How to save and load Random Forest from Scikit-Learn in Python?

June 24, 2020 by Piotr Płoński **Random forest**

In this post I will show you how to save and load Random Forest model trained with scikit-learn in Python. The method presented here can be applied to any algorithm from sckit-learn (this is amazing about scikit-learn!).

Additionally, I will show you, how to compress the model and get smaller file.

For saving and loading I will be using `joblib` package.

Let's load scikit-learn and joblib

```
import os
import joblib
import numpy as np
from sklearn.datasets import load_iris
from sklearn.ensemble import RandomForestClassifier
```

Create some dataset (I will use Iris dataset which is built-in in sklearn):

```
iris = load_iris()
X = iris.data
y = iris.target
```

Train the Random Forest classifier:

```
rf = RandomForestClassifier()
```

```
rf.predict(X)
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

Let’s save the Random Forest. I’m using `joblib.dump` method. The first argument of the method is variable with the model. The second argument is the path and the file name where the resulting file will be created.

```
# save
joblib.dump(rf, "./random_forest.joblib")
```

To load the model back I use `joblib.load` method. It takes as argument the path and file name. I will load the forest to new variable `loaded_rf`. Please notice that I don’t need to initilize this variable, just load the model into it.

```
# load, no need to initialize the loaded_rf
loaded_rf = joblib.load("./random_forest.joblib")
```

Let’s check if it works, by computing predictions, they should be exactly the same as from the `rf` model.

```
loaded_rf.predict(X)
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

They are the same. We successfully save and loaded back the Random Forest.

Extra tip for saving the Scikit-Learn Random

parameter to save the disk space. In the joblib docs there is information that `compress=3` is a good compromise between size and speed. Example below:

```
joblib.dump(rf, "RF_uncompressed.joblib", compress=0)
print(f"Uncompressed Random Forest: {np.round(os.path.getsize('RF_

>>> Uncompressed Random Forest: 0.17 MB
```

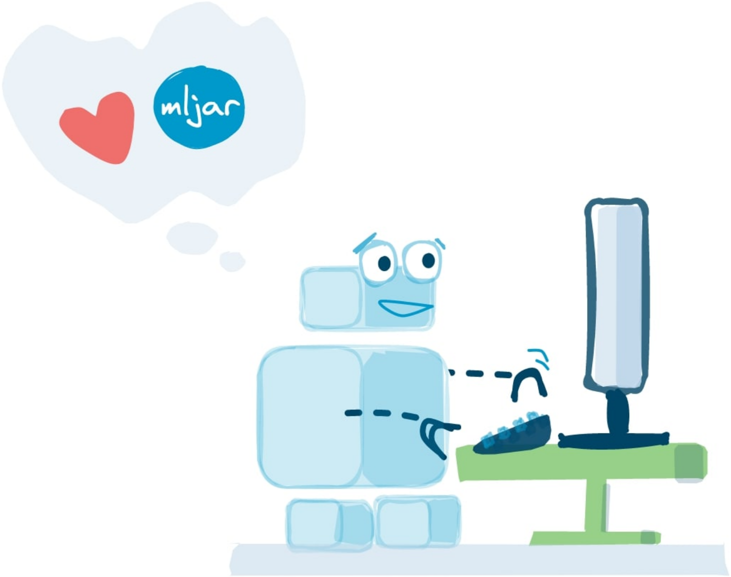
```
joblib.dump(rf, "RF_compressed.joblib", compress=3)  # compression
print(f"Compressed Random Forest: {np.round(os.path.getsize('RF_co

>>> Compressed Random Forest: 0.03 MB
```

Compressed Random Forest is `5.6` times smaller! The compression can be used to any sckit-learn model (sklearn is amazing!).

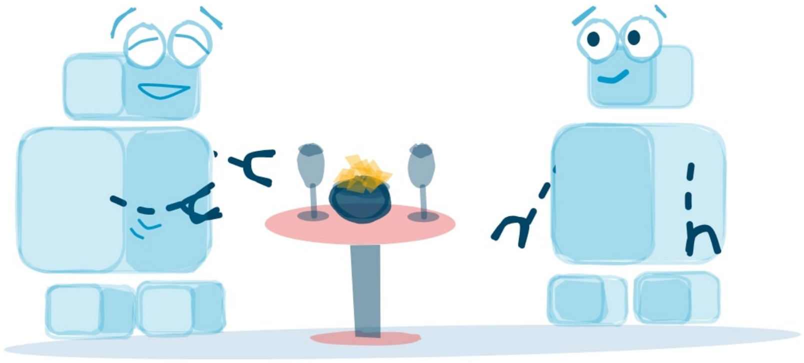
« How to reduce memory used by Random Forest from Scikit-Learn in Python?

Random Forest Feature Importance Computed in 3 Ways with Python »



Articles you might find interesing

1. 8 surprising ways how to use Jupyter Notebook
2. Create a dashboard in Python with Jupyter Notebook
3. Build Computer Vision Web App with Python
4. Develop NLP Web App from Python Notebook
5. Build dashboard in Python with updates and email notifications
6. Share Jupyter Notebook with non-technical users



Join our newsletter

Subscribe to our newsletter to receive product updates

Subscribe

mljar

Outstanding Data Science Tools

Blog

About

Brand Assets

GitHub

Twitter

Mercury

AutoML

Pricing

Compare Algorithms

AutoML Comparison

Random Forest vs Xgboost

Golden Features

Xgboost vs LightGBM

K-Means Features

CatBoost vs Xgboost

Feature Selection