

Want to learn machine learning? Use my **machine learning flashcards**.

Random Forest Classifier Example

20 Dec 2017

This tutorial is based on Yhat's 2013 tutorial on [Random Forests in Python](#). If you want a good summary of the theory and uses of random forests, I suggest you check out their guide. In the tutorial below, I annotate, correct, and expand on a short code example of random forests they present at the end of the article. Specifically, I 1) update the code so it runs in the latest version of pandas and Python, 2) write detailed comments explaining what is happening in each step, and 3) expand the code in a number of ways.

Let's get started!

A Note About The Data

The data for this tutorial is famous. Called, [the iris dataset](#), it contains four variables measuring various parts of iris flowers of three related species, and then a fourth variable with the species name. The reason it is so famous in machine learning and statistics communities is because the data requires very little preprocessing (i.e. no missing values, all features are floating numbers, etc.).

Preliminaries

```
# Load the library with the iris dataset
from sklearn.datasets import load_iris

# Load scikit's random forest classifier library
from sklearn.ensemble import RandomForestClassifier

# Load pandas
import pandas as pd

# Load numpy
import numpy as np
```

CHRIS ALBON

[ARTICLES](#) [ML/AI NOTES](#) [ABOUT](#)