# Variable Selection

**Recursive Feature elimination**

Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), recursive feature elimination ([**RFE**](http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html#sklearn.feature_selection.RFE)) is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and weights are assigned to each one of them. Then, features whose absolute weights are the smallest are pruned from the current set features. That procedure is recursively repeated on the pruned set until the desired number of features to select is eventually reached.

**ExtraTrees Classifier**

An extremely randomized tree classifier.Extra-trees differ from classic decision trees in the way they are built. When looking for the best split to separate the samples of a node into two groups, random splits are drawn for each of the max\_features randomly selected features and the best split among those is chosen. When max\_features is set 1, this amounts to building a totally random decision tree.Warning: Extra-trees should only be used within ensemble methods.

**Random Forest Regressor**

A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True (default).

**Compare output of each method**

**RFE output –**

Features sorted by their rank:

[(1, 'hit\_OBP'), (2, 'hit\_OPS'), (3, 'hit\_SLG'), (4, 'hit\_BA'), (5, 'pitch\_WHIP'), (6, 'pitch\_BB9'), (7, 'pitch\_H9'), (8, 'pitch\_HR9'), (9, 'pitch\_FIP'), (10, 'pitch\_WL'), (11, 'hit\_RG'), (12, 'pitch\_ERA'), (13, 'pitch\_SOW'), (14, 'pitch\_SO9'), (15, 'pitch\_W'), (16, 'pitch\_GS'), (17, 'pitch\_G'), (18, 'hit\_G'), (19, 'pitch\_RAG'), (20, 'pitch\_PAge'), (21, 'pitch\_CG'), (22, 'pitch\_GF'), (23, 'hit\_BatAge'), (24, 'pitch\_ER'), (25, 'pitch\_H'), (26, 'pitch\_BB'), (27, 'pitch\_IP'), (28, 'pitch\_cSho'), (29, 'pitch\_BK'), (30, 'hit\_H'), (31, 'hit\_AB'), (32, 'hit\_SF'), (33, 'hit\_SH'), (34, 'hit\_PA'), (35, 'pitch\_L'), (36, 'pitch\_SO'), (37, 'hit\_LOB'), (38, 'pitch\_LOB'), (39, 'pitch\_BF'), (40, 'pitch\_R\_allowed'), (41, 'hit\_TB'), (42, 'hit\_RBI'), (43, 'pitch\_tSho'), (44, 'hit\_HR'), (45, 'pitch\_numP'), (46, 'hit\_BB'), (47, 'hit\_HBP'), (48, 'hit\_GDP'), (49, 'year'), (50, 'hit\_CS'), (51, 'pitch\_WP'), (52, 'pitch\_HBP'), (53, 'hit\_R\_scored'), (54, 'hit\_2B'), (55, 'pitch\_HR'), (56, 'pitch\_ERAplus'), (57, 'pitch\_IBB'), (58, 'hit\_3B'), (59, 'hit\_OPS+'), (60, 'hit\_SB'), (61, 'hit\_SO'), (62, 'pitch\_SV'), (63, 'hit\_numBat'), (64, 'hit\_IBB')]

**ExtraTrees Classifier output -**

[(0.027300000000000001, 'hit\_BatAge'), (0.026800000000000001, 'hit\_3B'), (0.0253, 'pitch\_H9'), (0.025100000000000001, 'pitch\_WP'), (0.024899999999999999, 'hit\_SH'), (0.0246, 'pitch\_W'), (0.024, 'hit\_BA'), (0.023900000000000001, 'hit\_LOB'), (0.022599999999999999, 'pitch\_PAge'), (0.022499999999999999, 'pitch\_HBP'), (0.022499999999999999, 'hit\_TB'), (0.0224, 'hit\_2B'), (0.020899999999999998, 'hit\_BB'), (0.0201, 'pitch\_SOW'), (0.019599999999999999, 'pitch\_tSho'), (0.0189, 'pitch\_WHIP'), (0.0184, 'hit\_GDP'), (0.018100000000000002, 'pitch\_CG'), (0.017899999999999999, 'pitch\_R\_allowed'), (0.017600000000000001, 'pitch\_IBB'), (0.017600000000000001, 'hit\_IBB'), (0.017500000000000002, 'pitch\_SV'), (0.0172, 'pitch\_BK'), (0.017000000000000001, 'hit\_OPS+'), (0.016799999999999999, 'pitch\_BB'), (0.016400000000000001, 'pitch\_numP'), (0.016400000000000001, 'pitch\_H'), (0.016299999999999999, 'hit\_HBP'), (0.0161, 'pitch\_ERA'), (0.0161, 'hit\_HR'), (0.015900000000000001, 'pitch\_ERAplus'), (0.015699999999999999, 'pitch\_LOB'), (0.015699999999999999, 'hit\_OPS'), (0.015599999999999999, 'pitch\_SO'), (0.015299999999999999, 'pitch\_BB9'), (0.014500000000000001, 'pitch\_HR9'), (0.014500000000000001, 'pitch\_ER'), (0.0144, 'hit\_CS'), (0.014, 'hit\_SO'), (0.013299999999999999, 'pitch\_RAG'), (0.0132, 'hit\_RBI'), (0.0126, 'hit\_AB'), (0.0123, 'hit\_SB'), (0.0121, 'hit\_numBat'), (0.012, 'pitch\_SO9'), (0.012, 'pitch\_BF'), (0.012, 'hit\_RG'), (0.011599999999999999, 'hit\_SLG'), (0.0115, 'year'), (0.0115, 'hit\_H'), (0.0112, 'hit\_OBP'), (0.010500000000000001, 'pitch\_IP'), (0.010500000000000001, 'hit\_G'), (0.010200000000000001, 'pitch\_GF'), (0.0097999999999999997, 'pitch\_FIP'), (0.0092999999999999992, 'pitch\_G'), (0.0088000000000000005, 'hit\_SF'), (0.0086, 'pitch\_cSho'), (0.0086, 'hit\_PA'), (0.0077999999999999996, 'pitch\_WL'), (0.0077999999999999996, 'hit\_R\_scored'), (0.0071999999999999998, 'pitch\_L'), (0.0068999999999999999, 'pitch\_HR'), (0.0025000000000000001, 'pitch\_GS')]

**Random Forest Regressor output –**

Features sorted by their score:

[(0.077200000000000005, 'pitch\_HBP'), (0.061400000000000003, 'pitch\_PAge'), (0.059400000000000001, 'hit\_BB'), (0.0511, 'hit\_3B'), (0.039399999999999998, 'pitch\_WP'), (0.039100000000000003, 'pitch\_IBB'), (0.039100000000000003, 'hit\_BatAge'), (0.037499999999999999, 'hit\_CS'), (0.033700000000000001, 'hit\_SO'), (0.030099999999999998, 'pitch\_ER'), (0.0293, 'pitch\_WHIP'), (0.027400000000000001, 'pitch\_SV'), (0.0218, 'hit\_SH'), (0.020199999999999999, 'pitch\_IP'), (0.020199999999999999, 'pitch\_ERAplus'), (0.02, 'hit\_R\_scored'), (0.019199999999999998, 'hit\_SB'), (0.0189, 'pitch\_H'), (0.018499999999999999, 'pitch\_HR'), (0.017899999999999999, 'pitch\_H9'), (0.0178, 'pitch\_W'), (0.017500000000000002, 'pitch\_FIP'), (0.0154, 'hit\_IBB'), (0.0147, 'hit\_HR'), (0.0143, 'hit\_SLG'), (0.012800000000000001, 'pitch\_BF'), (0.012200000000000001, 'pitch\_SO'), (0.0117, 'hit\_H'), (0.0114, 'hit\_OPS+'), (0.0111, 'pitch\_WL'), (0.010699999999999999, 'pitch\_BK'), (0.0104, 'pitch\_SO9'), (0.010200000000000001, 'hit\_RBI'), (0.0101, 'pitch\_LOB'), (0.0092999999999999992, 'pitch\_tSho'), (0.0091000000000000004, 'pitch\_BB'), (0.0088000000000000005, 'year'), (0.0088000000000000005, 'pitch\_G'), (0.0086999999999999994, 'hit\_GDP'), (0.0083000000000000001, 'hit\_TB'), (0.0083000000000000001, 'hit\_SF'), (0.0082000000000000007, 'hit\_LOB'), (0.0077000000000000002, 'hit\_PA'), (0.0068999999999999999, 'pitch\_ERA'), (0.0064000000000000003, 'hit\_G'), (0.0061999999999999998, 'pitch\_numP'), (0.0061999999999999998, 'hit\_HBP'), (0.0047999999999999996, 'hit\_numBat'), (0.0041999999999999997, 'hit\_AB'), (0.0041000000000000003, 'hit\_OBP'), (0.0040000000000000001, 'hit\_2B'), (0.0032000000000000002, 'pitch\_SOW'), (0.0030000000000000001, 'pitch\_BB9'), (0.0022000000000000001, 'pitch\_R\_allowed'), (0.0018, 'pitch\_HR9'), (0.0018, 'hit\_BA'), (0.0012999999999999999, 'pitch\_L'), (0.0011999999999999999, 'hit\_OPS'), (0.0011000000000000001, 'pitch\_CG'), (0.001, 'pitch\_RAG'), (0.00080000000000000004, 'hit\_RG'), (0.00040000000000000002, 'pitch\_cSho'), (0.00040000000000000002, 'pitch\_GF'), (0.0, 'pitch\_GS')]