Contents

[Test Results – Random Forest Regression – Baseline Model 1](#_Toc101524003)

[Test 1 – Default RFR Parameters 1](#_Toc101524004)

# Test Results – Random Forest Regression – Baseline Model

Results from creating model, tuning, for only the features that are **non location-based**, i.e. the original data, but cleaned

RFR = ‘random forest regressor’ # acronym

The exact features in use:

['price\_per\_sq\_m\_per\_lease\_yr\_norm', # our target

'storey\_range\_min',

'storey\_range\_max',

'n\_rooms',

'town\_bedok',

'town\_bishan',

'town\_bukit\_batok',

'town\_bukit\_merah',

'town\_bukit\_panjang',

'town\_bukit\_timah',

'town\_central\_area',

'town\_choa\_chu\_kang',

'town\_clementi',

'town\_geylang',

'town\_hougang',

'town\_jurong\_east',

'town\_jurong\_west',

'town\_kallang\_whampoa',

'town\_lim\_chu\_kang',

'town\_marine\_parade',

'town\_pasir\_ris',

'town\_punggol',

'town\_queenstown',

'town\_sembawang',

'town\_sengkang',

'town\_serangoon',

'town\_tampines',

'town\_toa\_payoh',

'town\_woodlands',

'town\_yishun',

'flat\_type\_2\_room',

'flat\_type\_3\_room',

'flat\_type\_4\_room',

'flat\_type\_5\_room',

'flat\_type\_executive',

'flat\_type\_multi\_generation']

Note: Default RFR Parameters below

DEFAULT\_HYPS\_FOR\_RANDOM\_FOREST\_REGRESSOR =

'bootstrap': True,

'ccp\_alpha': 0.0,

'criterion': 'squared\_error',

'max\_depth': None,

'max\_features': 'auto',

'max\_leaf\_nodes': None,

'max\_samples': None,

'min\_impurity\_decrease': 0.0,

'min\_samples\_leaf': 1,

'min\_samples\_split': 2,

'min\_weight\_fraction\_leaf': 0.0,

'n\_estimators': 100,

'n\_jobs': None,

'oob\_score': False,

'random\_state': None,

'verbose': 0,

'warm\_start': False

## Test 1 – Default RFR Parameters

Base RFR model, with no modification of the parameters, i.e. parameter defaults across the board

regr = RandomForestRegressor(random\_state=42, n\_jobs=-1)

--- Test Set ---

Mean Absolute Error: ... 8.49765

Mean Squared Error:..... 135.27868

RMSE: .................. 11.63094

Coeff of det (R^2):..... 0.61458

--- Val Set ---

Mean Absolute Error: ... 7.63201

Mean Squared Error:..... 114.39955

RMSE: .................. 10.69577

Coeff of det (R^2):..... 0.71235

--- Train Set ---

Mean Absolute Error: ... 5.77682

Mean Squared Error:..... 62.74640

RMSE: .................. 7.92126

Coeff of det (R^2):..... 0.78757

--- Combined Test Set ---

Mean Absolute Error: ... 8.11715

Mean Squared Error:..... 126.10105

RMSE: .................. 11.22947

Coeff of det (R^2):..... 0.66091

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 100

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 11 seconds ---

## Test 2 – Default RFR Parameters + higher number of n\_estimators

Base RFR model, with no modification of the parameters, i.e. parameter defaults across the board

regr = RandomForestRegressor(random\_state=42,

n\_jobs=-1,

n\_estimators=300)

--- Test Set ---

Mean Absolute Error: ... 8.49564

Mean Squared Error:..... 135.22824

RMSE: .................. 11.62877

Coeff of det (R^2):..... 0.61472

--- Val Set ---

Mean Absolute Error: ... 7.63089

Mean Squared Error:..... 114.38442

RMSE: .................. 10.69507

Coeff of det (R^2):..... 0.71239

--- Train Set ---

Mean Absolute Error: ... 5.77664

Mean Squared Error:..... 62.74437

RMSE: .................. 7.92113

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11553

Mean Squared Error:..... 126.06613

RMSE: .................. 11.22792

Coeff of det (R^2):..... 0.66101

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 300

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 43 seconds ---

## Test 3 – Default RFR Parameters + incrementing higher number of n\_estimators

Base RFR model, with no modification of the parameters, i.e. parameter defaults across the board

regr = RandomForestRegressor(random\_state=42,

n\_jobs=-1,

n\_estimators=inc)

for inc in [100,300, 600, 900, 1200, 1500]:

RFR\_model\_3(X\_train,y\_train,X\_val,y\_val,X\_test,y\_test,X\_test\_combined, y\_test\_combined, inc)

Num\_Estimators: 100

--- Test Set ---

Mean Absolute Error: ... 8.49765

Mean Squared Error:..... 135.27868

RMSE: .................. 11.63094

Coeff of det (R^2):..... 0.61458

--- Val Set ---

Mean Absolute Error: ... 7.63201

Mean Squared Error:..... 114.39955

RMSE: .................. 10.69577

Coeff of det (R^2):..... 0.71235

--- Train Set ---

Mean Absolute Error: ... 5.77682

Mean Squared Error:..... 62.74640

RMSE: .................. 7.92126

Coeff of det (R^2):..... 0.78757

--- Combined Test Set ---

Mean Absolute Error: ... 8.11715

Mean Squared Error:..... 126.10105

RMSE: .................. 11.22947

Coeff of det (R^2):..... 0.66091

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 100

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 14 seconds ---

Num\_Estimators: 300

--- Test Set ---

Mean Absolute Error: ... 8.49564

Mean Squared Error:..... 135.22824

RMSE: .................. 11.62877

Coeff of det (R^2):..... 0.61472

--- Val Set ---

Mean Absolute Error: ... 7.63089

Mean Squared Error:..... 114.38442

RMSE: .................. 10.69507

Coeff of det (R^2):..... 0.71239

--- Train Set ---

Mean Absolute Error: ... 5.77664

Mean Squared Error:..... 62.74437

RMSE: .................. 7.92113

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11553

Mean Squared Error:..... 126.06613

RMSE: .................. 11.22792

Coeff of det (R^2):..... 0.66101

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 300

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 45 seconds ---

Num\_Estimators: 600

--- Test Set ---

Mean Absolute Error: ... 8.49451

Mean Squared Error:..... 135.21597

RMSE: .................. 11.62824

Coeff of det (R^2):..... 0.61476

--- Val Set ---

Mean Absolute Error: ... 7.63084

Mean Squared Error:..... 114.37887

RMSE: .................. 10.69481

Coeff of det (R^2):..... 0.71240

--- Train Set ---

Mean Absolute Error: ... 5.77660

Mean Squared Error:..... 62.74379

RMSE: .................. 7.92110

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11488

Mean Squared Error:..... 126.05681

RMSE: .................. 11.22750

Coeff of det (R^2):..... 0.66103

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 600

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 88 seconds ---

Num\_Estimators: 900

--- Test Set ---

Mean Absolute Error: ... 8.49390

Mean Squared Error:..... 135.20054

RMSE: .................. 11.62758

Coeff of det (R^2):..... 0.61480

--- Val Set ---

Mean Absolute Error: ... 7.63017

Mean Squared Error:..... 114.36184

RMSE: .................. 10.69401

Coeff of det (R^2):..... 0.71244

--- Train Set ---

Mean Absolute Error: ... 5.77655

Mean Squared Error:..... 62.74348

RMSE: .................. 7.92108

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11424

Mean Squared Error:..... 126.04068

RMSE: .................. 11.22678

Coeff of det (R^2):..... 0.66107

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 900

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 133 seconds ---

Num\_Estimators: 1200

--- Test Set ---

Mean Absolute Error: ... 8.49404

Mean Squared Error:..... 135.20539

RMSE: .................. 11.62779

Coeff of det (R^2):..... 0.61479

--- Val Set ---

Mean Absolute Error: ... 7.63020

Mean Squared Error:..... 114.36144

RMSE: .................. 10.69399

Coeff of det (R^2):..... 0.71245

--- Train Set ---

Mean Absolute Error: ... 5.77658

Mean Squared Error:..... 62.74342

RMSE: .................. 7.92107

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11433

Mean Squared Error:..... 126.04322

RMSE: .................. 11.22690

Coeff of det (R^2):..... 0.66107

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 1200

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 188 seconds ---

Num\_Estimators: 1500

--- Test Set ---

Mean Absolute Error: ... 8.49416

Mean Squared Error:..... 135.20244

RMSE: .................. 11.62766

Coeff of det (R^2):..... 0.61480

--- Val Set ---

Mean Absolute Error: ... 7.63025

Mean Squared Error:..... 114.35959

RMSE: .................. 10.69390

Coeff of det (R^2):..... 0.71245

--- Train Set ---

Mean Absolute Error: ... 5.77660

Mean Squared Error:..... 62.74338

RMSE: .................. 7.92107

Coeff of det (R^2):..... 0.78758

--- Combined Test Set ---

Mean Absolute Error: ... 8.11442

Mean Squared Error:..... 126.04075

RMSE: .................. 11.22679

Coeff of det (R^2):..... 0.66107

Model Hyperparameters Definitions:

bootstrap -> True

ccp\_alpha -> 0.0

criterion -> squared\_error

max\_depth -> None

max\_features -> auto

max\_leaf\_nodes -> None

max\_samples -> None

min\_impurity\_decrease -> 0.0

min\_samples\_leaf -> 1

min\_samples\_split -> 2

min\_weight\_fraction\_leaf -> 0.0

n\_estimators -> 1500

n\_jobs -> -1

oob\_score -> False

random\_state -> 42

verbose -> 0

warm\_start -> False

--- 228 seconds ---

## Test 4 – GridsearchCV to tune RFR model

regr = RandomForestRegressor(random\_state=42,

n\_jobs=-1,

n\_estimators=inc)

['accuracy',

'adjusted\_mutual\_info\_score',

'adjusted\_rand\_score',

'average\_precision',

'balanced\_accuracy',

'completeness\_score',

'explained\_variance',

'f1',

'f1\_macro',

'f1\_micro',

'f1\_samples',

'f1\_weighted',

'fowlkes\_mallows\_score',

'homogeneity\_score',

'jaccard',

'jaccard\_macro',

'jaccard\_micro',

'jaccard\_samples',

'jaccard\_weighted',

'max\_error',

'mutual\_info\_score',

'neg\_brier\_score',

'neg\_log\_loss',

'neg\_mean\_absolute\_error',

'neg\_mean\_absolute\_percentage\_error',

'neg\_mean\_gamma\_deviance',

'neg\_mean\_poisson\_deviance',

'neg\_mean\_squared\_error',

'neg\_mean\_squared\_log\_error',

'neg\_median\_absolute\_error',

'neg\_root\_mean\_squared\_error',

'normalized\_mutual\_info\_score',

'precision',

'precision\_macro',

'precision\_micro',

'precision\_samples',

'precision\_weighted',

'r2',

'rand\_score',

'recall',

'recall\_macro',

'recall\_micro',

'recall\_samples',

'recall\_weighted',

'roc\_auc',

'roc\_auc\_ovo',

'roc\_auc\_ovo\_weighted',

'roc\_auc\_ovr',

'roc\_auc\_ovr\_weighted',

'top\_k\_accuracy',

'v\_measure\_score']

def grid\_search\_2(X\_train,y\_train,X\_val,y\_val,X\_test,y\_test,X\_test\_combined, y\_test\_combined):

start\_time = time.time()

rf = RandomForestRegressor(random\_state=42, n\_jobs=22)

param\_grid = [{'n\_estimators': [250],

'criterion': ['squared\_error'], # you want this !

# 'max\_features': [5,10,20, 25],

'max\_depth': [10, 15, 20, 30, 40],

# 'min\_samples\_leaf': [1,2,5,10],

'bootstrap': [True]}]

grid\_search\_rf = GridSearchCV(estimator = rf,

param\_grid = param\_grid,

cv=3,

n\_jobs = 22,

verbose=1,

return\_train\_score=True,

scoring='r2')

grid\_search\_rf.fit(X\_train, y\_train)

print("\nBest parameters set found on Cross Validation:")

print(grid\_search\_rf.best\_params\_, " <-\*\*\*")

print('\nBest Estimator:\n', grid\_search\_rf.best\_estimator\_)

print("\nCross Validation R\u00b2 score ....... ", grid\_search\_rf.best\_score\_.round(3))

print("\n--- %s seconds ---" % round((time.time() - start\_time)))

Fitting 3 folds for each of 5 candidates, totalling 15 fits

Best parameters set found on Cross Validation:

{'bootstrap': True, 'criterion': 'squared\_error', 'max\_depth': 30, 'n\_estimators': 250} <-\*\*\*

Best Estimator:

RandomForestRegressor(max\_depth=30, n\_estimators=250, n\_jobs=22,

random\_state=42)

Cross Validation R² score ....... 0.775

--- 336 seconds ---

## Test 5 – GridsearchCV to tune RFR model

Bbbb