Pick-up + Machine Learning: A Proved Efficient Approach to Forecast Hotel Demand

Rachel Zhang

05/29/2020

Data

Data Importing and Cross-Validation
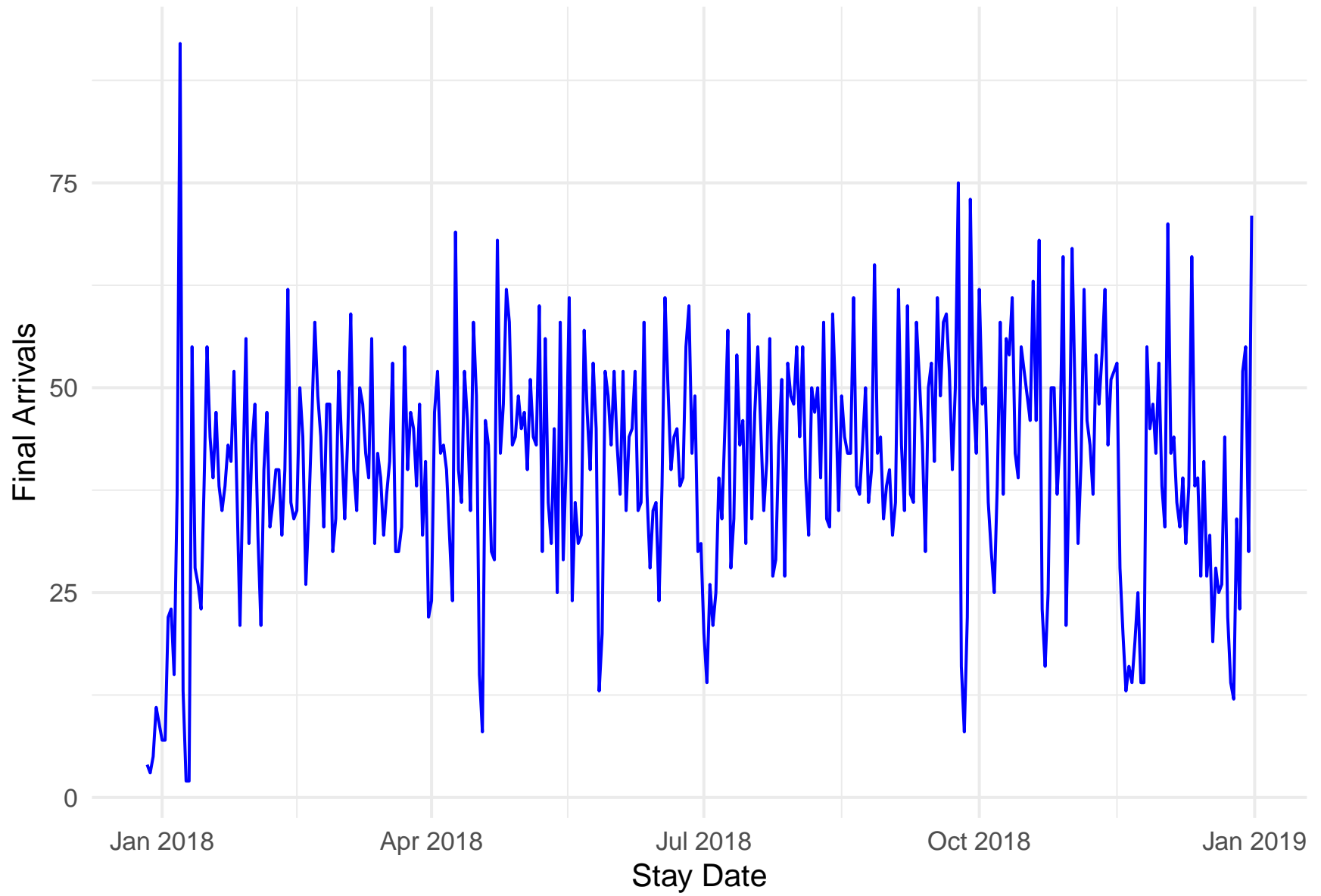
```
## [1] 370
```

There are trasactions with DBA up to 90 days.

Table 1: Training Set Overview

	ROH0	DOW	ROH1	ROH2	ROH3	ROH4	ROH5	ROH6	ROH7	ROH14	ROH21	ROH30	ROH60	ROH90
2018-11-15	52	Thursday	49	46	42	41	41	41	38	29	25	22	7	1
2018-06-11	58	Monday	58	56	53	51	47	47	47	40	37	30	9	4
2018-05-04	51	Friday	48	47	46	44	41	38	38	30	28	21	4	1
2018-10-21	68	Sunday	67	67	66	66	64	59	54	44	20	18	10	5
2018-09-22	40	Saturday	38	36	36	35	33	32	31	27	20	8	5	5
2018-07-01	20	Sunday	17	16	16	14	13	13	13	10	10	6	2	2
2018-10-29	66	Monday	61	56	55	53	51	49	43	32	26	25	7	2
2018-03-21	30	Wednesday	28	25	21	21	21	20	20	9	5	5	3	0
2018-09-29	49	Saturday	45	45	42	39	37	36	36	31	26	23	17	12
2018-11-21	14	Wednesday	14	13	12	12	12	12	11	10	10	9	6	2

Modeling

Additive Pick-up

Table 2: Additive Pick Ups

DOW	ROH1	ROH2	ROH3	ROH4	ROH5	ROH6	ROH7	ROH14	ROH21	ROH30	ROH60	ROH90
Sunday	2.11	3.04	4.40	5.58	6.51	8.18	9.22	13.9	18.8	22.8	30.1	33.9
Monday	2.84	4.34	5.16	7.59	9.57	11.14	12.89	21.8	27.4	33.2	42.6	45.9
Tuesday	2.42	4.28	4.58	4.97	6.17	7.25	8.92	16.6	21.7	27.3	33.5	36.7
Wednesday	1.95	3.58	5.37	5.77	6.16	7.30	8.51	15.5	20.5	25.3	32.6	35.8
Thursday	3.40	5.11	6.84	8.31	8.84	9.67	10.84	17.0	21.4	25.8	34.1	37.3
Friday	3.32	5.76	7.61	9.27	10.83	11.54	12.24	17.9	21.6	25.5	33.6	37.5
Saturday	3.71	5.69	7.74	9.55	10.98	12.14	12.88	17.3	20.3	23.7	30.3	33.3

Multiplicative Pick-up

Regression

Table 3: Multiplicative Pick Ups

DOW	ROH1	ROH2	ROH3	ROH4	ROH5	ROH6	ROH7	ROH14	ROH21	ROH30	ROH60	ROH90
Sunday	0.939	0.910	0.874	0.841	0.817	0.772	0.746	0.623	0.500	0.395	0.215	0.113
Monday	0.940	0.907	0.890	0.841	0.801	0.773	0.739	0.571	0.459	0.343	0.149	0.082
Tuesday	0.942	0.892	0.884	0.874	0.844	0.821	0.787	0.599	0.477	0.320	0.158	0.081
Wednesday	0.942	0.901	0.855	0.846	0.837	0.810	0.782	0.622	0.506	0.390	0.191	0.097
Thursday	0.914	0.873	0.835	0.802	0.789	0.766	0.738	0.585	0.482	0.380	0.181	0.106
Friday	0.921	0.864	0.822	0.785	0.748	0.733	0.717	0.590	0.507	0.419	0.236	0.145
Saturday	0.902	0.852	0.798	0.753	0.717	0.687	0.668	0.552	0.473	0.382	0.197	0.120

Neural Network

```
##                    Length Class      Mode
## call                   6  -none-     call
## response             296  -none-     numeric
## covariate           2368  -none-     numeric
## model.list             2  -none-     list
## err.fct                1  -none-     function
## act.fct                1  -none-     function
## linear.output          1  -none-     logical
## data                   9  data.frame list
## exclude                0  -none-     NULL
## net.result             1  -none-     list
## weights                1  -none-     list
## generalized.weights    1  -none-     list
## startweights           1  -none-     list
## result.matrix         34  -none-     numeric

##      Length Class Mode
## [1,] 14     nn    list
## [2,] 14     nn    list
## [3,] 14     nn    list
## [4,] 14     nn    list
## [5,] 14     nn    list
## [6,] 14     nn    list
```

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
(Intercept)	1.88***	2.54***	3.51***	4.17***	4.58***	5.92***	6.69***	11.59***	17.74***	21.88***	27.68***	32.50***
	(0.51)	(0.68)	(0.77)	(0.87)	(0.92)	(0.98)	(1.11)	(1.61)	(1.80)	(1.87)	(2.00)	(1.93)
DOWMonday	0.66	1.16	0.49	1.62	2.57**	2.35*	3.01**	7.60***	8.52***	10.36***	12.89***	12.19***
	(0.49)	(0.66)	(0.76)	(0.85)	(0.90)	(0.98)	(1.11)	(1.62)	(1.93)	(2.10)	(2.39)	(2.52)
DOWTuesday	0.30	1.23	0.15	−0.69	−0.44	−1.09	−0.44	2.91	3.02	4.68*	3.96	3.24
	(0.51)	(0.68)	(0.78)	(0.88)	(0.94)	(1.01)	(1.15)	(1.70)	(2.03)	(2.23)	(2.53)	(2.67)
DOWWednesday	−0.17	0.53	0.97	0.15	−0.43	−1.02	−0.86	1.65	1.77	2.61	2.90	2.19
	(0.48)	(0.65)	(0.74)	(0.84)	(0.89)	(0.97)	(1.10)	(1.62)	(1.93)	(2.12)	(2.41)	(2.53)
DOWThursday	1.28**	2.05**	2.42**	2.71**	2.28*	1.36	1.48	3.09	2.66	2.91	4.18	3.46
	(0.48)	(0.64)	(0.73)	(0.83)	(0.88)	(0.96)	(1.08)	(1.60)	(1.91)	(2.09)	(2.37)	(2.50)
DOWFriday	1.18*	2.68***	3.16***	3.62***	4.26***	3.22**	2.83*	3.87*	2.70	2.53	2.99	3.09
	(0.49)	(0.65)	(0.75)	(0.85)	(0.90)	(0.98)	(1.11)	(1.64)	(1.96)	(2.15)	(2.44)	(2.58)
DOWSaturday	1.62**	2.69***	3.45***	4.17***	4.78***	4.32***	4.04***	3.83*	1.63	0.94	0.49	−0.60
	(0.49)	(0.65)	(0.75)	(0.85)	(0.91)	(0.98)	(1.11)	(1.64)	(1.95)	(2.13)	(2.42)	(2.55)
ROH1	1.01***											
	(0.01)											
ROH2		1.01***										
		(0.01)										
ROH3			1.03***									
			(0.02)									
ROH4				1.04***								
				(0.02)								
ROH5					1.06***							
					(0.02)							
ROH6						1.07***						
						(0.02)						
ROH7							1.09***					
							(0.03)					
ROH14								1.09***				
								(0.05)				
ROH21									1.05***			
									(0.06)			
ROH30										1.06***		
										(0.07)		
ROH60											1.29***	
											(0.13)	
ROH90												1.31***
												(0.17)
R^2	0.97	0.95	0.93	0.92	0.91	0.89	0.86	0.69	0.56	0.47	0.32	0.24
Adj. R^2	0.97	0.95	0.93	0.91	0.90	0.89	0.85	0.68	0.55	0.46	0.30	0.22
Num. obs.	296	296	296	296	296	296	296	296	296	296	296	296
RMSE	2.27	3.03	3.48	3.94	4.19	4.53	5.13	7.60	9.07	9.92	11.25	11.87

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

```
## [7,] 14      nn     list
## [8,] 14      nn     list
## [9,] 14      nn     list
## [10,] 14     nn     list
## [11,] 14     nn     list
## [12,] 14     nn     list
```

K-Nearest Neighbor

For Weighted K-Nearest Neighbors:

```
## [[1]]
## [[1]]$kernel
## [1] "triangular"
##
## [[1]]$k
## [1] 3
##
##
## [[2]]
## [[2]]$kernel
## [1] "optimal"
##
## [[2]]$k
## [1] 3
##
##
## [[3]]
## [[3]]$kernel
## [1] "triangular"
##
## [[3]]$k
## [1] 3
##
##
## [[4]]
## [[4]]$kernel
## [1] "triangular"
##
```

```
## [[4]]$k
## [1] 4
##
##
## [[5]]
## [[5]]$kernel
## [1] "rank"
##
## [[5]]$k
## [1] 3
##
##
## [[6]]
## [[6]]$kernel
## [1] "optimal"
##
## [[6]]$k
## [1] 3
##
##
## [[7]]
## [[7]]$kernel
## [1] "triangular"
##
## [[7]]$k
## [1] 3
##
##
## [[8]]
## [[8]]$kernel
## [1] "optimal"
##
## [[8]]$k
## [1] 6
##
##
## [[9]]
## [[9]]$kernel
## [1] "triangular"
```

```
##
## [[9]]$k
## [1] 4
##
##
## [[10]]
## [[10]]$kernel
## [1] "optimal"
##
## [[10]]$k
## [1] 7
##
##
## [[11]]
## [[11]]$kernel
## [1] "triangular"
##
## [[11]]$k
## [1] 6
##
##
## [[12]]
## [[12]]$kernel
## [1] "triangular"
##
## [[12]]$k
## [1] 8
```

Tree

For Random Forest, this following chart shows the number of variables selected when building each tree:

```
## [[1]]
## [1] 7
##
## [[2]]
## [1] 7
##
## [[3]]
```

```
## [1] 7
##
## [[4]]
## [1] 7
##
## [[5]]
## [1] 7
##
## [[6]]
## [1] 7
##
## [[7]]
## [1] 4
##
## [[8]]
## [1] 4
##
## [[9]]
## [1] 4
##
## [[10]]
## [1] 4
##
## [[11]]
## [1] 2
##
## [[12]]
## [1] 2
```

Support Vector Machine

This following chart shows the gamma selected for each model

```
## [[1]]
## [1] 0.0312
##
## [[2]]
## [1] 0.0312
##
```

```
## [[3]]
## [1] 0.0312
##
## [[4]]
## [1] 0.0312
##
## [[5]]
## [1] 0.0312
##
## [[6]]
## [1] 0.0312
##
## [[7]]
## [1] 0.0312
##
## [[8]]
## [1] 0.0625
##
## [[9]]
## [1] 0.1
##
## [[10]]
## [1] 0.125
##
## [[11]]
## [1] 0.1
##
## [[12]]
## [1] 0.0312
```

Results

Table 5: Parameter selection of K for K-NN model

X

Table 6: Mean Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	-0.344	-0.629	-0.380	-3.17	0.913	-0.085	-0.075	-0.591	-0.636
DBA2	-0.294	-0.806	-0.373	-4.87	0.977	0.225	-0.335	-0.609	-0.715
DBA3	-0.299	-1.045	-0.446	-6.26	0.927	0.143	-0.602	-0.436	-0.661
DBA4	-0.042	-0.943	-0.270	-7.29	1.169	-0.082	-0.236	-0.681	-0.555
DBA5	0.161	-0.888	-0.148	-8.25	1.171	-0.019	-0.581	-0.649	-0.337
DBA6	0.338	-0.964	-0.031	-9.21	1.263	0.436	-0.631	-0.319	-0.436
DBA7	0.485	-1.098	0.068	-10.34	1.141	0.512	-0.592	0.137	-0.706
DBA14	1.841	-0.756	1.521	-15.44	1.724	1.357	0.893	0.965	0.291
DBA21	2.354	-1.374	2.201	-19.46	2.598	0.530	0.442	0.871	1.379
DBA30	3.061	-1.315	2.926	-23.88	3.656	1.866	2.831	1.755	3.267
DBA60	3.938	-1.976	3.544	-29.92	4.627	3.489	3.594	3.930	4.278
DBA90	4.422	-2.626	4.151	-32.54	3.843	3.441	4.393	4.455	5.299
13	1.302	-1.202	1.064	-14.22	2.001	0.984	0.758	0.736	0.872

Table 7: Mean Absolute Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	1.96	1.93	1.94	3.31	3.31	2.69	2.69	2.31	2.00
DBA2	2.57	2.60	2.54	4.91	3.85	3.07	3.19	2.70	2.57
DBA3	2.74	2.72	2.66	6.34	3.73	3.41	3.80	3.07	2.72
DBA4	3.06	3.19	2.97	7.37	4.12	3.70	3.75	3.40	2.98
DBA5	3.55	3.53	3.43	8.33	4.13	3.79	4.33	3.43	3.32
DBA6	4.04	3.98	3.88	9.24	4.57	4.38	4.55	3.82	3.72
DBA7	4.55	4.67	4.40	10.46	4.54	4.71	4.40	4.31	4.10
DBA14	6.03	6.58	5.84	15.62	6.03	5.57	5.65	5.42	5.17
DBA21	7.14	8.49	7.05	19.58	7.84	7.02	6.64	6.89	6.44
DBA30	8.33	10.56	8.22	23.93	8.87	8.95	8.03	7.92	7.77
DBA60	9.72	14.88	9.50	29.92	9.88	9.81	9.30	9.59	9.26
DBA90	10.77	22.73	10.70	32.54	10.85	11.94	11.14	11.11	11.20
13	5.37	7.15	5.26	14.29	5.97	5.75	5.62	5.33	5.10

Table 8: Standard Deviation Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	2.63	2.60	2.61	2.73	4.09	3.34	3.30	2.79	2.68
DBA2	3.39	3.31	3.35	3.35	4.97	3.96	3.92	3.48	3.33
DBA3	3.66	3.58	3.58	3.81	4.88	4.41	4.66	4.01	3.61
DBA4	4.06	4.08	3.96	4.18	5.06	4.87	4.67	4.37	3.87
DBA5	4.51	4.49	4.38	4.73	5.13	4.84	5.45	4.31	4.24
DBA6	4.98	5.04	4.83	5.31	5.44	5.40	5.51	4.68	4.66
DBA7	5.48	5.87	5.36	5.64	5.49	5.61	5.45	5.30	5.06
DBA14	7.37	8.73	7.19	8.24	7.53	7.03	7.19	6.90	6.71
DBA21	8.64	11.60	8.54	9.83	9.50	9.52	8.45	8.87	8.21
DBA30	10.34	12.99	10.22	11.08	10.50	10.89	10.11	10.15	9.57
DBA60	11.96	17.99	11.60	13.08	11.90	12.22	11.51	11.79	11.34
DBA90	13.05	29.42	12.93	14.16	13.38	14.76	13.32	13.46	13.36
13	6.67	9.14	6.55	7.18	7.32	7.24	6.96	6.68	6.39

Table 9: Mean Percentage Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.006	-0.025	0.003	-0.098	0.107	0.026	0.016	-0.026	-0.001
DBA2	0.019	-0.033	0.012	-0.150	0.124	0.035	0.002	-0.016	0.006
DBA3	0.028	-0.042	0.015	-0.189	0.122	0.038	-0.003	-0.008	0.016
DBA4	0.041	-0.045	0.020	-0.223	0.121	0.047	0.015	-0.021	0.020
DBA5	0.058	-0.043	0.031	-0.245	0.128	0.042	0.009	-0.021	0.034
DBA6	0.077	-0.043	0.044	-0.274	0.132	0.046	0.008	-0.013	0.036
DBA7	0.090	-0.051	0.053	-0.302	0.124	0.055	0.013	0.050	0.027
DBA14	0.193	-0.041	0.163	-0.419	0.174	0.130	0.062	0.106	0.095
DBA21	0.252	-0.060	0.238	-0.525	0.268	0.126	0.077	0.137	0.175
DBA30	0.325	-0.036	0.313	-0.655	0.335	0.237	0.237	0.200	0.295
DBA60	0.417	-0.035	0.388	-0.814	0.438	0.354	0.357	0.401	0.405
DBA90	0.463	-0.009	0.447	-0.881	0.447	0.444	0.458	0.471	0.503
13	0.164	-0.039	0.144	-0.398	0.210	0.132	0.104	0.105	0.134

Table 10: MAPE

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.060	0.061	0.059	0.105	0.158	0.103	0.098	0.080	0.064
DBA2	0.091	0.087	0.088	0.153	0.185	0.119	0.112	0.094	0.088
DBA3	0.100	0.094	0.094	0.196	0.180	0.127	0.128	0.106	0.097
DBA4	0.113	0.112	0.103	0.230	0.183	0.143	0.131	0.119	0.105
DBA5	0.137	0.121	0.123	0.253	0.189	0.145	0.146	0.120	0.122
DBA6	0.159	0.128	0.139	0.276	0.200	0.156	0.148	0.132	0.134
DBA7	0.180	0.146	0.156	0.313	0.196	0.174	0.147	0.156	0.144
DBA14	0.283	0.207	0.257	0.440	0.261	0.226	0.205	0.207	0.204
DBA21	0.352	0.266	0.341	0.540	0.380	0.276	0.247	0.280	0.283
DBA30	0.426	0.349	0.415	0.662	0.441	0.393	0.347	0.341	0.385
DBA60	0.529	0.499	0.506	0.814	0.539	0.486	0.465	0.507	0.501
DBA90	0.586	0.714	0.576	0.881	0.583	0.612	0.588	0.597	0.616
13	0.251	0.232	0.238	0.405	0.291	0.247	0.230	0.228	0.229

Table 11: Model Performances

	ME	MAE	SDE	MPE	MAPE	Time
Additive Pickup	1.302	5.37	6.67	0.164	0.251	0.136
Multiplicative Pickup	-1.202	7.15	9.14	-0.039	0.232	0.132
Regression	1.064	5.26	6.55	0.144	0.238	0.064
Neural Network	-14.218	14.29	7.18	-0.398	0.405	43.544
K-Nearest Neighbor	2.001	5.97	7.32	0.210	0.291	17.425
Weighted K-Nearest Neighbor	0.984	5.75	7.24	0.132	0.247	2.463
Decision Tree	0.758	5.62	6.96	0.104	0.230	0.160
Random Forest	0.736	5.33	6.68	0.105	0.228	94.332
Support Vector Machine	0.872	5.10	6.39	0.134	0.229	12.587

[1] Mean Error
[2] Mean Absolute Error
[3] Standard Deviation Error
[4] Mean Percentage Error
[5] Mean Absolute Percentage Error
[6] Time is calculated in seconds