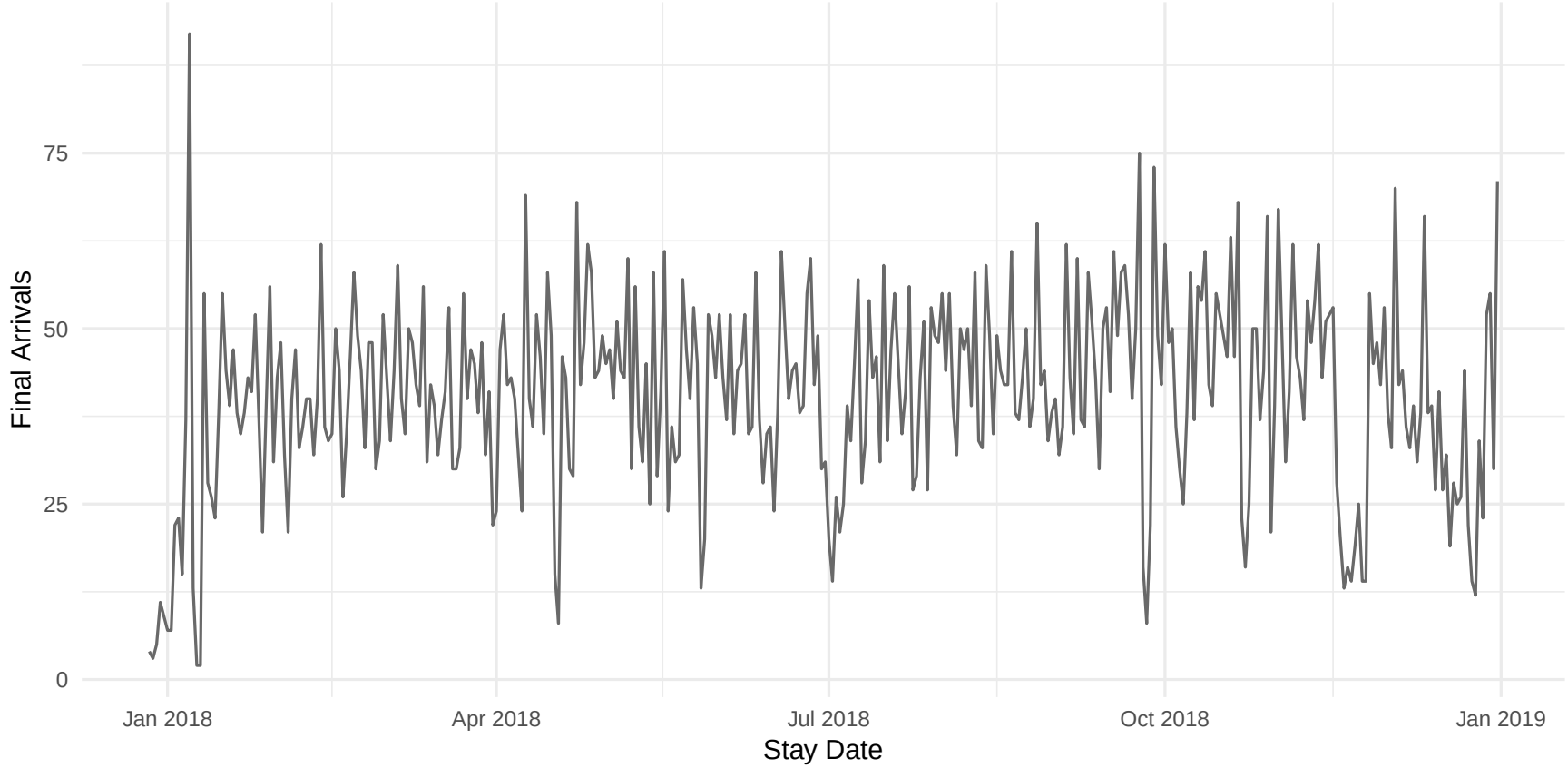
Pick-up method + machine learning: a proved efficient approach to forecast hotel demand

Rachel Zhang

06/18/2020

Data

Importe Dataset and Cross-Validation



This analysis is conducted under `R version 3.6.0` with `set.seed(0)`. The robust test in Section X is conducted under `set.seed(123)`.

We randomly selected 80% of the records as the training dataset to tune models, and the rest 20% records are used for model performance test. Here is a peek of the training set:

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

Note:
The pick-ups are calculated by taking the average of additive increments between current day and a future date by day of week.

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

Note:
The pick-ups are calculated by taking the average of ratio increments between current day and a future date by day of week.

Multiplicative Pick-up

Regression

The regression model uses the nearest ROH and the DOW of the target day.

Neural Network

When building model, the number of hidden units is set as 3. The dataset is scaled and DOW is converted to dummy variables.

Taking ROH=5 as the example, this plot provides a straightforward visualization of the relevant neural network.

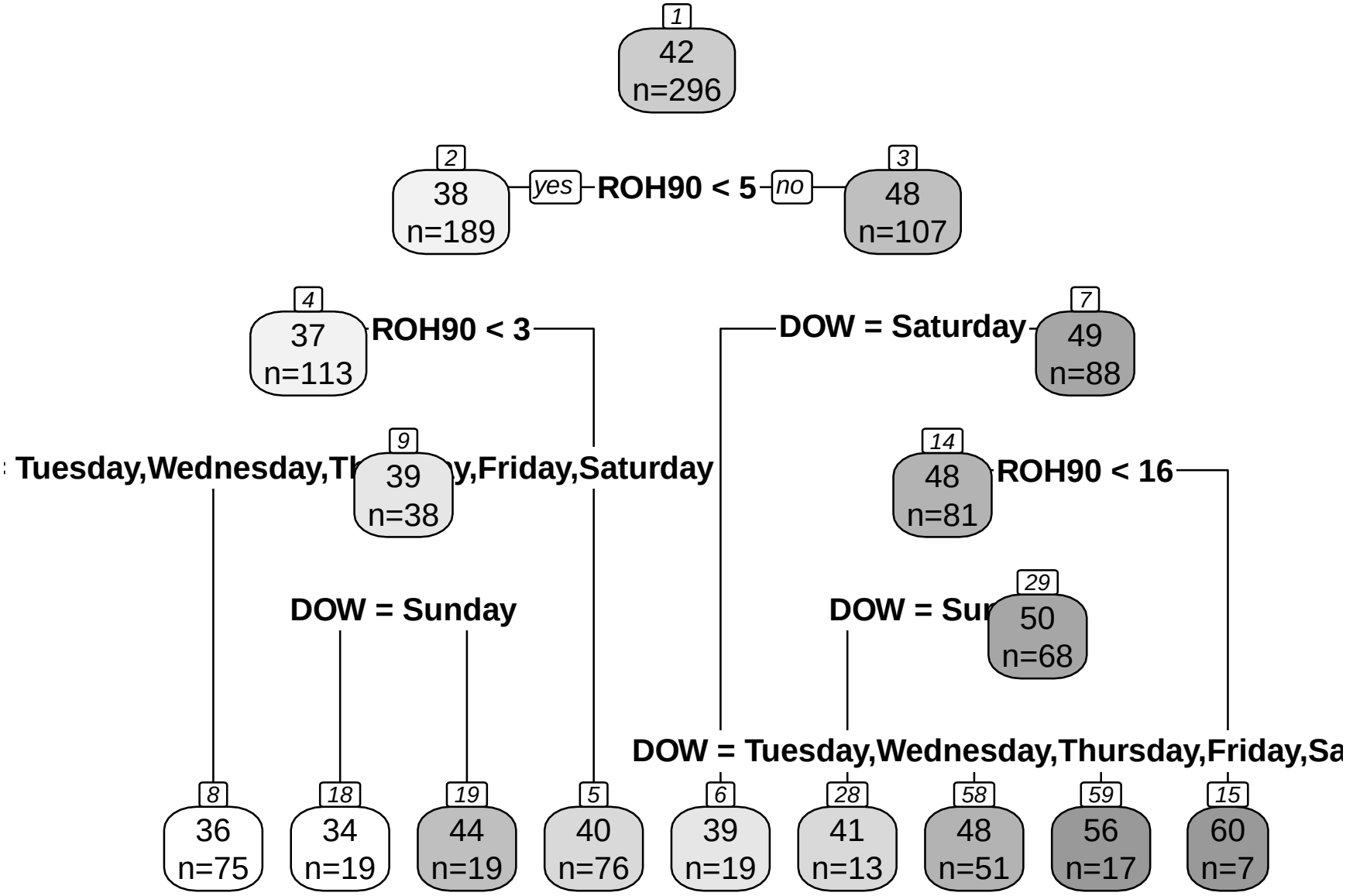
```
## pdf
##   2
```

K-Nearest Neighbor

```
##  [1]  5  5  7  5  5  5  5  5  5  9  9 13
```

```
##  [1]  5  5  7  5  5  5  5  5  5  9  9 13
```

Tree


1
42
n=296
2
38
n=189
yes
ROH90 < 5
no
3
48
n=107
4
37
n=113
ROH90 < 3
DOW = Saturday
7
49
n=88
9
39
n=38
Tuesday,Wednesday,Th
y,Friday,Saturday
14
48
n=81
ROH90 < 16
DOW = Sunday
DOW = Sun
29
50
n=68
DOW = Tuesday,Wednesday,Thursday,Friday,Sa
8
36
n=75
18
34
n=19
19
44
n=19
5
40
n=76
6
39
n=19
28
41
n=13
58
48
n=51
59
56
n=17
15
60
n=7

Support Vector Regression

After some mannual cross validation, we choose the `radial` kernel for this empirical study, then test different `gamma` values for the model. Here shows the selected γ values. Usually lower γ indicates more linear boundary.

```
## [[1]]
## [1] 0.0312
##
## [[2]]
## [1] 0.0312
##
## [[3]]
## [1] 0.0312
##
## [[4]]
## [1] 0.0312
##
## [[5]]
## [1] 0.0312
##
## [[6]]
## [1] 0.0625
##
## [[7]]
## [1] 0.0312
##
## [[8]]
## [1] 0.125
##
## [[9]]
## [1] 0.1
##
## [[10]]
## [1] 0.0625
##
## [[11]]
## [1] 0.0625
##
## [[12]]
```

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

Results

Table 4: Mean Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	-0.344	-0.629	-0.380	-3.26	0.354	-0.010	-0.075	-0.537	-0.755
DBA2	-0.294	-0.806	-0.373	-4.84	0.365	0.090	-0.335	-0.556	-0.848
DBA3	-0.299	-1.045	-0.446	-7.08	0.884	0.340	-0.768	-0.569	-0.838
DBA4	-0.042	-0.943	-0.270	-7.47	0.611	0.364	-0.417	-0.468	-0.817
DBA5	0.161	-0.888	-0.148	-8.99	0.973	0.316	-0.769	-0.665	-0.679
DBA6	0.338	-0.964	-0.031	-9.74	1.146	0.290	-0.631	-0.581	-0.507
DBA7	0.485	-1.098	0.068	-9.93	1.624	0.261	-0.592	-0.421	-0.645
DBA14	1.841	-0.756	1.521	-15.28	1.889	1.272	0.789	0.745	-0.200
DBA21	2.354	-1.374	2.201	-18.61	2.270	2.527	0.395	1.160	1.394
DBA30	3.061	-1.315	2.926	-23.45	3.981	2.977	2.266	2.588	2.793
DBA60	3.938	-1.976	3.544	-30.02	4.262	3.554	3.648	3.584	4.235
DBA90	4.422	-2.626	4.151	-32.65	3.843	3.441	4.213	4.310	5.299
13	1.302	-1.202	1.064	-14.28	1.850	1.285	0.644	0.716	0.703

Table 5: Mean Absolute Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	1.96	1.93	1.94	4.02	3.68	3.58	2.69	2.36	2.31
DBA2	2.57	2.60	2.54	5.19	3.75	3.98	3.19	2.86	2.88
DBA3	2.74	2.72	2.66	7.29	4.11	4.06	3.82	3.44	3.12
DBA4	3.06	3.19	2.97	7.97	4.13	4.28	3.82	3.60	3.32
DBA5	3.55	3.53	3.43	9.23	4.59	4.56	4.07	3.73	3.53
DBA6	4.04	3.98	3.88	9.93	5.21	4.86	4.55	3.96	4.17
DBA7	4.55	4.67	4.40	10.16	5.93	5.34	4.40	4.32	4.30
DBA14	6.03	6.58	5.84	15.72	7.26	6.30	5.73	5.39	5.46
DBA21	7.14	8.49	7.05	18.89	7.69	8.13	6.98	6.50	6.41
DBA30	8.33	10.56	8.22	23.55	9.24	8.73	7.78	8.01	7.84
DBA60	9.72	14.88	9.50	30.04	10.12	9.45	9.32	9.43	9.55
DBA90	10.77	22.73	10.70	32.65	10.85	11.94	11.01	10.96	11.20
13	5.37	7.15	5.26	14.55	6.38	6.27	5.61	5.38	5.34

Table 6: Standard Deviation Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	2.63	2.60	2.61	4.52	4.53	4.52	3.30	2.91	3.18
DBA2	3.39	3.31	3.35	3.97	4.60	4.93	3.92	3.58	3.75
DBA3	3.66	3.58	3.58	6.50	5.13	5.06	4.65	4.18	4.04
DBA4	4.06	4.08	3.96	5.81	5.21	5.24	4.82	4.40	4.22
DBA5	4.51	4.49	4.38	6.13	5.74	5.58	5.07	4.48	4.44
DBA6	4.98	5.04	4.83	5.59	6.47	5.94	5.51	4.74	5.24
DBA7	5.48	5.87	5.36	6.25	7.28	6.66	5.45	5.22	5.26
DBA14	7.37	8.73	7.19	9.77	8.77	8.00	7.18	7.06	7.30
DBA21	8.64	11.60	8.54	10.32	9.28	9.81	8.68	8.49	8.32
DBA30	10.34	12.99	10.22	11.10	11.01	10.63	9.98	10.14	9.81
DBA60	11.96	17.99	11.60	13.21	12.57	12.11	11.48	11.82	11.73
DBA90	13.05	29.42	12.93	14.21	13.38	14.76	13.27	13.38	13.36
13	6.67	9.14	6.55	8.11	7.83	7.77	6.94	6.70	6.72

Table 7: Mean Percentage Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.006	-0.025	0.003	-0.108	0.063	0.051	0.016	-0.023	-0.008
DBA2	0.019	-0.033	0.012	-0.143	0.065	0.057	0.002	-0.019	-0.007
DBA3	0.028	-0.042	0.015	-0.221	0.111	0.079	-0.007	-0.017	-0.005
DBA4	0.041	-0.045	0.020	-0.237	0.091	0.076	0.006	-0.016	0.000
DBA5	0.058	-0.043	0.031	-0.264	0.110	0.070	0.001	-0.018	0.009
DBA6	0.077	-0.043	0.044	-0.309	0.131	0.074	0.008	-0.017	0.020
DBA7	0.090	-0.051	0.053	-0.285	0.160	0.082	0.013	-0.008	0.028
DBA14	0.193	-0.041	0.163	-0.380	0.211	0.170	0.059	0.061	0.062
DBA21	0.252	-0.060	0.238	-0.476	0.241	0.279	0.078	0.113	0.174
DBA30	0.325	-0.036	0.313	-0.644	0.373	0.319	0.215	0.272	0.287
DBA60	0.417	-0.035	0.388	-0.817	0.426	0.373	0.361	0.380	0.414
DBA90	0.463	-0.009	0.447	-0.885	0.447	0.444	0.451	0.464	0.503
13	0.164	-0.039	0.144	-0.397	0.202	0.173	0.100	0.098	0.123

Table 8: MAPE

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.060	0.061	0.059	0.131	0.149	0.145	0.098	0.085	0.074
DBA2	0.091	0.087	0.088	0.158	0.151	0.159	0.112	0.100	0.093
DBA3	0.100	0.094	0.094	0.237	0.181	0.166	0.129	0.117	0.102
DBA4	0.113	0.112	0.103	0.256	0.170	0.170	0.134	0.124	0.108
DBA5	0.137	0.121	0.123	0.278	0.190	0.181	0.137	0.125	0.116
DBA6	0.159	0.128	0.139	0.321	0.222	0.191	0.148	0.130	0.138
DBA7	0.180	0.146	0.156	0.305	0.258	0.211	0.147	0.142	0.147
DBA14	0.283	0.207	0.257	0.425	0.329	0.278	0.203	0.192	0.195
DBA21	0.352	0.266	0.341	0.507	0.358	0.396	0.254	0.252	0.281
DBA30	0.426	0.349	0.415	0.657	0.475	0.433	0.328	0.383	0.385
DBA60	0.529	0.499	0.506	0.819	0.543	0.495	0.469	0.494	0.516
DBA90	0.586	0.714	0.576	0.885	0.583	0.612	0.583	0.591	0.616
13	0.251	0.232	0.238	0.415	0.301	0.286	0.228	0.228	0.231

Table 9: Model Performances

	ME	MAE	SDE	MPE	MAPE	Time
Additive Pickup	1.302	5.37	6.67	0.164	0.251	0.178
Multiplicative Pickup	-1.202	7.15	9.14	-0.039	0.232	0.173
Regression	1.064	5.26	6.55	0.144	0.238	0.096
Neural Network	-14.276	14.55	8.11	-0.397	0.415	153.194
K-Nearest Neighbor	1.850	6.38	7.83	0.202	0.301	49.106
Weighted K-Nearest Neighbor	1.285	6.27	7.77	0.173	0.286	2.139
Decision Tree	0.644	5.61	6.94	0.100	0.228	0.188
Random Forest	0.716	5.38	6.70	0.098	0.228	289.021
Support Vector Machine	0.703	5.34	6.72	0.123	0.231	27.428

[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