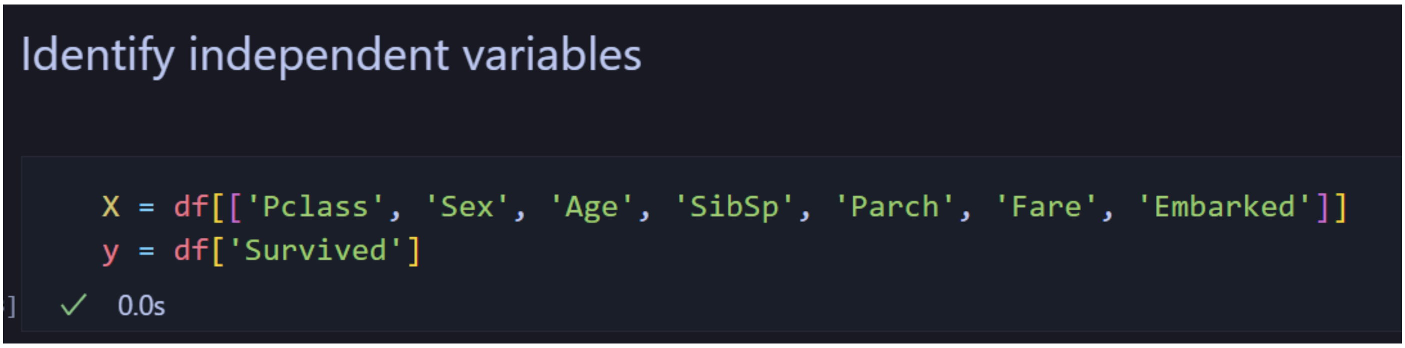
Lab 04 - Applied Machine Learning

Bui Minh Thuan - 104486358

For week 4 lab, we still try to predict Titanic survivor - the same task with lab 03 - but with Support Vector Machine (SVM) instead of Logistic Regression.

1. Selected columns

Here, the goal is to predict survival (**‘Survived’** column). Key features chosen as independent variables are: **‘Pclass’, ‘Sex’, ‘Age’, ‘SibSp’, ‘Parch’, ‘Fare’, and ‘Embarked’.** These features are selected based on their potential impact on survival. Below is the code for this task



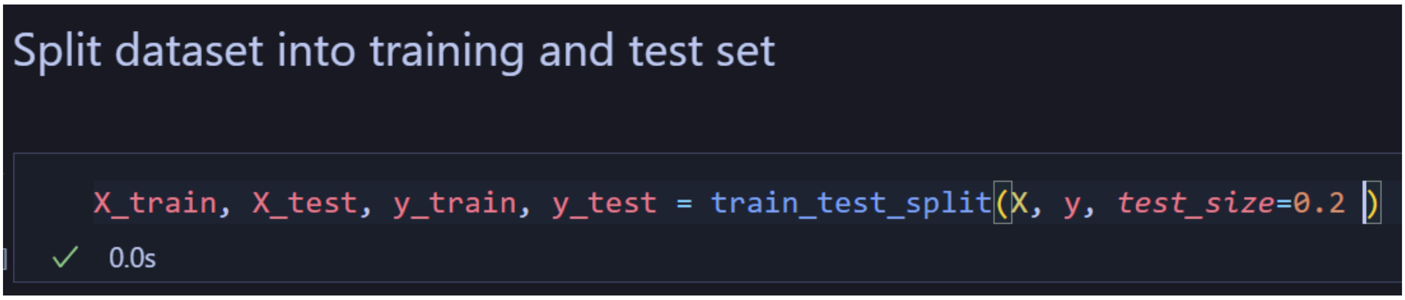
Identify independent variables

```
X = df[['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']]
y = df['Survived']
```

] ✓ 0.0s

2. Training and testing split

For training and testing split, I use the built-in train test split method from “SK-Learn” library, with 80% for train and 20% for test.



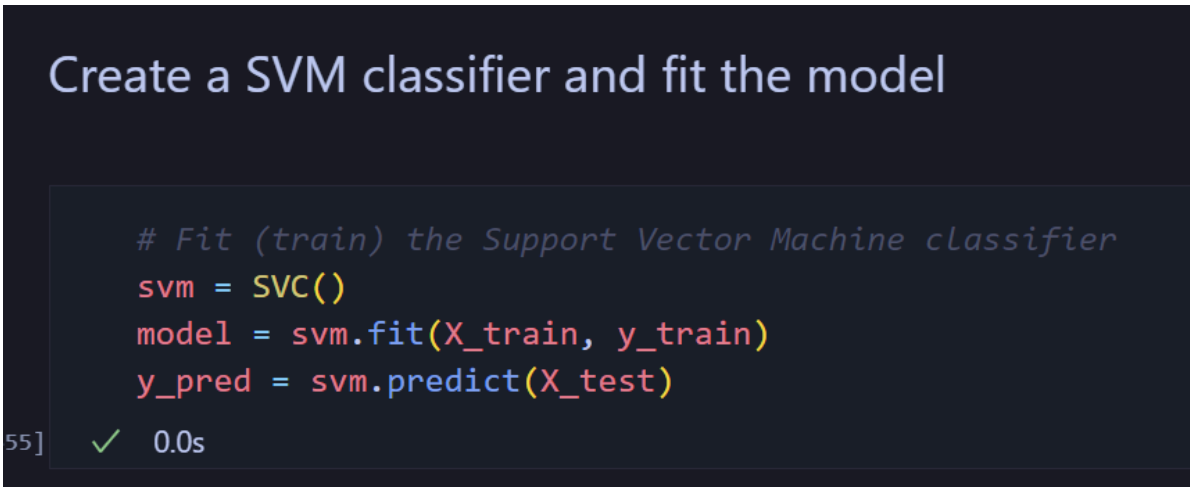
Split dataset into training and test set

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2 )
```

✓ 0.0s

3. Model building

About the SVM model, I also use the pre-built SVM classifier from ‘SKLearn’.



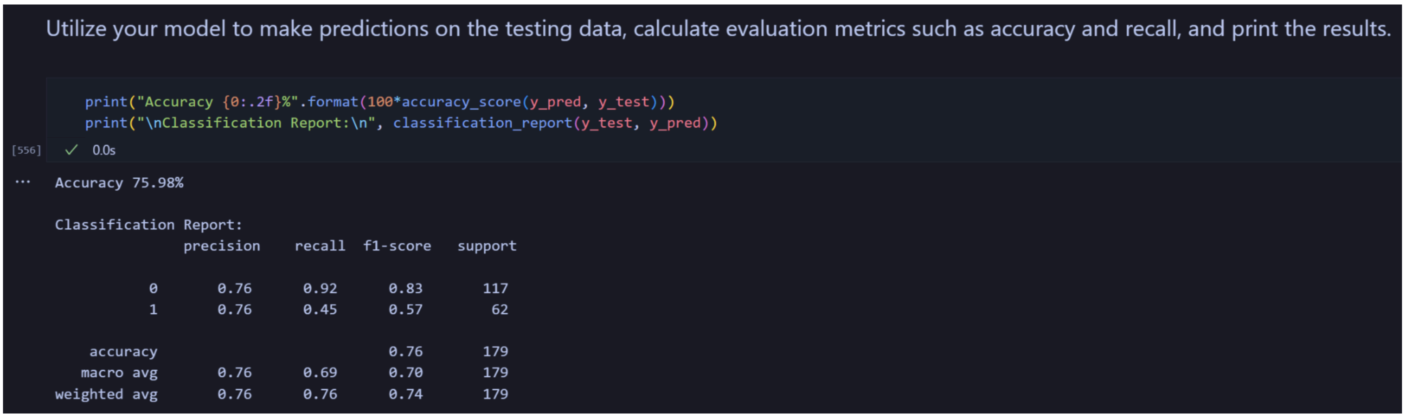
Create a SVM classifier and fit the model

```
# Fit (train) the Support Vector Machine classifier
svm = SVC()
model = svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)
```

55] ✓ 0.0s

4. Accuracy

Finally, we can evaluate the trained model. As you can see, SVM achieves around 76% accuracy, which is less than Logistic Regression we did last week (around 84%). This proves that choosing an effective model needs to based on the nature of the dataset, not about the complexity of the model itself.



Utilize your model to make predictions on the testing data, calculate evaluation metrics such as accuracy and recall, and print the results.

```
print("Accuracy {0:.2f}%".format(100*accuracy_score(y_pred, y_test)))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

[556] ✓ 0.0s

```
Accuracy 75.98%

Classification Report:
               precision    recall  f1-score   support

           0       0.76      0.92      0.83       117
           1       0.76      0.45      0.57        62

    accuracy                           0.76       179
   macro avg       0.76      0.69      0.70       179
weighted avg       0.76      0.76      0.74       179
```