

# Machine Learning - Assignment 3

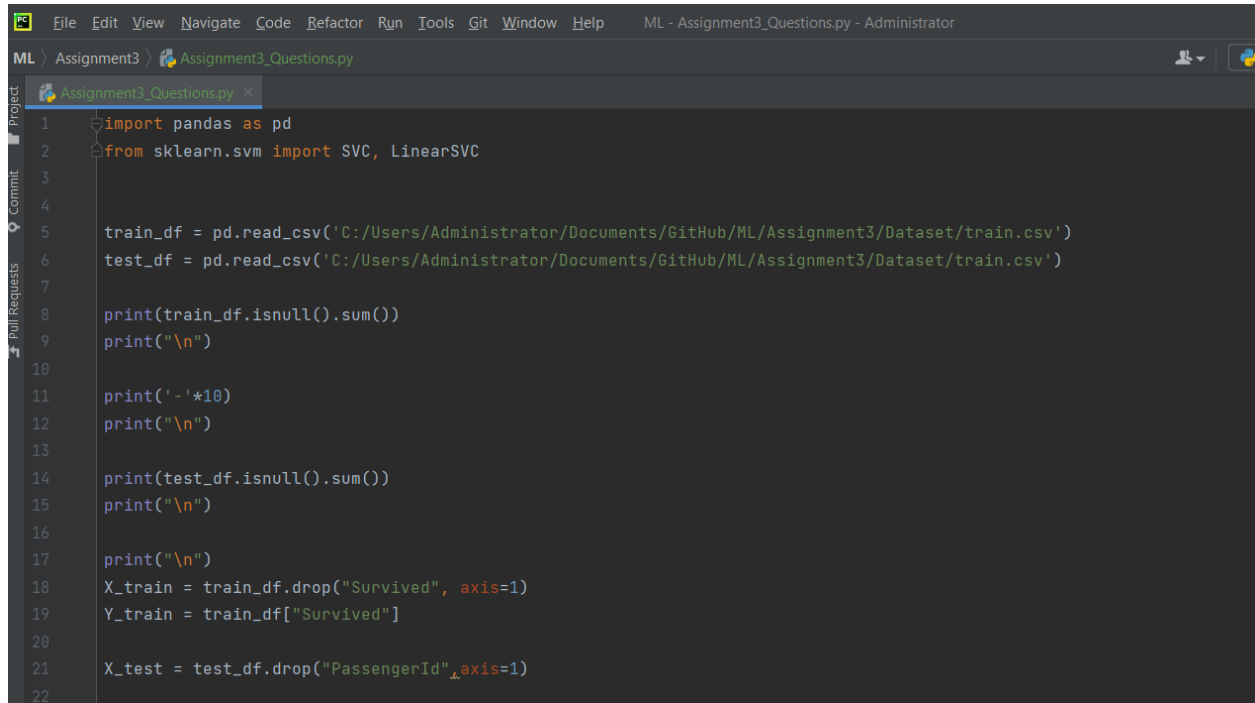
Name: Idavalapati Vijay Taraka Ramarao

ID: 700742485

CRN: 13428

## Question1

Here finding the correlation between 'survived' (target column) and 'sex' column for the Train use case in class.



```
File Edit View Navigate Code Refactor Run Tools Git Window Help ML - Assignment3_Questions.py - Administrator
ML > Assignment3 > Assignment3_Questions.py
Assignment3_Questions.py x
1 import pandas as pd
2 from sklearn.svm import SVC, LinearSVC
3
4
5 train_df = pd.read_csv('C:/Users/Administrator/Documents/GitHub/ML/Assignment3/Dataset/train.csv')
6 test_df = pd.read_csv('C:/Users/Administrator/Documents/GitHub/ML/Assignment3/Dataset/train.csv')
7
8 print(train_df.isnull().sum())
9 print("\n")
10
11 print('-'*10)
12 print("\n")
13
14 print(test_df.isnull().sum())
15 print("\n")
16
17 print("\n")
18 X_train = train_df.drop("Survived", axis=1)
19 Y_train = train_df["Survived"]
20
21 X_test = test_df.drop("PassengerId", axis=1)
22
```

```

23     print("\n")
24
25     svc = SVC(max_iter=1000)
26
27     svc.fit(X_train, Y_train)
28
29     Y_pred = svc.predict(X_test)
30
31     acc_svc = round(svc.score(X_train, Y_train) * 100, 2)
32
33     print("svm accuracy =", acc_svc)
34
35
36     # Adding the max_iter parameter and see the results
37     svc = LinearSVC()
38
39     svc.fit(X_train, Y_train)
40
41     Y_pred = svc.predict(X_test)
42
43     acc_svc = round(svc.score(X_train, Y_train) * 100, 2)
44
45     print("svm accuracy =", acc_svc)

```

## Outputs:

```

C:\Users\Administrator\Documents\GitHub\ML\venv\Scripts\python.exe C:/Users/Administrator/Documents/GitHub/ML/Assignment3/Question1_1.py
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch           0
Ticket          0
Fare             0
Cabin           687
Embarked         2
dtype: int64

```

```
PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64
```

Question2:

**Which algorithm you got better accuracy? Can you justify why?**

Naive Bayes algorithm got better accuracy, Naive Bayes is a classification algorithm which is suitable for binary and multiclass classification, the problem we dealt above is multiclass label problem.

whereas linear SVM doesn't support multiclass classification natively. It supports binary classification

**Related Links:**

SourceCode:

<https://github.com/VijayTarakaRamarao/ML/tree/main/Assignment3>

Recording:

[https://github.com/VijayTarakaRamarao/ML/blob/main/Assignment3/MachineLearning\\_Assignment3.mp4](https://github.com/VijayTarakaRamarao/ML/blob/main/Assignment3/MachineLearning_Assignment3.mp4)