-0.02
29	Google-TotalClicks_train.csv	0.41	-0.11	0.35	0.25	0.43	-0.40	-0.03
30	fpp_aus_departures_reslong_train.csv	0.75	0.78	0.53	0.70	0.68	0.74	-0.03
31	fpp_eu_electric_equipment_train.csv	0.88	0.78	0.79	0.91	0.83	0.81	-0.03
32	wunderground_Chicago_actual_min_temp_train.csv	0.79	0.79	0.82	0.35	-0.28	-0.85	-0.03
33	fpp_aus_departures_visshort_train.csv	0.92	0.43	0.92	0.91	0.96	0.95	-0.03
34	dr-commit-activity_train.csv	0.77	0.80	0.79	0.32	0.12	0.02	-0.04
35	fpp_antidiabetic_drugs_train.csv	0.84	0.82	0.88	0.83	0.81	0.80	-0.04
36	facebook_wp_with_outliers2_train.csv	0.56	0.60	0.58	0.41	0.08	0.05	-0.04
37	hyndman_turkish_electricity_demand_train.csv	0.54	0.58	0.48	0.19	-0.22	0.24	-0.05
38	facebook_peyton_manning_train.csv	0.75	0.81	0.73	-0.67	-0.27	-0.67	-0.06
39	Facebook-TotalSpend_train.csv	0.67	0.73	0.62	0.67	0.61	-0.62	-0.06
40	fpp_aus_departures_vislong_train.csv	0.88	0.71	0.95	0.78	0.96	0.91	-0.08
41	hyndman_usa_gasoline_train.csv	0.70	0.85	0.71	0.61	0.50	0.34	-0.15
42	fpp_aus_departures_permanent_train.csv	0.81	0.77	0.80	0.90	0.95	0.97	-0.16
43	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HWH _train.csv	-0.20	-0.05	-0.21	-0.32	-0.11	-0.32	-0.16
44	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF _train.csv	-0.07	0.08	0.02	-0.11	0.09	-0.18	-0.16
45	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_JUVENILES _train.csv	0.62	0.57	-0.64	0.80	-0.79	0.80	-0.18
46	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ HOPE_VILLAGE _train.csv	0.20	0.15	0.13	0.40	0.44	0.40	-0.24
47	internet_time_series_train.csv	0.04	0.29	0.05	0.09	-0.09	0.09	-0.24
48	fpp_aus_air_traffic_economy_train.csv	0.22	0.25	0.03	-0.10	0.49	-0.25	-0.28
49	fpp_us_personal_consumption_change_train.csv	-0.08	0.24	0.10	-0.26	-0.10	-0.26	-0.32
50	RecessionData_train.csv	-0.04	0.30	0.22	-0.13	-0.13	0.13	-0.34
51	BCHAIN-NTRAN_train.csv	0.25	0.59	0.53	0.49	0.23	-0.49	-0.34
52	google_trends_debt_searches_predict_market_returns_train.csv	-0.27	-0.14	0.11	-0.15	-0.15	-0.11	-0.38
53	fpp_us_personal_income_change_train.csv	0.01	0.02	-0.14	0.01	0.42	0.01	-0.41
54	facebook_data_train.csv	0.11	0.52	-0.13	0.13	-0.14	0.13	-0.41
55	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF_MALES _train.csv	0.21	0.39	0.68	-0.54	0.54	0.54	-0.46
56	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CTF_FEMALES _train.csv	-0.27	0.17	0.20	0.29	0.29	0.29	-0.56
57	SP500_URI_train.csv	0.17	0.74	0.64	-0.66	0.17	0.66	-0.57
58	fpp_aus_air_traffic_first_train.csv	0.22	0.80	0.60	0.64	0.30	-0.20	-0.59
59	Google-TotalImpressions_train.csv	-0.13	0.49	-0.03	0.29	0.22	-0.28	-0.62
60	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ CDF _train.csv	0.06	0.39	0.72	-0.58	0.51	0.58	-0.66
61	SP500_MSI_train.csv	0.14	0.58	0.83	0.83	-0.83	-0.83	-0.69
62	fpp_aus_air_traffic_business_train.csv	-0.08	0.47	0.53	-0.51	0.64	-0.51	-0.72
63	methane-input-into-gas-furnace-c_train.csv	0.00	0.37	-0.74	-0.49	-0.21	0.74	-0.74
64	twitter_volume_over_time_train.csv	-0.31	-0.37	-0.31	0.57	-0.57	-0.31	-0.88
65	SP500_DLTR_train.csv	-0.14	0.09	-0.88	0.42	-0.35	0.93	-1.07
66	SP500_VRSK_train.csv	-0.43	0.76	-0.86	0.87	-0.87	-0.87	-1.30
67	SWilliams_train.csv	-0.38	0.86	-0.86	-0.75	-0.95	0.95	-1.33
68	Federer_train.csv	-0.44	0.92	0.55	0.05	-0.44	-0.44	-1.36
69	SP500_SPG_train.csv	-0.50	0.88	-0.48	0.14	-0.49	-0.49	-1.37
70	incarceration-daily-counts-from-fy-2011-to-june-fy-2016_1_ FAIRVIEW _train.csv	-0.66	-0.19	-0.52	0.84	-0.18	-0.84	-1.50

Table 2: Full Results