Machine Learning Based Restaurant Sales Forecasting

(Supplementary Document)

Austin Schmidt[1], Md Wasi Ul Kabir[1], and Md Tamjidul Hoque[1,*]

[1]Department of Computer Science, University of New Orleans, New Orleans, LA 70148, USA; {sbaustin, mkabir4, thoque}@uno.edu

*Correspondence: thoque@uno.edu; Tel.: +1-504-280-2406

1. Holidays

Supplementary Table S1: Full List of Included Holidays. A holiday feature uses the index as a numerical identifier. Indexes labeled Single Feature roll for variable lengths of time, so a Boolean field is used separately for them. The holidays are an exhaustive list of any day which may affect sales. The restaurant location has a diverse population, so holidays from many cultures are included. The date of the holiday is included to show which dates are stable and which change yearly. In total, 28 events are considered in the Holiday feature, with four additional as their own feature.

Index	Holiday	Date
0	None	Any other day
1	New Year's Day	January 1st
2	Martin Luther King Jr. Day	3rd Monday of January
3	President's Day	3rd Monday in February
4	Super Bowl Sunday	First Sunday in February
5	Valentine's Day	February 14th
Single Feature	Carnival Season	January 6th to Mardi Gras Day
6	Mardi Gras Day	Day before Ash Wednesday
7	Ash Wednesday	46 Days before Easter Sunday. Between Feb 4th and March 22nd.
Single Feature	Lent Fasting	Starts Ash Wednesday and end 40 days later. Fasting stops on Sundays and St. Patrick's Day
8	Holika Dahan	Day before Holi
9	Holi	Early to mid-March
10	Saint Patrick's Day	March 17th
11	Easter Sunday	Sunday following the full moon. Late March to mid-April

12	Good Friday	Friday following Easter
13	Cinco de Mayo	May 5th
14	Mother's Day	2nd Sunday in May
15	Memorial Day	Last Monday in May
16	Father's Day	3rd Sunday in June
17	Independence Day	July 4th
18	Labor Day	1st Monday in September
19	Columbus Day	2nd Monday in October
20	Diwali	Late October to Early November. 5 Day celebration where day 3 is the largest.
21	Halloween	October 31st
22	Veteran's Day	November 11th
23	Thanksgiving Day	4th Thursday in November
27	Christmas Day	December 25th
25	New Year's Eve	December 31st
Single Feature	Ramadan	One full month out of the year.
26	Eid al Fitr	End of Ramadan
28	Christmas Eve	December 24th
Single Feature	Christmas Season	December 1st to December 25th

2. Model Specifics: Layer Structures & Tuning Parameters

Supplementary Table S2: Ensemble learning committee for Stacking Regression. The meta predictor is bolded and listed first, with the committee following. Notice the meta predictor is also a member of the committee.

Models
Ridge Regression
Multi-Layer Perceptron
K-Neighbors Regression
Linear Regression
Extra Trees Regression
Ridge Regression

Supplementary Table S3: Ensemble learning committee for Voting Regression. Voting has no meta regressor, and so all of the models share equal weight in decision making.

Models
Gradient Boosting Regression
Random Forest Regression
Linear Regression
Decision Tree Regression
Stochastic Gradient Decent Regression

1. Dimension Expansion
2. Simple RNN Layer (Number of nodes is 8, and we return sequences)
3. Dropout of 0.4
4. Simple RNN (Number of nodes is 4)
5. Dropout of 0.4
6. Dense Layer (1)
7. Re-Scale for Presentation (X_{New}=X*100)

Supplementary Figure S1: RNN Layer Structure. The implemented RNN structure for all tests.

1. Dimension Expansion
2. Convolutional 1D Layer (32 filters, kernel size 5, stride of 1)
3. Dropout of 0.4
4. LSTM Layer (Number of nodes is 32, and we return sequences)
5. Dropout of 0.4
6. LSTM Layer (Number of nodes is 32)
7. Dropout of 0.4
8. Dense Layer (1)
9. Re-Scale for Presentation (X_{New}=X*100)

Supplementary Figure S2: LSTM Layer Structure. The implemented LSTM structure for all tests.

1. Dimension Expansion
2. GRU Layer (Number of nodes is 32, and we return sequences)
3. Dropout of 0.4
4. GRU (Number of nodes is 32)
5. Dropout of 0.4
6. Dense Layer (1)
7. Re-Scale for Presentation (X_{New}=X*100)

Supplementary Figure S3: GRU Layer Structure. The implemented GRU structure for all tests.

1. Dimension Expansion
2. Convolutional 1D LayOer (32 filters, kernel size 5, stride of 1)
3. Dropout of 0.4
4. GRU Layer (Number of nodes is 32, and we return sequences)
5. Dropout of 0.4
6. GRU Layer (Number of nodes is 32)
7. Dropout of 0.4
8. Dense Layer (14)
9. Dense Layer (1)
10. Re-Scale for Presentation (X_{New}=X*100)

Supplementary Figure S4: GRU+ Layer Structure. The implemented GRU+ structure for all tests.

1. Quantile Loss
2. Logging Metrics
3. Input Embeddings
4. Prescalers
5. Static Variable Selection
6. Encoder Variable Selection
7. Decoder Variable Selection
8. Static Context Variable Selection
9. Static Context Initial Hidden LSTM
10. Static Context Initial Cell LSTM
11. Static Context Enrichment
12. LSTM Encoder
13. LSTM Decoder
14. Post-LSTM Gate Encoder
15. Post LSTM Add Norm Encoder
16. Static Enrichment
17. Multihead Attention
18. Post-Attention Gate Norm
19. Pre-Output Layer Norm
20. Output Layer

Supplementary Figure S5: General TFT Model Structure. We build six TFT models, but all use this same layer structure. Some layers are static and have no trainable parameters, while the rest will have a variable number of parameters defined by a hyperparameter tuning process.

Supplementary Table S4: TFT Hyperparameter Tuning Ranges. A grid search is completed on learning rate, gradient clipping, dropout, hidden size, hidden continuous size, and attention head size to determine the best possible parameters.

Parameter	Min Value	Max Value
Learning Rate	0.001	0.1
Gradient Clip	0.01	1
Dropout	0.1	0.3
Hidden Size	8	128
Hidden Continuous Size	8	128
Attention Head Size	1	4

Supplementary Table S5: Hyperparameter Tuning One-Day Results (Full Feature Set). Shown are the results for one-day forecasting from all three datasets using the full feature set. In addition, the best result's loss and the number of trainable parameters are included.

Parameter	Actual	Daily Difference	Weekly Difference
Learning Rate	0.097	0.084	0.066
Gradient Clip	0.529	0.044	0.020
Dropout	0.161	0.145	0.157
Hidden Size	93	101	22
Hidden Continuous Size	16	46	14
Attention Head Size	1	1	4
Loss	71.84	73.32	90.26
Trainable Parameters	1.1M	2.8M	229K

Supplementary Table S6: Hyperparameter Tuning One-Day Results (Reduced Feature Set). Shown are the results for one-day forecasting from all three datasets using the reduced feature set. Tuning results using the ranges defined in Supplementary Table S12. In addition, the best result's loss, the number of features in the reduced set, and the number of trainable parameters are included.

Parameter	Actual	Daily Difference	Weekly Difference
Learning Rate	0.099	0.098	0.079
Gradient Clip	0.193	0.792	0.075
Dropout	0.105	0.289	0.215
Hidden Size	123	116	121
Hidden Continuous Size	122	109	32
Attention Head Size	3	2	1
Loss	68.39	78.19	93.22
# Features	17	13	9
Trainable Parameters	3.0M	2.1M	968K

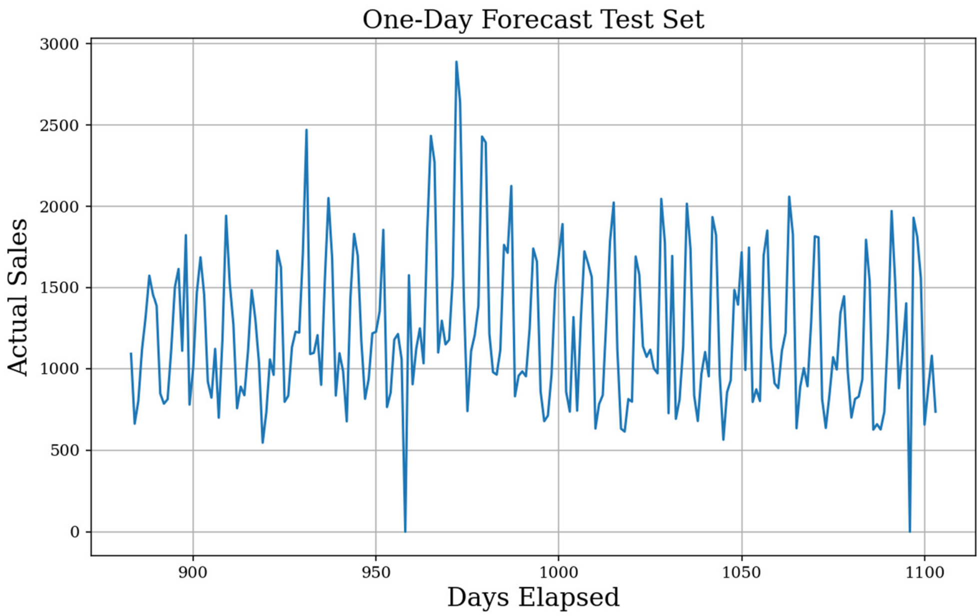
Supplementary Table S7: Hyperparameter Tuning One-Week Results (Full Feature Set). Shown are the results for one-week forecasting from all three datasets using the full feature set. Tuning results using the ranges defined in Supplementary Table S12. In addition, the best result's loss and the number of trainable parameters are included.

Parameter	Actual	Daily Difference	Weekly Difference
Learning Rate	0.098	0.099	0.087
Gradient Clip	0.651	0.060	0.035
Dropout	0.151	0.270	0.236
Hidden Size	119	128	127
Hidden Continuous Size	9	13	106
Attention Head Size	2	3	1
Loss	78.70	100.93	90.41
Trainable Parameters	1.2M	3.8M	7.9M

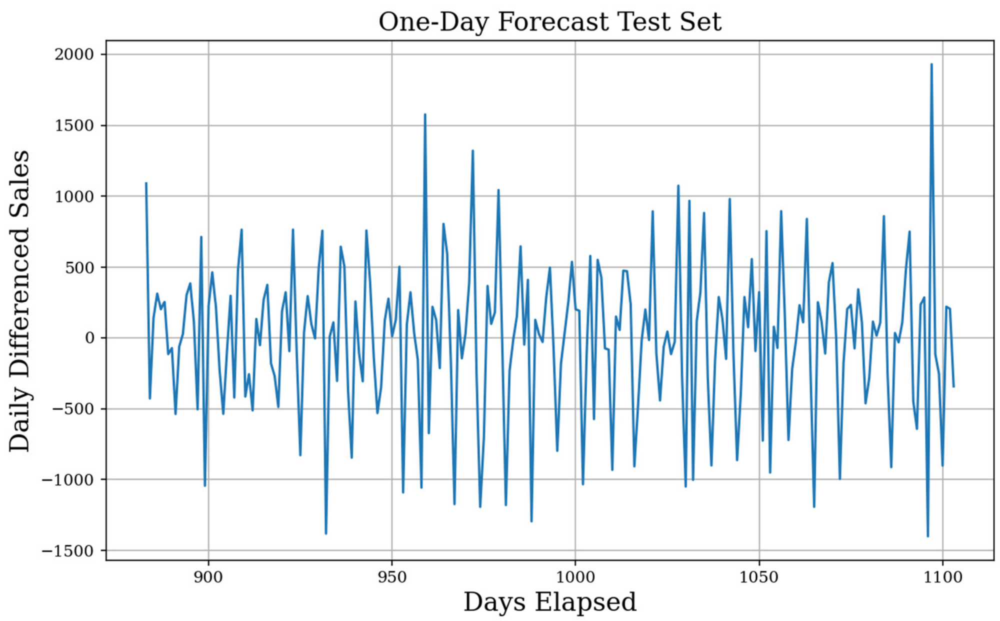
Supplementary Table S8: Hyperparameter Tuning One-Week Results (Reduced Feature Set). Shown are the results for one-week forecasting from all three datasets using the reduced feature set. Tuning results using the ranges defined in Supplementary Table S12. In addition, the best result's loss, the number of features in the reduced set, and the number of trainable parameters are included.

Parameter	Actual	Daily Difference	Weekly Difference
Learning Rate	0.098	0.099	0.093
Gradient Clip	0.911	0.574	0.173
Dropout	0.248	0.261	0.217
Hidden Size	122	123	123
Hidden Continuous Size	121	68	21
Attention Head Size	2	4	4
Loss	67.24	92.68	95.51
# Features	17	13	9
Trainable Parameters	3.1M	1.5M	896K

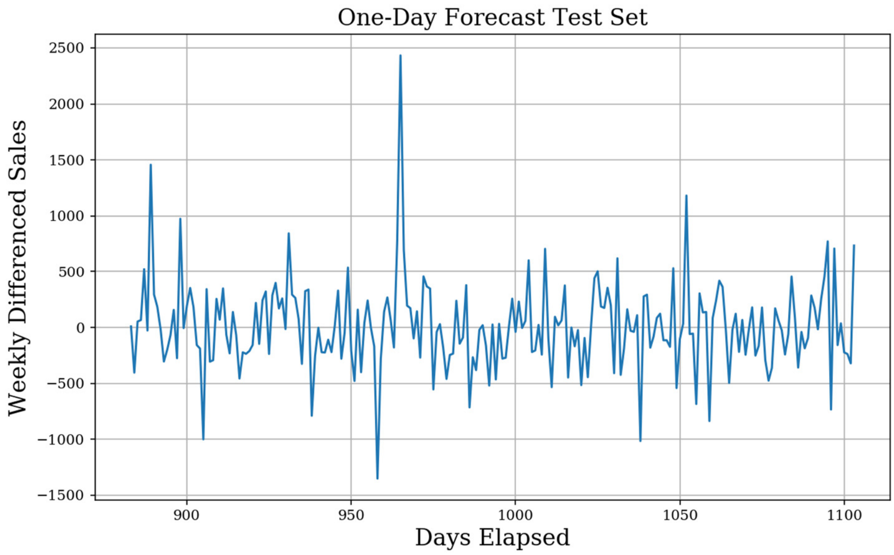
3. Forecast Test Shapes



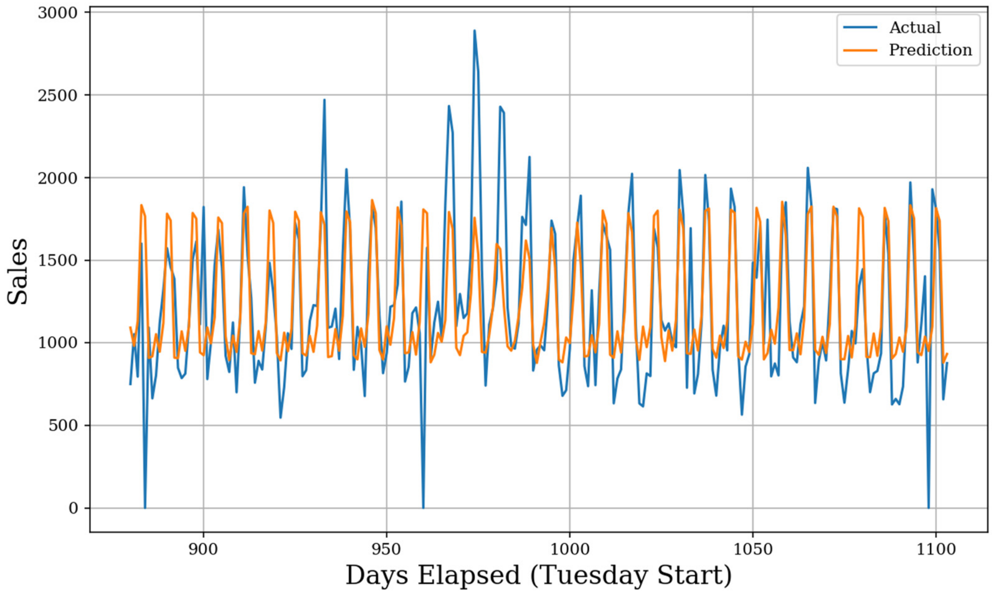
Supplementary Figure S6 – Shape of Test Data: Actual. The shape of our test dataset was used for all one-day actual forecasting test tasks. Included are 221 days of sales from the very end of the full dataset.



Supplementary Figure S7: Shape of Test Data: Daily Difference. The shape of our test dataset used for all one-day daily differenced forecasting test tasks. The test set includes 221 days of sales differenced (2) from the very end of the full dataset. Once a forecast is made, we easily transform it back to Figure S6.

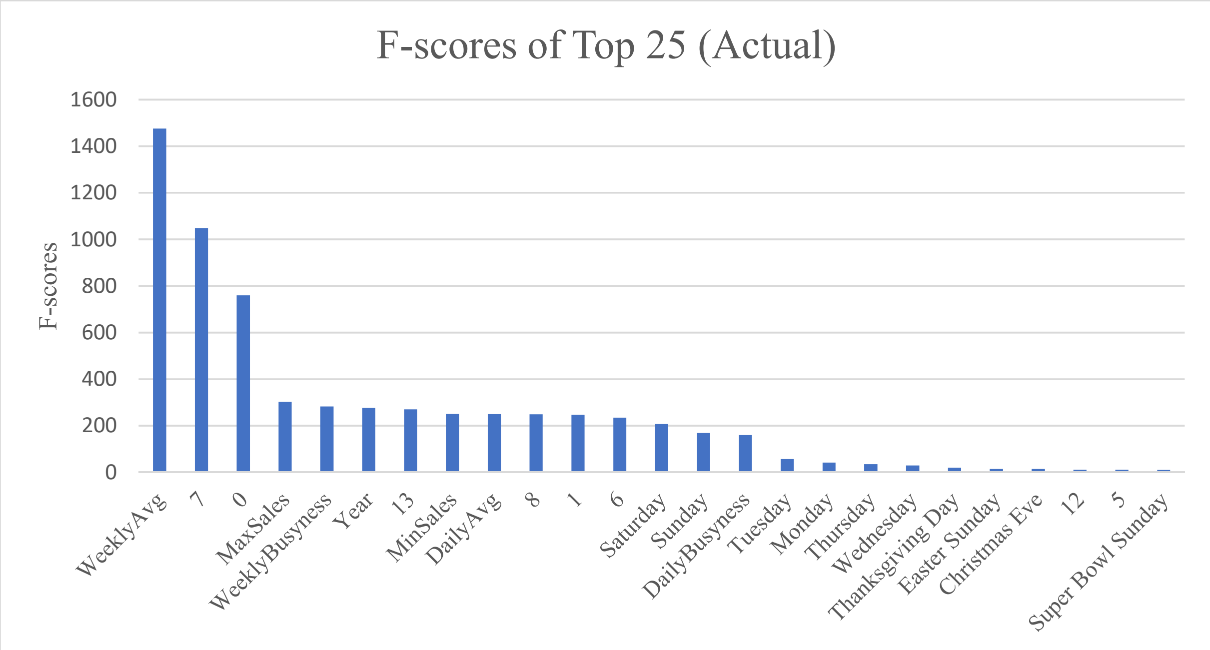


Supplementary Figure S8: Shape of Test Data: Weekly Difference. The shape of our test dataset was used for all one-day weekly differenced forecasting test tasks. The test set includes 221 days of sales differenced (3) from the very end of the full dataset. Once a forecast is made, we easily transform it back to Figure S6.

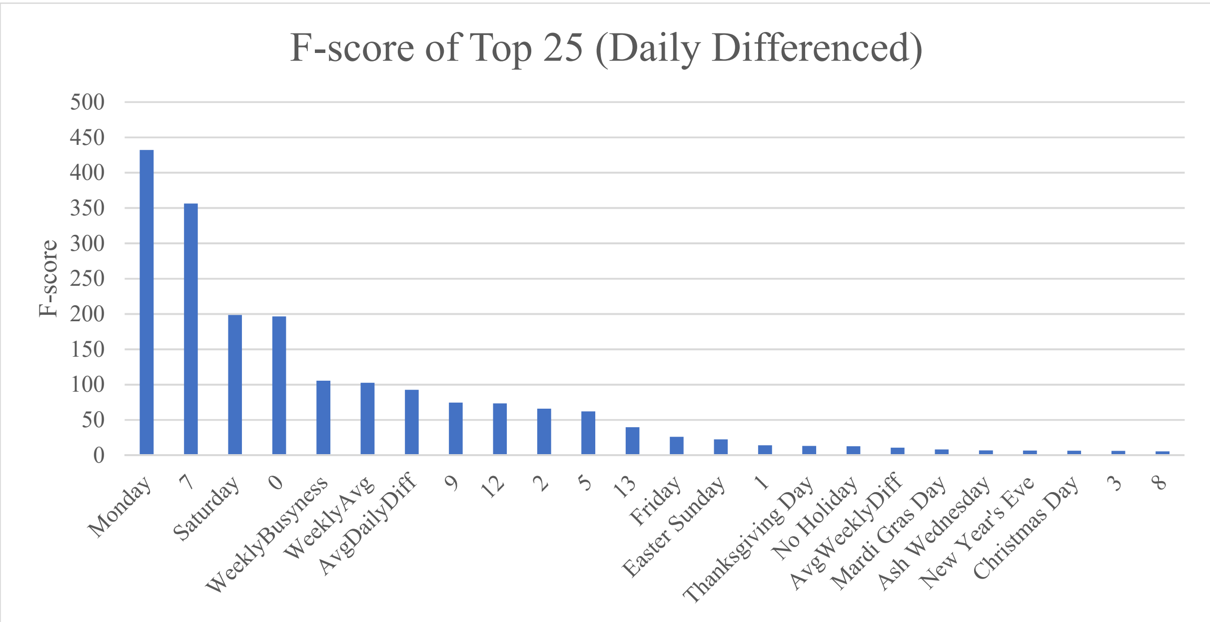


Supplementary Figure S9: One-Week Test Set (Start Day Tuesday). The one-week forecasting test set uses a sliding weekday start window. Each day of the week is used as the start for seven tests. The test sets are increased by three to a total of 224 instances, which allows for an exact number of one-week predictions. An additional zero sales holiday is shown at the beginning. Predictions starting on Tuesday are shown as an example.

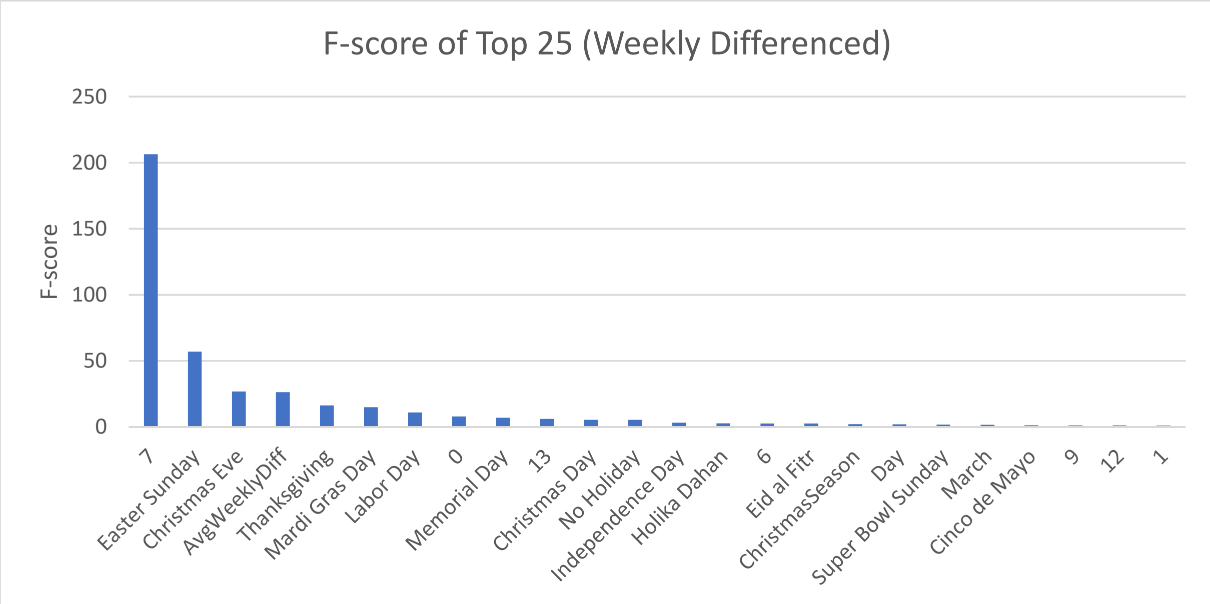
4. Feature Ranking and Feature Test Results



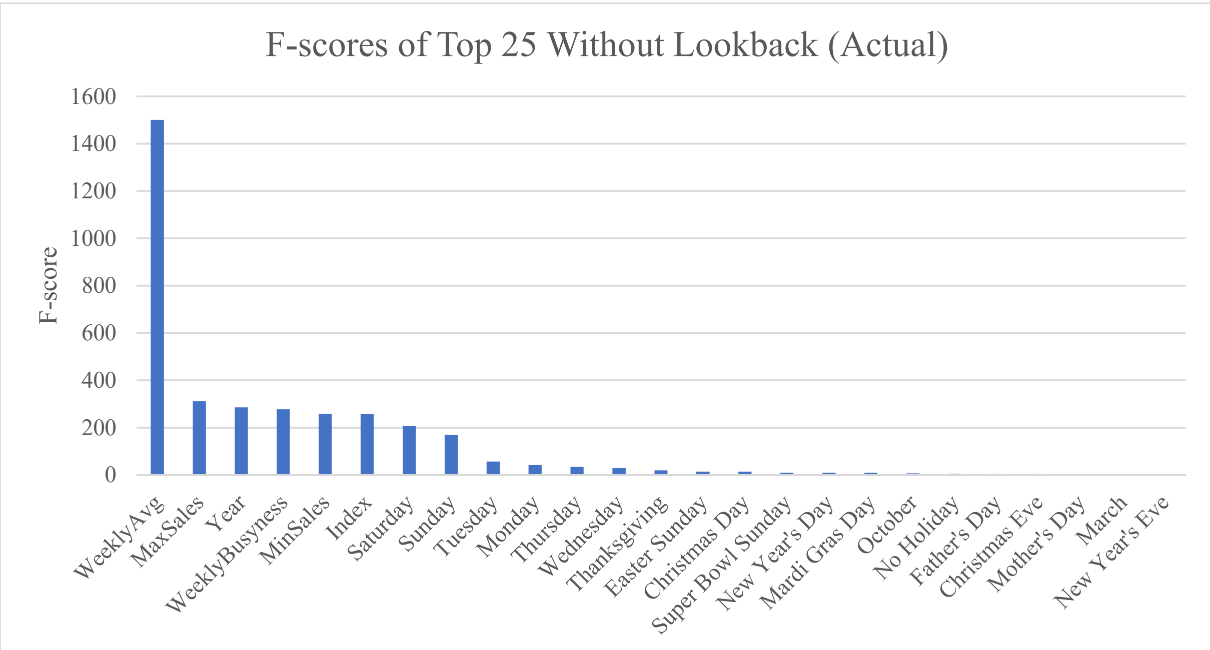
Supplementary Figure S10: F-score for Top Features (Actual). The top 25 features as ranked by their F-scores. Weekly sales average is the highest scoring feature by far with other statistical metrics and days of the week following.



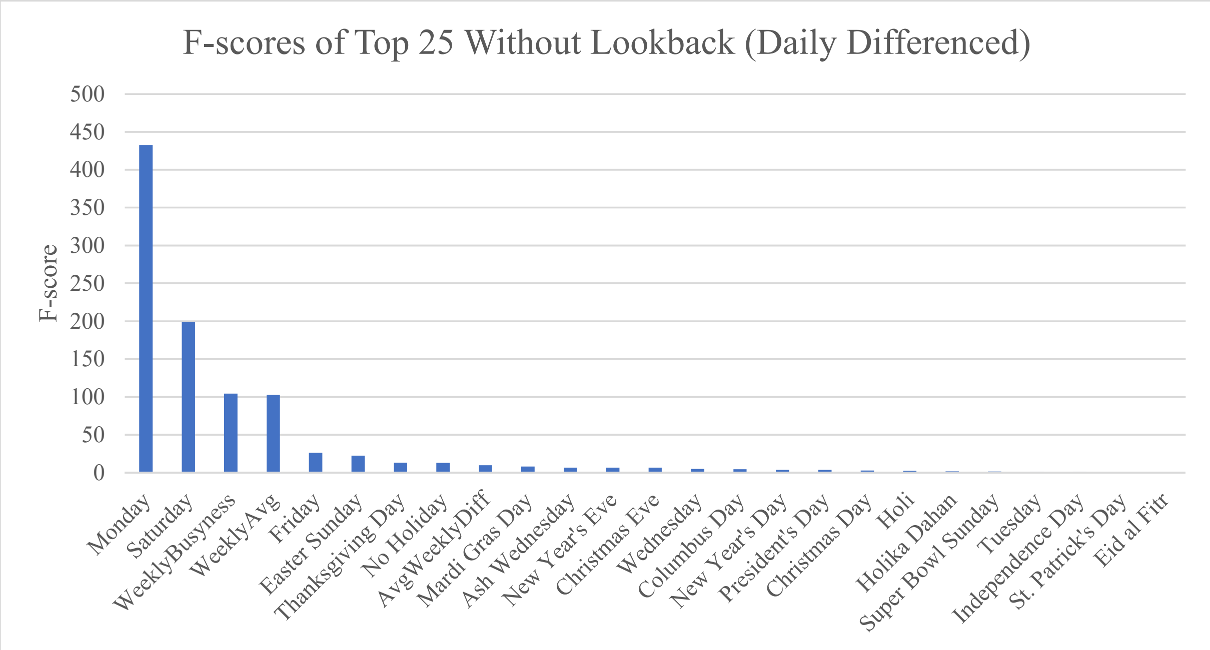
Supplementary Figure S11: F-score for Top Features (Daily Differenced). The top 25 features as ranked by their F-scores. The day of the week and some statistical measures are high-ranking features here. Some impactful holidays also are in the top 25.



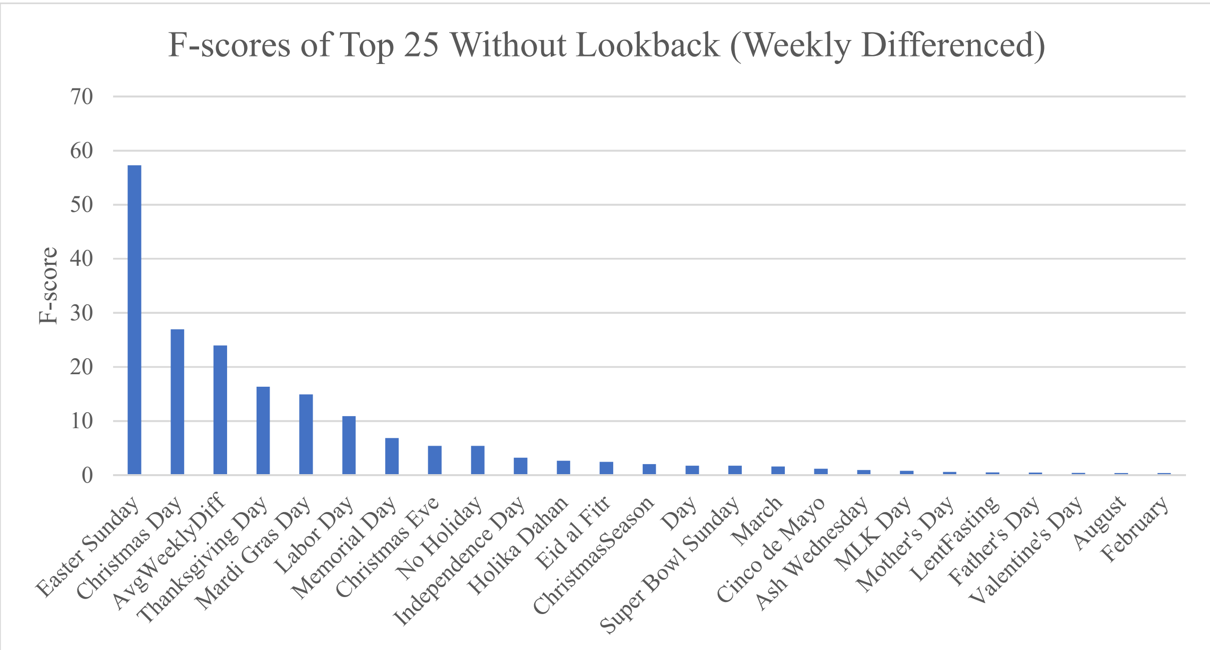
Supplementary Figure S12: F-score for Top Features (Weekly Differenced). The top 25 features as ranked by their F-scores. Only the previous week's sales have an F-score above 100. Only one statistical feature remains at the top, and the most relevant features are holidays.



Supplementary Figure S13: F-score for Transformer Top Features (Actual). The top 25 features as ranked by their F-scores. With temporal associations built by the model, feature selection can be completed without considering lookback days specifically.

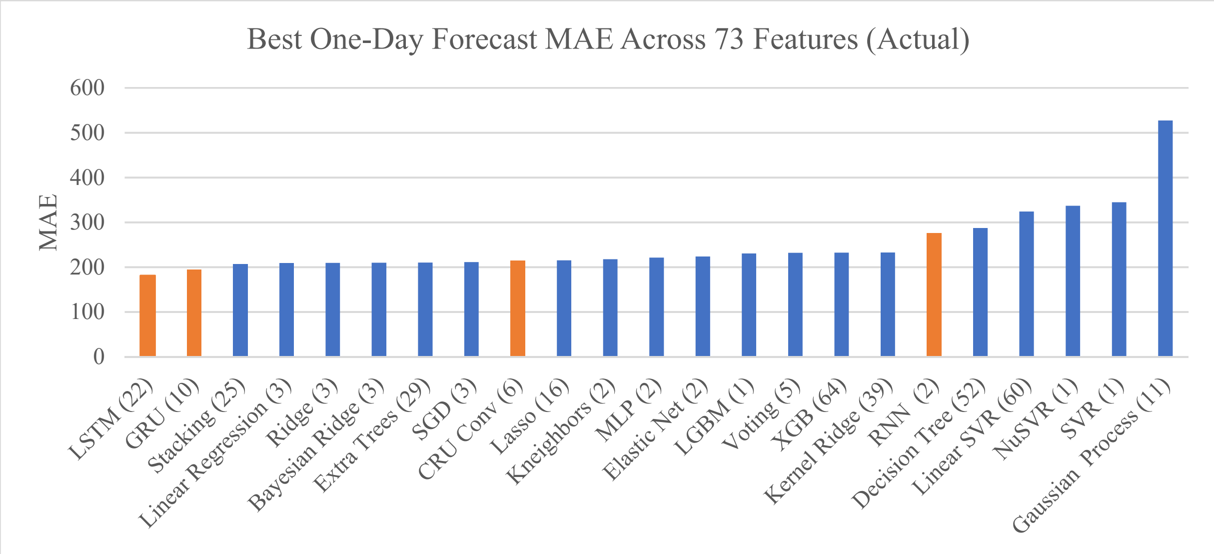


Supplementary Figure S14: F-score for Transformer Top Features (Daily Difference). The top 25 features as ranked by their F-scores. With temporal associations built by the model, feature selection can be completed without considering lookback days specifically.

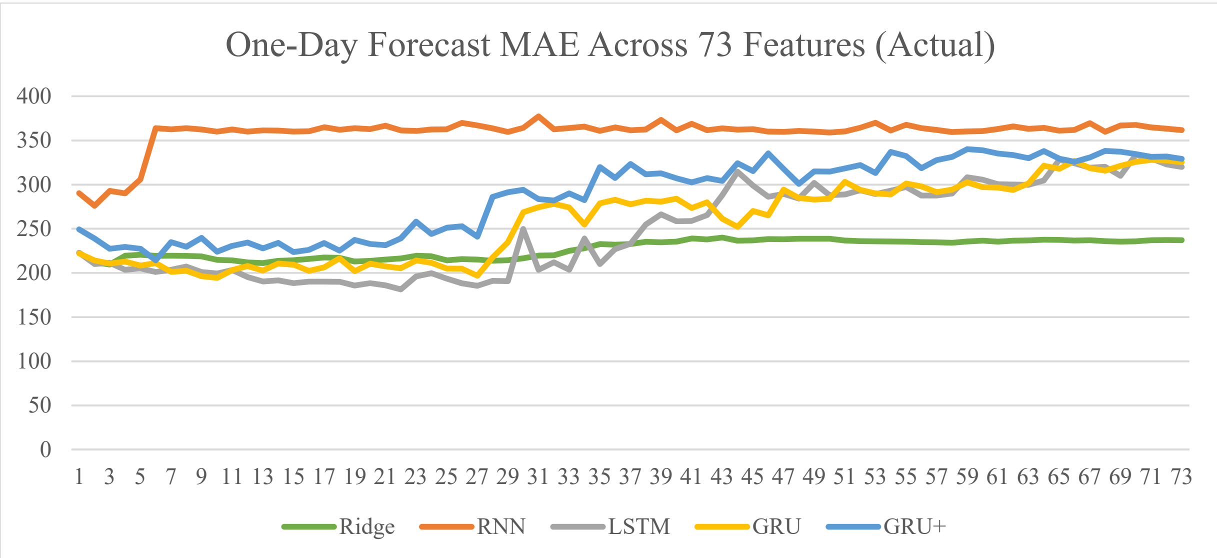


Supplementary Figure S15: F-score for Transformer Top Features (Weekly Difference). The top 25 features as ranked by their F-scores. With temporal associations built by the model, feature selection can be completed without considering lookback days specifically.

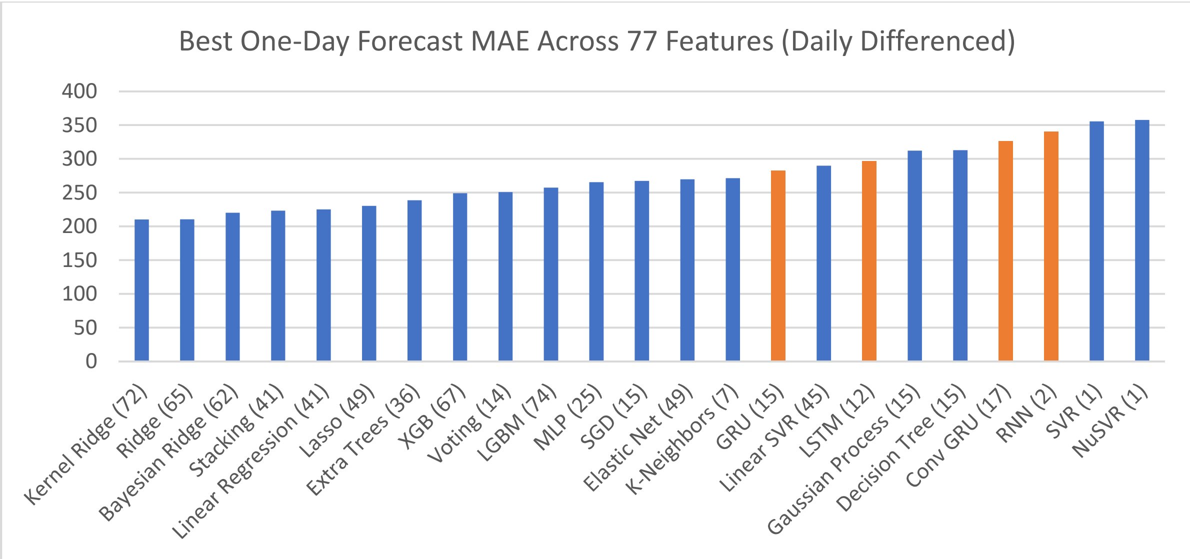
4.1 One-Day Feature Test Results



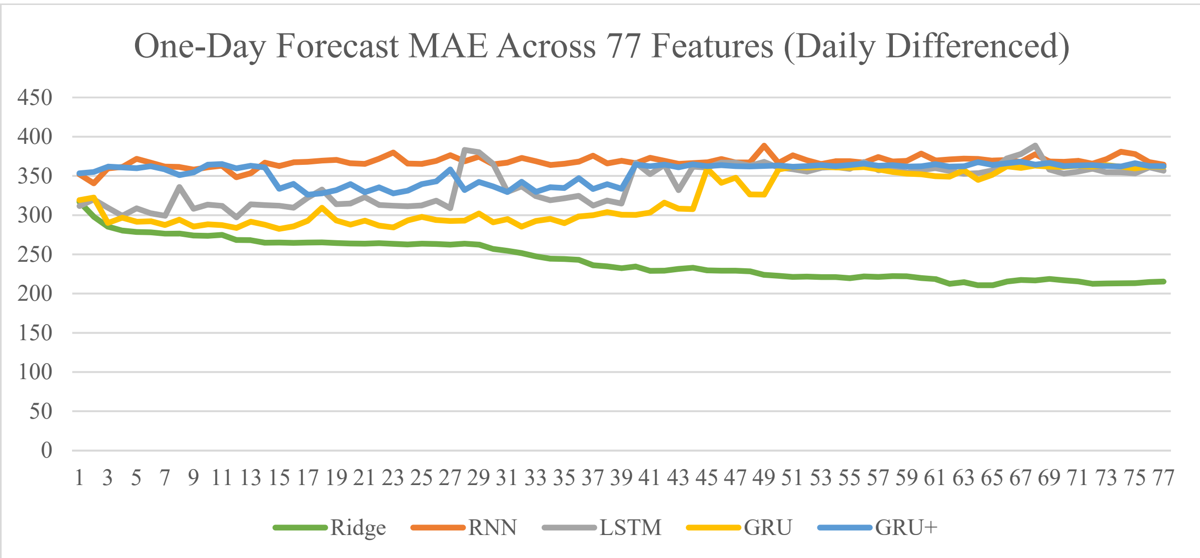
Supplementary Figure S16: Best One-Day Forecast MAE Found Across 73 Features (Actual). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S10, for one-day forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name.



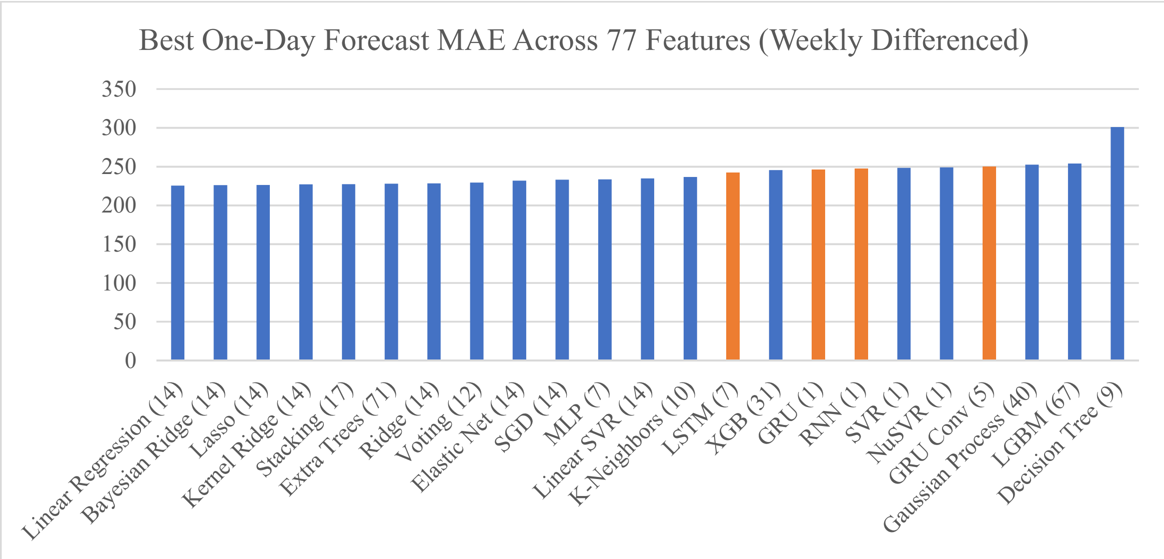
Supplementary Figure S17: All RNN Models and Ridge One-Day Forecast MAE Across 73 Features (Actual). We show how the number of features affects the MAE score for one-day forecasting in the actual dataset.



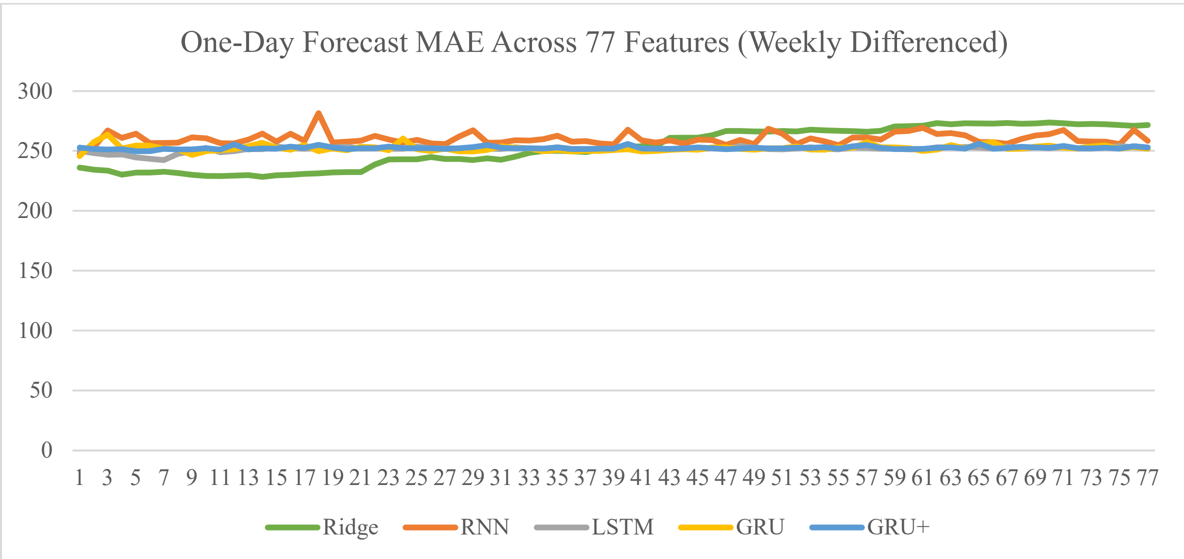
Supplementary Figure S18: Best One-Day Forecast MAE Found Across 77 Features (Daily Difference). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S11, for one-day forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name.



Supplementary Figure S19: All RNN Models and Ridge One-Day Forecast MAE Across 77 Features (Daily Difference). We show how the number of features affects the MAE score for one-day forecasting in the daily differenced dataset.

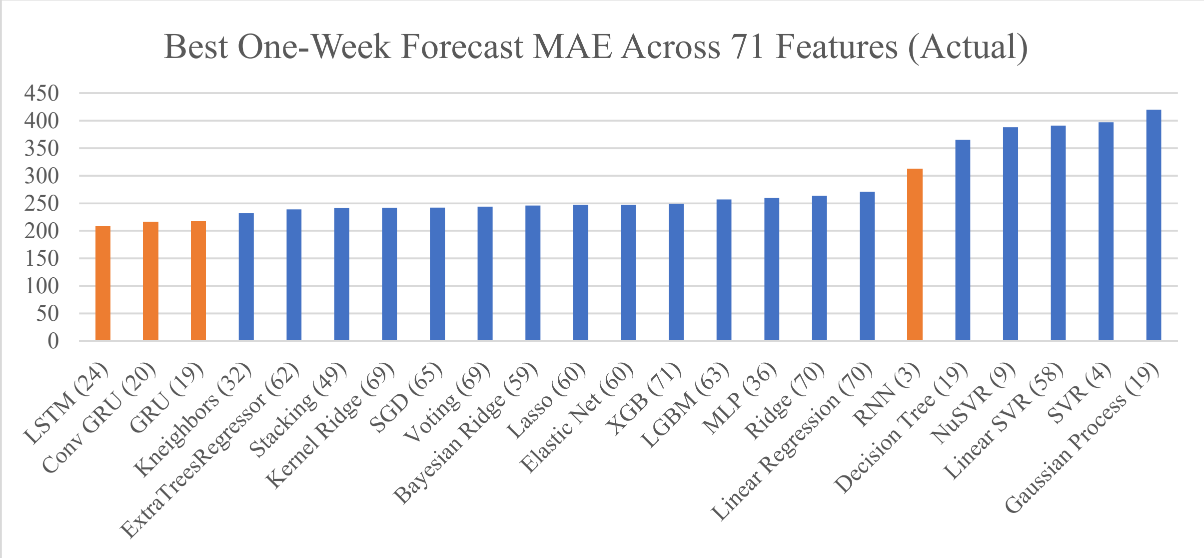


Supplementary Figure S20: Best One-Day Forecast MAE Found Across 77 Features (Weekly Difference). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S12, for one-day forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name.

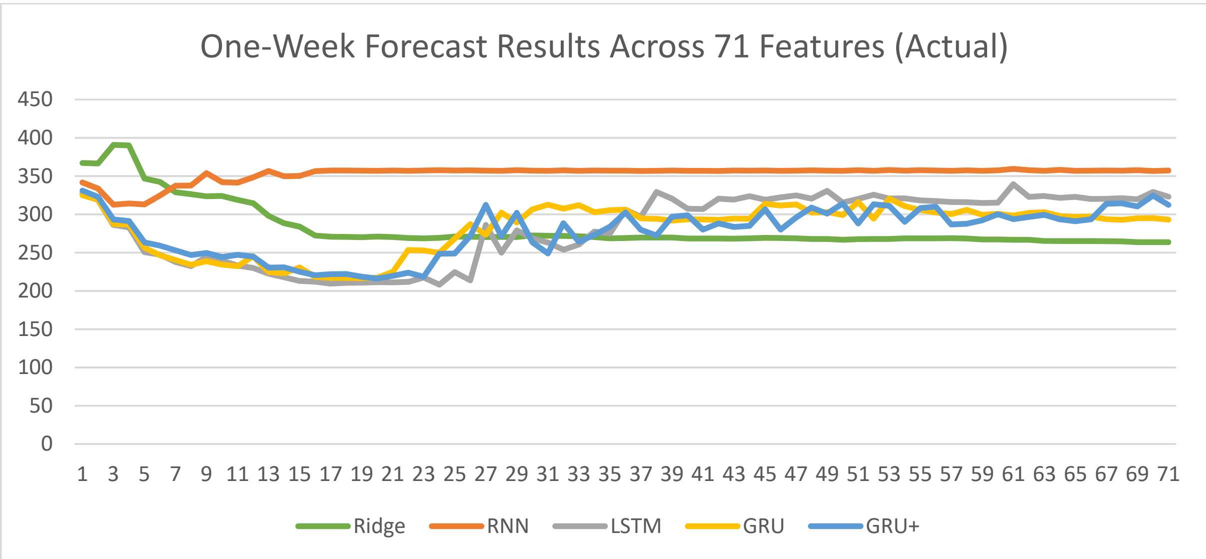


Supplementary Figure S21: All RNN Models and Ridge One-Day Forecast MAE Across 77 Features (Weekly Difference). We show how the number of features affects the MAE score for one-day forecasting in the weekly differenced dataset. The weekly differenced dataset shows little improvement as features are added.

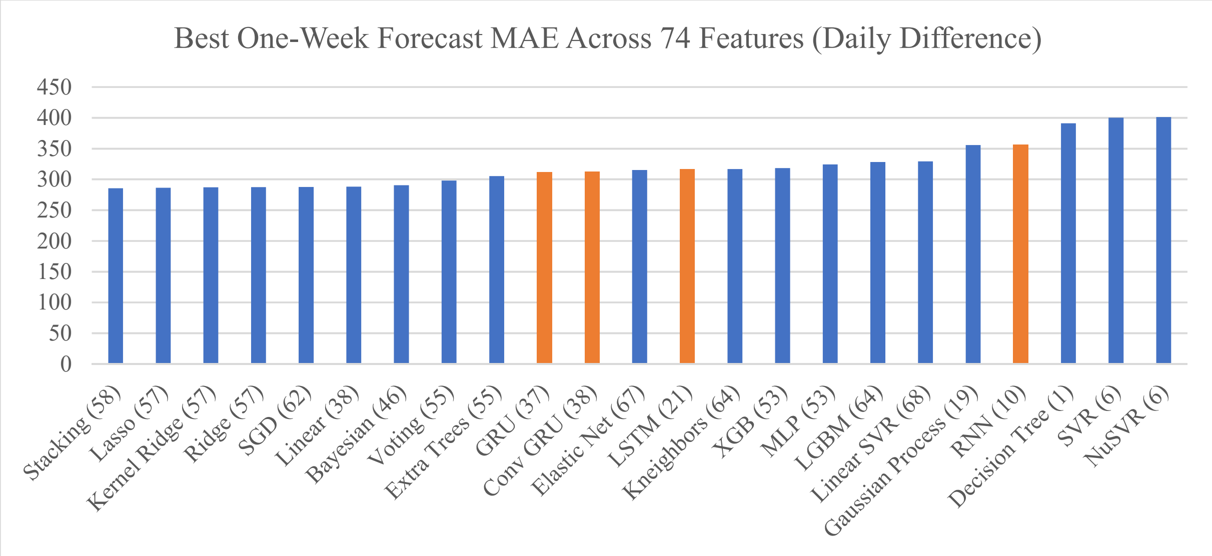
4.2 One-Week Feature Test Results



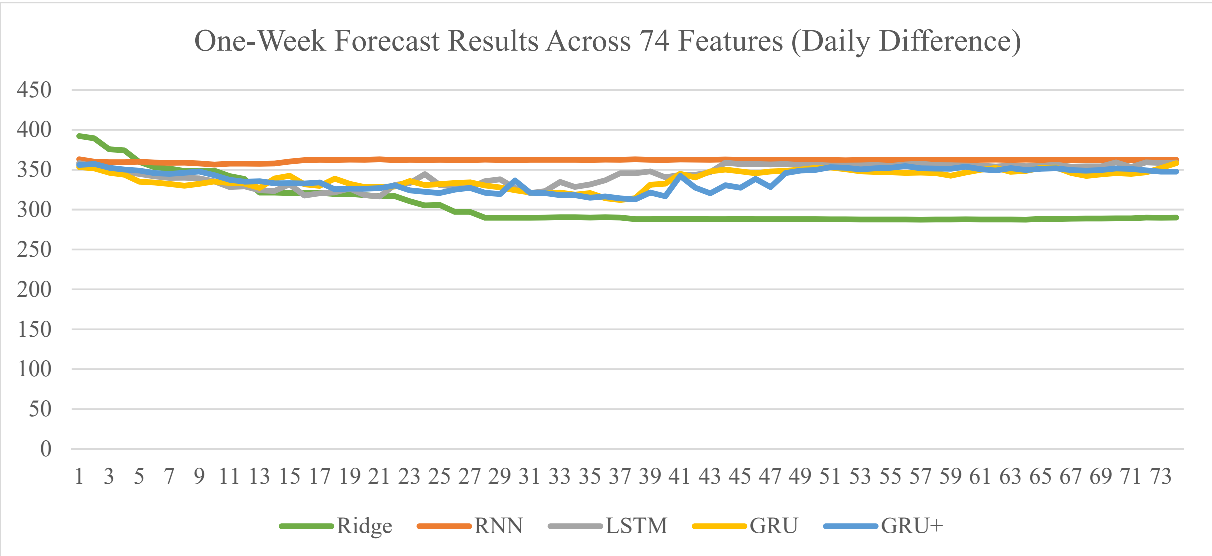
Supplementary Figure S22: Best One-Week Forecast MAE Found Across 71 Features (Actual). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S10, for one-week forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name. It is promising that most RNN models perform well in this stage, but we are wary of overfitting.



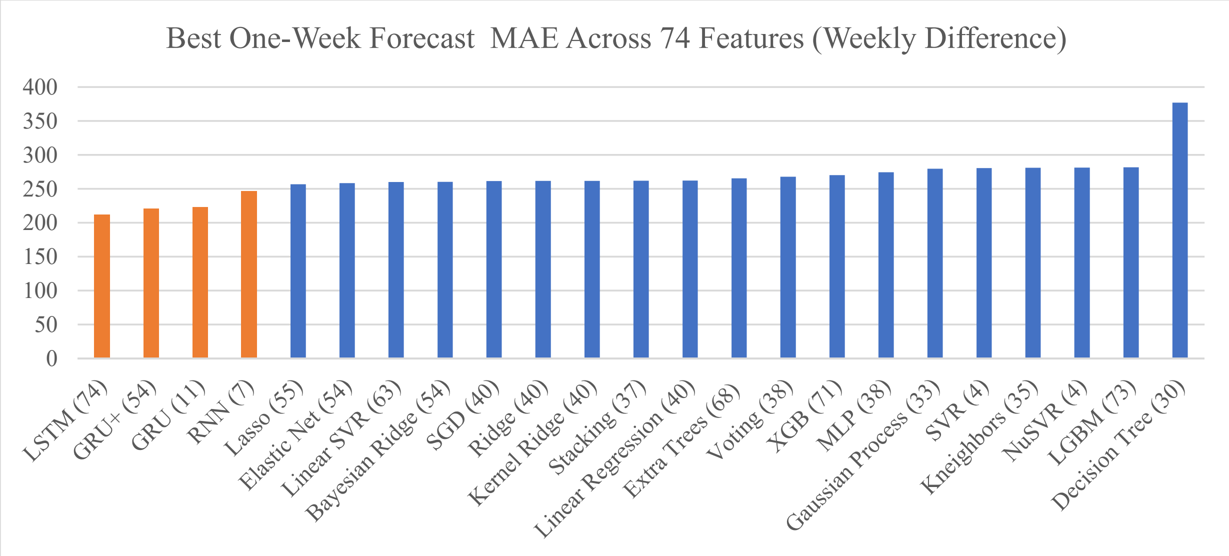
Supplementary Figure S23: All RNN Models and Ridge One-Day Forecast MAE Across 71 Features (Actual). We show how the number of features affects the MAE score for one-week forecasting in the actual dataset.



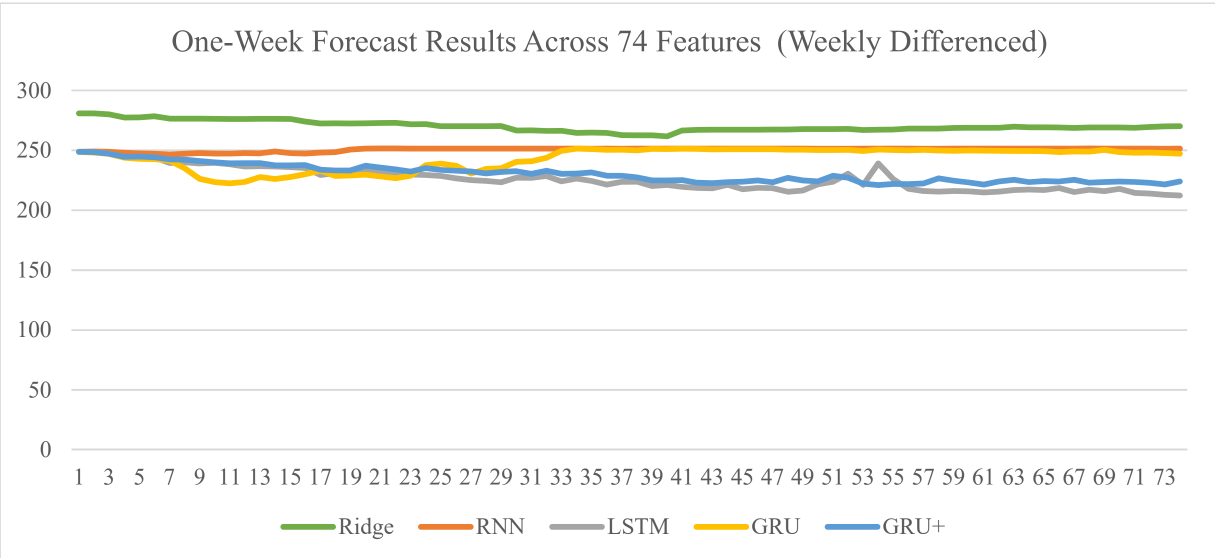
Supplementary Figure S24: Best One-Week Forecast MAE Found Across 74 Features (Daily Difference). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S11, for one-week forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name.



Supplementary Figure S25 All RNN Models and Ridge One-Week Forecast Across 74 Features (Daily Difference). The ridge example shows how the number of features affects the MAE score for one-week forecasting in the daily differenced dataset.



Supplementary Figure S26: Best One-Week Forecast MAE Found Across 74 Features (Weekly Difference). Recurrent (orange) and non-recurrent (blue) models are trained with an iteratively increasing number of ranked features, seen in Supplementary Figure S12, for one-week forecasting. The lowest MAE for each model is recorded with the number of features next to the model's name. Good results from RNN models are promising, but need to be verified in a fair final test.



Supplementary Figure S27: All RNN Models and Ridge One-Week Forecast Across 74 Features (Weekly Difference). The ridge example shows how the number of features affects the MAE score for one-week forecasting in the weekly differenced dataset.

5. Final Test Forecasting Results

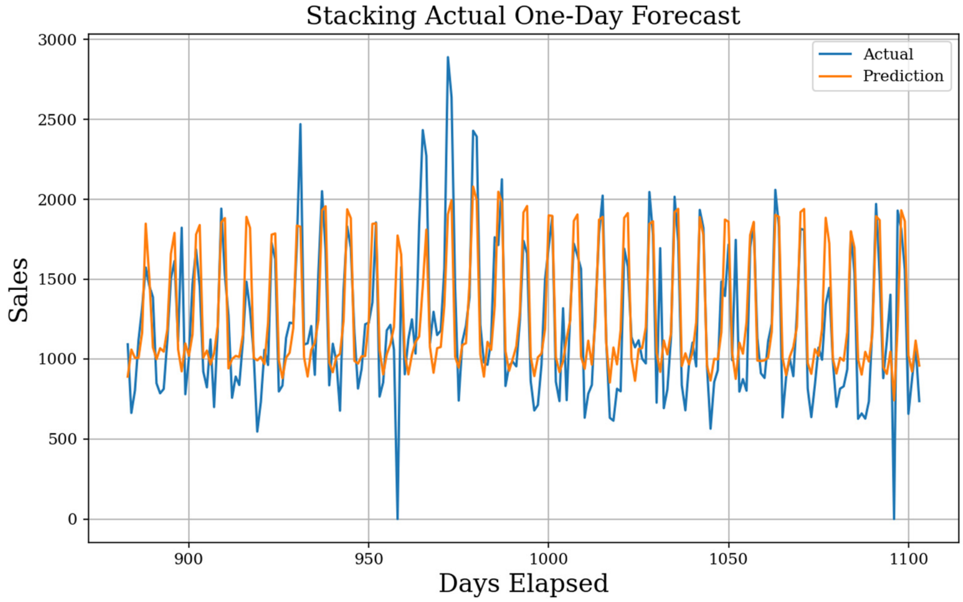
5.1 One-Day Final Test Results

Supplementary Table S9: One-Day Forecast Actual Full Results. The Supplementary Table displays the model, test MAE, sMAPE, gMAPE, and the dataset used to achieve the result. We display the full results of all models from all three datasets. Best results from each dataset and the baseline are bolded. Best results are seen in the Actual and Daily datasets. Many recurrent models have trouble scoring below the baseline.

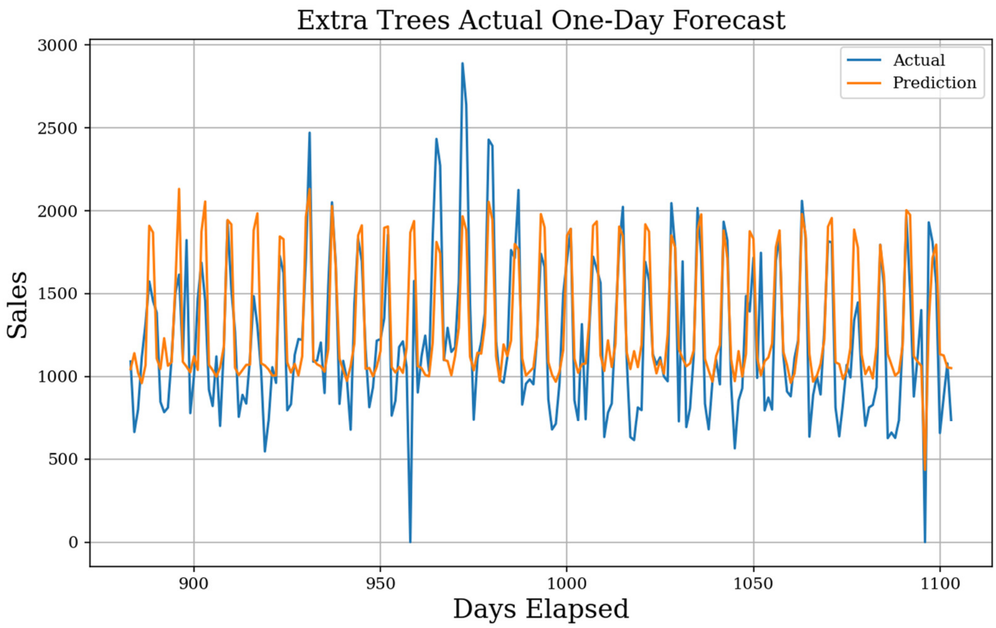
Model	Type	MAE	sMAPE	gMAE	Dataset
Kernel Ridge	**NR**	**214.230**	**0.196**	**126.258**	**Daily**
Ridge	NR	216.163	**0.195**	144.829	Daily
Bayesian Ridge	NR	217.896	0.196	146.150	Daily
Linear	NR	219.587	0.198	137.914	Daily
TFT Less Features	**R**	**220.085**	**0.196**	**133.816**	**Actual**
Stacking	NR	220.186	0.195	142.360	Actual
Bayesian Ridge	NR	221.625	0.195	144.142	Actual
Ridge	NR	221.719	0.195	144.238	Actual
Linear	NR	221.808	0.195	144.203	Actual
SGD	NR	221.823	0.195	144.079	Actual
LSTM	R	222.131	0.196	131.226	Actual
Lasso	NR	223.043	0.201	141.195	Daily
Stacking	NR	223.411	0.200	148.328	Daily
Lasso	NR	226.758	0.201	147.108	Actual
GRU	R	227.226	0.200	144.194	Actual
Extra Trees	NR	231.537	0.204	128.713	Actual
Voting	**NR**	**238.889**	**0.213**	**144.844**	**Weekly**
Use-Last-Week-Enhanced	**NR**	**239.756**	**0.215**	**150.068**	**Actual**
XGB	NR	241.949	0.214	152.700	Daily
Stacking	NR	242.788	0.215	139.604	Weekly
TFT All Features	R	244.656	0.215	159.264	Actual
Linear	NR	245.574	0.217	140.842	Weekly
Kernel Ridge	NR	245.686	0.219	144.613	Weekly

Bayesian Ridge	NR	245.871	0.218	142.508	Weekly
Lasso	NR	246.007	0.218	141.718	Weekly
Ridge	NR	246.901	0.218	144.810	Weekly
K-Neighbors	NR	248.726	0.215	149.253	Actual
Extra Trees	NR	253.214	0.224	171.946	Daily
GRU+	R	255.160	0.232	144.749	Weekly
Extra Trees	NR	256.181	0.229	166.908	Weekly
GRU	R	259.470	0.237	153.419	Weekly
RNN	R	262.974	0.237	159.277	Weekly
LSTM	R	263.990	0.238	159.088	Weekly
TFT All Features	R	267.929	0.225	169.346	Daily
TFT Less Features	R	270.420	0.231	176.263	Daily
TFT Less Features	R	272.283	0.249	162.680	Weekly
GRU	R	272.639	0.237	176.367	Daily
Use-Last-Week	**R**	**278.491**	**0.256**	**165.358**	**Actual**
GRU+	R	289.512	0.240	195.203	Actual
LSTM	R	294.707	0.247	185.289	Daily
GRU+	R	325.965	0.275	213.258	Daily
RNN	R	370.611	0.315	223.414	Daily
Use-Yesterday	**NR**	**392.659**	**1.205**	**228.724**	**Daily**
TFT All Features	R	456.367	0.355	**320.791**	Weekly
RNN	R	**481.217**	0.426	271.143	Actual

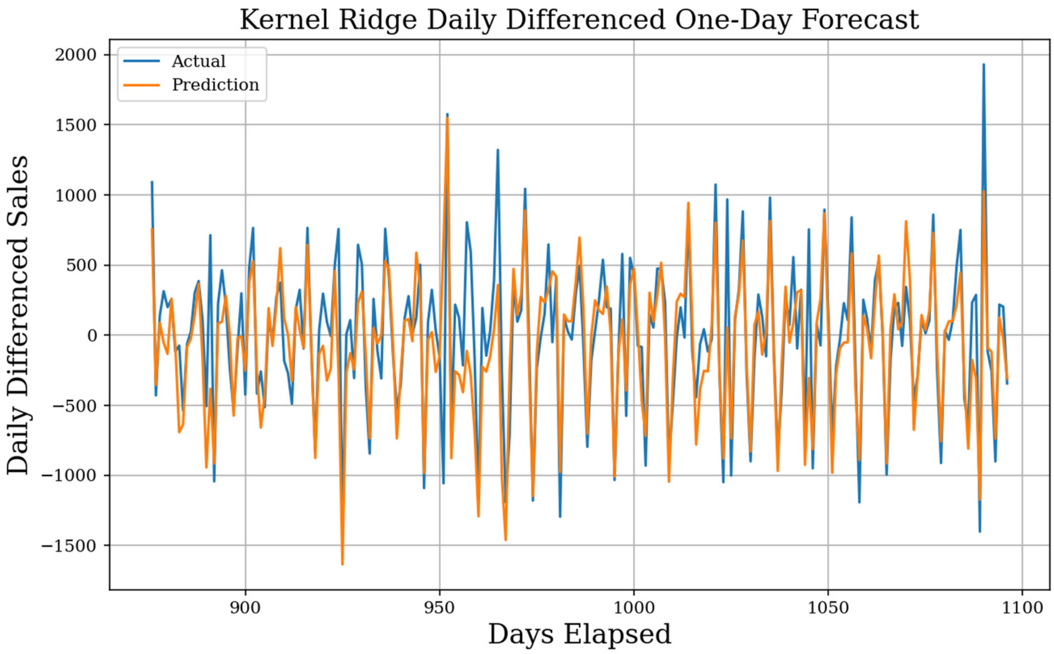
5.1.1 Non-Recurrent Algorithm Forecast Figures

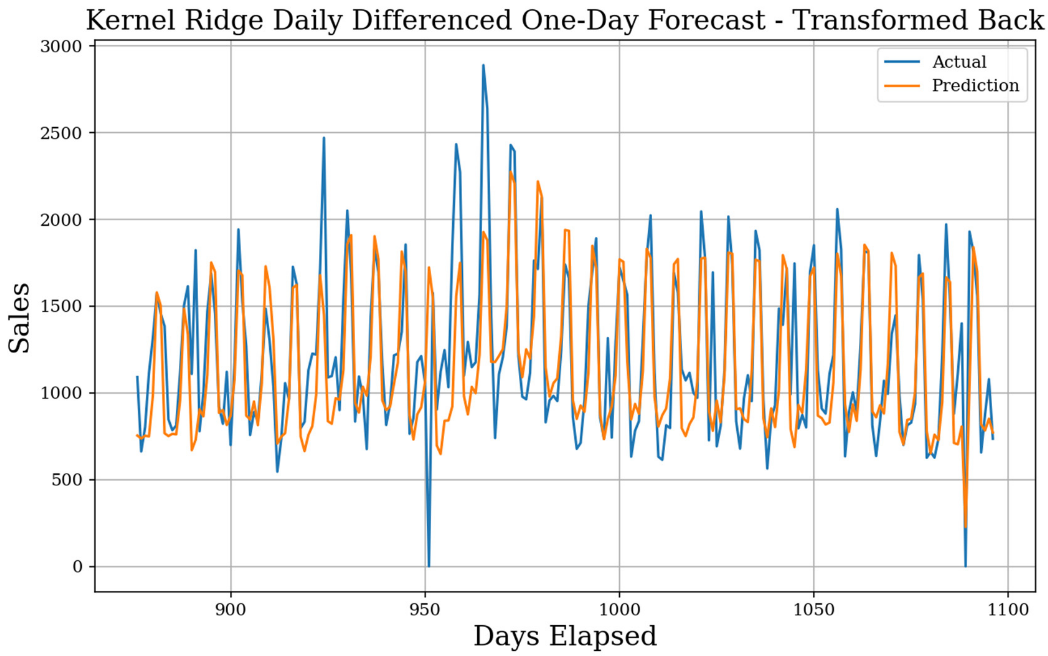


Supplementary Figure S28: Stacking Actual One-Day Forecast. MAE of 220, sMAPE of 19.5%, and a gMAE of 142, with 25 features

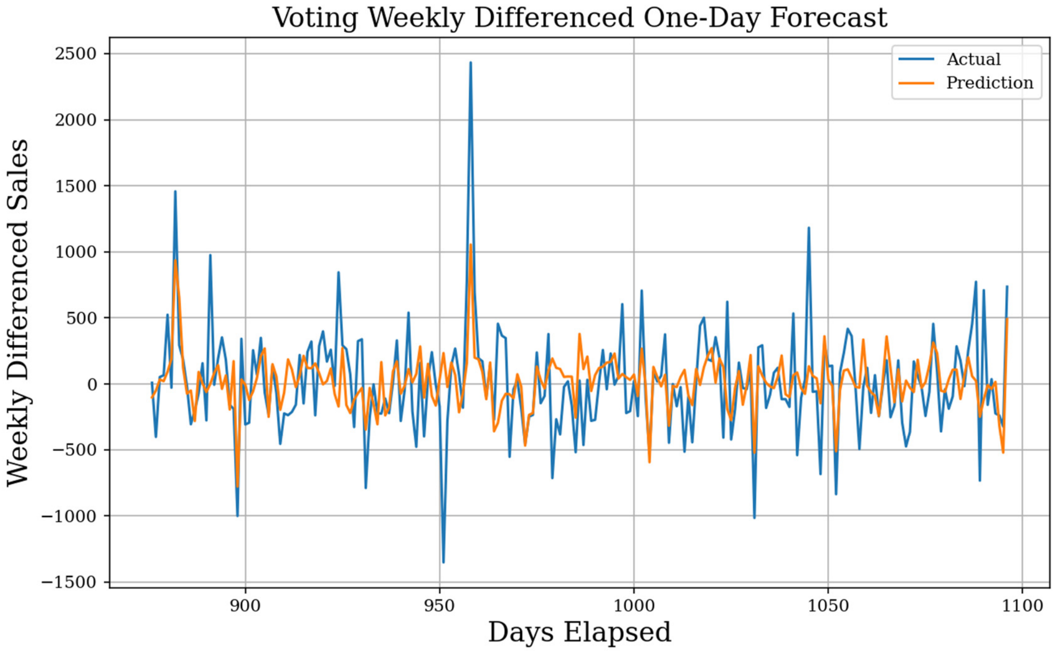


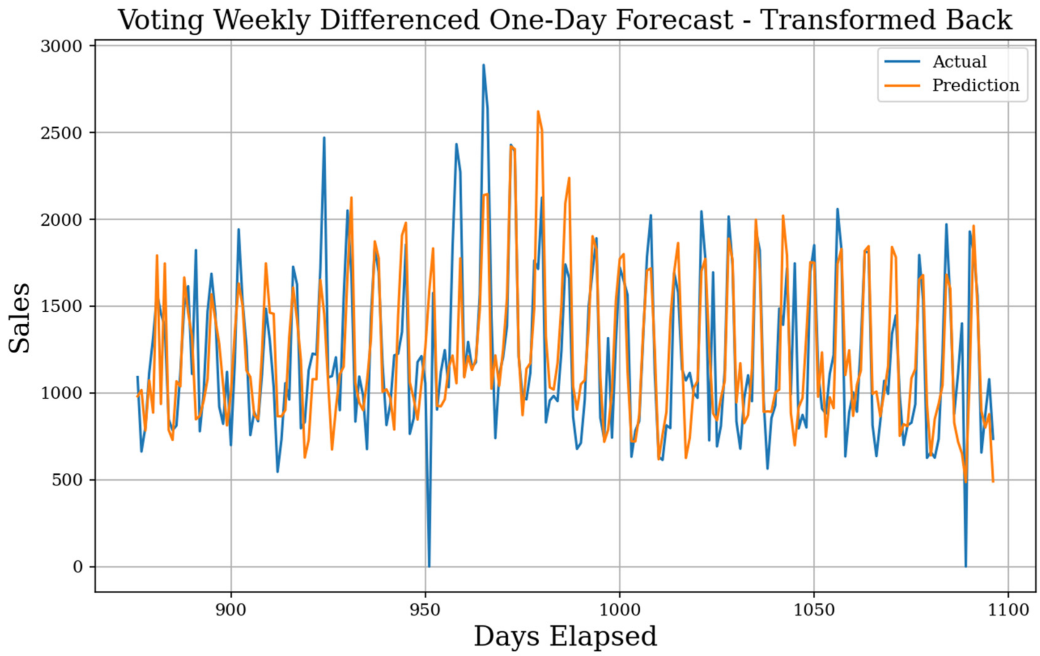
Supplementary Figure S29: Extra Trees Actual One-Day Forecast. MAE of 231, sMAPE of 20.4%, and a gMAE of 128, with 29 features.





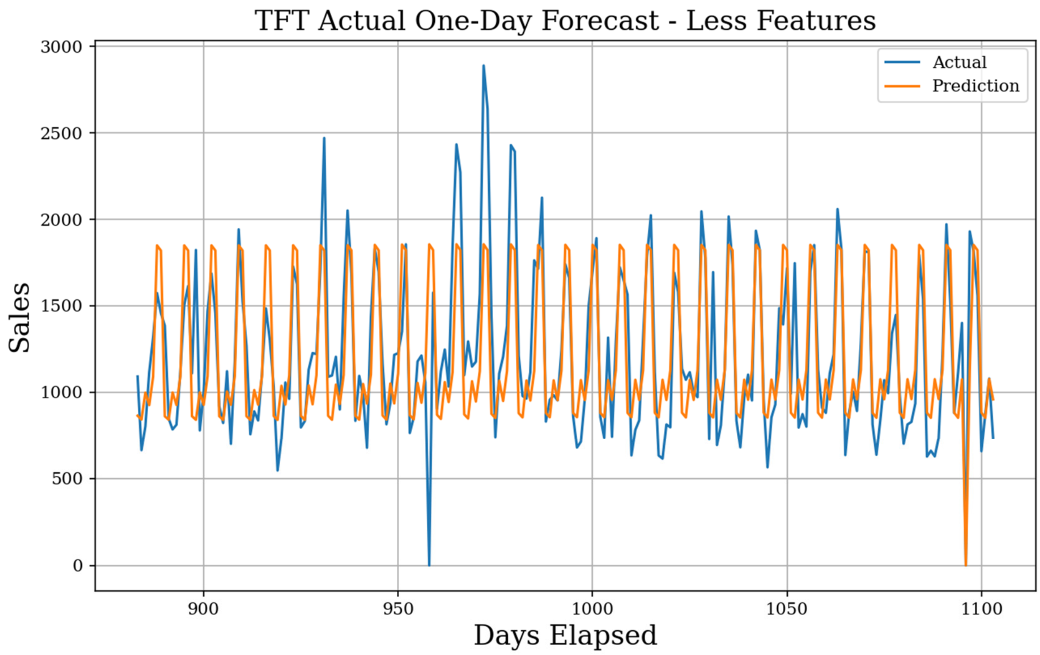
Supplementary Figure S30: Kernel Ridge Daily Differenced One-Day Forecast. MAE of 214, sMAPE of 19.6%, and a gMAE of 126, with 72 features. Original predictions (first) and then transformed back version (second) are both shown.



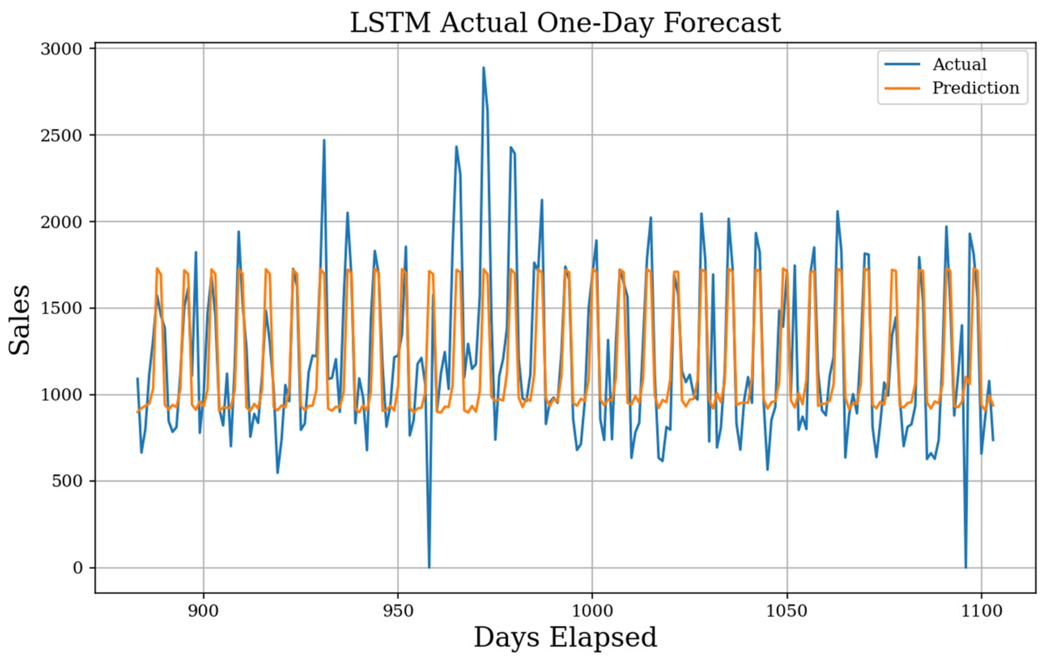


Supplementary Figure S31: Voting Weekly Differenced One-Day Forecast. MAE of 238, sMAPE of 21.3%, and a gMAE of 144, with 12 features.

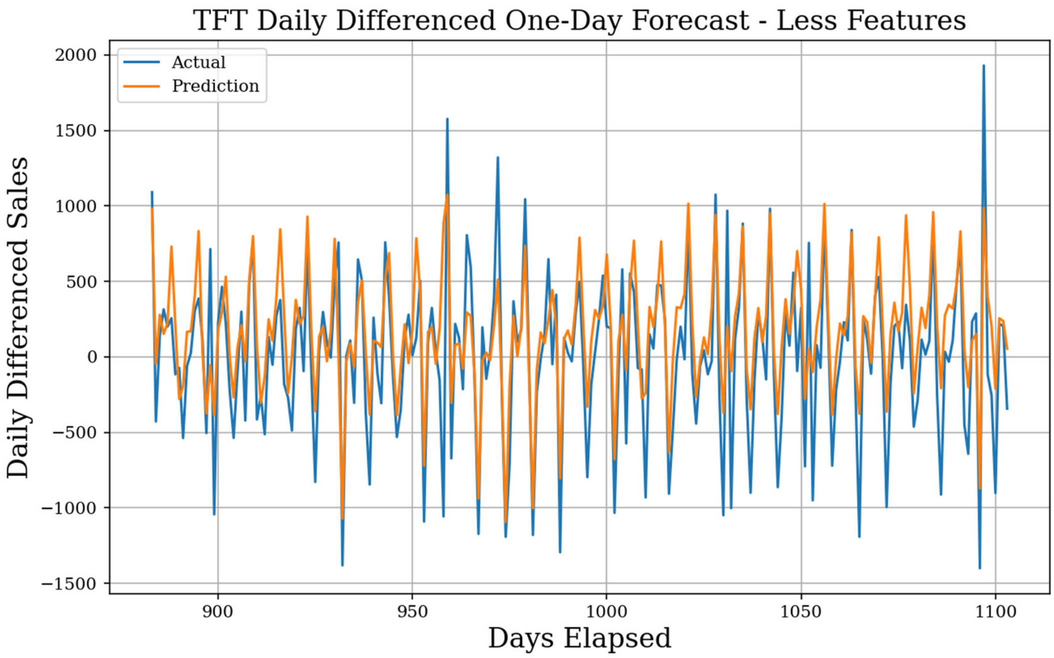
5.1.2 Recurrent Algorithms Forecast Figures

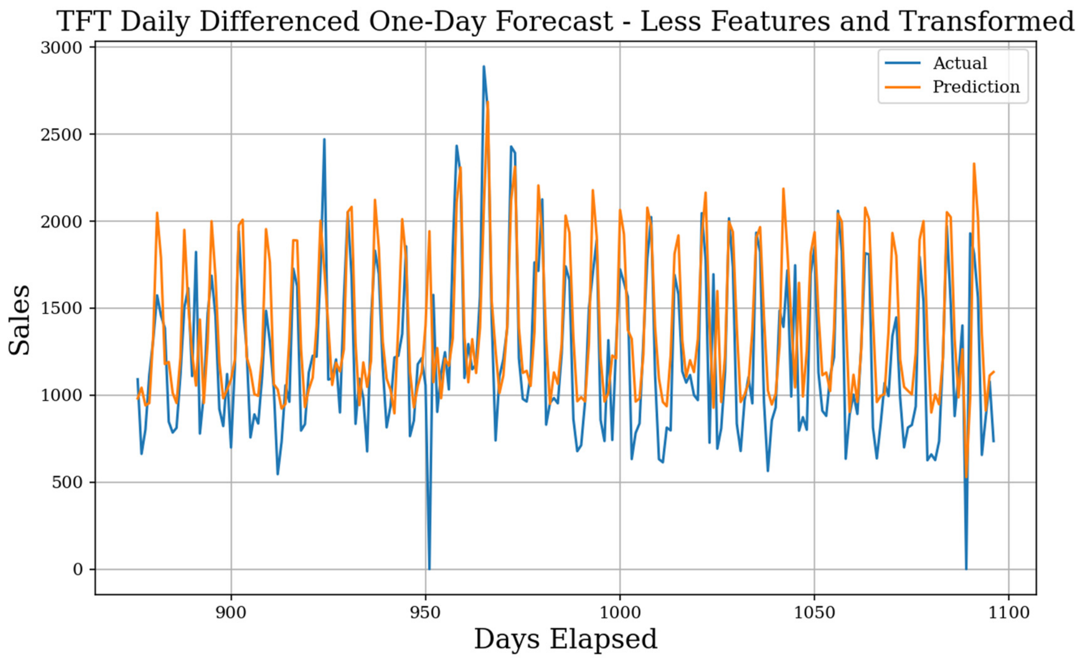


Supplementary Figure S32: Transformer Less Features Actual One-Day Forecast. MAE of 220, sMAPE of 19.6%, and a gMAE of 133, with 17 features.

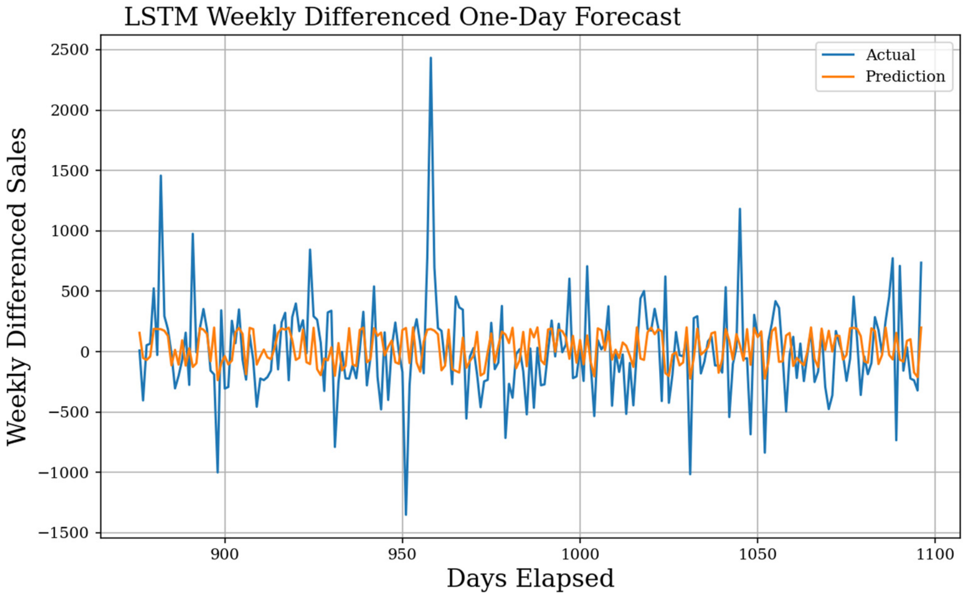


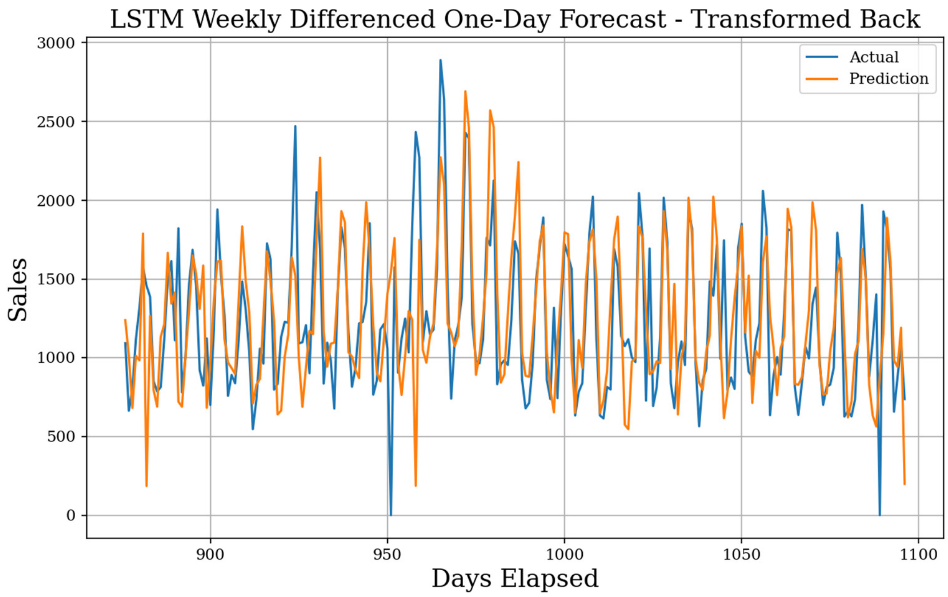
Supplementary Figure S33: LSTM Actual One-Day Forecast. MAE of 222, sMAPE of 19.6%, and a gMAE of 131, with 22 features.





Supplementary Figure S34: Transformer Less Features Daily Differenced One-Day Forecast. MAE of 270, sMAPE of 23.1%, and a gMAE of 176, with 13 features.





Supplementary Figure S35: LSTM Weekly Differenced One-Day Forecast. MAE of 263, sMAPE of 23.8%, and a gMAE of 159, with seven features.

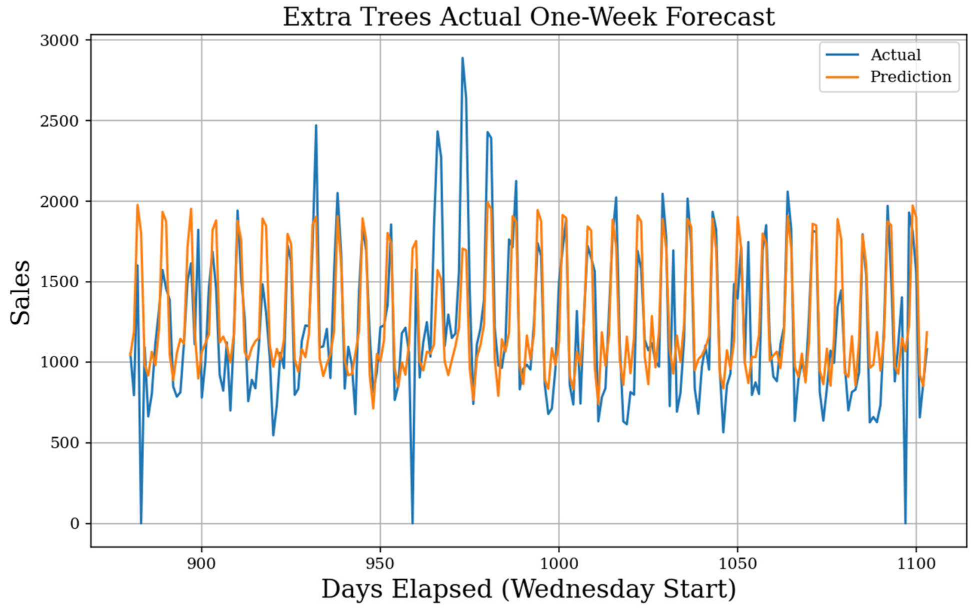
5.2 One-Week Feature Test Results

Supplementary Table S10: One-Week Forecast Actual Full Results. The Supplementary Table displays the model, test MAE, sMAPE, gMAPE, and the dataset used to achieve the result. One-week specific metrics like best start day, the mean of each weekday start, and the standard deviation between each start are also included. Best results, by MAE, for each dataset and the baseline are bolded. RNN models with the Actual dataset are the only results to beat the baseline Use-Last-Week-Enhanced. Different methodologies for extending non-RNN models to longer horizon windows are required.

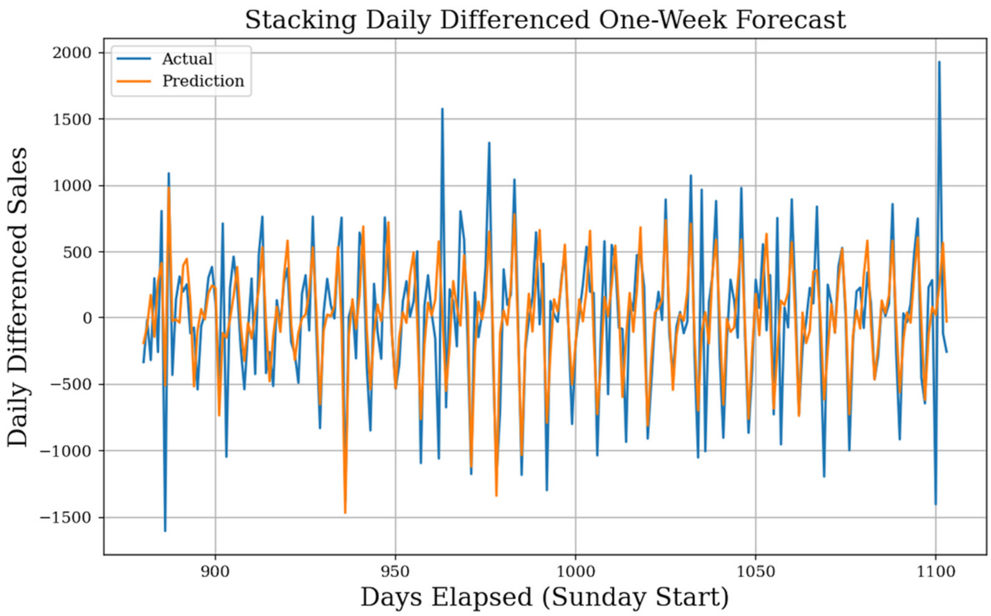
Model	Type	MAE	sMAPE	gMAE	Dataset	Weekday	Mean	Std Dev
TFT Less Features	**R**	**215.309**	**0.202**	**123.842**	**Actual**	**Friday**	**222.657**	**3.363**
GRU	R	218.806	**0.195**	**116.842**	Actual	Sunday	233.224	13.477
LSTM	R	222.183	0.197	134.022	Actual	Thursday	228.797	5.339
Use-Last-Week-Enhanced	**NR**	**230.865**	**0.203**	**139.891**	**Actual**	**Tuesday**	**232.964**	**2.437**
GRU+	R	233.485	0.204	136.278	Actual	Wednesday	246.925	14.612
Extra Trees	NR	235.671	0.206	145.267	Actual	Wednesday	240.753	4.085
Stacking	NR	237.524	0.208	146.056	Actual	Tuesday	243.917	4.634
Voting	NR	237.673	0.209	140.365	Actual	Friday	246.417	8.256
Kernel Ridge	NR	239.376	0.213	143.088	Actual	Wednesday	244.985	4.229
SGD	NR	240.525	0.214	140.116	Actual	Tuesday	249.395	7.712
Bayesian Ridge	NR	242.961	0.216	145.162	Actual	Wednesday	248.505	3.408
Lasso	NR	243.377	0.218	147.151	Actual	Thursday	248.388	2.979
Lasso	**NR**	**253.214**	**1.284**	**137.134**	**Weekly**	**Sunday**	**256.633**	**3.156**
Ridge	NR	256.788	1.274	144.496	Weekly	Sunday	261.872	3.403
Kernel Ridge	NR	257.239	1.274	146.107	Weekly	Sunday	262.033	3.436
Elastic Net	NR	257.627	1.327	153.369	Weekly	Sunday	259.266	1.495
SGD	NR	257.779	1.280	148.201	Weekly	Monday	261.827	2.978

LinSVR	NR	258.030	1.405	149.518	Weekly	Sunday	260.927	1.939
Bayesian Ridge	NR	259.125	1.304	151.530	Weekly	Sunday	260.490	1.210
Stacking	NR	260.480	1.281	151.251	Weekly	Monday	264.051	2.694
TFT Less Features	R	263.407	1.371	147.057	Weekly	Tuesday	278.024	9.849
TFT All Features	R	267.178	0.239	153.347	Actual	Wednesday	268.460	1.131
RNN	R	273.599	1.722	162.270	Weekly	Sunday	278.034	2.950
GRU	R	273.650	1.674	154.359	Weekly	Sunday	279.818	4.318
Lasso	**NR**	**280.306**	**1.016**	**162.255**	**Daily**	**Sunday**	**287.353**	**6.530**
LSTM	R	280.849	1.876	166.470	Weekly	Sunday	283.464	2.274
GRU+	R	280.853	**1.917**	165.276	Weekly	Sunday	283.244	2.129
Stack	NR	282.991	1.022	168.323	Daily	Sunday	289.068	4.932
Ridge	NR	283.704	1.007	166.965	Daily	Wednesday	289.434	6.562
Kernel Ridge	NR	283.880	1.009	165.939	Daily	Wednesday	289.372	6.364
Linear	NR	284.196	1.016	165.769	Daily	Wednesday	290.070	6.083
TFT Less Features	R	285.203	1.081	179.880	Daily	Monday	294.081	7.208
SGD	NR	286.235	1.046	162.893	Daily	Saturday	291.113	4.538
Voting	NR	287.070	1.029	164.446	Daily	Wednesday	299.976	8.864
Bayesian Ridge	NR	287.731	1.028	168.909	Daily	Sunday	292.999	5.247
LSTM	R	292.751	1.175	173.468	Daily	Sunday	310.536	15.681
TFT All Features	R	330.009	1.076	197.982	Daily	Wednesday	334.242	4.535
RN	R	361.654	1.570	223.516	Daily	Thursday	400.535	18.508
RNN	R	371.586	0.315	235.160	Actual	Saturday	380.430	6.642
GRU	R	405.286	1.719	232.294	Daily	Sunday	408.500	2.928
GRU+	R	406.992	1.725	240.084	Daily	Monday	408.595	1.202
TFT All Features	R	**452.629**	1.403	**302.849**	Weekly	Wednesday	473.895	13.427

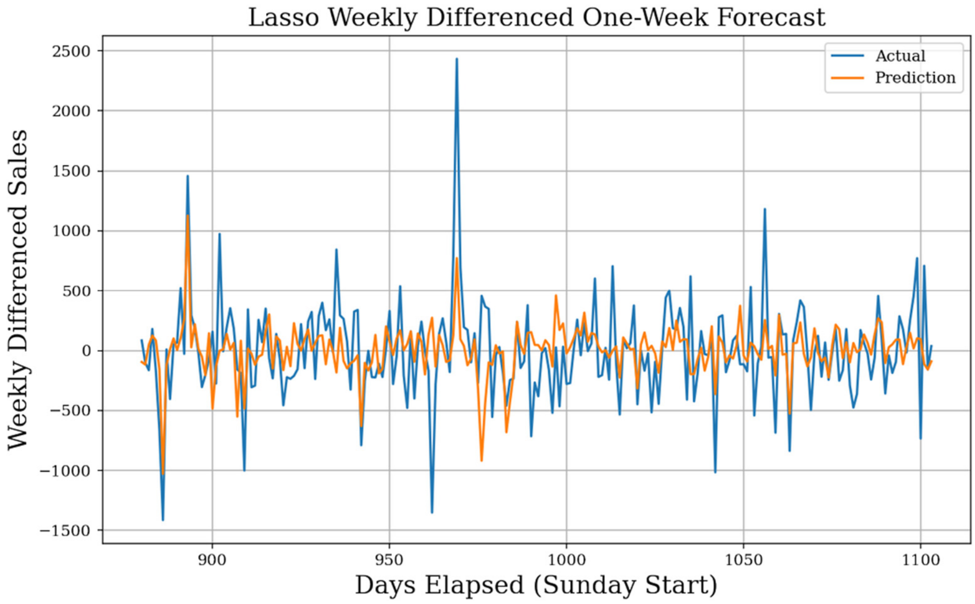
5.2.1 Non-Recurrent Algorithm Forecast Figures



Supplementary Figure S36: Extra Trees Actual One-Week Forecast. The best start day MAE of 235, sMAPE of 20.6%, and gMAE of 145 are found when starting predictions on Wednesday. A mean MAE of 240 and a standard deviation of 4.085 are found with 69 features.

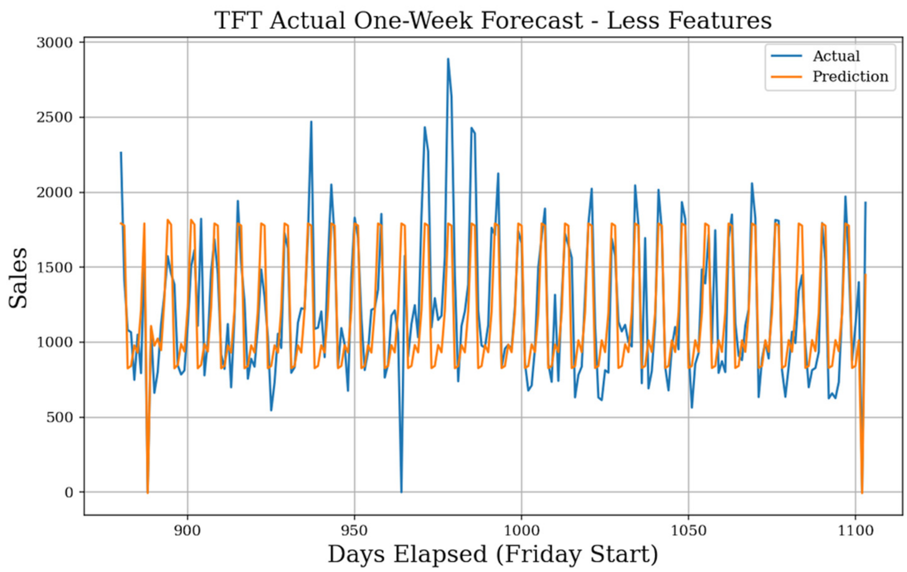


Supplementary Figure S37: Stacked Daily Differenced One-Week Forecast. The best start day MAE of 282, sMAPE of 100.02%, and gMAE of 168 are found when starting predictions on Sunday. A mean MAE of 289 and a standard deviation of 4.93 are found with 58 features.

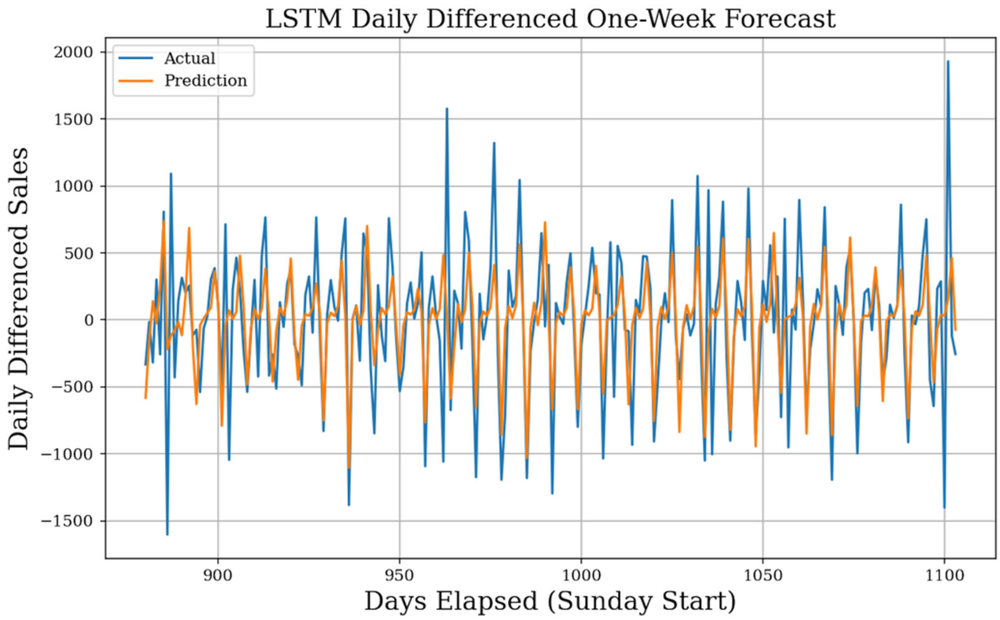


Supplementary Figure S38: Lasso Weekly Differenced One-Week Forecast. The best start day MAE of 253, sMAPE of 128.4%, and gMAE of 137 are found when starting predictions on Sunday. A mean MAE of 256 and a standard deviation of 3.15 are found with 55 features.

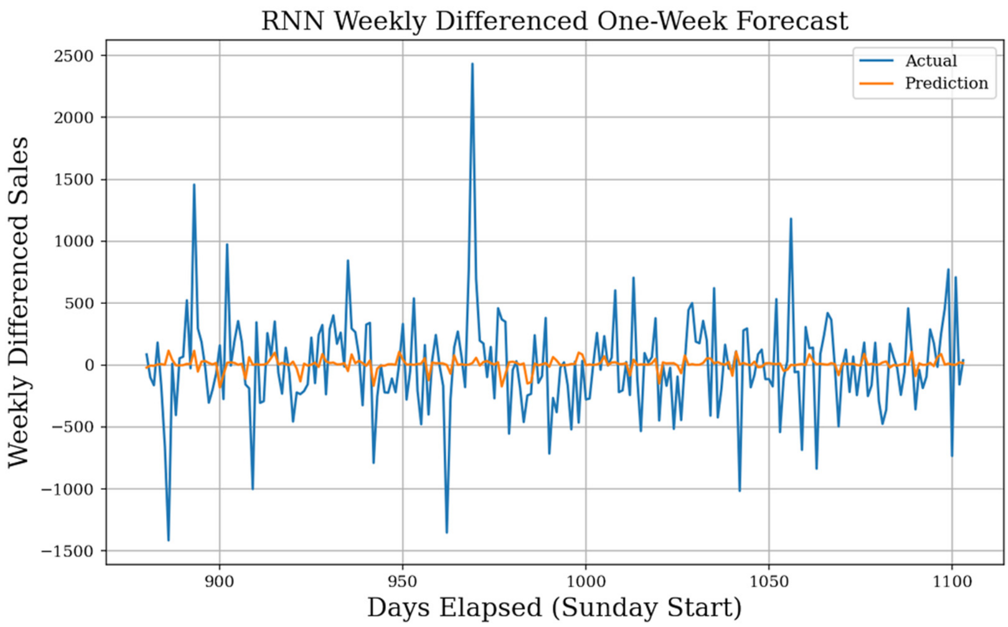
5.2.2 Recurrent Algorithm Forecast Figures



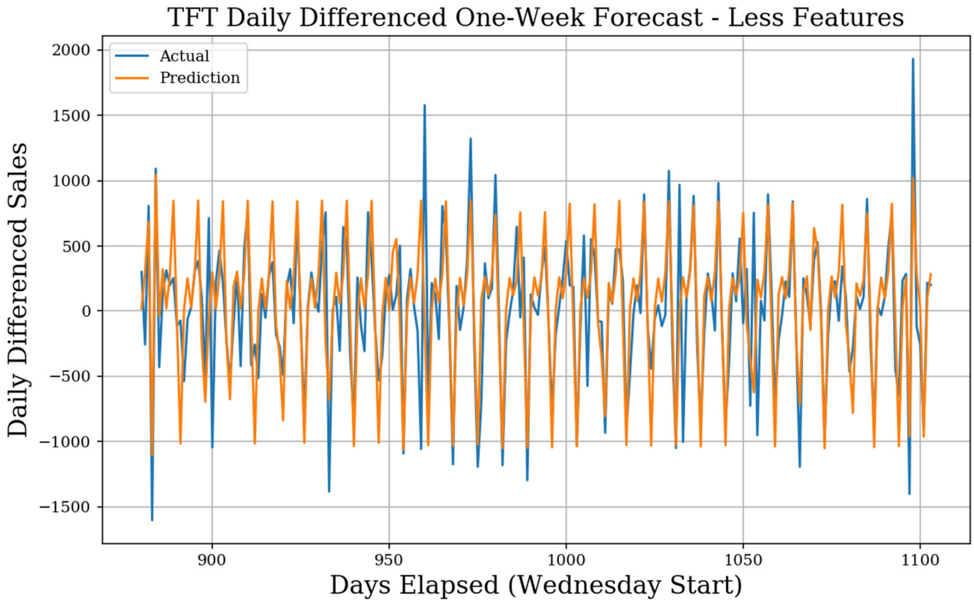
Supplementary Figure S39: Transformer Less Features Actual One-Week Forecast. The best start day MAE of 215, sMAPE of 20.2%, and gMAE of 123 are found when starting on Friday. A mean MAE of 222 and a standard deviation of 3.36 are found with 17 features.



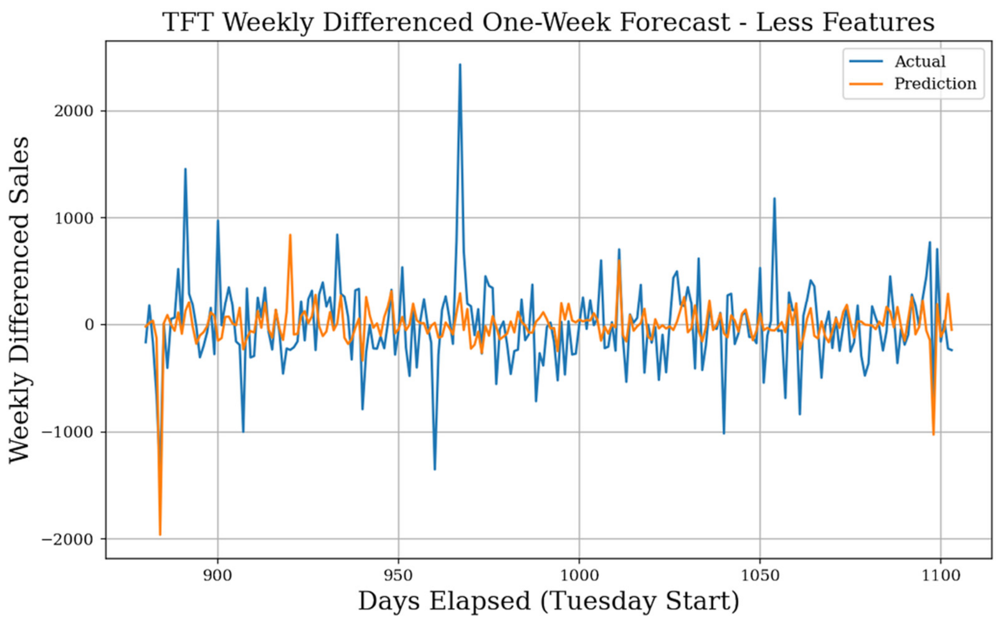
Supplementary Figure S40: LSTM Daily Differenced One-Week Forecast. The best start day MAE of 292, sMAPE of 117.5%, and gMAE of 173 are found when starting predictions on Sunday. A mean MAE of 310 and a standard deviation of 15.68 are found with 21 features.



Supplementary FigureS41: RNN Weekly Differenced One-Week Forecast. The best start day MAE of 273, sMAPE of 172.2%, and gMAE of 162, are found when starting predictions on Sunday. A mean MAE of 278 and a standard deviation of 2.95 are found with 7 features.



Supplementary Figure S42: Transformer Less Features Daily Differenced One-Week Forecast. The best start day MAE of 285, sMAPE of 108.1%, and gMAE of 179 are found when starting predictions on Wednesday. A mean MAE o 294 and a standard deviation of 7.2 are found with 13 features.



Supplementary Figure S43: Transformer Less Features Weekly Differenced One-Week Forecast. The best start day MAE of 263, sMAPE of 137.1%, and gMAE of 147 are found when starting predictions on Tuesday. A mean MAE of 278 and a standard deviation of 2.95 are found with less features..hgfhgf …h.