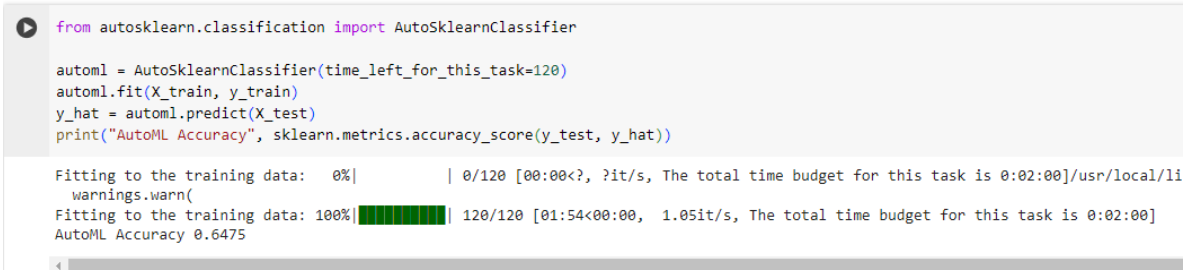
Charles Myers

I believe the reason why auto-sklearn is failing in this instance is because AutoSklearnClassifier is given the parameter called time_left_for_this_task and given the value of 300. Looking up the documentation for auto-sklearn I find on this page that parameter. There it says it will use that time in seconds to search for appropriate models and its default time is 3600. I believe that because the value given of 300 is so low auto-sklearn does not have anywhere near enough time to find a good model. If this time was increased, even to its default value it should increased performance.

Here is an image of it being a lower value of 120.



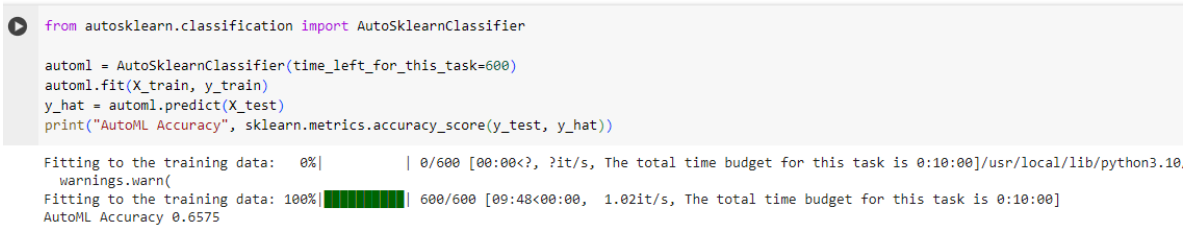
```
from autosklearn.classification import AutoSklearnClassifier

automl = AutoSklearnClassifier(time_left_for_this_task=120)
automl.fit(X_train, y_train)
y_hat = automl.predict(X_test)
print("AutoML Accuracy", sklearn.metrics.accuracy_score(y_test, y_hat))
```

```
Fitting to the training data:   0%|          | 0/120 [00:00<?, ?it/s, The total time budget for this task is 0:02:00]/usr/local/li
  warnings.warn(
Fitting to the training data: 100%|██████████| 120/120 [01:54<00:00,  1.05it/s, The total time budget for this task is 0:02:00]
AutoML Accuracy 0.6475
```

We can see the accuracy has decreased to 0.6475

Here is an image of doubling the value to 600.



```
from autosklearn.classification import AutoSklearnClassifier

automl = AutoSklearnClassifier(time_left_for_this_task=600)
automl.fit(X_train, y_train)
y_hat = automl.predict(X_test)
print("AutoML Accuracy", sklearn.metrics.accuracy_score(y_test, y_hat))
```

```
Fitting to the training data:   0%|          | 0/600 [00:00<?, ?it/s, The total time budget for this task is 0:10:00]/usr/local/lib/python3.10
  warnings.warn(
Fitting to the training data: 100%|██████████| 600/600 [09:48<00:00,  1.02it/s, The total time budget for this task is 0:10:00]
AutoML Accuracy 0.6575
```

We can see the accuracy raised to 0.6575.

Through these tests we learn that the reason why it's worse then the standard RF is because the time_left_for_this_task value is too low.

Source: https://automl.github.io/auto-sklearn/master/api.html