

Ensemble/Voting Classification in Python with Scikit-Learn

(<https://stackabuse.com/triplebyte/>)

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

By Dan Nelson

0 Comments (/ensemble-voting-classification-in-python-with-scikit-learn/#disqus_t)

- Coinbase, Plaid, Zoox, and Heap are still hiring!



(<https://stackabuse.com/triplebyte/>)

Introduction



classification models (<https://scikit-learn.org/stable/modules/ensemble.html>) can be powerful machine learning tools

achieving excellent performance and generalizing well to new, unseen data.



(<https://stackabuse.com/triplebyte/>)

The value of an ensemble classifier is that, in joining together the predictions of multiple classifiers, it can correct for errors made by any individual classifier,

leading to better accuracy overall. Let's take a look at the different ensemble

classification methods and see how these classifiers can be implemented in Scikit-Learn.

Prepping for an interview?

Improve your skills by solving one coding problem every day

(<https://stackabuse.com/daily-coding-problem/>)

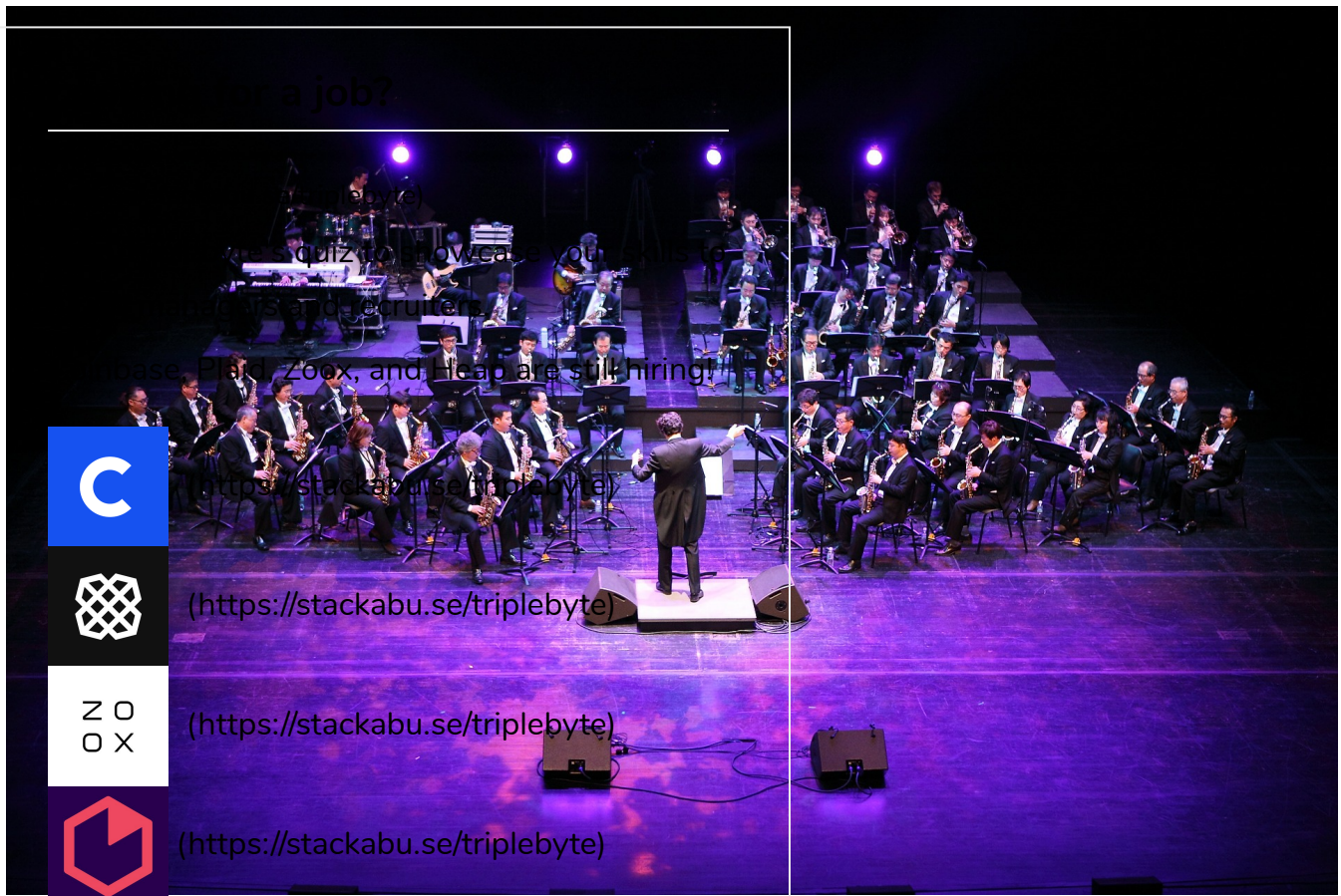
What are Ensemble Models in Machine Learning?

Searches starting the next morning via email

- Practice on **actual problems** asked by top companies, like:

Google facebook amazon.com Microsoft

</> Daily Coding Problem (<https://stackabuse.com/daily-coding-problem/>)



Credit: Pixabay (<https://pixabay.com/photos/music-played-saxophone-big-band-619256/>)

Ensemble models are an ensemble learning method that combines different

algorithms together. In this sense, it is a meta-algorithm rather than an algorithm

itself. Ensemble learning methods are valuable because they can improve the

performance of a predictive model.

- Improve your skills by solving one coding problem every day

Ensemble learning methods work off of the idea that tying the predictions of

multiple classifiers together will lead to better performance by either improving

prediction accuracy or reducing aspects like bias and variance.

Google facebook amazon.com Microsoft

In general, an ensemble model falls into one of two categories:

sequential approaches and parallel approaches (<https://medium.com/ml-research-lab/ensemble-learning-the-heart-of-machine-learning-b4f59a5f9777>).

A sequential ensemble model operates by having the base learners/models generated in sequence. Sequential ensemble methods are typically used to try and increase overall performance, as the ensemble model can compensate for

inaccurate predictions by re-weighting the examples that were previously

misclassified. A notable example of this is AdaBoost.

(<https://stackabu.se/triplebyte>)

A **parallel model** is, as you may be able to guess, methods that rely on creating and

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

training the base learners in parallel. Parallel methods aim to reduce the error rate

- Coinbase, Plaid, Zoox, and Heap are still hiring!

by training many models in parallel and averaging the results together. A notable

example of a parallel method is the Random Forest Classifier (</random-forest-algo-with-python-and-scikit-learn/>).



(<https://stackabu.se/triplebyte>)

Another way of thinking about this is a distinction between homogenous and



(<https://stackabu.se/triplebyte>)

heterogeneous learners. While most of the ensemble learning methods use

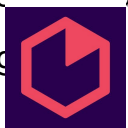
homogenous base learners (many of the same type of learners), some ensemble



(<https://stackabu.se/triplebyte>)

methods use heterogeneous learners (different learning algorithms joined

together).



(<https://stackabu.se/triplebyte>)

To recap, Triplebyte's Quiz (<https://stackabu.se/triplebyte>)

- Sequential models try to increase performance by re-weighting examples, and

Prepping for an interview?

models are generated in sequence.

- Parallel models work by averaging results together after training many models

(<https://stackabu.se/daily-coding-problem>)

at the same time.

- Improve your skills by solving one coding

problem every day.

We'll now cover different methods of employing these models to solve machine

learning classification problems. Get the solutions the next morning via email

- Practice on **actual problems** asked by top

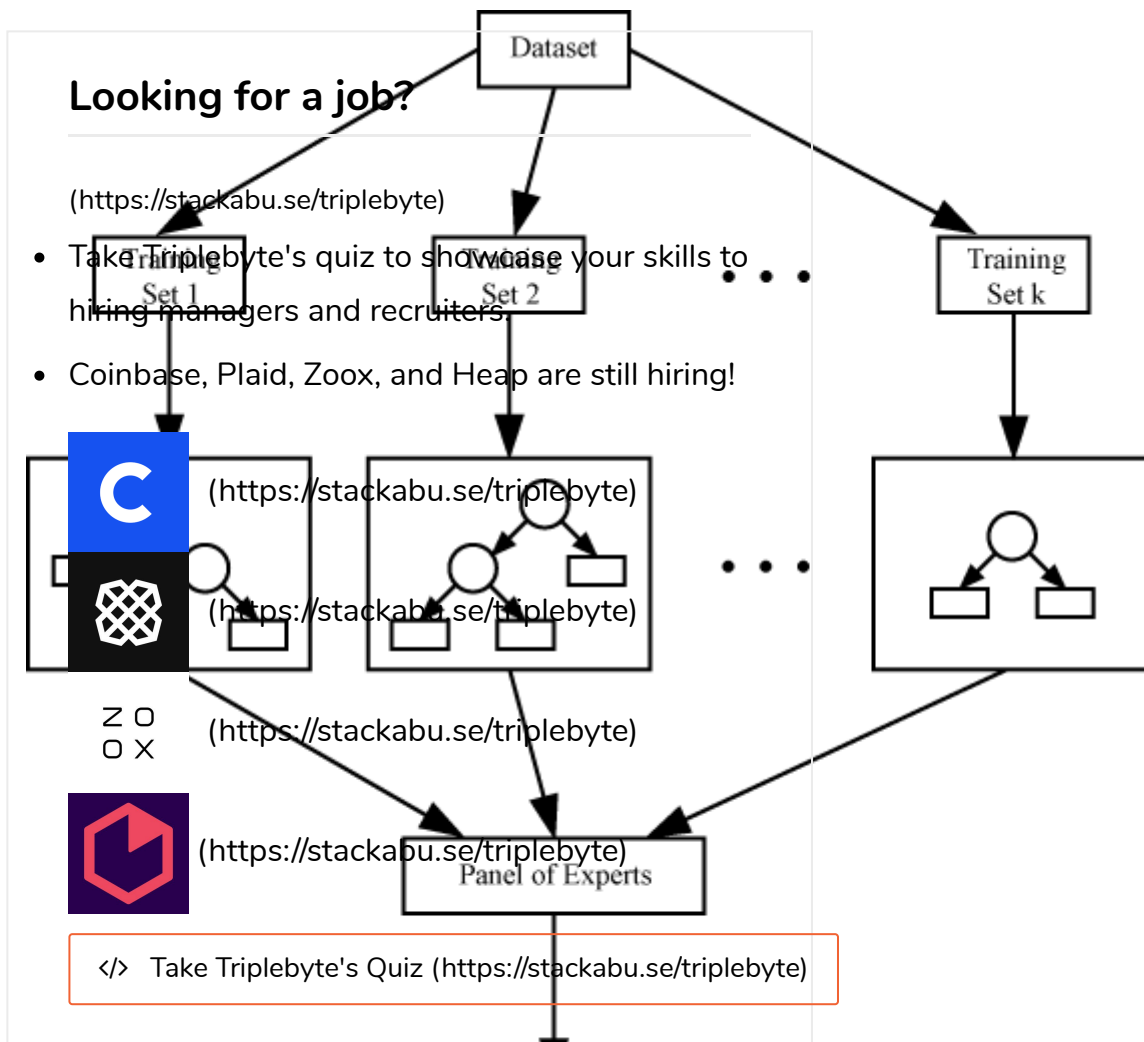
companies, like:

Different Ensemble Classification Methods



</> Daily Coding Problem (<https://stackabu.se/daily-coding-problem>)

Bagging



Prepping for an interview?

(<https://stackabu.se/daily-coding-problem>)

Credit: Wikimedia Commons (https://commons.wikimedia.org/wiki/File:DTE_Bagging.png)

- Improve your skills by solving one coding problem every day

Bagging, also known as bootstrap aggregating

- Get the solutions the next morning via email (https://en.wikipedia.org/wiki/Bootstrap_aggregating), is a classification method
- Practice on **actual problems** asked by top companies, like:

together. Bagging creates subsets from the main dataset that the learners are trained on.

Google facebook amazon.com Microsoft

</> Daily Coding Problem (<https://stackabu.se/daily-coding-problem>)

In order for the predictions of the different classifiers to be aggregated, either an averaging is used for regression, or a voting approach is used for classification (based on the decision of the majority).

One example of a bagging classification method is the *Random Forests Classifier*.

Looking for a job?

In the case of the random forests classifier, all the individual trees are trained on a different sample of the dataset.

(<https://stackabu.se/triplebyte>)

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

The tree is also trained using random selections of features. When the results are averaged together, the overall variance decreases and the model performs better as a result.



(<https://stackabu.se/triplebyte>)

Boosting



(<https://stackabu.se/triplebyte>)

Boosting algorithms (<https://www.analyticsvidhya.com/blog/2015/11/quick-introduction-to-boosting-algorithms-machine-learning/>) are capable of taking weak,

uninformative models and converting them into strong models. The idea behind

boosting algorithms is that you assign many weak learning models to the datasets,

and the weights for misclassified examples are tweaked during subsequent

rounds of learning.



(<https://stackabu.se/triplebyte>)

Take Triplebyte's Quiz (<https://stackabu.se/triplebyte>)

The predictions of the classifiers are aggregated and then the final predictions are

made through a weighted sum (in the case of regressions), or a weighted majority

vote (in the case of classification).

(<https://stackabu.se/daily-coding-problem>)

- Improve your skills by solving one coding

AdaBoost is one example of a boosting classifier method, as is Gradient Boosting,

which was derived from the aforementioned algorithm.

- Get the solutions the next morning via email

- Practice on **actual problems** asked by top

If you'd like to read more about Gradient Boosting (/gradient-boosting-classifiers-

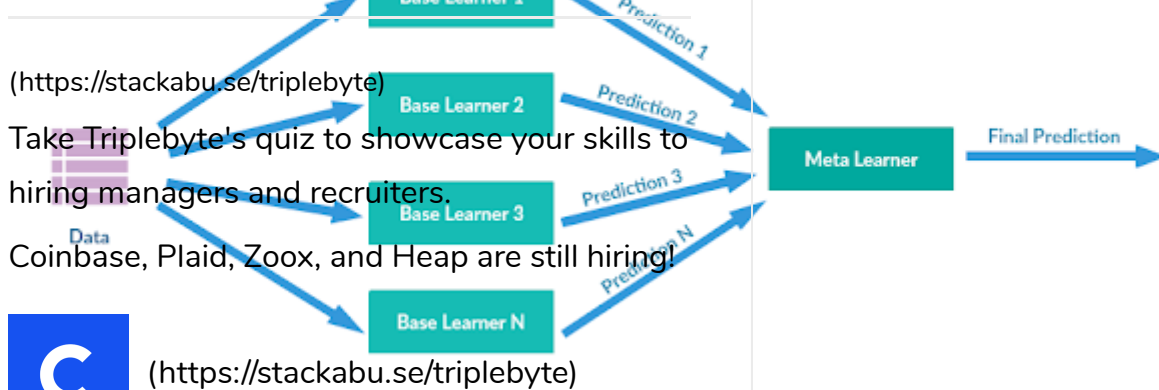
in-python-with-scikit-learn/) and the theory behind it, we've already covered that

in a previous article.    

</> Daily Coding Problem (<https://stackabu.se/daily-coding-problem>)

Stacking

Looking for a job?



(<https://stackabuse.com/triplebyte/>)

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

- Coinbase, Plaid, Zoex, and Heap are still hiring!



(<https://stackabuse.com/triplebyte/>)

Created by <https://stackabuse.com/triplebyte/> (https://commons.wikimedia.org/wiki/File:Stacking.png)

Stack Overflow (<https://stackoverflow.com/>)

(<https://blogs.sas.com/content/subconsciousmusings/2017/05/18/stacked-ensemble-models-win-data-science-competitions/>)

Stacked ensemble models (https://stackabuse.com/triplebyte/) are an ensemble learning method that combines the decision of different regression or classification

algorithms. The component models are trained on the entire training dataset. After these component models are trained, a meta-model is assembled from the different

models and then it's trained on the outputs of the component models. This

Prepping for an interview?

approach typically creates a heterogeneous ensemble because the component

models are usually different algorithms.

- Improve your skills by solving one coding problem every day

Example Implementations

Now that we've explored different methods we can use to create ensemble models, let's take a look at how we could implement a classifier using the different

methods. [google](https://www.google.com/) [facebook](https://www.facebook.com/) [amazon.com](https://www.amazon.com/) [Microsoft](https://www.microsoft.com/)

</> Daily Coding Problem (<https://stackabuse.com/daily-coding-problem/>)

Though, before we can take a look at different ways of implementing ensemble classifiers, we need to select a dataset to use and do some preprocessing of the dataset.

We'll be using the Titanic dataset, which can be downloaded here

(<https://www.kaggle.com/c/titanic>). Let's do some preprocessing of the data in

order to get rid of missing values and scale the data to a uniform range. Then we

(<https://stackabu.se/triplebyte>)


can go about setting up the ensemble classifiers.

- Take Triplebyte's quiz to showcase your skills to

hiring managers and recruiters.

Coinbase, Plaid, Zooy, and Heap are still hiring!

Data Preprocessing

To  with, we'll start by importing all functions we need from their respective

(<https://stackabu.se/triplebyte>)

libraries. We'll be using Pandas (<https://pandas.pydata.org/>) and Numpy

(<https://numpy.org/>) to load and transform the data, as well as the LabelEncoder

(<https://stackabu.se/triplebyte>)

and StandardScaler tools.

Z O
O X

(<https://stackabu.se/triplebyte>)

We'll need the machine learning metrics and the `train_test_split` function.

First, we need the classifiers we want to use:



```
import warnings
import numpy as np
```

```
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.metrics import accuracy_score, f1_score, log_loss
from sklearn.model_selection import train_test_split, KFold, cross_val_score

from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import AdaBoostClassifier, RandomForestClassifier, ExtraTreesClassifier
```

companies, like:

We'll start by loading in the training and testing data and then creating a function

to check for the presence of any null values:

```
</> Daily Coding Problem (https://stackabu.se/daily-coding-problem)
```

```
training_data = pd.read_csv("train.csv")
testing_data = pd.read_csv("test.csv")
```

Looking for a job?


```
def get_nulls(training, testing):
```

```
    print("Training Data:")
    (https://stackabu.se/triplebyte)
    print(pd.isnull(training).sum())
```

- Take Triplebyte's Quiz to Showcase your skills to hiring managers and recruiters.

```
    print(pd.isnull(testing).sum())
```

- Coinbase, Plaid, Zook, and Heap are still hiring!

As  ens, there are a lot of missing values in the Age and Cabin categories.

```
Training Data:
PassengerId (https://stackabu.se/triplebyte)
```

```
Survived 0
```

```
Pclass 0
```

```
Name Z O (https://stackabu.se/triplebyte)
```

```
Sex O X 0
```

```
Age 177
```

```
SibSp (https://stackabu.se/triplebyte)
```

```
Par 0
```

```
Ticket 0
```

```
Fare 0
Cabin 687
Embarked 2
```

```
dtype: int64
```

```
Testing Data:
```

```
PassengerId 0
```

```
Pclass 0
```

```
Name 0
```

```
Sex (https://stackabu.se/daily-coding-problem)
```

```
Age 86
```

```
SibSp 0
```

```
Par 0
```

```
Ticket 0
```

```
Fare 1
```

```
Cabin 327
```

```
Embarked 0
```

```
dtype: int64
```

companies, like:

We're going to start by dropping some of the columns that will likely be useless - the Cabin column and the Ticket column. The Cabin column has far too many missing values and the Ticket column is simply comprised of too many categories to be useful.

After that we will need to impute some missing values. When we do so, we must account for how the dataset is slightly right skewed (young ages are slightly more prominent than older ages). We'll use the median values when we impute the data[^]

because due to large outliers taking the average values would give us imputed values that are far from the center of the dataset:

(<https://stackabu.se/triplebyte>)

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters

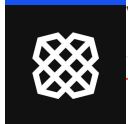
Subscribe to our Newsletter

- Coinbase, Plaid, Zoex, and Heap are still hiring!
Get occasional tutorials, guides, and jobs in your inbox. No spam ever.



Unsubscribe at any time.

(<https://stackabu.se/triplebyte>)



your email...

(<https://stackabu.se/triplebyte>)

Subscribe

Z O
O X

(<https://stackabu.se/triplebyte>)



(<https://stackabu.se/triplebyte>)

```
# Drop the cabin column, as there are too many missing values
# Drop the ticket numbers too, as there are too many categories
# Drop Take Triplebyte's Quiz (https://stackabu.se/triplebyte) to help predict survivors
```

```
training_data.drop(labels=['Cabin', 'Ticket', 'Name'], axis=1, inplace=True)
testing_data.drop(labels=['Cabin', 'Ticket', 'Name'], axis=1, inplace=True)
```

Prepping for an interview?

```
# Taking the mean/average value would be impacted by the skew
# so we should use the median value to impute missing values
```

```
(https://stackabu.se/daily-coding-problem)
training_data["Age"].fillna(training_data["Age"].median(), inplace=True)
training_data["Age"].fillna(training_data["Age"].median(), inplace=True)
training_data["Embarked"].fillna("S", inplace=True)
testing_data["Fare"].fillna(testing_data["Fare"].median(), inplace=True)
```

- Get the solutions the next morning via email
get_nulls(training_data, testing_data)
- Practice on **actual problems** asked by top

companies, like:

Now we can see there's no more missing values:

Google facebook amazon.com Microsoft

</> Daily Coding Problem (<https://stackabu.se/daily-coding-problem>)

Training Data:

PassengerId 0
Survived 0

Pclass 0

Name 0
(https://stackabu.se/triplebyte)

Sex 0

Age 0
Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

SibSp 0

Parch 0
Coinbase, Plaid, Zoex, and Heap are still hiring!

Fare 0

Embarked 0

dt 64

Te 0

Pa 0

PC 0

Na 0

Se 0

Ag 0

Si 0

Pa 0

Fa 0

Em 0

dt 64

Te 0

Pa 0

PC 0

Na 0

Se 0

Ag 0

Si 0

Pa 0

Fa 0

Em 0

dt 64

Te 0

Pa 0

PC 0

Na 0

Se 0

Ag 0

Si 0

Pa 0

Fa 0

Em 0

dt 64

Te 0

Pa 0

PC 0

Na 0

Se 0

Ag 0

Si 0

Pa 0

Fa 0

Em 0

dt 64

Te 0

Pa 0

PC 0

Na 0

Se 0

Ag 0

Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.

Coinbase, Plaid, Zoex, and Heap are still hiring!



We're now going to need to encode the non-numerical data. Let's set up a

</> Take Triplebyte's Quiz (https://stackabu.se/triplebyte)

LabelEncoder and fit it on the Sex feature and then transform the data with the

encoder. We'll then replace the values in the Sex feature with those that have

been encoded and then do the same for the Embarked feature.

Prepping for an interview?

Finally, let's scale the data using the StandardScaler, so there aren't huge

fluctuations in values.

- Improve your skills by solving one coding problem every day
- Get the solutions the next morning via email
- Practice on **actual problems** asked by top companies, like:

Google facebook amazon.com Microsoft

</> Daily Coding Problem (https://stackabu.se/daily-coding-problem)

```

encoder_1 = LabelEncoder()
# Fit the encoder on the data
encoder_1.fit(training_data["Sex"])

# Transform and replace training data
training_sex_encoded = encoder_1.transform(training_data["Sex"])
test_sex_encoded = encoder_1.transform(testing_data["Sex"])
testing_data["Sex"] = test_sex_encoded

# Coinbase, Plaid, Zook, and Heap are still hiring!
encoder_2 = LabelEncoder()
encoder_2.fit(training_data["Embarked"])
training_embarked_encoded = encoder_2.transform(training_data["Embarked"])
training_data["Embarked"] = training_embarked_encoded
testing_embarked_encoded = encoder_2.transform(testing_data["Embarked"])
testing_data["Embarked"] = testing_embarked_encoded

# Since we want to reshape needs be turned into array first
ages_train = np.array(training_data["Age"]).reshape(-1, 1)
fares_train = np.array(training_data["Fare"]).reshape(-1, 1)
ages_test = np.array(testing_data["Age"]).reshape(-1, 1)
fares_test = np.array(testing_data["Fare"]).reshape(-1, 1)

# Takes arrays
scaler = StandardScaler()
training_data["Age"] = scaler.fit_transform(ages_train)
training_data["Fare"] = scaler.fit_transform(fares_train)
testing_data["Age"] = scaler.fit_transform(ages_test)
testing_data["Fare"] = scaler.fit_transform(fares_test)

```

Prepping for an interview?

Now that our data has been preprocessed, we can select our features and labels and then use the `train_test_split` function to divide our entire training data up into training and testing sets:

- Get the solutions the next morning via email

Practice on actual problems asked by top companies like:

Google, Facebook, Amazon, Microsoft

```

# Make the train/test data from validation
X_train, X_val, y_train, y_val = train_test_split(X_features, y_labels, test_size=0.1, random_state=27)

```

We're now ready to start implementing ensemble classification methods.

Simple Averaging Approach

Before we get into the big three ensemble methods we covered earlier, let's cover a very quick and easy method of using an ensemble approach -

Looking for a job?

averaging predictions

(<https://stackabu.se/triplebyte>)

([https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-](https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/)

• Take Triplebyte's quiz to showcase your skills to ensemble-models/). We simply add the different predicted values of our chosen hiring managers and recruiters.

classifiers together and then divide by the total number of classifiers, using floor division to get a whole value.



(<https://stackabu.se/triplebyte>)

In this case we'll be using logistic regression, a Decision Tree Classifier

(/decision-trees-in-python-with-scikit-learn/), and the Support Vector Classifier



(<https://stackabu.se/triplebyte>)

(/implementing-svm-and-kernel-svm-with-pythons-scikit-learn/). We fit the

classifier on the data and then save the predictions as variables. Then we simply add



(<https://stackabu.se/triplebyte>)

predictions together and divide:



(<https://stackabu.se/triplebyte>)

```
LogReg_clf = LogisticRegression()
DTree_clf = DecisionTreeClassifier()
SVC_clf = SVC()
```

```
LogReg_clf.fit(X_train, y_train)
```

```
DTree_clf.fit(X_train, y_train)
```

```
SVC_clf.fit(X_train, y_train)
```

Prepping for an interview?

```
LogReg_pred = LogReg_clf.predict(X_val)
```

```
DTree_pred = DTree_clf.predict(X_val)
```

```
SVC_pred = SVC_clf.predict(X_val)
```

• Improve your skills by solving one coding

problem every day

```
averaged_preds = (LogReg_pred + DTree_pred + SVC_pred) // 3
```

```
acc = accuracy_score(y_val, averaged_preds)
```

• Get the solutions the next morning via email

• Practice on **actual problems** asked by top

Here's the accuracy we got from this method:

companies, like:

0.8444444444444444



</> Daily Coding Problem (<https://stackabu.se/daily-coding-problem>)

Voting\Stacking Classification Example

When it comes to creating a stacking/voting classifier, Scikit-Learn provides us with some handy functions that we can use to accomplish this.

The VotingClassifier ([https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html)

Looking for a job? [learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html)) takes in a list of different estimators as arguments and a voting method. The hard voting method uses the predicted labels and a majority rules system, while the soft

• Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.
voting method predicts a label based on the argmax/largest predicted value of the

sum of the predicted probabilities.
• Coinbase, Paid, Zox, and Heap are still hiring!

After we provide the desired classifiers, we need to fit the resulting ensemble classifier object. We can then get predictions and use accuracy metrics:

```
from sklearn.ensemble import VotingClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, log_loss, f1_score

# Create individual classifiers
svm = SVC()
tree = DecisionTreeClassifier()
logit = LogisticRegression()

# Fit the classifiers on training data
svm.fit(X_train, y_train)
tree.fit(X_train, y_train)
logit.fit(X_train, y_train)

# Create the VotingClassifier
voting_clf = VotingClassifier(estimators=[('SVC', svm), ('DTree', tree), ('Log', logit)], voting='hard')

# Predict on validation data
preds = voting_clf.predict(X_val)

# Calculate accuracy, log loss, and F1 score
acc = accuracy_score(y_val, preds)
l_loss = log_loss(y_val, preds)
f1 = f1_score(y_val, preds)
```

print("Accuracy is: " + str(acc))
</> Take Triplebyte's Quiz (<https://stackabu.se/triplebyte>)
print("Log Loss is: " + str(l_loss))
print("F1 Score is: " + str(f1))

Here's what the metrics have to say about the VotingClassifier's performance:

```
Accuracy is: 0.8888888888888888
Log Loss is: 0.3376684749044165
F1 Score is: 0.8484848484848486
```

problem every day

- Get the solutions the next morning via email

Bagging Classification Example

Practice on actual problems asked by top companies, like:

Here's how we can implement bagging classification with Scikit-Learn. Sklearn's

BaggingClassifier ([https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html)

[learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html)) takes in a chosen classification model as well as the number of estimators that you want to use - you can use a model like Logistic Regression or Decision Trees.

Sklearn also provides access to the RandomForestClassifier and the ExtraTreesClassifier, which are modifications of the decision tree classification. These classifiers can also be used alongside the K-folds cross-

validation tool.

Looking for a job?

We'll compare several different bagging classification approaches here, printing out

(<https://stackabuse.com/triplebyte/>)

the mean results of the K-fold cross validation score:

- Take Triplebyte's quiz to showcase your skills to

hiring managers and recruiters.

logreg_bagging_model = BaggingClassifier(base_estimator=LogReg_clf, n_estimators=50,
• Coinbase, Plaid, Zoox, and Heap are still hiring!

dtree_bagging_model = BaggingClassifier(base_estimator=DTree_clf, n_estimators=50,
n_estimators=12)

random_forest = RandomForestClassifier(n_estimators=100, random_state=12)

extra_trees = ExtraTreesClassifier(n_estimators=100, random_state=12)

def cross_val_score(model):

(<https://stackabuse.com/triplebyte/>)

cv = KFold(n_splits=20, random_state=12)

score = cross_val_score(model, X_train, y_train, cv=cv)

mean_score = results.mean()

(<https://stackabuse.com/triplebyte/>)

logreg_bagging_model = BaggingClassifier(base_estimator=LogReg_clf, n_estimators=50,

dtree_bagging_model = BaggingClassifier(base_estimator=DTree_clf, n_estimators=50,

random_forest = RandomForestClassifier(n_estimators=100, random_state=12)

extra_trees = ExtraTreesClassifier(n_estimators=100, random_state=12)

</> Take Triplebyte's Quiz (<https://stackabuse.com/triplebyte/>)

Here's the results we got from the classifiers:

0.7865853658536585

0.8162439024390244

0.8002439024390245

0.7902439024390244

(<https://stackabuse.com/daily-coding-problem/>)

- Improve your skills by solving one coding

problem every day

Boosting Classification Example

- Get the solutions the next morning via email

Finally, we'll take a look at how to use a boosting classification method. As

- Practice on actual problems asked by top

mentioned, there's a separate article on the topic of Gradient Boosting you can

companies, like:

read here (</gradient-boosting-classifiers-in-python-with-scikit-learn/>).

Google facebook amazon.com Microsoft

Scikit-Learn has a built-in AdaBoost classifier (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html>),

which takes in a given number of estimators as the first argument. We can try

using a for loop to see how the classification performance changes at different

values, and we can also combine it with the K-Folds cross validation tool:

```
k_folds = KFold(n_splits=20, random_state=12)
```

Looking for a job?

```
num_estimators = [20, 40, 60, 80, 100]
```

```
for i in num_estimators:
```

```
(https://stackabu.se/triplebyte)
```

```
ada_boost = AdaBoostClassifier(n_estimators=i, random_state=12)
```

- Takes Triplebyte's quiz to show case your skills to hiring managers and recruiters.

```
print("Results for {} estimators:".format(i))
```

```
print(results.mean())
```

- Coinbase, Plaid, Zoex, and Heap are still hiring!

Here's the results we got:



(https://stackabu.se/triplebyte)

```
Results for 20 estimators:
```

```
0.902439024
```

```
(https://stackabu.se/triplebyte)
```

```
Results for 40 estimators:
```

```
0.902439025
```

```
Results for 60 estimators:
```

```
0.902439024 (https://stackabu.se/triplebyte)
```

```
Results for 80 estimators:
```

```
0.902439024
```

```
Results for 100 estimators:
```

```
0.902439024 (https://stackabu.se/triplebyte)
```



</> Take Triplebyte's Quiz (https://stackabu.se/triplebyte)

Summing Up

Prepping for an interview?

We've seen three different ensemble classification techniques:

voting, stacking, bagging, and boosting.

(https://stackabu.se/daily-coding-problem)

- Improve your skills by solving one coding problem every day
- Scikit-Learn allows you to easily create instances of the different ensemble classifiers. These ensemble objects can be combined with other Scikit-Learn tools like K-Folds cross validation.
- Get the solutions the next morning via email
- Practice on **actual problems** asked by top

companies, like:

If you'd like to learn more about appropriate uses for ensemble classifiers, and the

theories behind them, I suggest checking out the links found here



(https://blog.statshot.co/ensemble-learning-d1dcd548e936) or here

</> Daily Coding Problem (https://stackabu.se/daily-coding-problem)

(https://analyticsindiamag.com/primer-ensemble-learning-bagging-boosting/).!

python (/tag/python/), scikit-learn (/tag/scikit-learn/), data science (/tag/data-science/), machine learning (/tag/machine-learning/), artificial intelligence (/tag/artificial-intelligence/)





(https://twitter.com/share?

Looking for a job?

text=Ensemble%20Voting%20Classification%20in%20Python%20with%20Scikit-Learn&url=https://stackabuse.com/ensemble-voting-classification-in-python-with-scikit-learn/)

- Take Triplebyte's quiz to showcase your skills to



(https://www.facebook.com/sharer/sharer.php?

u=https://stackabuse.com/ensemble-voting-classification-in-python-with-scikit-learn/)



(https://www.linkedin.com/shareArticle?

mi%26url=https://stackabuse.com/ensemble-voting-classification-in-

py%26url=https://stackabuse.com/ensemble-voting-classification-in-python-with-scikit-learn/%26source=https://stackabuse.com)

ZO
OX

(https://stackabu.se/triplebyte/author/daniel/)

About Dan Nelson (/author/daniel/)

(https://stackabu.se/triplebyte/)

Arizona



</> Take Triplebyte's quiz (https://stackabuse.com/triplebyte/)

use my multiple talents and skillsets to teach others about the

transformative power of computer programming and data science.

Prepping for an interview?

(https://stackabu.se/daily-coding-problem)

- Improve your skills by solving one coding problem every day
- Get the solutions the next morning via email

Subscribe to our Newsletter

Get occasional tutorials, guides, and jobs in your inbox. No spam ever. Unsubscribe

- Practice on actual problems asked by top companies, like: at any time.

Google facebook amazon.com Microsoft

Enter your email...



</> Daily Coding Problem (https://stackabu.se/daily-coding-problem)

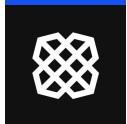
Subscribe

[< Previous Post \(/adding-a-postgresql-database-to-a-node-js-app-on-heroku/\)](#)

Looking for a job?

[Next Post > \(/string-formatting-with-python-3s-f-strings/\)](#)
<https://stackabu.se/triplebyte>

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.
- Coinbase, Plaid, Zoex, and Heap are still hiring!


<https://stackabu.se/triplebyte>

<https://stackabu.se/triplebyte>

<https://stackabu.se/triplebyte>

<https://stackabu.se/triplebyte>
[</> Take Triplebyte's Quiz \(https://stackabu.se/triplebyte\)](#)

Follow Us


<https://twitter.com/StackAbuse>

<https://www.facebook.com/stackabuse/>

<https://stackabuse.com/rss/>

Prepping for an interview?

- Improve your skills by solving one coding problem every day
- Get the solutions the next morning via email

Newsletter

- Practice on **actual problems** asked by top companies, like.
- Subscribe to our newsletter! Get occasional tutorials, guides, and reviews in your inbox.

[Google](#)
[facebook](#)
[amazon.com](#)
[Microsoft](#)

Enter your email...

[</> Daily Coding Problem \(https://stackabu.se/daily-coding-problem\)](#)

Subscribe

No spam ever. Unsubscribe at any time.

Ad

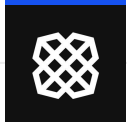
Looking for a job?

(<https://stackabu.se/triplebyte>)

- Take Triplebyte's quiz to showcase your skills to hiring managers and recruiters.
- Coinbase, Plaid, Zoox, and Heap are still hiring!



(<https://stackabu.se/triplebyte>)



(<https://stackabu.se/triplebyte>)

Z O
O X

(<https://stackabu.se/triplebyte>)

a remote job?



(<https://stackabu.se/triplebyte>)

Sold 18 days ago

~~\$100K - \$135K~~

(<https://stackabu.se/triplebyte>)
(<https://hiredremote.io/remote-job/1539-software-engineer-in-test-at-commentsold>)

php

(<https://hiredremote.io/remote-php-jobs>)

redis

(<https://hiredremote.io/remote-redis-jobs>)

mysql

(<https://hiredremote.io/remote-mysql-jobs>)

javascript

(<https://hiredremote.io/remote-javascript-jobs>)

Prepping for an interview?

(<https://stackabu.se/daily-coding-problem>)

(<https://stackabu.se/daily-coding-problem>)

- **TrueCar** 18 days ago (<https://hiredremote.io/remote-job/1538-software-engineer-back-end-at-truecar>)
Improve your skills by solving one coding problem every day

(<https://hiredremote.io/remote-aws-jobs>)

react

(<https://hiredremote.io/remote-react-jobs>)

- **Geely** (<https://hiredremote.io/remote-javascript-jobs>)

javascript

(<https://hiredremote.io/remote-javascript-jobs>)

- **Practice on actual problems** asked by top

companies, like:

FULL-STACK DJANGO DEVELOPER AND PROJECT LEADE...

Scalable Path 23 hours ago (<https://hiredremote.io/remote-job/1670-full-stack-django-developer-and-project-leader-at-scalable-path>)
Google facebook amazon.com Microsoft

django

(<https://hiredremote.io/remote-django-jobs>)

python

(<https://hiredremote.io/remote-python-jobs>)

(<https://stackabu.se/daily-coding-problem>)

python-jobs

scrum

(<https://hiredremote.io/remote-scrum-jobs>)

javascript

(<https://hiredremote.io/remote-javascript-jobs>)



More jobs (<https://hiredremote.io>)

Jobs via HireRemote.io (<https://hiredremote.io>)

Looking for a job?

<https://stackoverflow.com/questions/57697920/formatting-strings-in-java/>

- Take Triplebyte's quiz to showcase your skills to Hiring Managers and Recruiters. [\(https://triplebyte.com/\)](https://triplebyte.com/)
- Using SCP to Copy and Securely Transfer Files and Folders [\(/using-scp-to-copy-and-securely-transfer-files-and-folders/\)](https://triplebyte.com/)
- Coinbase, Plaid, Zoox, and Heap are still hiring!

Writing to a File with Python's print() Function [\(/writing-to-a-file-with-pythons-print-function/\)](https://triplebyte.com/) [\(https://triplebyte.com/\)](https://triplebyte.com/)



[\(https://triplebyte.com/\)](https://triplebyte.com/)

[algorithms \(/tag/algorithms/\)](/tag/algorithms/)
[amqp \(/tag/amqp/\)](/tag/amqp/)
[angular \(/tag/angular/\)](/tag/angular/)

[\(https://triplebyte.com/\)](https://triplebyte.com/)

[announcements \(/tag/announcements/\)](/tag/announcements/)
[apache \(/tag/apache/\)](/tag/apache/)
[api \(/tag/api/\)](/tag/api/)

[arduino \(/tag/arduino/\)](/tag/arduino/)
[artificial intelligence \(/tag/artificial-intelligence/\)](/tag/artificial-intelligence/)

[asynchronous \(/tag/asynchronous/\)](/tag/asynchronous/)
[aws \(/tag/aws/\)](/tag/aws/)

</> Take Triplebyte's Quiz [\(https://triplebyte.com/\)](https://triplebyte.com/)

Follow Us

Prepping for an interview?

<https://twitter.com/StackAbuse> <https://www.facebook.com/stackabuse/> <https://stackabuse.com/rss/>
<https://stackabu.se/daily-coding-problem>

- Improve your skills by solving one coding

problem every day

Copyright © 2020, Stack Abuse [\(https://stackabuse.com/\)](https://stackabuse.com/). All Rights Reserved.

- Get the solutions the next morning via email
- Disclosure [\(/disclosure\)](/disclosure) • Privacy Policy [\(/privacy-policy\)](/privacy-policy) • Terms of Service [\(/terms-of-service\)](/terms-of-service)
- Practice on **actual problems** asked by top companies, like:

Google facebook amazon.com Microsoft

</> Daily Coding Problem <https://stackabu.se/daily-coding-problem>