

Navigation



Want help with Python and scikit-learn? Take the FREE Mini-Course

Search...

Save and Load Machine Learning Models in Python with scikit-learn

by Jason Brownlee on June 8, 2016 in Python Machine Learning



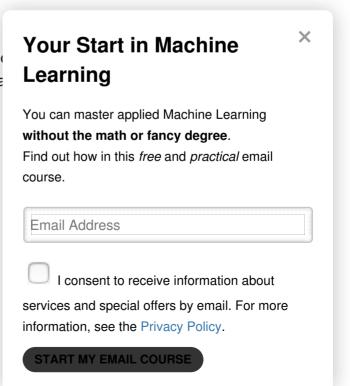
Finding an accurate machine learning model is not the end of the project.

In this post you will discover how to save and load your machine learning model in Python using scikit-learn.

This allows you to save your model to file and load it later in order to make predictions.

Let's get started.

- Update Jan/2017: Updated to reflect
- Update March/2018: Added alternate appears to have been taken down.





Save and Load Machine Learning Models in Python with scikit-learn Photo by Christine, some rights reserved.

Need help with Machine Learning in Python?

Take my free 2-week email course and discover data prep, algorithms and more (with code).

Click to sign-up now and also get a free PDF Ebook version of the course.

Start Your F **Your Start in Machine** Learning You can master applied Machine Learning without the math or fancy degree. **Finalize Your Model wit** Find out how in this free and practical email course. Pickle is the standard way of serializing You can use the pickle operation to seria ìе serialized format to a file. I consent to receive information about Later you can load this file to deserialize services and special offers by email. For more information, see the Privacy Policy. The example below demonstrates how y ma Indians onset of diabetes dataset, save START MY EMAIL COURSE the unseen test set (update: download fr

```
# Save Model Using Pickle
   import pandas
3
   from sklearn import model_selection
  from sklearn.linear_model import LogisticRegression
  import pickle
  url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-di
  names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
dataframe = pandas.read_csv(url, names=names)
7
9 array = dataframe.values
10 X = array[:,0:8]
11 Y = array[:,8]
12 test_size = 0.33
13 \text{ seed} = 7
14 X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y, test_siz
15 # Fit the model on 33%
16 model = LogisticRegression()
17 model.fit(X_train, Y_train)
18 # save the model to disk
19 filename = 'finalized_model.sav'
20 pickle.dump(model, open(filename, 'wb'))
22 # some time later...
23
24 # load the model from disk
25 loaded model = pickle.load(open(filename, 'rb'))
26 result = loaded_model.score(X_test, Y_test)
27 print(result)
```

Running the example saves the model to **finalized_model.sav** in your local working directory. Load the saved model and evaluating it provides an estimate of accuracy of the model on unseen data.

1 0.755905511811

Finalize Your Model with joblib

Joblib is part of the SciPy ecosystem an

It provides utilities for saving and loading structures, efficiently.

This can be useful for some machine leastore the entire dataset (like K-Nearest)

The example below demonstrates how y Indians onset of diabetes dataset, saves predictions on the unseen test set.

1	# Save Model Using joblib
2	import pandas
3	from sklearn import model_selecti
4	<pre>from sklearn.linear_model import</pre>
5	from sklearn externals import jok
6	<pre>url = "https://raw.githubusercont</pre>
	·

Your	Start	ın	Machine
Learr	ning		

You can master applied Machine Learning without the math or fancy degree.

X

na

s-di

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

```
7 names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
8 dataframe = pandas.read_csv(url, names=names)
9 array = dataframe.values
10 X = array[:,0:8]
11 Y = array[:,8]
12 test_size = 0.33
13 \text{ seed} = 7
14 X_train, X_test, Y_train, Y_test = model_selection.train_test_split(X, Y, test_siz
15 # Fit the model on 33%
16 model = LogisticRegression()
17 model.fit(X_train, Y_train)
18 # save the model to disk
19 filename = 'finalized_model.sav'
20 joblib.dump(model, filename)
21
22 # some time later...
23
24 # load the model from disk
25 loaded_model = joblib.load(filename)
26 result = loaded_model.score(X_test, Y_test)
27 print(result)
```

Running the example saves the model to file as **finalized_model.sav** and also creates one file for each NumPy array in the model (four additional files). After the model is loaded an estimate of the accuracy of the model on unseen data is reported.

```
1 0.755905511811
```

Tips for Finalizing Your Model

This section lists some important considerations when finalizing your machine learning models.

- Python Version. Take note of the p major (and maybe minor) version of load it and deserialize it.
- Library Versions. The version of all project almost certainly need to be the not limited to the version of NumPy
- Manual Serialization. You might like model so that you can use them dired.
 Often the algorithms used by maching simpler than those used to learn the custom code that you have control of

Take note of the version so that you can cannot reload your model on another ma

Summary

		ıme
Your Start in Machine Learning	×	r
You can master applied Machine Learning without the math or fancy degree.		is
Find out how in this free and practical email		ed
course.		€.
		ot
I consent to receive information about		
services and special offers by email. For more		
information, see the Privacy Policy.		
START MY EMAIL COURSE		

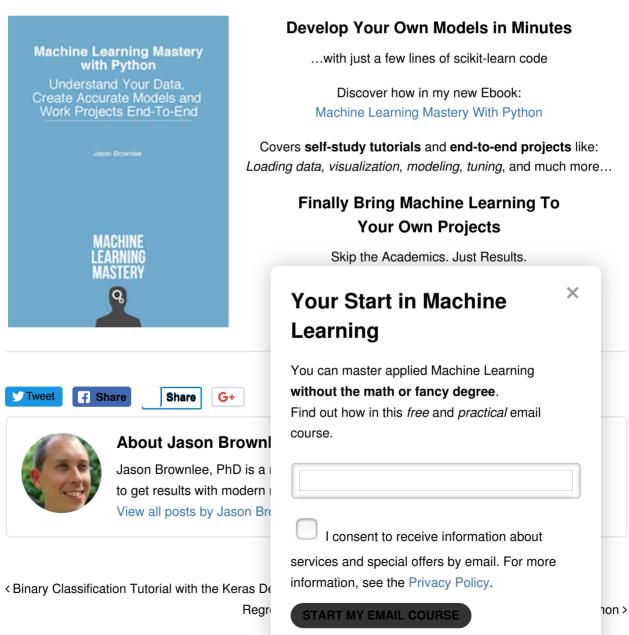
In this post you discovered how to persist your machine learning algorithms in Python with scikit-learn.

You learned two techniques that you can use:

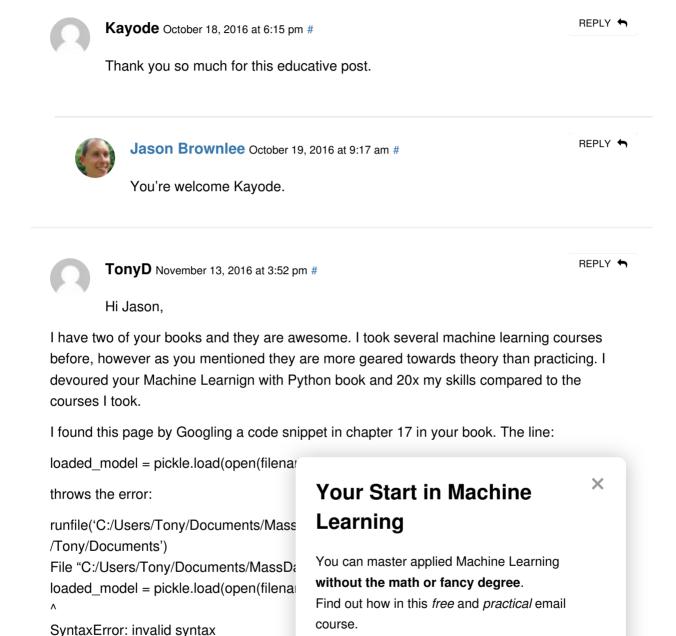
- The pickle API for serializing standard Python objects.
- The joblib API for efficiently serializing Python objects with NumPy arrays.

Do you have any questions about saving and loading your machine learning algorithms or about this post? Ask your questions in the comments and I will do my best to answer them.

Frustrated With Python Machine Learning?



146 Responses to Save and Load Machine Learning Models in Python with scikit-learn



6 of 44 1/27/19, 12:58 AM

I consent to receive information about

services and special offers by email. For more

information, see the Privacy Policy.

START MY EMAIL COURSE



Jason Brownlee November 14, 2016 at 7:36 am #

REPLY 5

Thanks TonyD.

I wonder if there is a copy-paste error, like an extra space or something?

Does the code example (.py file) provided with the book for that chapter work for you?



William January 7, 2019 at 9:37 pm #

REPLY 🦴

As Jason already said, this is a copy paste problem. In your line specifically, the quotes are the problem.

loaded model = pickle.load(open(filename, 'rb'))

It should be

loaded model = pickle.load(open(filename, 'rb'))

Try to understand the difference:).



Jason Brownlee January 8, 2019 at 6:49 am #

REPLY 🦴

X

Thanks.

This might help:

https://machinelearningmastery.com/foo/single_foo/bow_doi.com/code_from

a-tutorial

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this free and practical email course.

I consent to receive information about services and special offers by email. For more

START MY EMAIL COURSE

information, see the Privacy Policy.



Konstantin November 19, 2016 at 6:01 am #

REPLY 🦴

Hello, Jason

Where we can get X_test, Y_test "sometime later"? It is "garbag collected"! X_test, Y_test not pickled In your example you pickle classifier only but you keep refer to x and y. Real applications is not single flow I found work around and get Y from clf.classes_object.

What is correct solution? Should we pickle decorator class with X and Y or use pickled classifier to pull Ys values? I didn't find legal information from documentation on KNeighborclassifier(my example) as well; how to pull Y values from classifier.

Can you advise?



Jason Brownlee November 19, 2016 at 8:51 am #

REPLY 🦴

X

Hi Konstantin,

I would not suggest saving the data. The idea is to show how to load the model and use it on new data – I use existing data just for demonstration purposes.

You can load new data from file in the future when you load your model and use that new data to make a prediction.

If you have the expected values also (y), you can compare the predictions to the expected values and see how well the model performed.



Guangping Zhang N

I'm newer Pythoner, yc used windows 10.



Jason Brownlee

Thanks Guangping

The save file is in your curre commandline.

If you're using a notebook of

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about
services and special offers by email. For more
information, see the Privacy Policy.

START MY EMAIL COURSE

Mohammed Alnemari December 13, 2016 at 2:45 pm

REPLY

Hi Jason,

I am just wondering if can we use Yaml or Json with sklearn library. I tried to do it many times but I could not reach to an answer. I tried to do it as your lesson of Kares, but for some reason is not working . hopefully you can help me if it is possible

Jason Brownlee December 14, 2016 at 8:24 am #

REPLY

Hi Mohammed, I believe the serialization of models to yaml and ison is specific to the Keras library.

sklearn serialization is focused on binary files like pickle.



Normando Zubia December 29, 2016 at 9:55 am

REPLY 🖛

Hi, my name is Normando Zubia and I have been reading a lot of your material for my school lessons.

I'm currently working on a model to predict user behavoir in a production environment. Due to several situations I can not save the model in a pickle file. Do you know any way to save the model in a json file?

I have been playing a little with sklearn classes and I noticed that if I save some parameters

for example: n values, feature indices can reproduce the results. Could this be each model's parameters to load each r

PS: Sorry for my bad english and thanks

Your Start in Machine Learning

and active features in a OneHatEncoding model X

You can master applied Machine Learning without the math or fancy degree.

Find out how in this free and practical email course.



Jason Brownlee Decembe

Hi Normando,

If you are using a simple model, you then try and put them back in a new algorithm yourself (very easy for mc

Let me know how you go.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

Samuel February 6, 2017 at 3:14 pm

REPLY 🦴

X

Hello Jason.

I am new to machine learning. I am your big fan and read a lot of your blog and books. Thank you very much for teaching us machine learning.

I tried to pickle my model but fail. My model is using VGG16 and replace the top layer for my classification solution. I further narrowed down the problem and find that it is the VGG16 model failed to pickle. Please find my simplified code below and error log below:

It will be highly appreciated if you can give me some direction on how to fix this error.

Thank you very much

Save Model Using Pickle from keras.applications.vgg16 import VGG16 import pickle

model = VGG16(weights='imagenet', include_top=False)

filename = 'finalized_model.sav' pickle.dump(model, open(filename, 'wb'))

/Library/Frameworks/Python.framework/Versions/2.7/bin/python2.7 /Users/samueltin/Projects /bitbucket/share-card-ml/pickle_test.py

Using TensorFlow backend.

Traceback (most recent call last):

File "/Users/samueltin/Projects/bitbucket/share-card-ml/pickle test.py", line 8, in

pickle.dump(model, open(filename, 'wb' File "/Library/Frameworks/Python.frame in dump

Pickler(file, protocol).dump(obj)

File "/Library/Frameworks/Python.frame dump

self.save(obj)

File "/Library/Frameworks/Python.frame save

self.save_reduce(obj=obj, *rv)

File "/Library/Frameworks/Python.frame save_reduce

save(state)

File "/Library/Frameworks/Python.frame save

f(self, obj) # Call unbound method with e File "/Library/Frameworks/Python.frame save_dict

Your	Start	in	Machine
Learr	ning		

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about			
ervices and special offers by email. For more			
formation, see the Privacy Policy.			

START MY EMAIL COURSE



self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py",

line 669, in _batch_setitems

save(v)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 606, in save list

self._batch_appends(iter(obj))

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 621, in _batch_appends

save(x)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 331, in save

self.save reduce(obj=obj, *rv)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 425, in save_reduce

save(state)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 655, in save dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 669, in

_batch_setitems

save(v)

File "/Library/Frameworks/Python.frame save

f(self, obj) # Call unbound method with & File "/Library/Frameworks/Python.frame save list

self._batch_appends(iter(obj))

File "/Library/Frameworks/Python.frame

_batch_appends

save(x)

File "/Library/Frameworks/Python.frame save

self.save_reduce(obj=obj, *rv)

File "/Library/Frameworks/Python.frame

save_reduce

save(state)

File "/Library/Frameworks/Python.frame

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

X

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 655, in save dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 669, in batch setitems

save(v)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 331, in save

self.save_reduce(obj=obj, *rv)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 425, in save_reduce

save(state)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 655, in save dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 669, in _batch_setitems

save(v)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 331, in save

self.save_reduce(obj=obj, *rv)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 425, in

save reduce

save(state)

File "/Library/Frameworks/Python.frame save

f(self, obj) # Call unbound method with & File "/Library/Frameworks/Python.frame save_dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.frame

_batch_setitems

save(v)

File "/Library/Frameworks/Python.frame save

f(self, obj) # Call unbound method with 6 File "/Library/Frameworks/Python.frame save_dict

self. batch setitems(obj.iteritems())

File "/Library/Frameworks/Python.frame

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

X

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

batch setitems save(v) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in f(self, obj) # Call unbound method with explicit self File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 606, in save list self. batch appends(iter(obj)) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 621, in batch appends save(x) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 331, in save self.save reduce(obi=obi, *rv) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 425, in save reduce save(state) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save f(self, obj) # Call unbound method with explicit self File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 655, in save dict self. batch setitems(obj.iteritems()) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 669, in batch setitems save(v) File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 331, in X self.save reduce(obj=obj, *rv) Your Start in Machine File "/Library/Frameworks/Python.frame Learning save reduce save(state) You can master applied Machine Learning File "/Library/Frameworks/Python.frame without the math or fancy degree. save Find out how in this free and practical email f(self, obj) # Call unbound method with course. File "/Library/Frameworks/Python.frame save_dict self. batch setitems(obj.iteritems()) File "/Library/Frameworks/Python.frame _batch_setitems I consent to receive information about save(v) services and special offers by email. For more File "/Library/Frameworks/Python.frame information, see the Privacy Policy. save self.save reduce(obj=obj, *rv) START MY EMAIL COURSE File "/Library/Frameworks/Python.frame

save reduce

save(state)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 655, in save dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 669, in _batch_setitems

save(v)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 606, in save_list

self._batch_appends(iter(obj))

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 621, in _batch_appends

save(x)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in save

f(self, obj) # Call unbound method with explicit self

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 568, in save_tuple

save(element)

File "/Library/Frameworks/Python.framework/Versions/2.7/lib/python2.7/pickle.py", line 286, in

save

f(self, obj) # Call unbound method with 6 File "/Library/Frameworks/Python.frame save_dict

self._batch_setitems(obj.iteritems())

File "/Library/Frameworks/Python.frame _batch_setitems

save(v)

File "/Library/Frameworks/Python.frame save

rv = reduce(self.proto)

File "/Library/Frameworks/Python.frame in reduce ex

raise TypeError, "can't pickle %s objects
TypeError: can't pickle module objects

Process finished with exit code 1

Your Start in Machine Learning

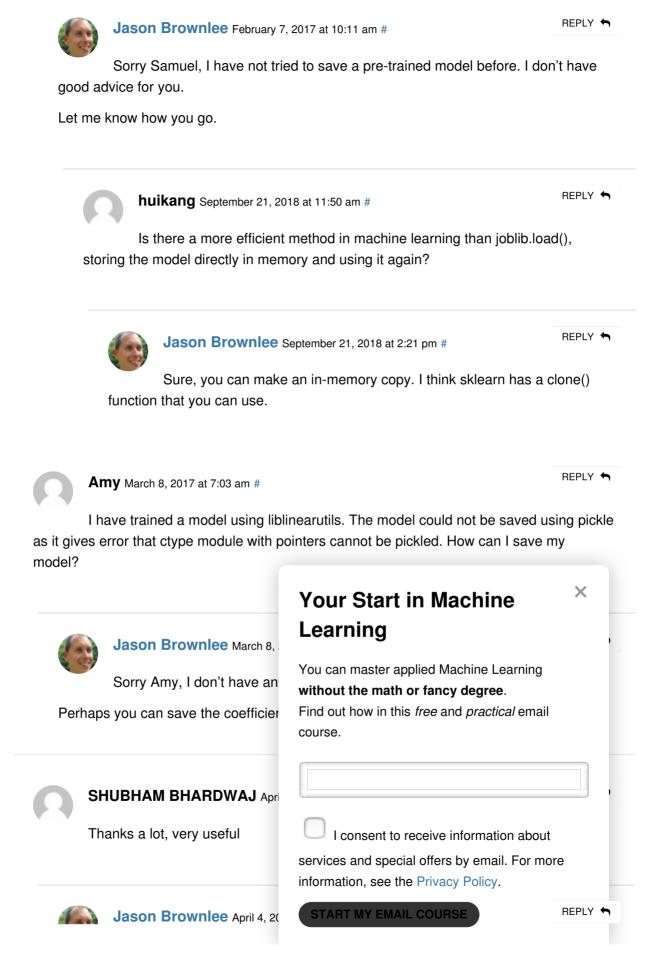
You can master applied Machine Learning without the math or fancy degree.

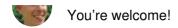
X

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE







Benju April 11, 2017 at 1:35 am #

REPLY 🦴

My saved modells are 500MB+ Big....is that normal?



Jason Brownlee April 11, 2017 at 9:34 am #

REPLY 🦴

Ouch, that does sound big.

If your model is large (lots of layers and neurons) then this may make sense.



REPLY 🖴

How to use model file ("finalized_model.sav") to test unknown data. Like, if the model is for tagger, how this model will tag the text file data? Is there any example?



REPLY 🖴

X

You can load the saved model and start making predictions (e.g. yhat = model.predict(X)).

See this post on finalizing models: http://machinelearningmastery.com/

Your Start in Machine Learning



Oss Mps April 21, 2017 at 3:09 pm

Dear Sir, please advice on how

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.



Jason Brownlee April 22, 2

I would suggest extracting in your preferred format.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

Suhas May 24, 2017 at 4:44 am #

REPLY 🦴

Hi I love your website; it's very useful!

Are there any examples showing how to save out the training of a model after say 100 epochs/iterations? It's not immediately clear from looking at joblib or scikit learn.

This is esp. useful when dealing with large datasets and/or computers or clusters which may be unreliable (e.g., subject to system reboots, etc.)

Jason Brownlee May 24, 2017 at 4:59 am #



I'm not sure how to do this with sklearn. You may need to write something custom. Consider posting to stackoverflow.

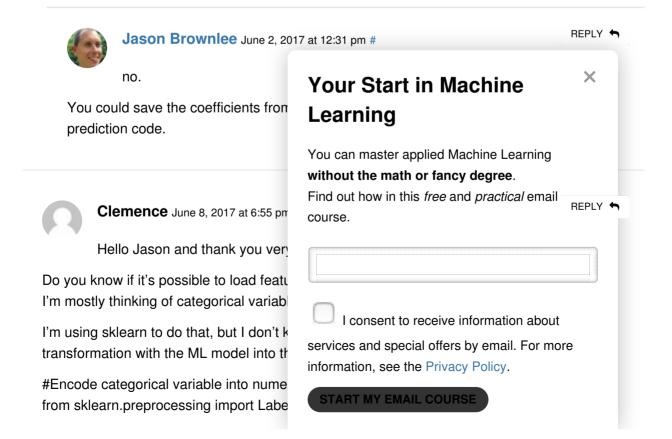


Viktor May 30, 2017 at 8:52 am #

REPLY 🦴

Hev

Is it possible to open my saved model and make a prediction on cloud server where is no sklearn installed?



list_var = ['country', 'city']
encoder = LabelEncoder()
for i in list_var:
df[i] = encoder.fit transform(df[i])

Then I fit the model on the training dataset...

And I need to save this transformation with the model. Do you know if that's possible? Thank you!



Jason Brownlee June 9, 2017 at 6:23 am #

REPLY 🦴

I'm not sure I follow sorry.

You can transform your data for your model, and you can apply this same transform in the future when you load your model.

You can save the transform objects using pickle. Is that what you mean?



Bhavani Shanker June 22, 2017 at 1:24 am

REPLY 🦴

Hi Jason,

Kindly accept my encomiums for the illustrative lecture that you have delivered on Machine Learning using Python.

*********** # save the model to disk X Your Start in Machine filename = 'finalized model.sav' joblib.dump(model, filename) Learning # sometime later... You can master applied Machine Learning # load the model from disk without the math or fancy degree. loaded_model = joblib.load(filename) Find out how in this free and practical email result = loaded_model.score(X_test, Y_ course. print(result) After saving the model 'finalized model. session at later date? I consent to receive information about I would appreciate if you can advice on services and special offers by email. For more information, see the Privacy Policy. **START MY EMAIL COURSE** REPLY +



Jason Brownlee June 22, 2017 at 6:11 am #

The code after "sometime later" would be in a new session.



jinsh June 28, 2017 at 8:57 pm #

REPLY 🦴

Hello sir.

The above code saves the model and later we can check the accuracy also but what i have to do for making predicting the class of unknown data?

I mean which function have to be called?

eg: 2,132,40,35,168,43.1,2.288,33

can you suggest how to get the class of above data through prediction? thank you

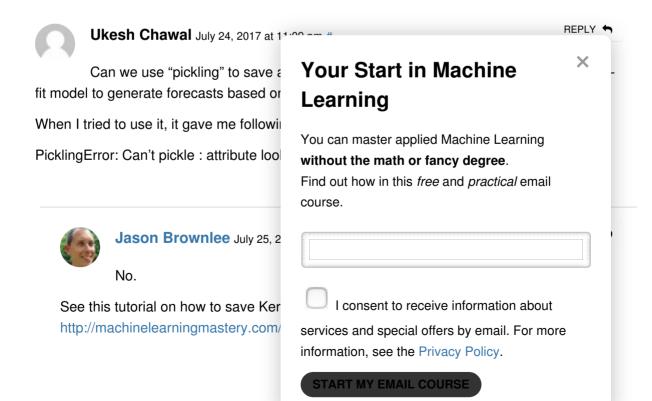


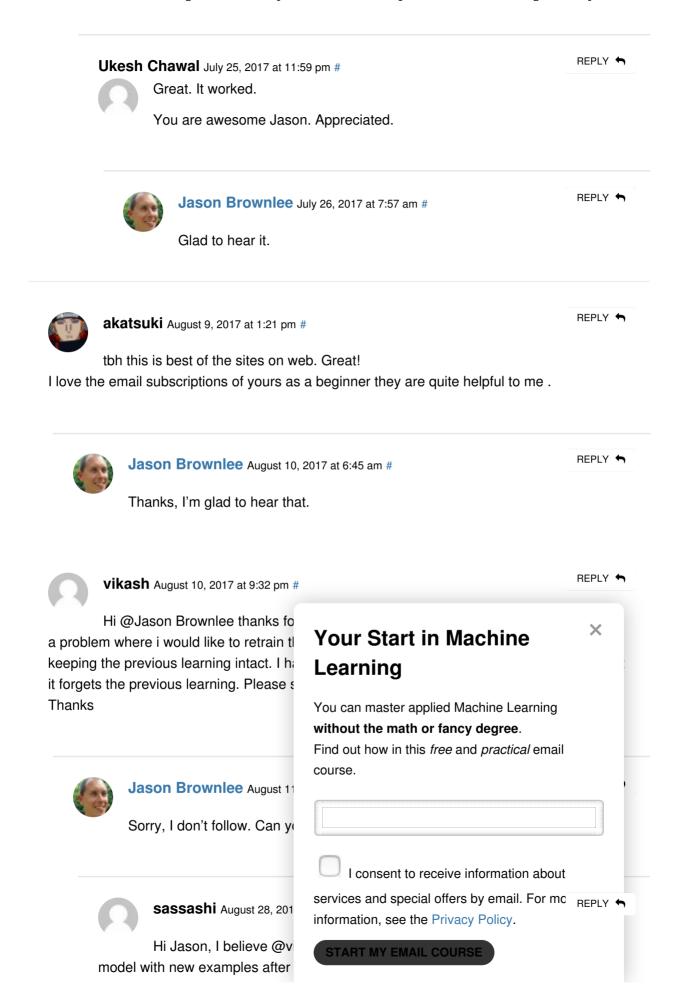
Jason Brownlee June 29, 2017 at 6:35 am #

REPLY 🦴

Pass in input data to the predict function and use the result.

1 yhat = model.predict(X)





searching for as well. I know it is possible to retrain a model in tensorflow with new examples but I am not sure if it's possible with sklearn.

to expand the question some more: 1-you train a model with sklearn 2-save it with pickle or joblib

3-then you get your hands on some new examples that were not available at the time of initial training "step 1" 4-you load the previous model 5-and now you try to train the model again using the new data without losing the previous knowledge... is step 5 possible with sklearn?



Jason Brownlee August 28, 2017 at 6:52 am #

REPLY 🦴

I have not updated a model in sklearn, but I would expect you can.

Here is an example of updating a model in Keras which may help in general principle:

https://machinelearningmastery.com/update-lstm-networks-training-time-series-forecasting/



Navdeep Singh August 22, 2017 at 8:30 pm #

REPLY 🦴

X

Hi Json,

I need your guidance on Updation of saved pickle files with new data coming in for training

I recall 3 methods, Online Learning which is train one every new observation coming in and in

this case model would always be biased

Second is, Whenever some set of n obs do retraining again from scratch, that i d time

Third is Mini batch learning, i know som method and do same but I have other a logistic regression. I wana ask can i upo

I am doing it in text classification, I read take new features of new data (made u help.

Also as domain is same, and If client(Pr sharing old data with new client (new pr update it with training in new client data

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about
services and special offers by email. For more
information, see the Privacy Policy.

START MY EMAIL COURSE



Jason Brownlee August 23, 2017 at 6:48 am #

REPLY 숙

Great question.

This is a challenging problem to solve. Really, the solution must be specific to your project requirements.

A flexible approach may be to build-in capacity into your encodings to allow for new words in the future.

The simplest approach is to ignore new words.

These, and other strategies are testable. See how performance degrades under both schemes with out-of-band test data.



Merari September 11, 2017 at 7:59 am #

REPLY 🖴

X

Gracias por compartir,

Existe alguna forma en la que pueda realizar predicciones con nuevos datos solo con el modelo guardado? llamando este modelo desde un archivo nuevo? lo he intentado con la instruccion final:

load the model from disk
loaded_model = pickle.load(open(filename, 'rb'))
result = loaded_model.score(X_test, Y_test)
print(result)

pero no lo he logrado

373/5000

Thanks for sharing,

Is there any way I can make predictions model from a new file? I have tried with

load the model from disk loaded_model = pickle.load (open (filenaresult = loaded_model.score (X_test, Y_print (result)

but I have not achieved it

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE



Jason Brownlee September 11, 2017 at 12:11 pm #

REPLY 🦴

That is exactly what we do in this tutorial.

What is the problem exactly?



AP September 29, 2017 at 6:36 am #

REPLY 🦴

Hi Jason, I learn a lot reading your python books and blogs. Thank you for everything.

I'm having an issue when I work on text data with loaded model in a different session. I fit and transform training data with countvectorizer and tfidf. Then I only transform the test data with the fitted instances as usual. But, when work on loaded pretrained model in a different session, I am having problem in feature extraction. I can't just transform the test data as it asks for fitted instance which is not present in the current session. If I fit and transform on test data only, model prediction performance drastically decreases. I believe that is wrong way of doing machine learning. So, how can I do the feature extraction using countvectorizer, tfidf or other cases while working with previously trained model?

I'm using spark ML but I think it would be the same for scikit-learn as well.

	Jason Brownlee September 30, 2017 at 7:31 am #		
the sec	Perhaps you can pickle your and session?	data transform objects as well. and re-use	them in
the second session?		Your Start in Machine	×
		Learning	
Bh	navya Chugh October 29, 2017	You can master applied Machine Learning)
Hi	Jason,	without the math or fancy degree. Find out how in this <i>free</i> and <i>practical</i> emains	
I trained a random forest model and save then copied that pickle file to my remote giving incorrect predictions. I am using provided the version of scikit-learn are seconds.		course.	
		I consent to receive information about	
	Jason Brownlee October 2	services and special offers by email. For morinformation, see the Privacy Policy.	REPLY 🦴
	No idea, perhaps see if the	START MY EMAIL COURSE	

or different machines with the same version of Python?



Berkin Albert Antony November 10, 2017 at 5:45 pm

REPLY 🦴

Hi Jason Brownlee,

I have a LogisticRegression model for binary classification. I wish to find a similar data points in a trained model for a given test data points. So that I can show these are similar data points predicted with these same class.

Could you please suggest your thoughts for the same. I am using scikit learn logistic regression

Thanks



Jason Brownlee November 11, 2017 at 9:18 am #



Perhaps you could find data points with a low Euclidean distance from each other?



James November 16, 2017 at 8:47 am #

REPLY

Hi Jason -

If you pickle a model trained on a subse loading the pickled model in a different f You train the model on a,c,e. Is it possik see the model was trained on a,c,e?

Thanks, James

Your Start in Machine Learning



You can master applied Machine Learning without the math or fancy degree.

Find out how in this free and practical email course.

Jason Brownlee Novembe

Yes, you can save your mo on new data.



Mrinal Mitra November 22, 2017 a

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

REPLY +



Hi Jason.

Thanks for explaining it so nicely. I am new to this and will be needing your guidance. I have data using which I have trained the model. Now I want this model to predict an untested data set. However, my requirement is an output which will have the data and corresponding prediction by the model. For example, record 1 – type a, record 2 – type a, record 3 – type c and so on. Could you please guide me on this?

Jason Brownlee November 22, 2017 at 11:16 am #



You can provide predictions one at a time or in a group to the model and the predictions will be in the same order as the inputs.

Does that help?



Niranjan December 3, 2017 at 3:22 pm #

REPLY 🖴

X

Hi,

I am using chunks functionality in the read csv method in pandas and trying to build the model iteratively and save it. But it always saves the model that is being built in the last chunk and not the entire model. Can you help me with it

clf_SGD = SGDClassifier(loss='modified_huber', penalty='l2', alpha=1e-3, max_iter=500, random_state=42)

pd.read_csv("file_name",chunksize = 1000):

data preparation and cleaning

hashing = hv.fit_transform(X_train['desc clf_SGD.partial_fit(hashing, y_train, class

joblib.dump(clf_SGD, source_folder + o



You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.



Jason Brownlee Decembe

Sorry, I'm not sure I follow,



Shabbir December 8, 2017 at 8:50

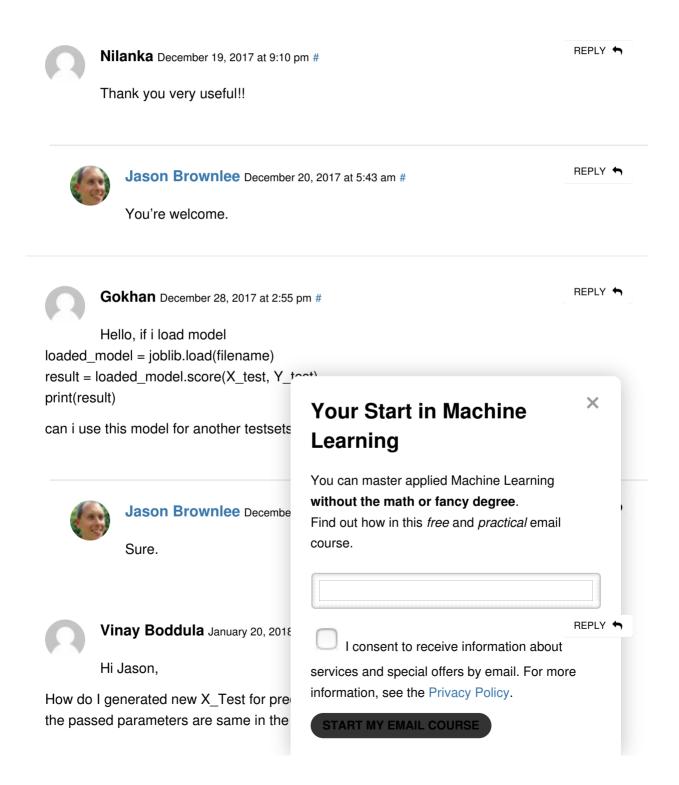
I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE

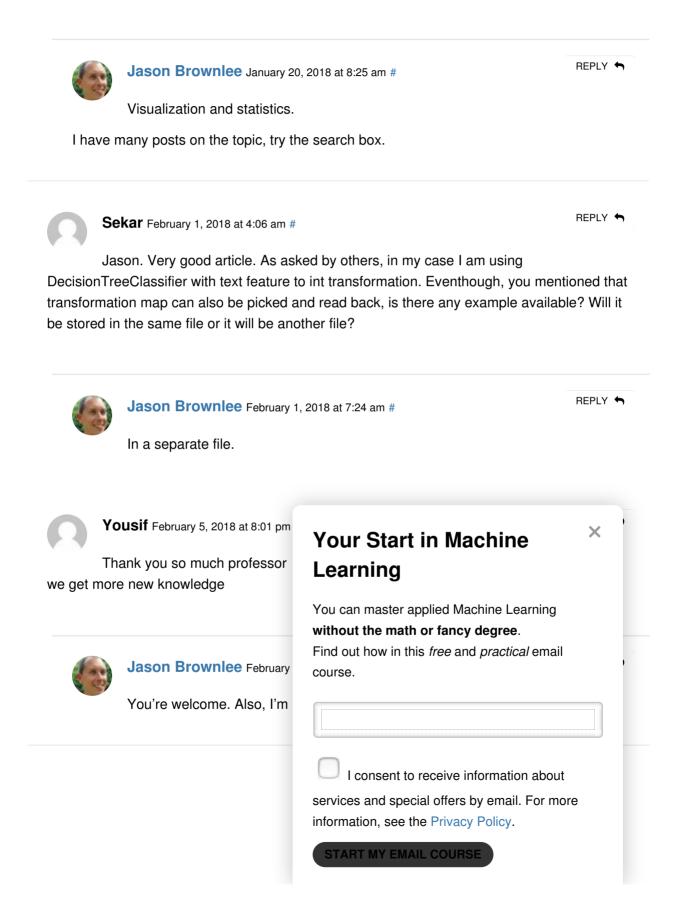
Hi Jason,

This is extremely helpful and saved me quite a bit of processing time.

I was training a Random Forest Classifier on a 250MB data which took 40 min to train everytime but results were accurate as required. The joblib method created a 4GB model file but the time was cut down to 7 Minutes to load. That was helpful but the results got inaccurate or atleast varied quite a bit from the original results. I use average of 2 Decision Tree and 1 Random Forest for the model. Decision Tree Models have kept there consistency loading vs training but RF hasn't. Any ideas?



Background: I am basically saving the model and predicting with new values from time to time. How do we check whether the new values have all the parameters and correct data type.





Adarsh C February 8, 2018 at 12:29 pm #

REPLY 👆

HI sir.

I would like to save predicted output as a CSV file. After doing ML variable I would like to save "y_predicted". And I'm using python ide 3.5.x I have pandas, sklearn, tensorflow libraries



Jason Brownlee February 9, 2018 at 8:58 am #

REPLY 🖴

You can save the numpy array as a csv.

https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.savetxt.html



Atul March 11, 2018 at 6:45 am #

REPLY 🖴

Hi Jason,

I would like to save predicted output as a CSV file. After doing ML variable I would like to save "y_predicted". How I can save Naive Bayes, SVM, RF and DT Classification for final predictions for all samples saved as a .csv with three columns namely Sample, Actual value, Prediction values

REPLY + Jason Brownlee March 12, 2019 at 6:04 am X **Your Start in Machine** Perhaps create a datafram dataframe directly via to csv(): Learning https://pandas.pydata.org/pandas-d You can master applied Machine Learning without the math or fancy degree. Find out how in this free and practical email Tommy March 22, 2018 at 11:14 pn course. I have a list of regression coeff coefficients into the sklearn logistic regre Thanks! Tommy I consent to receive information about services and special offers by email. For more information, see the Privacy Policy. REPLY + Jason Brownlee March 23 START MY EMAIL COURSE



No model is needed, use each coefficient to weight the inputs on the data, the weighted sum is the prediction.



Vincent April 10, 2018 at 10:25 am #

REPLY 🖴

Hi,all

I am using scikit 0.19.1

I generated a training model using random forest and saved the model. These were done on ubuntu 16.01 x86_64.

I copied the model to a windows 10 64 bit machine and wanted to reuse the saved model. But unfortunately i get the following

Traceback (most recent call last):

File "C:\Users\PC\Documents\Vincent\nicholas\feverwizard.py.py", line 19, in rfmodel=joblib.load(modelfile)

 $\label{limits} File \ ``C:\Python27\lib\site-packages\sklearn\externals\joblib\numpy_pickle.py", line 578, in load obj = _unpickle(fobj, filename, mmap_mode)$

 $\label{limits} File \ \ ``C:\Python27\lib\site-packages\sklearn\externals\joblib\numpy_pickle.py", line 508, in _unpickle$

obj = unpickler.load()

File "C:\Python27\lib\pickle.py", line 864, in load dispatchkey

File "C:\Python27\lib\pickle.py", line 1139, in load_reduce value = func(*args)

File "sklearn\tree tree.pyx", line 601, in sklearn.tree. tree.Tree.cinit

ValueError: Buffer dtype mismatch, expected 'SIZE_t' but got 'long long'

Jason Brownlee April 11, 2
Perhaps the pickle file is not

Pramod April 17, 2018 at 9:03 pm #

Can we load model trained on

Jason Brownlee April 18, 2

What could be happening? Is it because

Your Start in Machine Learning You can master applied Machine Learning without the math or fancy degree. Find out how in this free and practical email course. I consent to receive information about services and special offers by email. For more information, see the Privacy Policy. START MY EMAIL COURSE



I'm skeptical that it would work. Try it and see. Let me know how you go.



Arnaud April 17, 2018 at 9:29 pm #

REPLY 5

Dear Jason:

Thank you for 'le cours' which is very comprehensive.

I have a maybe tricky but 'could be very usefull' question about my newly created standard Python object.

Is it possible to integrate a call to my Python object in a Fortran program?

Basically I have a deterministic model in which I would like to make recursive calls to my Python object at every time step.

Do I need some specific libraries?

Thank you Best regards



Jason Brownlee April 18, 2018 at 8:06 am #

REPLY 5

X

You're welcome.

I suspect it is possible. It's all just code at the end of the day. You might need some kind of Python-FORTRAN bridge software. I have not done this, sorry.



Pratip April 23, 2018 at 4:32 pm #

Hi Sir,

I wanted to know if its possible to combi datasets to get more training data to get loaded dataset and then save model us Which method will be correct? Please help.

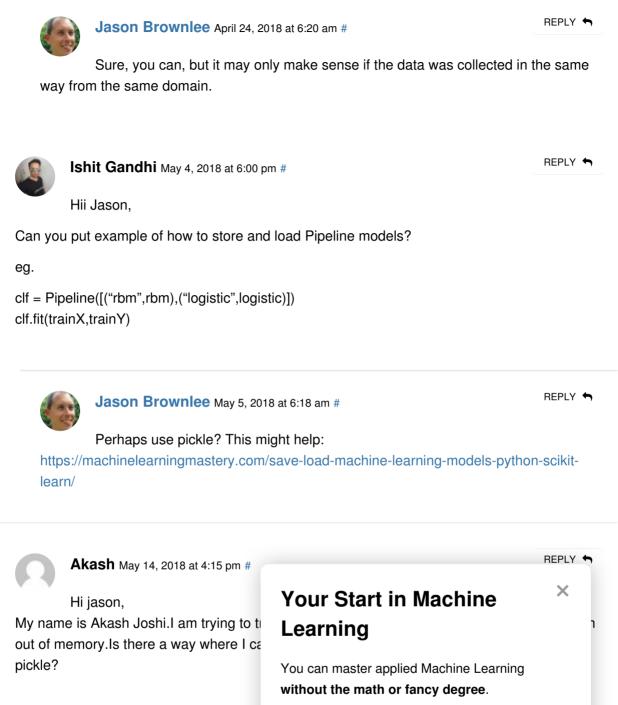
Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this free and practical email course.

I consent to receive information about
services and special offers by email. For more
nformation, see the Privacy Policy.

START MY EMAIL COURSE



Learning

You can master applied Machine Learning without the math or fancy degree.
Find out how in this free and practical email course.

Jason Brownlee May 15, 2

Perhaps try running on a m

Perhaps try using a sample of your
Perhaps use a generator to progres

Perhaps use a generator to progres

Samarth May 14, 2018 at 4:54 pm

Learning

You can master applied Machine Learning without the math or fancy degree.
Find out how in this free and practical email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE



Hi Jason

I want to know how can presist a minmax transjformation? There are ways to persist the final model but to persist the transformations?

Thanks



Jason Brownlee May 15, 2018 at 7:51 am #



Save the min and max values for each variable.

Or save the whole object.



SOORAJ T S May 16, 2018 at 12:30 am #



thank you the post, it is very informative but i have a doubt about the labels or names of the dataset can specify each.



Jason Brownlee May 16, 2018 at 6:05 am #



What do you mean exactly?



SOORAJ T S May 16, 2018 at 4:1

names = ['preg', 'plas', 'pres', 's

in the above code what are these "preg'



Jason Brownlee May 17, 2

You can learn about these https://github.com/jbrownlee/Datase

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE



SOORAJ T S May 17, 2018 at 4:23 pm #

REPLY 🦴

thank you sir...



Aniko June 7, 2018 at 12:13 am #

REPLY 🦴

HI Jason!

I created a machine learning (GBM) model to predict house prices and a Django application to usability. This model has more than 1000 n_estimators and it takes more than 1 minutes to load before getting the prediction in every request.

I would like to load joblib dump file just once and store the model in memory, avoiding loading the model in every get requests.

What is your best practice for this?

Thanks

Jason Brownlee June 7, 2018 at 6:31 am #

REPLY 🦴

This sounds like a web application software engineering question rather than a machine learning question.

Perhaps you can host the model behind a web service?



Aniko June 7, 2018 at 6:5

thank you, meanwhile documentation, this perhaps so



Jason Brownlee

Glad to hear it.



LamaOS223 June 9, 2018 at 2:00

okay what if i had 2 datasets fo

Your Start in Machine Learning

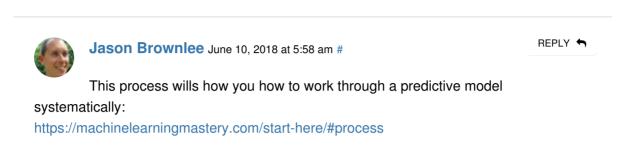
You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about			
services and special offers by email. For more			
information, see the Privacy Policy.	REPLY	+	

START MY EMAIL COURSE

the first dataset has a Loan_Status attribute and the second one does not have a Loan_Status attribute if i trained the model on the first dataset and i want to predict the Loan_Status for the second dataset, how to do that? please make it simple for me i'm beginner



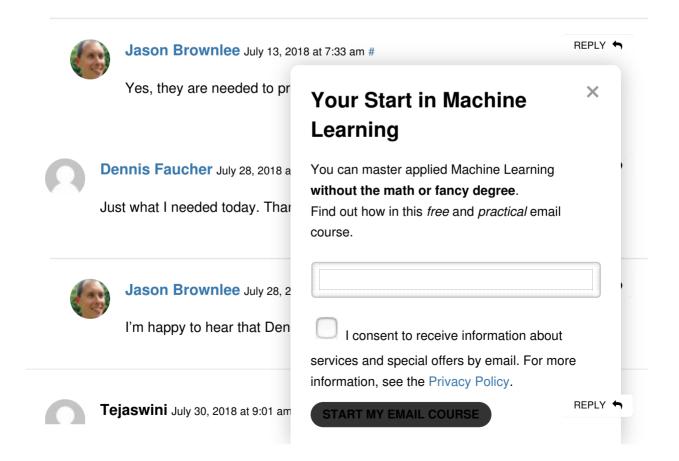


Imti July 12, 2018 at 4:55 pm #

REPLY 👆

Hey Jason, I am working on a model to classify text files. I am using the CountVectorizer, TfidfTransformer and SGDClassifier in the same sequence on a set of data files. I am saving the SGDClassifier object via the joblib.dump method you have mentioned in this article.

Do I also need to save the vectorizer and transformer objects/models? Since when i take a new file for classification I will need to go through these steps again.



Hi Jason,

Appreciate for the article. when i am saving the model and loading it in different page. Then it is showing different accuracy.

Problem trying to solve: I am using one classsvm model and detecting outliers in sentences.

Jason Brownlee July 30, 2018 at 2:15 pm #

REPLY 🦴

I have not seen that, are you sure you are evaluating the model on exactly the same data?



Tejaswini August 2, 2018 at 2:10 pm #

REPLY 🦴

Yes Jason i am using gensim word2vec to convert text into feature vectors and then performing classification task.after saving model and reloading in another session its giving different results.



Jason Brownlee August 2, 2018 at 2:11 pm #

REPLY 🦴

That is odd. I have not seen this.

Perhaps report a fault/bug?



EvapStudent August 7, 2018 at 1

Hi Jason.

I am training a neural network using ML different geometries of heat exchangers low MRE, but I can't figure out how to us call again, I am getting an error when us about bringing in new data for the netwo

Jason Brownlee August 7,

I don't recommend using p your model.

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

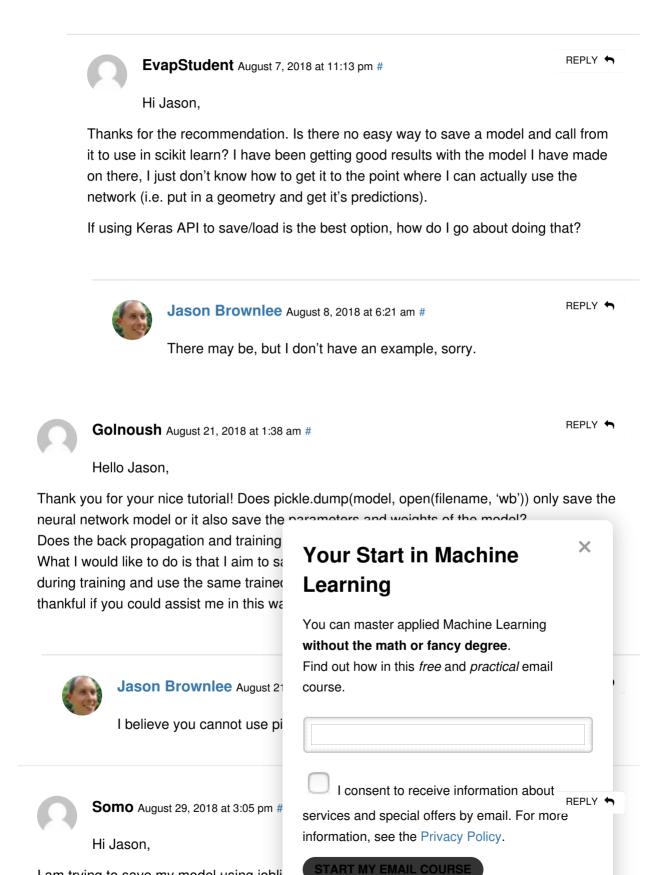
Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more

information, see the Privacy Policy.

START MY EMAIL COURSE

Once you find a config that works for your problem, perhaps switch from the sklearn wrappers to the Keras API directly.



36 of 44 1/27/19, 12:58 AM

I am trying to save my model using jobli

another .py file model = joblib.load('model.pkl') but then the accuracy dropped and each time I run it the accuracy differs a lot. I coefficient and the intercept and the same for both models. Any ideas why this might happen. Thanks in advance.



Jason Brownlee August 30, 2018 at 6:26 am #

REPLY 🖴

Perhaps this will help:

https://machinelearningmastery.com/faq/single-faq/why-do-i-get-different-results-each-time-i-run-the-code

Dhrumil September 1, 2018 at 3:11 pm #

REPLY 🖛

Hey man I am facing a trouble with pickle, when I try to load my .pkl model I am getting following error :

UnicodeDecodeError: 'ascii' codec can't decode byte 0xbe in position 3: ordinal not in range(128)

Can you please tell me something since I have tried all fixes I could find..



Jason Brownlee September 2, 2018 at 5:30 am #

REPLY 🖴

X

Perhaps post your error on stackoverflow?

Ω

Aakash Aggarwal September 8

I want to develop to train my m onwards, when i want to train the model append mode that reduces the time of timodel.

Any helps would be greatly appreciated

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more information, see the Privacy Policy.

START MY EMAIL COURSE



Jason Brownlee September 8, 2018 at 6:17 am #

REPLY 👆

This post shows how:

https://machinelearningmastery.com/save-load-machine-learning-models-python-scikit-learn/



Aakash Aggarwal October 2, 2018 at 12:45 am

REPLY 🦴

This article shows how to save a model that is built from scratch. But I am looking to train the model by including additional data so as to achieve high prediction performance and accuracy for unseen data. Is there any leads or approach you can think?



Jason Brownlee October 2, 2018 at 6:26 am #

REPLY 🖴

I don't understand, sorry. Training a model and saving it are separate tasks.



My3 October 15, 2018 at 10:11 pm #

REPLY 🖴

X

Hi Jason,

I have some requirement to integrate py

I have a ML model which is trained as s to load this one time using java and ther python. So my workflow is like:

- 1. Read Randomforestclassifier.pkl file (
- Send this model as input to function of for each request
- python_file.py has prediction code an code

Please provide suggestions for this worl to execute python_file.py and everything activity.

Can you help me with some client serve time model loading?

Your Start in Machine Learning

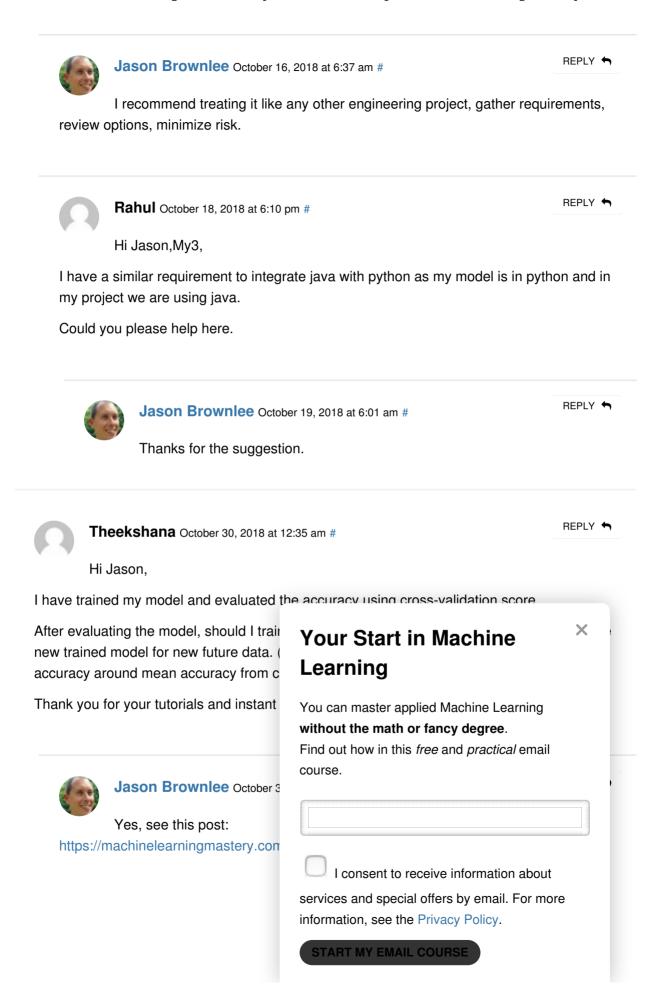
You can master applied Machine Learning without the math or fancy degree.

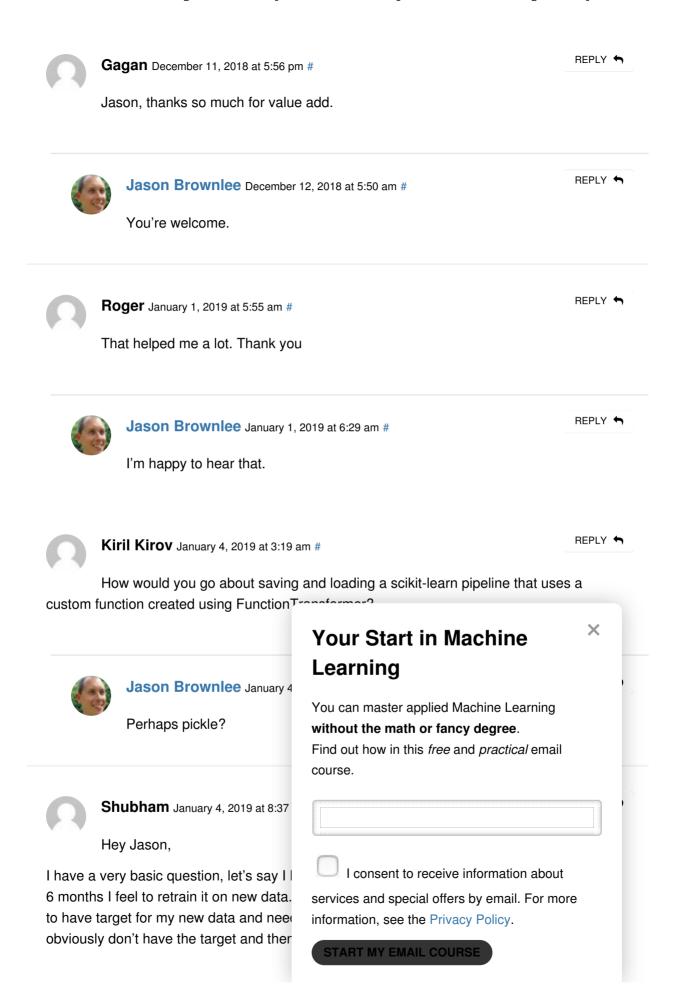
Find out how in this *free* and *practical* email course.

I consent to receive information about services and special offers by email. For more

START MY EMAIL COURSE

information, see the Privacy Policy.

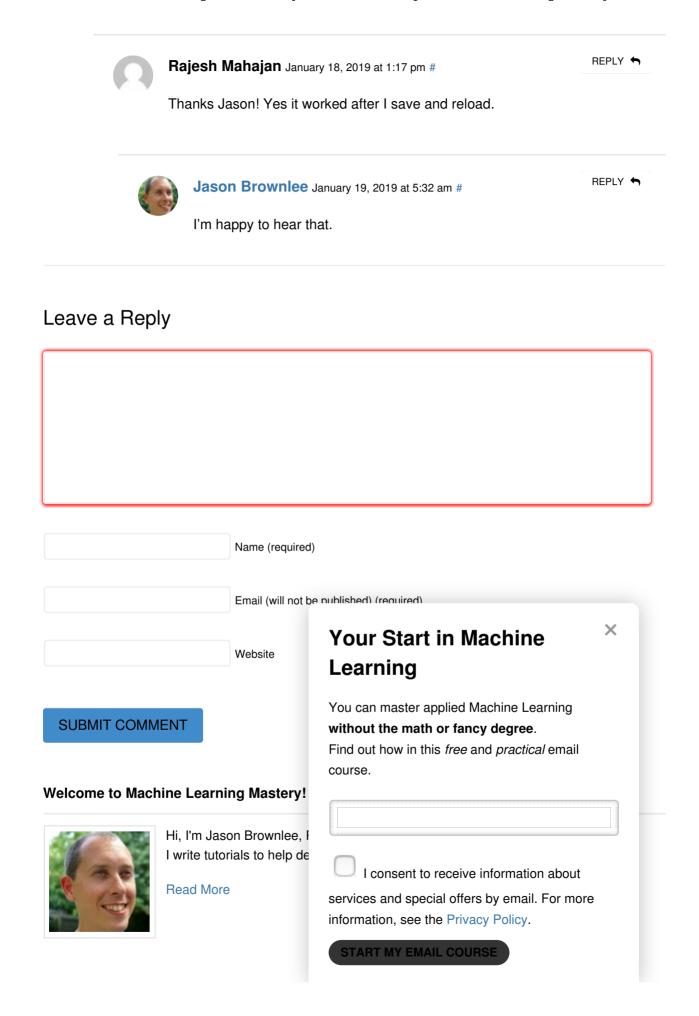




Jason Brownlee January 4, 2019 at 11:01 am #

REPLY 5

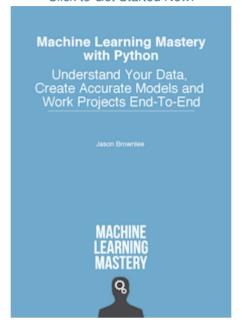
You have many options, e.g. develop a new model, update the old model, some mixture of the two with an ensemble. REPLY Rajesh Mahajan January 18, 2019 at 7:44 am # Hi Jason. I am new to this.. So pardon, if I am asking something incorrect... I have two stages. Build model and predict. For Build model: I am using vectorizer.fit transform(data) and building the logistic model. My data is a bunch of comments and the target is a set of categories. In order for me to use that model for predicting categories for new comments, I am using the vector created earlier during building of model to predict So, when I do the save model joblib.dump(log model, "model.sav") Foe Predict: When I try to re-run the model (saved) at a later point of time, I don't have the original vectorizer anymore with the original data set log_model = joblib.load("model.sav") inputfeatures nd = vectorizer.transform(newComment); pred = log_model.predict(inputfeatures_nd) I get this error - sklearn.exceptions.Not X Your Start in Machine fitted. Learning What do you suggest I should do? Sho You can master applied Machine Learning without the math or fancy degree. Find out how in this free and practical email Jason Brownlee January 1 course. You must use the same ve it along with your model. I consent to receive information about services and special offers by email. For more information, see the Privacy Policy. START MY EMAIL COURSE

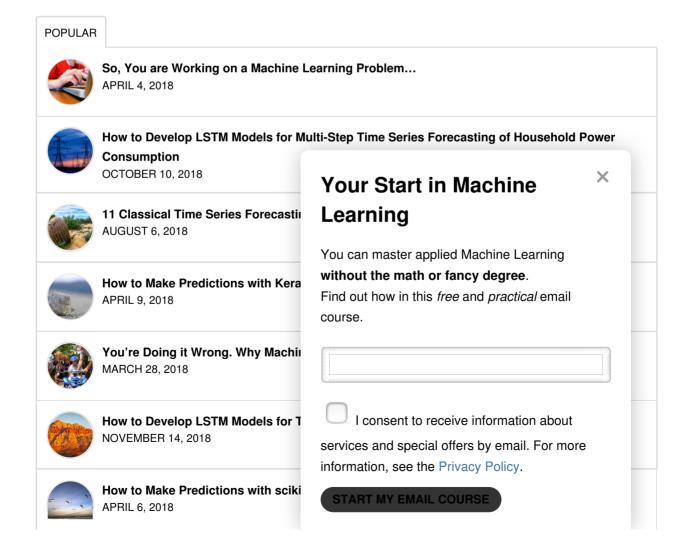


Machine Learning with Python

Develop predictive models with scikit-learn.

Click to Get Started Now!







You might also like...

- How to Install Python for Machine Learning
- Your First Machine Learning Project in Python
- Your First Neural Network in Python
- Your First Classifier in Weka
- Your First Time Series Forecasting Project

© 2019 Machine Learning Mastery. All Rights Reserved.

Privacy | Disclaimer | Terms | Contact

Your Start in Machine Learning You can master applied Machine Learning without the math or fancy degree. Find out how in this free and practical email course. I consent to receive information about services and special offers by email. For more information, see the Privacy Policy. START MY EMAIL COURSE