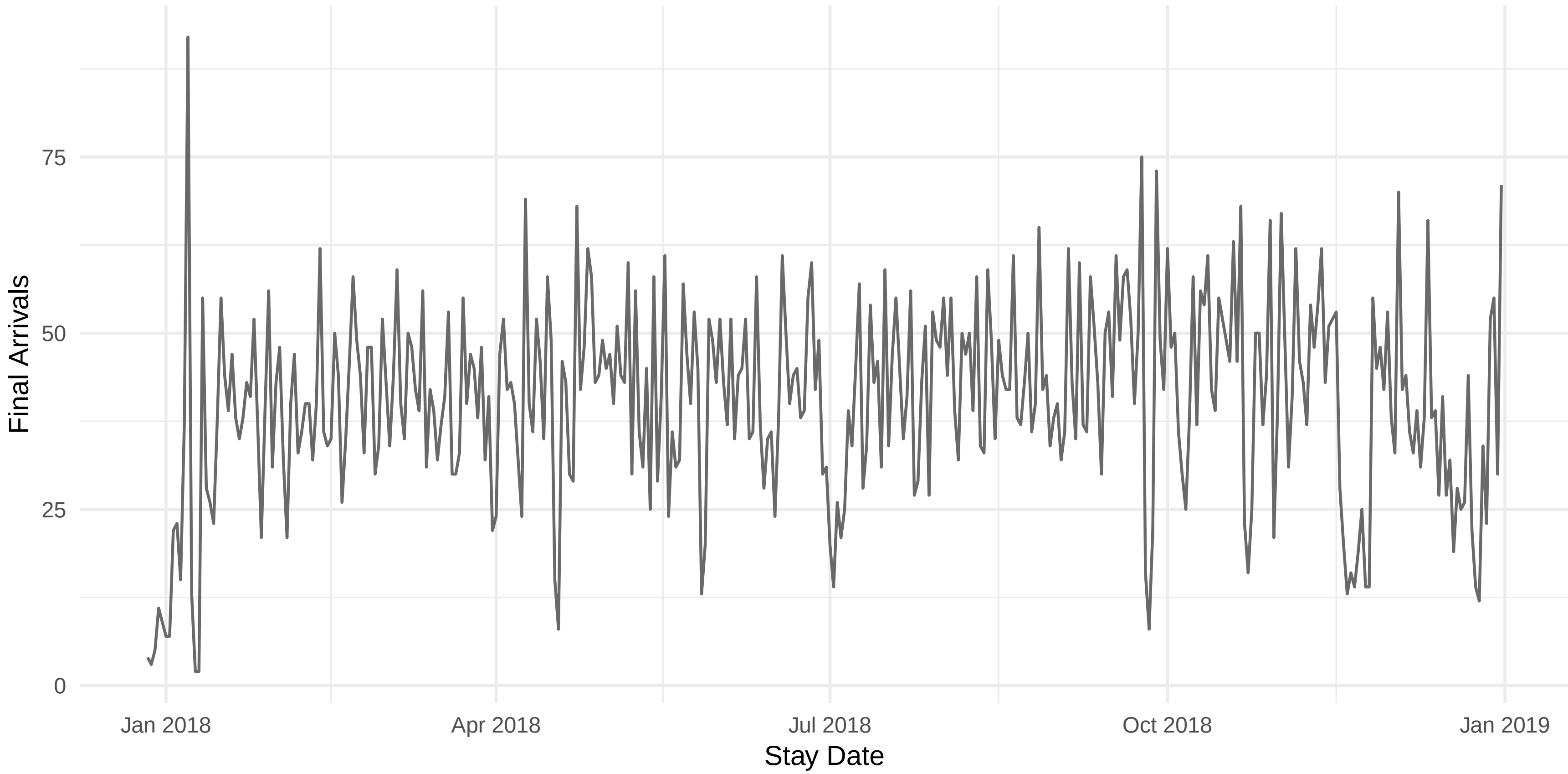
# Pick-up method + machine learning: a proved efficient approach to forecast hotel demand

*Rachel Zhang*

*06/18/2020*

# Data

## Import Dataset and Cross-Validation




This 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2018-06-23 | 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.

Call: lm(formula = ROH0 ~ ., data = this.train)

Coefficients: (Intercept) DOWMonday DOWTuesday DOWWednesday DOWThursday
2.00699 0.33631 -0.48085 -0.37901 1.02595
DOWFriday DOWSaturday ROH1 ROH2 ROH3
1.53349 1.48209 1.09756 -0.07892 -0.00319
ROH4 ROH5 ROH6 ROH7 ROH14
0.07200 -0.10278 0.24261 -0.18147 -0.06584
ROH21 ROH30 ROH60 ROH90
0.07009 -0.09603 0.01900 -0.00188

### 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.

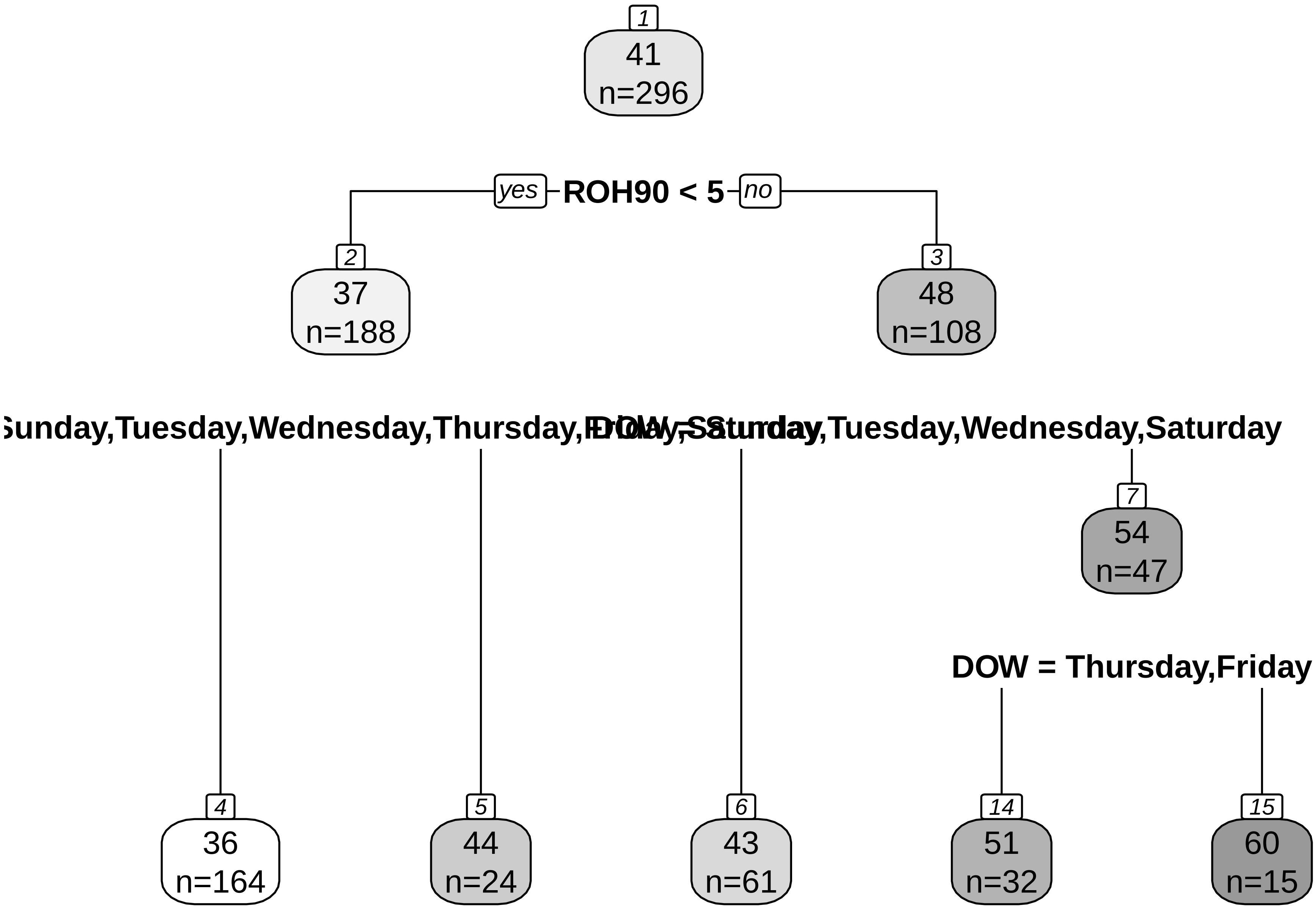
```
## pdf
##   2
```

### K-Nearest Neighbor

```
##  [1]  5  7  5  5  5  5  5  5  5 13 11 11
```

```
##  [1]  5  7  5  5  5  5  5  5  5 13 11 11
```

**Tree**


1
41
n=296
yes
ROH90 < 5
no
2
37
n=188
3
48
n=108
Sunday,Tuesday,Wednesday,Thursday,Friday,Saturday
DOW = Sunday,Tuesday,Wednesday,Saturday
7
54
n=47
DOW = Thursday,Friday
4
36
n=164
5
44
n=24
6
43
n=61
14
51
n=32
15
60
n=15

### 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 $\gamma$ values. Usually lower $\gamma$ 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] 0.0625
##
## [[8]]
## [1] 0.0312
##
## [[9]]
## [1] 0.0312
##
## [[10]]
## [1] 0.1
##
## [[11]]
## [1] 0.0312
##
## [[12]]
```

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

## Results

Table 4: Mean Errors

| | apk | mpk | reg | nn | knn | wknn | dtree | rf | svm |
|---|---|---|---|---|---|---|---|---|---|
| DBA1 | 0.142 | 0.093 | 0.179 | -2.52 | 0.324 | 0.458 | 0.398 | 0.050 | 0.194 |
| DBA2 | 0.225 | 0.102 | 0.300 | -4.37 | 0.834 | 0.677 | 0.518 | 0.140 | 0.162 |
| DBA3 | 0.275 | 0.046 | 0.362 | -6.11 | 0.727 | 0.649 | 0.877 | 0.296 | 0.400 |
| DBA4 | 0.548 | 0.272 | 0.726 | -7.10 | 0.976 | 0.994 | 1.026 | 0.713 | 0.550 |
| DBA5 | 0.557 | 0.162 | 0.630 | -7.61 | 1.370 | 1.104 | 0.686 | 0.581 | 0.663 |
| DBA6 | 0.954 | 0.517 | 1.096 | -8.99 | 1.345 | 1.343 | 0.975 | 0.802 | 0.951 |
| DBA7 | 1.182 | 0.671 | 1.375 | -9.86 | 1.314 | 1.469 | 0.916 | 0.896 | 1.264 |
| DBA14 | 0.657 | -1.159 | 0.429 | -16.91 | 1.459 | 0.799 | 0.437 | 0.128 | 0.069 |
| DBA21 | 1.233 | -1.307 | 1.239 | -20.24 | 2.065 | 2.101 | 1.211 | 0.903 | 1.318 |
| DBA30 | 1.420 | -2.278 | 1.323 | -24.73 | 2.440 | 2.105 | 0.556 | 1.505 | 1.245 |
| DBA60 | 2.542 | -0.909 | 2.330 | -31.32 | 2.906 | 2.985 | 2.427 | 2.406 | 2.353 |
| DBA90 | 2.573 | -3.569 | 2.304 | -34.67 | 1.805 | 2.079 | 2.383 | 2.387 | 3.361 |
| 13 | 1.025 | -0.613 | 1.024 | -14.54 | 1.464 | 1.397 | 1.034 | 0.901 | 1.044 |

Table 5: Mean Absolute Errors

| | apk | mpk | reg | nn | knn | wknn | dtree | rf | svm |
|---|---|---|---|---|---|---|---|---|---|
| DBA1 | 2.02 | 2.10 | 1.97 | 3.01 | 3.38 | 3.47 | 2.44 | 2.30 | 2.42 |
| DBA2 | 2.57 | 2.73 | 2.77 | 4.57 | 3.93 | 3.81 | 3.60 | 2.95 | 2.79 |
| DBA3 | 2.81 | 3.10 | 2.99 | 6.25 | 4.08 | 3.91 | 3.53 | 3.04 | 3.02 |
| DBA4 | 3.03 | 3.42 | 3.20 | 7.28 | 4.32 | 4.24 | 3.15 | 3.39 | 3.25 |
| DBA5 | 3.30 | 3.51 | 3.17 | 7.86 | 4.67 | 4.38 | 3.72 | 3.50 | 3.42 |
| DBA6 | 3.73 | 3.69 | 3.52 | 8.99 | 4.99 | 4.52 | 3.98 | 3.60 | 3.66 |
| DBA7 | 4.29 | 4.15 | 3.92 | 9.99 | 5.56 | 4.81 | 4.43 | 4.12 | 4.02 |
| DBA14 | 6.38 | 6.33 | 6.12 | 16.91 | 6.92 | 6.62 | 5.74 | 5.60 | 5.94 |
| DBA21 | 7.47 | 7.22 | 7.31 | 20.34 | 8.53 | 7.89 | 7.51 | 6.98 | 7.14 |
| DBA30 | 8.89 | 9.64 | 8.87 | 24.82 | 9.27 | 8.80 | 8.97 | 8.87 | 8.31 |
| DBA60 | 9.94 | 14.94 | 9.75 | 31.40 | 9.75 | 9.86 | 9.60 | 9.84 | 10.04 |
| DBA90 | 10.46 | 22.18 | 10.55 | 34.67 | 10.41 | 10.79 | 10.67 | 10.56 | 10.78 |
| 13 | 5.41 | 6.92 | 5.34 | 14.67 | 6.32 | 6.09 | 5.61 | 5.39 | 5.40 |

Table 6: Standard Deviation Errors

| | apk | mpk | reg | nn | knn | wknn | dtree | rf | svm |
|---|---|---|---|---|---|---|---|---|---|
| DBA1 | 2.52 | 2.63 | 2.53 | 2.87 | 4.30 | 4.61 | 3.02 | 2.84 | 3.16 |
| DBA2 | 3.18 | 3.38 | 3.43 | 3.54 | 4.86 | 4.98 | 4.46 | 3.69 | 3.64 |
| DBA3 | 3.66 | 3.95 | 3.95 | 4.38 | 5.22 | 5.11 | 4.45 | 3.85 | 4.03 |
| DBA4 | 3.89 | 4.18 | 4.22 | 5.13 | 5.56 | 5.45 | 3.83 | 4.22 | 4.25 |
| DBA5 | 4.09 | 4.31 | 4.15 | 4.49 | 6.17 | 5.68 | 4.62 | 4.24 | 4.37 |
| DBA6 | 4.34 | 4.48 | 4.43 | 4.75 | 6.54 | 5.81 | 4.80 | 4.34 | 4.42 |
| DBA7 | 4.88 | 4.95 | 4.86 | 5.16 | 7.17 | 6.24 | 5.25 | 4.90 | 4.89 |
| DBA14 | 8.02 | 8.44 | 7.67 | 8.03 | 8.61 | 8.60 | 7.45 | 7.37 | 7.66 |
| DBA21 | 9.39 | 9.70 | 9.25 | 9.77 | 10.92 | 9.89 | 9.51 | 8.80 | 8.86 |
| DBA30 | 10.97 | 12.25 | 10.91 | 11.82 | 11.94 | 11.09 | 11.31 | 10.93 | 10.23 |
| DBA60 | 12.75 | 19.42 | 12.44 | 13.72 | 13.13 | 12.75 | 12.16 | 12.44 | 12.70 |
| DBA90 | 13.68 | 28.31 | 13.72 | 14.25 | 13.68 | 14.07 | 13.88 | 13.64 | 13.85 |
| 13 | 6.78 | 8.83 | 6.80 | 7.32 | 8.18 | 7.86 | 7.06 | 6.77 | 6.84 |

Table 7: Mean Percentage Errors

| | apk | mpk | reg | nn | knn | wknn | dtree | rf | svm |
|---|---|---|---|---|---|---|---|---|---|
| DBA1 | 0.028 | -0.009 | 0.021 | -0.076 | 0.082 | 0.097 | 0.089 | 0.015 | 0.063 |
| DBA2 | 0.051 | -0.014 | 0.038 | -0.134 | 0.162 | 0.127 | 0.101 | 0.022 | 0.066 |
| DBA3 | 0.071 | -0.018 | 0.051 | -0.170 | 0.171 | 0.127 | 0.111 | 0.027 | 0.071 |
| DBA4 | 0.089 | -0.016 | 0.061 | -0.212 | 0.201 | 0.120 | 0.118 | 0.038 | 0.080 |
| DBA5 | 0.103 | -0.018 | 0.063 | -0.224 | 0.225 | 0.120 | 0.114 | 0.039 | 0.086 |
| DBA6 | 0.133 | -0.007 | 0.089 | -0.282 | 0.229 | 0.125 | 0.122 | 0.044 | 0.095 |
| DBA7 | 0.161 | 0.002 | 0.115 | -0.281 | 0.263 | 0.186 | 0.134 | 0.055 | 0.127 |
| DBA14 | 0.249 | -0.034 | 0.199 | -0.483 | 0.298 | 0.254 | 0.132 | 0.092 | 0.179 |
| DBA21 | 0.344 | -0.022 | 0.333 | -0.492 | 0.415 | 0.384 | 0.247 | 0.188 | 0.297 |
| DBA30 | 0.421 | -0.024 | 0.406 | -0.642 | 0.526 | 0.445 | 0.260 | 0.372 | 0.342 |
| DBA60 | 0.554 | 0.063 | 0.515 | -0.763 | 0.597 | 0.568 | 0.409 | 0.520 | 0.541 |
| DBA90 | 0.593 | 0.002 | 0.567 | -0.871 | 0.566 | 0.524 | 0.575 | 0.583 | 0.635 |
| 13 | 0.233 | -0.008 | 0.205 | -0.386 | 0.311 | 0.256 | 0.201 | 0.166 | 0.215 |

Table 8: MAPE

| | apk | mpk | reg | nn | knn | wknn | dtree | rf | svm |
|---|---|---|---|---|---|---|---|---|---|
| DBA1 | 0.071 | 0.066 | 0.064 | 0.101 | 0.151 | 0.167 | 0.147 | 0.085 | 0.119 |
| DBA2 | 0.106 | 0.087 | 0.101 | 0.155 | 0.224 | 0.201 | 0.179 | 0.103 | 0.134 |
| DBA3 | 0.127 | 0.098 | 0.118 | 0.181 | 0.243 | 0.204 | 0.180 | 0.108 | 0.139 |
| DBA4 | 0.142 | 0.112 | 0.127 | 0.229 | 0.273 | 0.197 | 0.173 | 0.117 | 0.151 |
| DBA5 | 0.160 | 0.117 | 0.132 | 0.239 | 0.292 | 0.199 | 0.185 | 0.121 | 0.157 |
| DBA6 | 0.189 | 0.119 | 0.147 | 0.282 | 0.302 | 0.200 | 0.191 | 0.123 | 0.158 |
| DBA7 | 0.224 | 0.129 | 0.169 | 0.298 | 0.349 | 0.253 | 0.213 | 0.140 | 0.186 |
| DBA14 | 0.363 | 0.178 | 0.314 | 0.483 | 0.408 | 0.369 | 0.240 | 0.204 | 0.297 |
| DBA21 | 0.464 | 0.203 | 0.449 | 0.533 | 0.541 | 0.495 | 0.368 | 0.304 | 0.410 |
| DBA30 | 0.568 | 0.276 | 0.555 | 0.649 | 0.657 | 0.574 | 0.429 | 0.513 | 0.481 |
| DBA60 | 0.696 | 0.462 | 0.657 | 0.804 | 0.728 | 0.700 | 0.542 | 0.661 | 0.688 |
| DBA90 | 0.746 | 0.675 | 0.728 | 0.871 | 0.733 | 0.696 | 0.739 | 0.741 | 0.777 |
| 13 | 0.321 | 0.210 | 0.297 | 0.402 | 0.408 | 0.355 | 0.299 | 0.268 | 0.308 |

Table 9: Model Performances

| | ME | MAE | SDE | MPE | MAPE | Time |
|---|---|---|---|---|---|---|
| Additive Pickup | 1.025 | 5.41 | 6.78 | 0.233 | 0.321 | 0.277 |
| Multiplicative Pickup | -0.613 | 6.92 | 8.83 | -0.008 | 0.210 | 0.372 |
| Regression | 1.024 | 5.34 | 6.80 | 0.205 | 0.297 | 0.229 |
| Neural Network | -14.535 | 14.67 | 7.32 | -0.386 | 0.402 | 197.364 |
| K-Nearest Neighbor | 1.464 | 6.32 | 8.18 | 0.311 | 0.408 | 24.257 |
| Weighted K-Nearest Neighbor | 1.397 | 6.09 | 7.86 | 0.256 | 0.355 | 2.537 |
| Decision Tree | 1.034 | 5.61 | 7.06 | 0.201 | 0.299 | 0.156 |
| Random Forest | 0.901 | 5.39 | 6.77 | 0.166 | 0.268 | 330.700 |
| Support Vector Machine | 1.044 | 5.40 | 6.84 | 0.215 | 0.308 | 21.096 |

[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