Evolution results: Geographically weighted Random Forest

```
java version "1.8.0_381"
Java(TM) SE Runtime Environment (build 1.8.0_381-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.381-b09, mixed mode)

Starting H2O JVM and connecting:  Connection successful!

R is connected to the H2O cluster:
    H2O cluster uptime:         4 seconds 44 milliseconds
    H2O cluster timezone:       Europe/Berlin
    H2O data parsing timezone:  UTC
    H2O cluster version:        3.42.0.1
    H2O cluster version age:    1 month and 24 days
    H2O cluster name:           H2O_started_from_R_omkarp_ujt891
    H2O cluster total nodes:    1
    H2O cluster total memory:   42.65 GB
    H2O cluster total cores:    12
    H2O cluster allowed cores:  12
    H2O cluster healthy:        TRUE
    H2O Connection ip:          localhost
    H2O Connection port:        54321
    H2O Connection proxy:       NA
    H2O Internal Security:      FALSE
    R Version:                  R version 4.3.1 (2023-06-16 ucrt)
```

Random Forest Optimizations

```
Bandwidth: 153
R2 of Local Model: 0.465286918683426
Bandwidth: 154
R2 of Local Model: 0.477267902317772
Bandwidth: 155
R2 of Local Model: 0.425406185453502
Bandwidth: 156
R2 of Local Model: 0.492977094721852

> bandwidth$Best.BW
[1] 156

Number of Observations: 2990
Number of Independent Variables: 22
Kernel: Adaptive
Neightbours: 156

--------------- Global Model Summary ---------------

Ranger result

Call:
 ranger(trip_count ~ Income + AreaProx + LUP_Other + Resi_lur +      Edu_l
ur + Recre_lur + Public_lur + Health_lur + Other_lur +      Comm_lur + Rd_
Pri + Rd_Cyc + Rd_Ter + Rd_Pedes + T_Enable +      T_Hinder + Rd_Sec + Wee
kend + Summer + Winter + Spring +      Dist_center, data = combined.df, nu
m.trees = 120, mtry = 2,      importance = "impurity", num.threads = 1)

Type:                             Regression
Number of trees:                  120
Sample size:                      2990
Number of independent variables:  22
Mtry:                             2
Target node size:                 5
Variable importance mode:         impurity
Splitrule:                        variance
OOB prediction error (MSE):       888630.5
R squared (OOB):                  0.6694221
```

Importance:

```
     Income    AreaProx   LUP_Other    Resi_lur     Edu_lur   Recre_lur  P
ublic_lur  Health_lur   Other_lur
  600034888   476167718   254804178   394012175   142303689   214300449
277140819    64081458   411074515
   Comm_lur      Rd_Pri      Rd_Cyc      Rd_Ter    Rd_Pedes    T_Enable
T_Hinder      Rd_Sec     Weekend
  698889011    64347706   151337426   197152082   706713259   208354176
263945485   149563334   561098345
     Summer      Winter      Spring Dist_center
  446479432   107263835    76044946   503283054
```

Mean Square Error (Not OBB): 369514.067
R-squared (Not OBB) %: 86.249
AIC (Not OBB): 38377.633
AICc (Not OBB): 38378.005

--------------- Local Model Summary ---------------

Residuals OOB:

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-8905.00  -184.72   -25.67    16.53    92.91 21696.53
```

Residuals Predicted (Not OBB):

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-5248.210   -99.838   -15.800    -1.414    30.094  8720.420
```

Local Variable Importance:

```
                 Min       Max     Mean      StD
Income      33634.028 187356562 10490159 22656342
AreaProx    27538.719 286022765 10132564 24561922
LUP_Other   29218.121 151979072  6573634 15011845
Resi_lur    24413.616 361887070  7741327 23333328
Edu_lur      3055.181 157294575  4919025 12222957
Recre_lur   36582.346 176316635  5846427 14363240
Public_lur  23623.688 154291228  6779973 14631440
Health_lur   1623.978  91486929  2187085  5507973
Other_lur   20883.026 168717207  9235835 19668573
Comm_lur    40352.128 263841711 11045758 26905912
Rd_Pri       3952.034 117754778  3293668  8551114
Rd_Cyc      27596.057 126733209  6658524 14278068
Rd_Ter      61975.463 143343772  7722322 16175554
Rd_Pedes    35916.830 177410509  7893528 16163326
T_Enable    23473.852 202996235  6858239 18703610
T_Hinder    15073.188 159999633  6517921 17659465
Rd_Sec       8973.755 163334667  5721281 15527729
Weekend     39808.534 910345117 31750008 92997650
Summer      87961.938 780780829 25574543 74256013
Winter       9471.914 207441361  7225730 20292920
Spring       7845.662 219925826  4713773 17064052
Dist_center 52305.594 187059336  8222395 19194384
```

Mean squared error (OOB): 1715534.444
R-squared (OOB) %: 36.159
AIC (OBB): 42968.153
AICc (OBB): 42968.526
Mean squared error Predicted (Not OBB): 325691.282
R-squared Predicted (Not OBB) %: 87.88
AIC Predicted (Not OBB): 38000.179
AICc Predicted (Not OBB): 38000.551

```
$l.MSE.OOB          $l.r.OOB            $l.MSE.Pred         $l.r.Pred
[1] 1715534         [1] 0.3615935       [1] 325691.3        [1] 0.8787996
```

```
Optimised mtry

2990 samples
  22 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 2690, 2693, 2690, 2691, 2690, 2692, ...
Resampling results across tuning parameters:

  mtry  RMSE       Rsquared   MAE
   1    1014.7364  0.6462216  432.9113
   2     904.7358  0.7144355  369.2696
   3     839.8657  0.7585481  336.4581
   4     788.9484  0.7903991  310.8240
   5     748.1370  0.8147868  291.1577
   6     720.0772  0.8287369  277.7368
   7     696.2963  0.8404869  266.9760
   8     679.7170  0.8478784  260.6591
   9     668.1136  0.8536085  256.2150
  10     657.3245  0.8576944  253.3054
  11     650.8930  0.8597799  251.3660
  12     645.5373  0.8614532  250.4052
  13     642.7958  0.8623187  250.7714
  14     638.2154  0.8645787  250.0918
  15     638.6197  0.8636113  250.1876
  16     635.7922  0.8642213  250.2629
  17     633.8188  0.8647965  250.5460
  18     637.0717  0.8628315  251.0545
  19     634.6186  0.8634810  251.1167
  20     635.2328  0.8628686  251.5156
  21     636.5527  0.8617272  252.3859
  22     636.2312  0.8618458  252.7026

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 17.

> bandwidth <- grf.bw(trip_count ~ Income+AreaProx+LUP_Other+Resi_lur+Edu_
lur+Recre_lur+Public_lur+Health_lur+Other_lur+Comm_lur+Rd_Pri+Rd_Cyc+Rd_Te
r+Rd_Pedes+T_Enable+T_Hinder+Rd_Sec+Weekend+Summer++Winter+Spring+Dist_cen
ter, dataset = combined.df, kernel="adaptive", coords=Coords, bw.min = 153
, bw.max = 156, step = 1, trees=500, mtry=mtry, importance="impurity", nth
reads = 1, forests = TRUE, weighted = TRUE)

Bandwidth: 153
R2 of Local Model: 0.336708606072047
Bandwidth: 154
R2 of Local Model: 0.312668252790397
Bandwidth: 155
R2 of Local Model: 0.331870726416428
Bandwidth: 156
R2 of Local Model: 0.311300781031856
Best Bandwidth (Based on the Local Model): 153

Number of Observations: 2990
Number of Independent Variables: 22
Kernel: Adaptive
Neightbours: 153

--------------- Global Model Summary ---------------

Ranger result

Call:
 ranger(trip_count ~ Income + AreaProx + LUP_Other + Resi_lur +      Edu_l
ur + Recre_lur + Public_lur + Health_lur + Other_lur +      Comm_lur + Rd_
Pri + Rd_Cyc + Rd_Ter + Rd_Pedes + T_Enable +      T_Hinder + Rd_Sec + Wee
kend + Summer + Winter + Spring +      Dist_center, data = combined.df, nu
m.trees = 120, mtry = mtry,      importance = "impurity", num.threads = 1)
```

```
Type:                             Regression
Number of trees:                  120
Sample size:                      2990
Number of independent variables:  22
Mtry:                             17
Target node size:                 5
Variable importance mode:         impurity
Splitrule:                        variance
OOB prediction error (MSE):       460506.8
R squared (OOB):                  0.8286877

Importance:

Mean Square Error (Not OBB): 98111.55
R-squared (Not OBB) %: 96.349
AIC (Not OBB): 34412.643
AICc (Not OBB): 34413.015

--------------- Local Model Summary ---------------

Residuals OOB:

      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-10324.500   -179.661     -9.494     36.647    110.110  20844.000

Residuals Predicted (Not OBB):

    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-568.992  -11.199   -1.716    1.195    5.531 1040.255
Local Variable Importance:

                    Min         Max        Mean       StD
Income         9161.38938  306060275 13428669.0  28026858
AreaProx       5069.11488  762415573 14352580.4  50192822
LUP_Other      4999.31489  110505530  3660726.2   8906879
Resi_lur       6009.00541  694859835  6953957.6  35337951
Edu_lur         145.95851   99281685  2166201.5   7002047
Recre_lur      2412.17304  176724635  2931166.6  10655868
Public_lur     2199.73210  105583782  4339415.4  10738686
Health_lur       26.93786   18610950   773168.8   1705177
Other_lur      2579.06654  355103147 11239430.1  29857138
Comm_lur       9660.04390  670689509 18220743.4  59512977
Rd_Pri           37.12727   66856820  1266791.4   4085721
Rd_Cyc         6839.05424  114238446  4560167.1  11481909
Rd_Ter        20437.80071   89590387  4836292.8   8337127
Rd_Pedes       4152.97716  320024101  7464628.6  20614544
T_Enable       2786.18852  251262988  3870844.9  15695025
T_Hinder       1637.26562  136679572  3334908.8  10946174
Rd_Sec           98.49008  230186484  3782659.7  15802830
Weekend     118221.80844 1729890825 71813989.2 199831351
Summer      136419.50499 2256400006 71241697.9 223789495
Winter         387.09136  227892451  7855987.1  23068270
Spring        3878.41366   75751909  2100947.7   5213006
Dist_center  17194.87710  245968381  7798557.9  22957463

Mean squared error (OOB): 1867035.658
R-squared (OOB) %: 30.521
AIC (OBB): 43221.189
AICc (OBB): 43221.561
Mean squared error Predicted (Not OBB): 3880.005
R-squared Predicted (Not OBB) %: 99.856
AIC Predicted (Not OBB): 24754.139
AICc Predicted (Not OBB): 24754.512
                                        $l.MSE.Pred
$l.MSE.OOB                              [1] 3880.005
[1] 1867036
$l.r.OOB                                $l.r.Pred
[1] 0.305215                            [1] 0.9985561
```

```
grf.model <- grf(trip_count ~ Income+AreaProx+LUP_Other+Resi_lur+Edu_lur+R
ecre_lur+Public_lur+Health_lur+
+                    Other_lur+Comm_lur+Rd_Pri+Rd_Cyc+Rd_Ter+Rd_Pedes+T_En
able+T_Hinder+Rd_Sec+Weekend+
+                     Summer+Winter+Spring+Dist_center,
+                  bw=bw, importance = "impurity", nthreads = 1, dframe=co
mbined.df,
+                  ntree=120, mtry=2, kernel=kernel, forests = TRUE, coord
s=Coords, weighted=TRUE)

Number of Observations: 2990
Number of Independent Variables: 22
Kernel: Adaptive
Neightbours: 153

--------------- Global Model Summary ---------------

Ranger result

Call:
 ranger(trip_count ~ Income + AreaProx + LUP_Other + Resi_lur +      Edu_l
ur + Recre_lur + Public_lur + Health_lur + Other_lur +      Comm_lur + Rd_
Pri + Rd_Cyc + Rd_Ter + Rd_Pedes + T_Enable +      T_Hinder + Rd_Sec + Wee
kend + Summer + Winter + Spring +      Dist_center, data = combined.df, nu
m.trees = 120, mtry = 2,      importance = "impurity", num.threads = 1)

Type:                             Regression
Number of trees:                  120
Sample size:                      2990
Number of independent variables:  22
Mtry:                             2
Target node size:                 5
Variable importance mode:         impurity
Splitrule:                        variance
OOB prediction error (MSE):       870626.2
R squared (OOB):                  0.6761199

Importance:

     Income    AreaProx   LUP_Other    Resi_lur     Edu_lur   Recre_lur  P
ublic_lur  Health_lur   Other_lur    Comm_lur      Rd_Pri      Rd_Cyc
Rd_Ter    Rd_Pedes    T_Enable
  574244320   408536666   278279157   339315704   134270020   199838591
241969382    64756989   387301883   737835675    82109603   136581111   19
3389896   719310902   227773514
   T_Hinder      Rd_Sec     Weekend      Summer      Winter      Spring Di
st_center
  247808896   131214740   568033369   442182398   126392210    71418225
502487547

Mean Square Error (Not OBB): 363601.914
R-squared (Not OBB) %: 86.469
AIC (Not OBB): 38329.407
AICc (Not OBB): 38329.779

--------------- Local Model Summary ---------------


Residuals OOB:

     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-10824.62   -200.28    -25.09     14.71     95.53  22412.42

Residuals Predicted (Not OBB):

     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-4966.770  -103.227   -14.499     0.788    30.142  9505.151
```

```
Local Variable Importance:

                  Min       Max     Mean       StD
Income      31955.886 231059479 10301397 22465098
AreaProx    36602.484 329307601 10047359 25009012
LUP_Other   32336.494 151062023  6507844 14907180
Resi_lur    26219.097 309267258  7764246 23465931
Edu_lur      4285.194 150775509  4921833 12592242
Recre_lur   23485.284 152119413  5642646 13638414
Public_lur  20978.811 138057406  6753421 14680106
Health_lur   2054.532  52247499  2079824  4835036
Other_lur   22811.047 158585600  8987236 19429465
Comm_lur    34564.768 250757439 10661861 25726285
Rd_Pri       2558.972 101453655  3205594  8270406
Rd_Cyc      29676.195 163600948  6469881 14251376
Rd_Ter      52949.512 152330993  7261809 14917885
Rd_Pedes    18973.309 154438929  7778642 16043452
T_Enable    20873.618 168100422  6661672 17919573
T_Hinder    18605.883 170937369  6396452 17581652
Rd_Sec       9963.081 182447853  5693661 15952424
Weekend     42961.432 861532460 31022613 90672739
Summer      89117.137 805313732 24775527 70697746
Winter       7302.016 219838848  7069391 19703846
Spring       7610.603 221601039  4647024 17142880
Dist_center 45208.773 186794024  8017233 18860541

Mean squared error (OOB): 1619171.476
R-squared (OOB) %: 39.745
AIC (OBB): 42795.301
AICc (OBB): 42795.673
Mean squared error Predicted (Not OBB): 325326.353
R-squared Predicted (Not OBB) %: 87.894
AIC Predicted (Not OBB): 37996.827
AICc Predicted (Not OBB): 37997.199
```

Github:: parishwadomkar/Usage-analysis-e-scooter: https://github.com/parishwadomkar/Usage-analysis-e-scooter/tree/main

References

Marvin N. Wright, Andreas Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(1), 1-17. doi:10.18637/jss.v077.i01

Blas M. Benito (2021). R package spatialRF: Easy Spatial Regression with Random Forest. doi: 10.5281/zenodo.4745208. url: https://blasbenito.github.io/spatialRF/

Github:: cran/SpatialML : https://github.com/cran/SpatialML/tree/master/R

Github:: dizhu-gis/SRGCNN: Zhu, D., Liu, Y., Yao, X. et al. Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. Geoinformatica 26, 645–676 (2022). https://doi.org/10.1007/s10707-021-00454-x

Feature Engineering

Interaction	Importance (% of max)	R-squared improvement	Max cor with predictors
Weekend..x..Other lur	100.0	0.015	0.74
Weekend..x..Comm_lur	100.0	0.019	0.71
Weekend..pca..Resi lur	99.7	0.012	0.71
Weekend..pca..T_Hinder	83.5	0.010	0.71
Dist center..pca..Other_lur	86.9	0.010	0.60
Comm_lur..x..Summer	59.5	0.011	0.75
Summer..x..Resi_lur	72.4	0.010	0.62
Other lur..x..Resi_lur	60.5	0.013	0.68
Weekend..x..Summer	58.1	0.010	0.64
Rd Pedes..x..Resi_lur	46.0	0.012	0.71
Comm lur..x..T Enable	62.2	0.011	0.58
Other_lur..pca..T Enable	45.0	0.010	0.71
Dist center..pca..T_Enable	60.3	0.012	0.57
Comm lur..x..Other_lur	48.1	0.014	0.564

Global Random Forest Model performance

Model performance
- R squared (oob): 0.832325
- R squared (cor(obs, pred)^2): 0.9542827
- Pseudo R squared (cor(obs, pred)):0.976874
- RMSE (oob): 671.3637
- RMSE: 403.0246
- Normalized RMSE: 0.5704524

Local/SpatialRF Model performance > spatialRF::print_performance(model.spatial.repeat)

Model performance (median +/- mad)
- R squared (oob): 0.827 +/- 0.0038
- R squared (cor(obs, pred)^2): 0.953 +/- 0.002
- Pseudo R squared: 0.976 +/- 0.001
- RMSE (oob): 681.28 +/- 7.4243
- RMSE: 408.351 +/- 6.1675
- Normalized RMSE: 0.578 +/- 0.0087

Spatial evaluation
- Training fraction: 0.80
- Spatial folds: 28