

[Start Here](#)[Blog](#)[Training](#)[About](#)

# Feature Selection in Python with Scikit-Learn

by **Jason Brownlee** on July 14, 2014 in **Uncategorized**

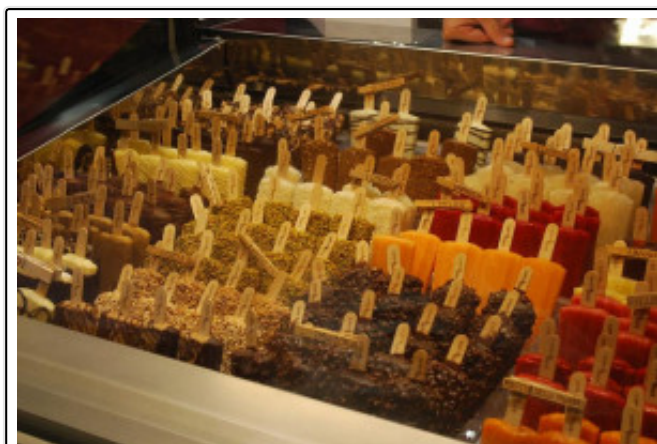
5

4

20

Not all data attributes are created equal. More is not always better when it comes to attributes or columns in your dataset.

In this post you will discover how to select attributes in your data before creating a machine learning model using the [scikit-learn library](#).



Cut Down on Your Options with Feature Selection  
Photo by [Josh Friedman](#), some rights reserved

## Select Features

Feature selection is a process where you automatically select those features in your data that contribute most to the prediction variable or output in which you are interested.

Having too many irr

[Machine Learning Crash Course](#)

acy of the models. Three

benefits of performing

- **Reduces Overfitting** on noise.
- **Improves Accuracy**
- **Reduces Training Time**

Two different feature selection methods are available in the `sklearn.feature_selection` module: `Recursive Feature Elimination` and `Recursive Feature Selection`.

## Recursive Feature Elimination

The Recursive Feature Elimination (RFE) method is a wrapper around a classifier that recursively removes attributes and building a model on those attributes that remain. It uses the model accuracy to identify which attributes (and combination of attributes) contribute the most to predicting the target attribute.

## Machine Learning Crash Course

You can learn and apply Machine Learning even without a background in maths or a fancy degree. Find out how in this ridiculously practical (and totally free) email course.

**SIGN ME UP**

to make decisions based on the data. The accuracy improves. The process is iterative. The `sklearn.feature_selection` module on library are Recursive Feature Elimination and Recursive Feature Selection.

This recipe shows the use of RFE on the Iris floweres dataset to select 3 attributes.

```
1 # Recursive Feature Elimination
2 from sklearn import datasets
3 from sklearn.feature_selection import RFE
4 from sklearn.linear_model import LogisticRegression
5 # load the iris datasets
6 dataset = datasets.load_iris()
7 # create a base classifier used to evaluate a subset of attributes
8 model = LogisticRegression()
9 # create the RFE model and select 3 attributes
10 rfe = RFE(model, 3)
11 rfe = rfe.fit(dataset.data, dataset.target)
12 # summarize the selection of the attributes
13 print(rfe.support_)
14 print(rfe.ranking_)
```

For more information see the [RFE method](#) in the API documentation.

## Feature Importance

Methods that use ensembles of decision trees (like Random Forest or Extra Trees) can also compute the relative importance of each attribute. These importance values can be used to inform a feature selection process.

This recipe shows the construction of an Extra Trees ensemble of the iris flowers dataset and the display of the relative feature importance.

```
1 # Feature Importance
2 from sklearn import datasets
3 from sklearn import metrics
4 from sklearn.ensemble import ExtraTreesClassifier
5 # load the iris datasets
6 dataset = datasets.load_iris()
7 # fit an Extra Trees model to the data
8 model = ExtraTreesClassifier()
```

```
9 model.fit(dataset.data, dataset.target)
10 # display the relative importance of each attribute
11 print(model.feature_importances_)
```

For more information, see the [ExtraTreesClassifier method](#) in the API documentation.

## Your Guide to Essential Scikit-Learn Resources



I've hand-picked the essential Scikit-Learn resources and created an exclusive Resource Kit.

Included are the best Scikit-Learn books, videos and MORE.

I can send it to you right now, *for FREE!*

[Download For Free](#)

You will also get emails with my best Tips and Tricks for getting started in machine learning in Python.

### Are you ready to Jump Ahead?

Checkout my [Jump-Start Scikit-Learn Playbook](#).

In it, you will discover 35 standalone Scikit-Learn recipes that you can copy-paste into your project.

## Summary

Feature selection methods can give you useful information on the relative importance or relevance of features for a given problem. You can use this information to create filtered versions of your dataset and increase the accuracy of your models.

In this post you discovered two feature selection methods you can apply in Python using the scikit-learn library.



### About Jason Brownlee

The editor-in-chief at MachineLearningMastery.com. Jason is a husband, father, researcher, author, professional programmer and a machine learning practitioner. [Learn more about him.](#)

[View all posts by Jason Brownlee →](#)

---

### 3 Responses to *Feature Selection in Python with Scikit-Learn*



**Harsh** October 9, 2014 at 4:51 pm #

REPLY ↩

Nice post, how does RFE and Feature selection like chi2 are different. I mean, finally they are achieving the same goal, right?



**jasonb** October 10, 2014 at 6:52 am #

REPLY ↩

Both seek to reduce the number of features, but they do so using different methods. chi squared is a univariate statistical measure that can be used to rank features, whereas RFE tests different subsets of features.



**Bozhidar** June 26, 2015 at 11:04 pm #

REPLY ↩

Hello,

Can you tell me which feature selection methods you suggest for time-series data?

### Leave a Reply

Name (required)

Email (will not be published) (required)

Website

[SUBMIT COMMENT](#)

## Welcome!



Hi, my name is Jason and welcome to Machine Learning Mastery where we make programmers awesome at machine learning.

If you have any questions or you just want to chat about machine learning, [send me an email](#).

## Free Guide to the Best Resources

Discover the best videos, books, courses and MUCH MORE in my Machine Learning Resource Guide.

Get your copy right now!

[Download For Free](#)

### POPULAR



#### A Tour of Machine Learning Algorithms

NOVEMBER 25, 2013



#### How to Run Your First Classifier in Weka

FEBRUARY 17, 2014



#### Tutorial To Implement k-Nearest Neighbors in Python From Scratch

SEPTEMBER 12, 2014



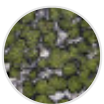
#### Machine Learning for Programmers: Leap from developer to machine learning practitioner

AUGUST 17, 2015



#### How To Implement Naive Bayes From Scratch in Python

DECEMBER 8, 2014



#### Discover Feature Engineering, How to Engineer Features and How to Get Good at It

SEPTEMBER 26, 2014



#### 4 Self-Study Machine Learning Projects

JANUARY 3, 2014

#### Best Machine Learning Resources for Getting Started

NOVEMBER 27, 2013



**Find Your Machine Learning Tribe: Get Started And Avoid Getting The Wrong Advice**

AUGUST 31, 2015



**Feature Selection with the Caret R Package**

SEPTEMBER 22, 2014

**Looking for something? Search!**



© 2016 Machine Learning Mastery. All Rights Reserved.

[Privacy](#) | [Contact](#) | [About](#)