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

Reference: https://www.kaggle.com/code, https://machinelearningmastery.com/save-load-machine-learning-models-python-scikit-learn/

```
!pip install --upgrade scikit-learn==1.0.2
!pip install --upgrade numpy==1.21.5
```

```
import time
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('seaborn')

# Load dataset
from sklearn.datasets import load_iris
iris = load_iris()
iris_df =  pd.DataFrame(iris.data, columns=iris.feature_names)
iris_df['target_names'] = [iris.target_names[t] for t in iris.target]

feature_col = iris.feature_names
iris_df.head()
```

Build Simple Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier()
dtree.fit(iris_df[feature_col], iris_df['target_names'])
```

```
y_pred = dtree.predict(iris_df[feature_col])
y_pred[:10]
```

`pickle` is library to save/load most of variables in python.

```
import pickle
```

```
# Open file - Write binary mode
model_file = open('dtree.model', 'wb')

# Save Decision tree model
pickle.dump(dtree, model_file)

# Close file
model_file.close()
```

Restart your runtime

- Menu > Runtime > Restart runtime

Run Code below. `dtree` was gone.

```
dtree
```

Check your file

```
# List file
!ls
```

Load Saved Model

```
import time
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('seaborn')

# Load dataset
from sklearn.datasets import load_iris
iris = load_iris()
iris_df =  pd.DataFrame(iris.data, columns=iris.feature_names)

feature_col = iris.feature_names
```

```
import pickle

# Open file - Read binary mode
model_file = open('dtree.model', 'rb')

# Load your model
dtree = pickle.load(model_file)

# Close file
model_file.close()
```

```
dtree
```

```
y_pred = dtree.predict(iris_df[feature_col])
y_pred[:10]
```