Name: Yoshi Miyamoto
GTID: ymiyamoto3

1. After I've tested with different leaf sizes from 1 to 20 and above(Exhibit A) with Istanbul.csv, I noticed that out of samples's RMSE kept decreasing until the leaf size passing 10 and it stayed the same until the leaf size became more than 20. I consider that Bigger leaf size is similar to the one having less polynomial degree. so around before 20 leaves where overfitting occurs.

2. From my experiment(ExhibitB), 20 of DTLearner in the bagging sampling Istanbul.csv is neutralizing the RMSE and correlation better that there is no huge difference between In and Out samples. It reduces overfitting.

3. I timed the process speed of Random Tree and Decision Tree. Random Tree is faster to process for bigger data (Exhibit C). While DT has more distinct difference between different number of leaves, RT doesn't. It means DT could more overfitting than RT. (Exhibit D)

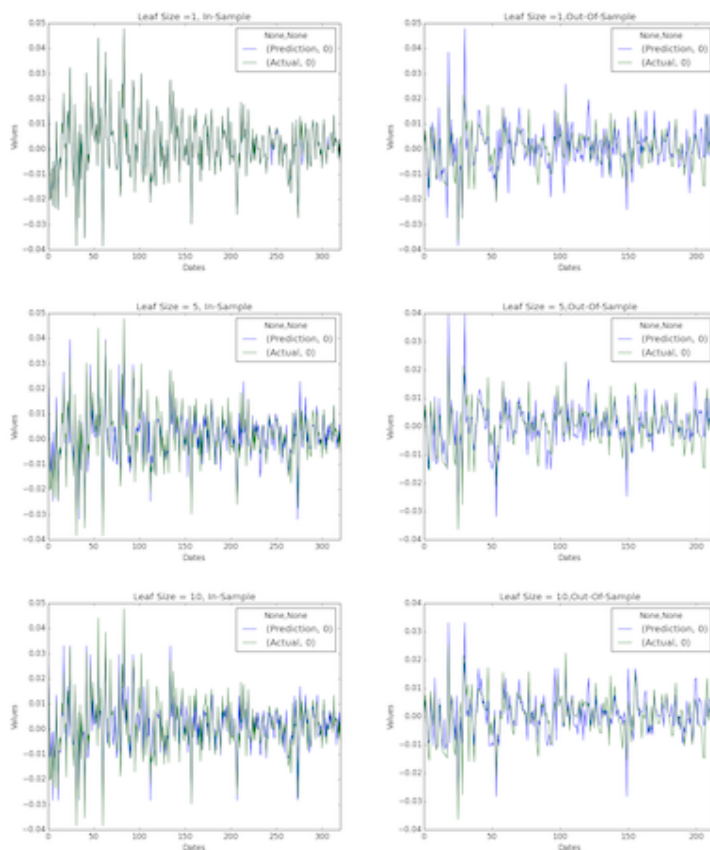Exhibit A   Exhibit B

Leaf Size 1

In sample results
RMSE: 0.00107543562706
corr: 0.995848875176

Out of sample results
RMSE: 0.00736486015624
corr: 0.677234532996

Leaf Size 1, Bag Size 20

In sample results
RMSE: 0.00254563079778
corr: 0.980417990392

Out of sample results
RMSE: 0.00473649479295
corr: 0.826365404871

Leaf Size 5

In sample results
RMSE: 0.00507093799414
corr: 0.90244537809

Out of sample results
RMSE: 0.00602474229028
corr: 0.742139561791

Leaf Size 5, Bag Size 20

In sample results
RMSE: 0.00412269492157
corr: 0.944746449578

Out of sample results
RMSE: 0.00455429748417
corr: 0.836282373792

Leaf Size 10

In sample results
RMSE: 0.00578326999341
corr: 0.870962945799

Out of sample results
RMSE: 0.00590412249351
corr: 0.734214336961

Leaf Size 10, Bag Size 20

In sample results
RMSE: 0.00498904942455
corr: 0.913122454967

Out of sample results
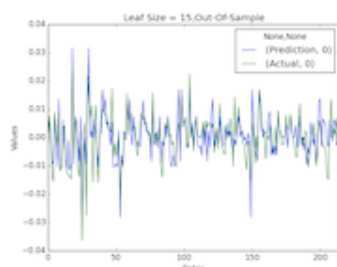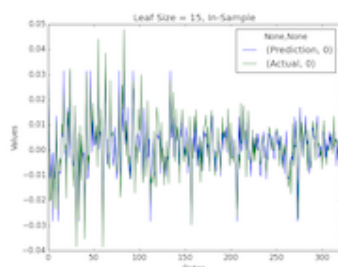RMSE: 0.00456417119175
corr: 0.830699388512
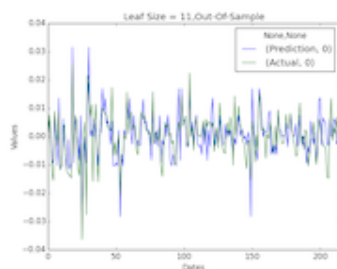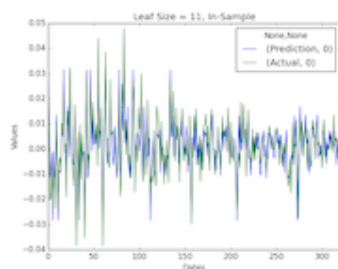
Leaf Size 11

In sample results
RMSE: 0.00579236131618
corr: 0.870526739018

Out of sample results
RMSE: 0.00587624929777
corr: 0.73486419647

Leaf Size 11, Bag Size 20

In sample results
RMSE: 0.00515825839091
corr: 0.909614831222

Out of sample results
RMSE: 0.0045094506132
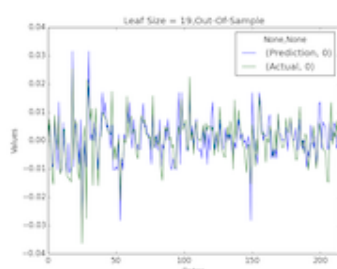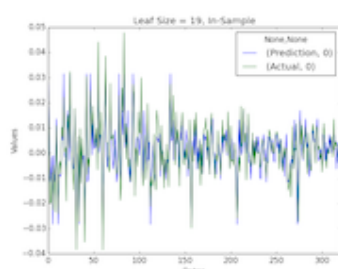corr: 0.836227561767

Leaf Size 15

In sample results
RMSE: 0.00579236131618
corr: 0.870526739018

Out of sample results
RMSE: 0.00587624929777
corr: 0.73486419647

Leaf Size 15, Bag Size 20

In sample results
RMSE: 0.00506576284925
corr: 0.912420528169

Out of sample results
RMSE: 0.00465957282345
corr: 0.825870985227

Leaf Size 19

In sample results
RMSE: 0.00579236131618
corr: 0.870526739018

Out of sample results
RMSE: 0.00587624929777
corr: 0.73486419647

Leaf Size 19, Bag Size 20

In sample results
RMSE: 0.00512037691161
corr: 0.907954382428

Out of sample results
RMSE: 0.0044365007905
corr: 0.842416814905
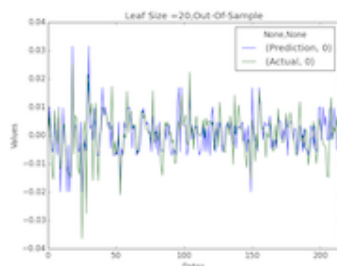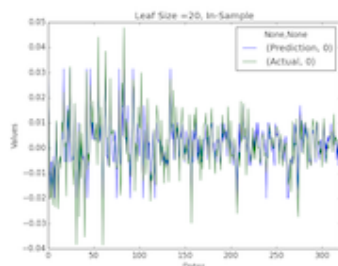
Leaf Size 20

In sample results
RMSE: 0.00667752731375
corr: 0.823493909472

Out of sample results
RMSE: 0.00612170826575
corr: 0.693504268339

Leaf Size 20, Bag Size 20

In sample results
RMSE: 0.00573735071993
corr: 0.886219656747

Out of sample results
RMSE: 0.00462570434767
corr: 0.824551516662
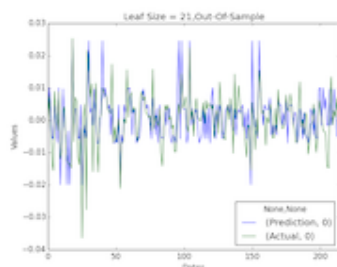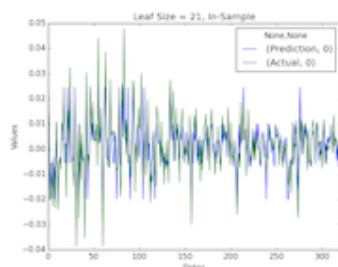
Leaf Size 21

In sample results
RMSE: 0.00693198519339
corr: 0.808173590566

Out of sample results
RMSE: 0.00634440636318
corr: 0.682157965544

Leaf Size 21, Bag Size 20

In sample results
RMSE: 0.00591921457392
corr: 0.875957481018

Out of sample results
RMSE: 0.00474690153229
corr: 0.816030714358

Leaf Size = 25, In-Sample   Leaf Size = 25, Out-Of-Sample
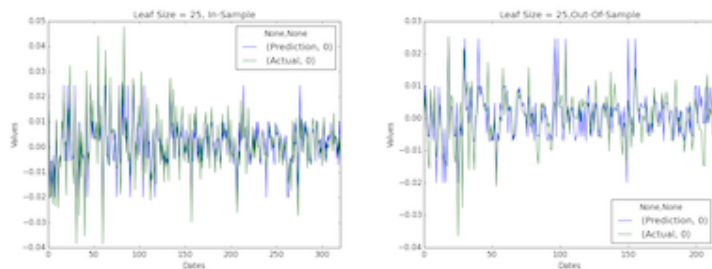
Leaf Size 25

In sample results
RMSE: 0.00693198519339
corr: 0.808173590566

Out of sample results
RMSE: 0.00634440636318
corr: 0.682157965544

Leaf Size 25, Bag Size 20

In sample results
RMSE: 0.00606013799835
corr: 0.866319180809

Out of sample results
RMSE: 0.00487129790794
corr: 0.804075275891

Exhibit C with ripple.csv

```
Leaf Size 1, Learner = <DTLearner.DTLearner object at 0x10eb77290>

In sample results
RMSE:  0.0
corr:  1.0

Out of sample results
RMSE:  0.524586987682
corr:  0.714750708135
--- 0.107620954514 seconds ---

Leaf Size 1, Learner = <RTLearner.RTLearner object at 0x10ae0c290>

In sample results
RMSE:  0.0
corr:  1.0

Out of sample results
RMSE:  0.557568076985
corr:  0.692279478193
--- 0.0676109790802 seconds ---
```

Exhibit D with Istanbul.csv

DT	RT
Leaf Size 1, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>	Leaf Size 1, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>
In sample results RMSE: 0.00264620117941 corr: 0.97925215988	In sample results RMSE: 0.00284437678878 corr: 0.977964899974
Out of sample results RMSE: 0.00458032762769 corr: 0.833159186616 --- 1.04606199265 seconds ---	Out of sample results RMSE: 0.0046747005697 corr: 0.823559335314 --- 0.511961936951 seconds ---
Leaf Size 5, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>	Leaf Size 5, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>
In sample results RMSE: 0.00407147726878 corr: 0.944355111168	In sample results RMSE: 0.00582620411624 corr: 0.887667629486
Out of sample results RMSE: 0.00469495037215 corr: 0.82403553644 --- 0.363982200623 seconds ---	Out of sample results RMSE: 0.00483156738797 corr: 0.813820148551 --- 0.361968994141 seconds ---
Leaf Size 10, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>	Leaf Size 10, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>
In sample results RMSE: 0.00496956609634 corr: 0.912892430171	In sample results RMSE: 0.00634112212508 corr: 0.863418941938
Out of sample results RMSE: 0.00471548162597 corr: 0.824723820941 --- 0.302423000336 seconds ---	Out of sample results RMSE: 0.00472190770187 corr: 0.818204755101 --- 0.206550036563 seconds ---
Leaf Size 11, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>	Leaf Size 11, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>
In sample results RMSE: 0.00497901337772 corr: 0.914548766235	In sample results RMSE: 0.00639038311979 corr: 0.860785103247
Out of sample results RMSE: 0.00464850181142 corr: 0.829049010795 --- 0.270176887512 seconds ---	Out of sample results RMSE: 0.00471592354931 corr: 0.81693794595 --- 0.256266117096 seconds ---
Leaf Size 15, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>	Leaf Size 15, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>
In sample results RMSE: 0.00512118539666 corr: 0.908692335879	In sample results RMSE: 0.00625457390022 corr: 0.865593811305
Out of sample results RMSE: 0.00466406448744 corr: 0.826117035432 --- 0.25381398201 seconds ---	Out of sample results RMSE: 0.0046506007747 corr: 0.829487156836 --- 0.192044973373 seconds ---

```
Leaf Size 19, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>    Leaf Size 19, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>

In sample results                                                     In sample results
RMSE:  0.00499009297869                                               RMSE:  0.00657202200432
corr:  0.91346186102                                                  corr:  0.848610967246

Out of sample results                                                 Out of sample results
RMSE:  0.00456946507545                                               RMSE:  0.00473221267011
corr:  0.834321044194                                                 corr:  0.819339159159
---- 0.260789871216 seconds ----                                      ---- 0.21045589447 seconds ----


Leaf Size 20, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>    Leaf Size 20, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>

In sample results                                                     In sample results
RMSE:  0.00586354875888                                               RMSE:  0.00695847249483
corr:  0.878498383806                                                 corr:  0.824027094122

Out of sample results                                                 Out of sample results
RMSE:  0.00453717326769                                               RMSE:  0.00477558336126
corr:  0.831878773018                                                 corr:  0.8137188493353
---- 0.176883935928 seconds ----                                      ---- 0.218217134476 seconds ----


Leaf Size 21, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>    Leaf Size 21, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>

In sample results                                                     In sample results
RMSE:  0.00591642674398                                               RMSE:  0.00727328580875
corr:  0.877038127909                                                 corr:  0.81180637923

Out of sample results                                                 Out of sample results
RMSE:  0.00470072640943                                               RMSE:  0.00485066695485
corr:  0.8194565512                                                   corr:  0.806941604496
---- 0.18244600296 seconds ----                                       ---- 0.144994020462 seconds ----


Leaf Size 25, Bag Size 20, Learner = <class 'DTLearner.DTLearner'>    Leaf Size 25, Bag Size 20, Learner = <class 'RTLearner.RTLearner'>

In sample results                                                     In sample results
RMSE:  0.00587234433743                                               RMSE:  0.00722686560723
corr:  0.878834148976                                                 corr:  0.811441735819

Out of sample results                                                 Out of sample results
RMSE:  0.00480451159421                                               RMSE:  0.00472652560465
corr:  0.811363510833                                                 corr:  0.816497077821
---- 0.170121192932 seconds ----                                      ---- 0.160471916199 seconds ----
```