**Feature Selection**

Snega

AP21110010474

* **Feature selection** or **variable selection** is the process of selecting a subset of relevant features or variables from the total features of a level in a data set to build machine learning algorithms.

**Advantages of selecting features**[¶](https://www.kaggleusercontent.com/kf/47174422/eyJhbGciOiJkaXIiLCJlbmMiOiJBMTI4Q0JDLUhTMjU2In0..MpwQTsjuppqBkHsiYuPayQ.GuVPdBym4VGM8mB-XZFPmnoDVtnoYRpSSBRb0wPVKj0CfSq7MyzsuxVOVz5OogvGe8YXu-HKXFIkhPGGy8gG3IeBMsKkIM9xPWlX0wZeOFk5I7-hqmY7HLD4SbkHTl18jtFhCq2maplla7qHpoBKF5ADxDwK310xwSvAVb4y_JUaQ5Hn025SkOyXwsNHszkl2k0OA3q49AkcQ8nVQdIdw8aB-Anc3eo-TN52JHomaDrYIi-N9WL4ktZiH1iFyKpT-EHzQfV9405WtLEsJXug6o-k9Ex2QbUgYN0HN2Cj9jYnpxdSiKFUYAO-bQt5cjqEFpJlKBmO8r3MsCeEcXnz52Ys1k6P1jT2LG4SEapxqINwWewi0QNqQlHJOUCtZ7eA8Hz6p8AecLagDPY8Ufqu9jZmoIi16xuKCDEWyK9-oqzoeCnF3Pj-Dis_dmo_quuvjIyd_zwbUumPeJ-61gpCWXU-0aYU_NJKfmvVsLYGUL9rTcFahJnk-w1thM7mEu2Qnc7H_i57c8SfOOSBfWJeqB4J9WIapIGvUFhGwiufp6hhfe8dM2yTEAn0hw59s5443AlMrgKWAivnlubnTEwWOzwoyldIr3o7O_X9MeYi3JQrDS2ENQ1CmJjVZaof58WQa9ei93LQYxRM4omEczs3Sr2RKndGwVBLCfQHbVxO-ts.AMsHdsSuTU9LAeb76AhDeA/__resultx__.html?sharingControls=true#Advantages-of-selecting-features)

* There are various advantages of feature selection process. These are as follows:-
  1. Improved accuracy
  2. Simple models are easier to interpret.
  3. Shorter training times
  4. Enhanced generalization by reducing Overfitting
  5. Easier to implement by software developers
  6. Reduced risk of data errors by model use
  7. Variable redundancy
  8. Bad learning behaviour in high dimensional spaces

**Feature Selection – Techniques**

* Feature selection techniques are categorized into 3 typers. These are as follows:-
  1. Filter methods
  2. Wrapper methods
  3. Embedded methods

**Filter Methods**

* Filter methods consists of various techniques as given below:-
  1. Basic methods
  2. Univariate methods
  3. Information gain
  4. Fischer score
  5. Correlation Matrix with Heatmap

**Wrapper Methods**

* 1. Forward Selection
  2. Backward Elimination
  3. Exhaustive Feature Selection
  4. Recursive Feature Elimination
  5. Recursive Feature Elimination with Cross-Validation

**Embedded Methods**

* Embedded methods consists of the following techniques:-
  1. LASSO
  2. RIDGE
  3. Tree Importance
* Now, we will discuss these methods in detail.

**2. Filter Methods**

* Filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithms. Instead, features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. The characteristics of these methods are as follows:-
  + These methods rely on the characteristics of the data (feature characteristics)
  + They do not use machine learning algorithms.
  + These are model agnostic.
  + They tend to be less computationally expensive.
  + They usually give lower prediction performance than wrapper methods.
  + They are very well suited for a quick screen and removal of irrelevant features.
* Filter methods consists of various techniques as given below:-
  + 2.1. Basic methods
  + 2.2. Univariate feature selection
  + 2.3. Information gain
  + 2.4. Fischer score
  + 2.5. ANOVA F-Value for Feature Selection
  + 2.6. Correlation Matrix with Heatmap
* Filter methods can be explained with the help of following graphic:

