

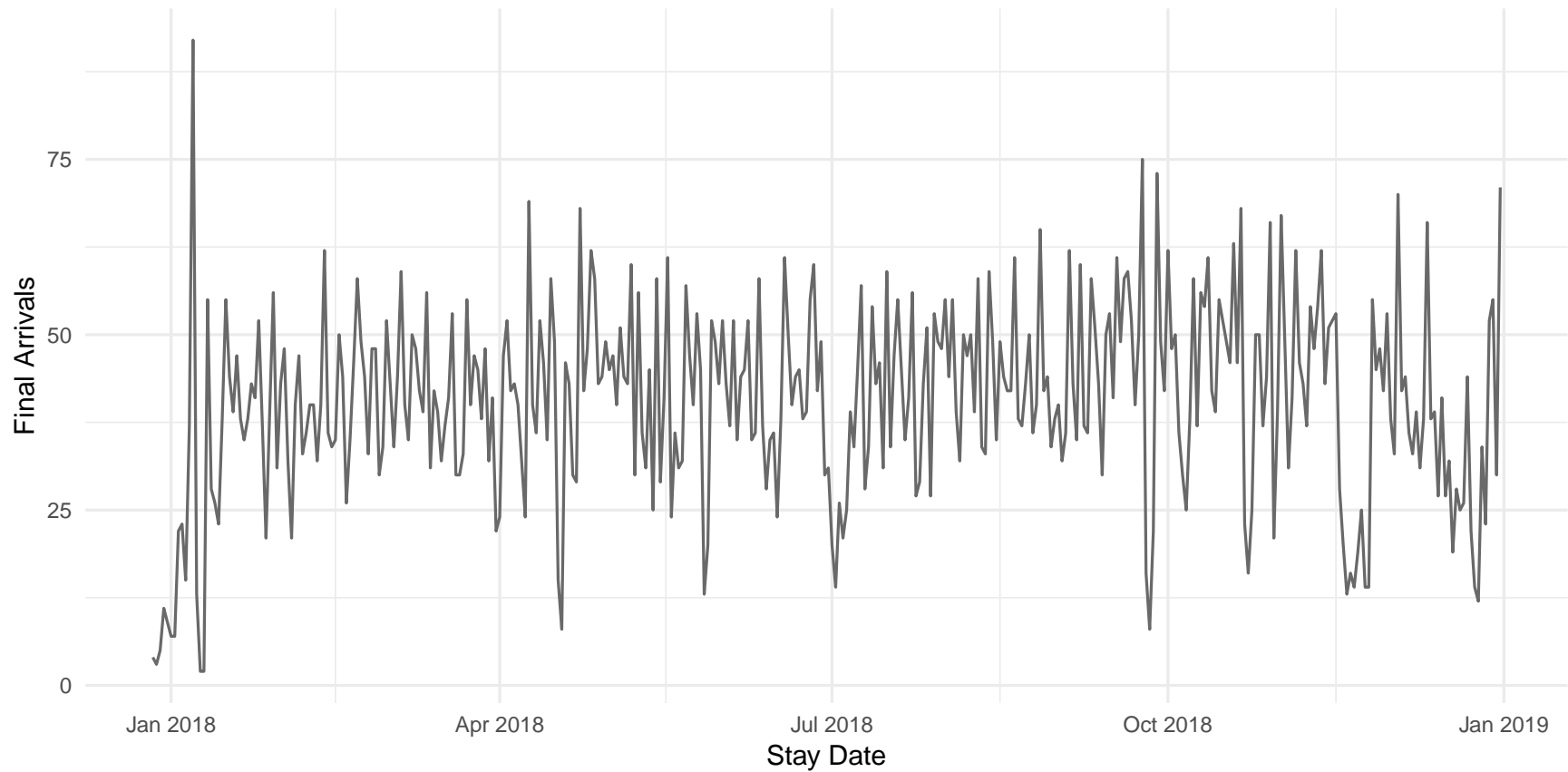
Pick-up method + machine learning: a proved efficient approach to forecast hotel demand (Robust test for newest ROH)

Rachel Zhang

06/18/2020

Data

Import Dataset and Cross-Validation



This robust analysis is conducted under R version 3.6.0 with `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
2018-06-23	38	Saturday	38	38	33	32	31	29	29	19	14	14	10	8
2018-01-09	2	Tuesday	2	2	2	2	2	2	2	2	2	1	0	0
2018-07-09	57	Monday	54	51	51	46	44	41	39	31	8	5	5	4
2018-10-28	44	Sunday	42	40	37	33	31	26	24	22	18	13	5	0
2018-04-23	68	Monday	67	67	66	61	59	58	58	42	34	23	9	5
2018-10-21	68	Sunday	67	67	66	66	64	59	54	44	20	18	10	5
2018-08-12	33	Sunday	32	29	28	28	25	22	21	18	17	13	7	3
2018-08-27	65	Monday	60	60	58	55	53	51	47	44	38	26	16	10
2018-12-30	30	Sunday	24	22	20	19	19	19	18	16	13	8	7	4
2018-05-28	20	Monday	19	18	18	17	17	17	16	12	11	9	5	2

Modeling

Additive Pick-up

Table 2: Additive Pick Ups

DOW	ROH1	ROH2	ROH3	ROH4	ROH5	ROH6	ROH7	ROH14	ROH21	ROH30	ROH60	ROH90
Sunday	2.16	3.09	4.43	5.68	6.57	8.21	9.25	13.8	18.4	22.1	29.4	32.9
Monday	2.95	4.41	5.33	7.67	9.33	10.97	12.56	20.6	26.4	33.1	42.6	46.0
Tuesday	2.24	4.00	4.42	4.78	5.95	6.98	8.88	15.7	20.6	25.8	32.1	35.0
Wednesday	2.10	3.90	5.55	6.08	6.60	7.92	9.30	16.1	21.7	26.2	33.1	36.1
Thursday	3.33	5.14	6.95	8.43	8.98	9.81	10.95	16.8	21.0	25.0	33.6	37.2
Friday	3.89	6.33	8.16	9.78	11.44	12.27	13.00	18.7	22.4	26.4	34.5	38.4
Saturday	3.76	5.64	7.67	9.44	10.73	11.91	12.53	16.6	19.6	22.9	29.6	32.3

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	0.936	0.906	0.870	0.831	0.808	0.764	0.738	0.612	0.494	0.397	0.213	0.120
Monday	0.937	0.905	0.888	0.842	0.812	0.780	0.748	0.594	0.482	0.358	0.161	0.088
Tuesday	0.941	0.894	0.883	0.874	0.842	0.819	0.769	0.605	0.479	0.329	0.158	0.087
Wednesday	0.948	0.902	0.858	0.846	0.833	0.801	0.769	0.611	0.477	0.368	0.175	0.094
Thursday	0.913	0.870	0.828	0.795	0.781	0.758	0.731	0.583	0.484	0.387	0.184	0.102
Friday	0.905	0.849	0.807	0.770	0.731	0.713	0.695	0.566	0.482	0.390	0.204	0.119
Saturday	0.899	0.851	0.797	0.752	0.720	0.690	0.673	0.563	0.478	0.388	0.199	0.124

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
```

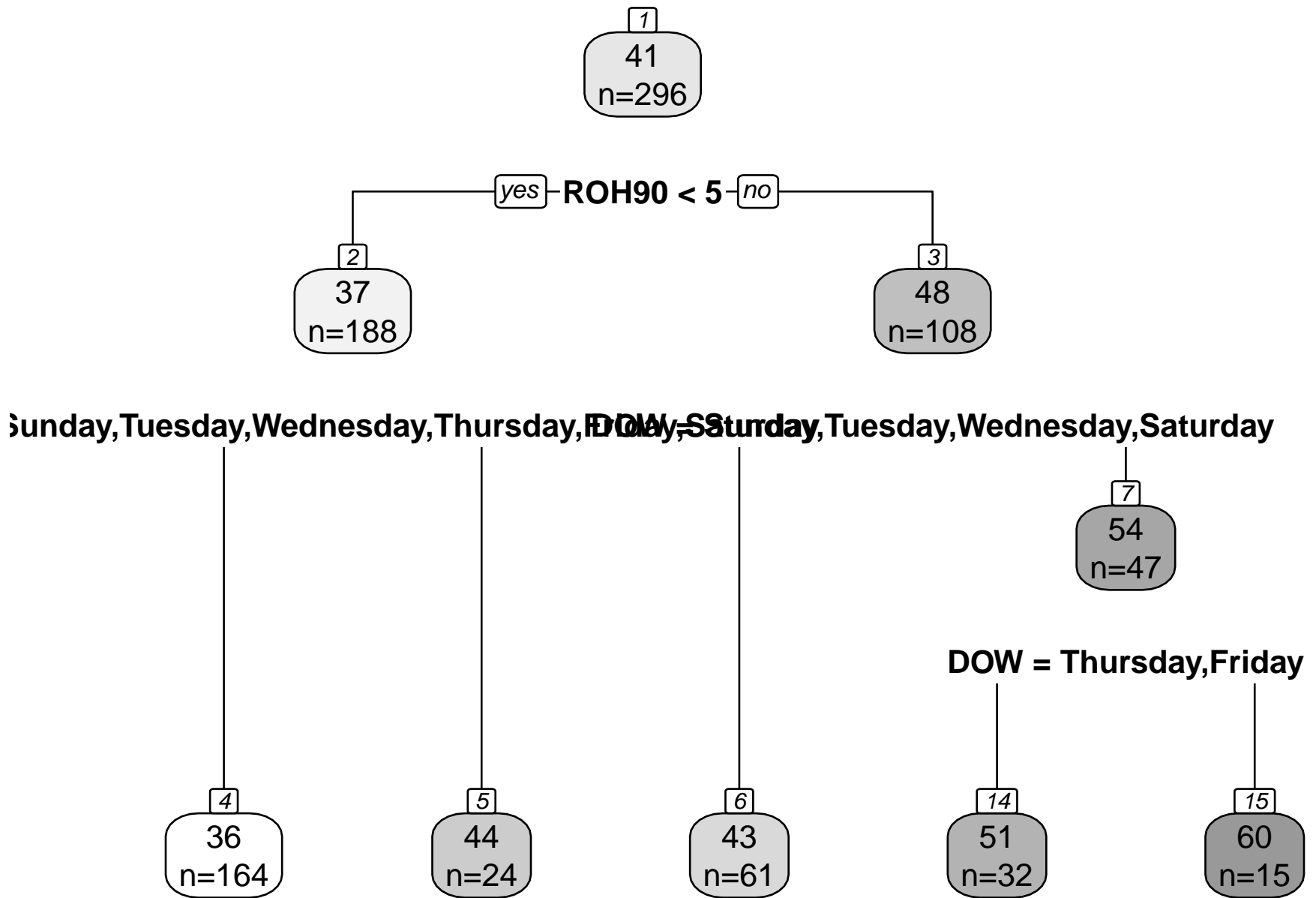
K-Nearest Neighbor

```
## [1] 5 5 5 5 5 5 5 5 5 5 7 11 11
## [1] 5 5 5 5 5 5 5 5 5 5 7 11 11
```

	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)	1.89*** (0.51)	2.51*** (0.68)	3.39*** (0.76)	4.27*** (0.86)	4.75*** (0.93)	6.24*** (1.00)	7.30*** (1.12)	11.81*** (1.54)	17.79*** (1.73)	20.94*** (1.85)	26.22*** (1.97)	30.78*** (1.94)
DOWMonday	0.70 (0.52)	1.13 (0.69)	0.54 (0.79)	1.52 (0.90)	2.19* (0.96)	2.12* (1.05)	2.68* (1.19)	6.29*** (1.66)	7.80*** (1.98)	10.87*** (2.16)	13.46*** (2.42)	13.36*** (2.58)
DOWTuesday	0.08 (0.50)	0.92 (0.67)	-0.03 (0.75)	-0.96 (0.86)	-0.67 (0.93)	-1.33 (1.01)	-0.42 (1.14)	2.07 (1.62)	2.24 (1.94)	3.93 (2.14)	3.70 (2.41)	2.90 (2.56)
DOWWednesday	-0.07 (0.50)	0.79 (0.67)	1.09 (0.76)	0.32 (0.87)	-0.08 (0.94)	-0.43 (1.02)	-0.08 (1.15)	2.28 (1.63)	3.29 (1.95)	4.26* (2.14)	4.37 (2.42)	3.82 (2.57)
DOWThursday	1.16* (0.50)	2.03** (0.66)	2.49** (0.75)	2.71** (0.85)	2.34* (0.92)	1.47 (1.01)	1.57 (1.14)	2.91 (1.61)	2.48 (1.92)	2.88 (2.11)	4.42 (2.38)	4.59 (2.53)
DOWFriday	1.70*** (0.49)	3.19*** (0.65)	3.65*** (0.74)	4.00*** (0.84)	4.80*** (0.91)	3.92*** (0.99)	3.58** (1.12)	4.78** (1.58)	3.85* (1.89)	4.20* (2.08)	4.74* (2.34)	5.18* (2.49)
DOWSaturday	1.62** (0.49)	2.62*** (0.65)	3.37*** (0.74)	3.98*** (0.85)	4.47*** (0.91)	4.03*** (1.00)	3.58** (1.12)	3.15* (1.59)	1.29 (1.89)	0.94 (2.08)	0.71 (2.34)	-0.35 (2.48)
ROH1	1.01*** (0.01)											
ROH2		1.02*** (0.01)										
ROH3			1.03*** (0.02)									
ROH4				1.04*** (0.02)								
ROH5					1.06*** (0.02)							
ROH6						1.07*** (0.02)						
ROH7							1.07*** (0.03)					
ROH14								1.08*** (0.04)				
ROH21									1.03*** (0.06)			
ROH30										1.07*** (0.07)		
ROH60											1.38*** (0.13)	
ROH90												1.43*** (0.17)
R ²	0.97	0.95	0.94	0.92	0.90	0.89	0.85	0.71	0.58	0.50	0.37	0.28
Adj. R ²	0.97	0.95	0.94	0.92	0.90	0.88	0.85	0.70	0.57	0.49	0.35	0.27
Num. obs.	296	296	296	296	296	296	296	296	296	296	296	296
RMSE	2.29	3.07	3.47	3.96	4.28	4.67	5.27	7.47	8.91	9.79	11.02	11.71

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

Tree



Support Vector Regression

After some manual 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] 0.0312
##
## [[7]]
## [1] 0.0312
##
## [[8]]
## [1] 0.0312
##
## [[9]]
## [1] 0.0625
##
## [[10]]
## [1] 0.0625
##
## [[11]]
## [1] 0.0625
##
## [[12]]
```


[1] 0.0312

Results

Table 5: Mean Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.142	0.093	0.119	-2.87	0.745	0.246	0.398	-0.086	0.088
DBA2	0.225	0.102	0.175	-4.51	1.108	0.444	0.518	0.043	-0.040
DBA3	0.275	0.046	0.183	-5.80	1.205	1.170	0.825	0.285	0.015
DBA4	0.548	0.272	0.431	-6.97	1.591	1.155	1.174	0.701	0.187
DBA5	0.557	0.162	0.402	-8.15	1.582	0.685	0.883	0.522	0.068
DBA6	0.954	0.517	0.804	-8.92	1.612	0.608	0.975	0.604	0.429
DBA7	1.182	0.671	1.043	-9.81	1.617	1.256	0.916	0.689	0.361
DBA14	0.657	-1.159	0.446	-16.53	1.350	1.207	0.437	0.877	-0.068
DBA21	1.233	-1.307	1.168	-20.36	2.263	2.092	1.211	1.733	1.545
DBA30	1.420	-2.278	1.291	-24.74	1.769	1.506	0.359	1.048	1.321
DBA60	2.542	-0.909	2.294	-31.64	2.908	2.454	2.114	2.550	2.545
DBA90	2.573	-3.569	2.304	-34.74	1.805	2.079	2.383	2.334	3.361
13	1.025	-0.613	0.888	-14.59	1.630	1.242	1.016	0.942	0.818

Table 6: Mean Absolute Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	2.02	2.10	2.00	3.07	2.98	2.65	2.44	2.14	2.05
DBA2	2.57	2.73	2.56	4.59	3.48	3.35	3.60	2.88	2.56
DBA3	2.81	3.10	2.77	5.82	3.85	4.01	3.50	3.33	2.74
DBA4	3.03	3.42	2.97	7.02	4.07	4.26	3.27	3.67	2.91
DBA5	3.30	3.51	3.20	8.17	4.15	4.05	3.97	3.69	3.17
DBA6	3.73	3.69	3.45	8.92	4.22	3.87	3.98	3.67	3.33
DBA7	4.29	4.15	4.01	9.87	4.69	4.43	4.43	4.04	3.75
DBA14	6.38	6.33	6.19	16.53	6.19	6.37	5.74	5.86	5.97
DBA21	7.47	7.22	7.40	20.42	7.21	7.17	7.51	6.83	6.93
DBA30	8.89	9.64	8.82	24.77	8.15	8.44	8.69	8.06	8.39
DBA60	9.94	14.94	9.81	31.67	9.90	10.83	9.47	9.60	9.85
DBA90	10.46	22.18	10.55	34.75	10.41	10.79	10.67	10.55	10.78
13	5.41	6.92	5.31	14.63	5.78	5.85	5.61	5.36	5.20

Table 7: Standard Deviation Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	2.52	2.63	2.51	2.56	4.28	3.52	3.02	2.67	2.59
DBA2	3.18	3.38	3.16	3.22	4.57	4.38	4.46	3.62	3.21
DBA3	3.66	3.95	3.62	3.65	5.20	5.44	4.44	4.36	3.63
DBA4	3.89	4.18	3.81	3.94	5.24	5.57	3.97	4.62	3.76
DBA5	4.09	4.31	3.95	4.17	5.35	5.08	4.89	4.65	3.92
DBA6	4.34	4.48	4.13	4.39	5.31	5.01	4.80	4.52	4.06
DBA7	4.88	4.95	4.64	5.11	5.71	5.53	5.25	5.01	4.55
DBA14	8.02	8.44	7.80	8.22	8.06	7.94	7.45	7.32	7.69
DBA21	9.39	9.70	9.30	10.13	9.45	9.37	9.51	8.88	8.69
DBA30	10.97	12.25	10.84	11.86	10.62	11.07	10.67	10.40	10.34
DBA60	12.75	19.42	12.53	13.57	12.97	13.87	11.85	12.09	12.53
DBA90	13.68	28.31	13.72	14.32	13.68	14.07	13.88	13.63	13.85
13	6.78	8.83	6.67	7.10	7.54	7.57	7.01	6.81	6.57

Table 8: Mean Percentage Errors

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.028	-0.009	0.023	-0.099	0.197	0.089	0.089	0.015	0.050
DBA2	0.051	-0.014	0.041	-0.143	0.205	0.101	0.101	0.019	0.057
DBA3	0.071	-0.018	0.053	-0.197	0.228	0.124	0.111	0.027	0.070
DBA4	0.089	-0.016	0.064	-0.233	0.234	0.127	0.125	0.039	0.069
DBA5	0.103	-0.018	0.071	-0.261	0.238	0.106	0.123	0.035	0.065
DBA6	0.133	-0.007	0.099	-0.290	0.222	0.103	0.122	0.040	0.091
DBA7	0.161	0.002	0.127	-0.289	0.225	0.123	0.134	0.046	0.109
DBA14	0.249	-0.034	0.216	-0.441	0.261	0.189	0.132	0.152	0.189
DBA21	0.344	-0.022	0.333	-0.488	0.350	0.322	0.247	0.252	0.305
DBA30	0.421	-0.024	0.404	-0.598	0.379	0.379	0.253	0.281	0.368
DBA60	0.554	0.063	0.516	-0.795	0.592	0.585	0.401	0.496	0.541
DBA90	0.593	0.002	0.567	-0.869	0.566	0.524	0.575	0.578	0.635
13	0.233	-0.008	0.210	-0.392	0.308	0.231	0.201	0.165	0.212

Table 9: MAPE

	apk	mpk	reg	nn	knn	wknn	dtree	rf	svm
DBA1	0.071	0.066	0.068	0.115	0.249	0.148	0.147	0.077	0.099
DBA2	0.106	0.087	0.099	0.154	0.257	0.166	0.179	0.095	0.119
DBA3	0.127	0.098	0.115	0.200	0.286	0.191	0.180	0.113	0.134
DBA4	0.142	0.112	0.126	0.243	0.291	0.200	0.179	0.126	0.134
DBA5	0.160	0.117	0.137	0.264	0.296	0.189	0.196	0.123	0.140
DBA6	0.189	0.119	0.153	0.291	0.279	0.181	0.191	0.126	0.155
DBA7	0.224	0.129	0.188	0.299	0.293	0.199	0.213	0.137	0.180
DBA14	0.363	0.178	0.331	0.441	0.357	0.297	0.240	0.251	0.311
DBA21	0.464	0.203	0.454	0.522	0.443	0.418	0.368	0.348	0.410
DBA30	0.568	0.276	0.552	0.615	0.502	0.515	0.420	0.418	0.508
DBA60	0.696	0.462	0.661	0.800	0.725	0.749	0.537	0.631	0.681
DBA90	0.746	0.675	0.728	0.874	0.733	0.696	0.739	0.737	0.777
13	0.321	0.210	0.301	0.402	0.393	0.329	0.299	0.265	0.304

Table 10: Model Performances

	ME	MAE	SDE	MPE	MAPE	Time
Additive Pickup	1.025	5.41	6.78	0.233	0.321	0.136
Multiplicative Pickup	-0.613	6.92	8.83	-0.008	0.210	0.137
Regression	0.888	5.31	6.67	0.210	0.301	0.064
Neural Network	-14.586	14.63	7.10	-0.392	0.402	33.730
K-Nearest Neighbor	1.630	5.78	7.54	0.308	0.393	18.631
Weighted K-Nearest Neighbor	1.242	5.85	7.57	0.231	0.329	1.830
Decision Tree	1.016	5.61	7.01	0.201	0.299	0.096
Random Forest	0.942	5.36	6.81	0.165	0.265	174.460
Support Vector Machine	0.818	5.20	6.57	0.212	0.304	40.203

¹ Mean Error² Mean Absolute Error³ Standard Deviation Error⁴ Mean Percentage Error⁵ Mean Absolute Percentage Error⁶ Time is calculated in seconds