**ICP4 Report**

**Harika Gurram**

**Class ID :17**

**Question 1. find the correlation between ‘survived’(target column) and ‘sex’ column for the Titanic use case in class. Do you think we should keep this feature?**

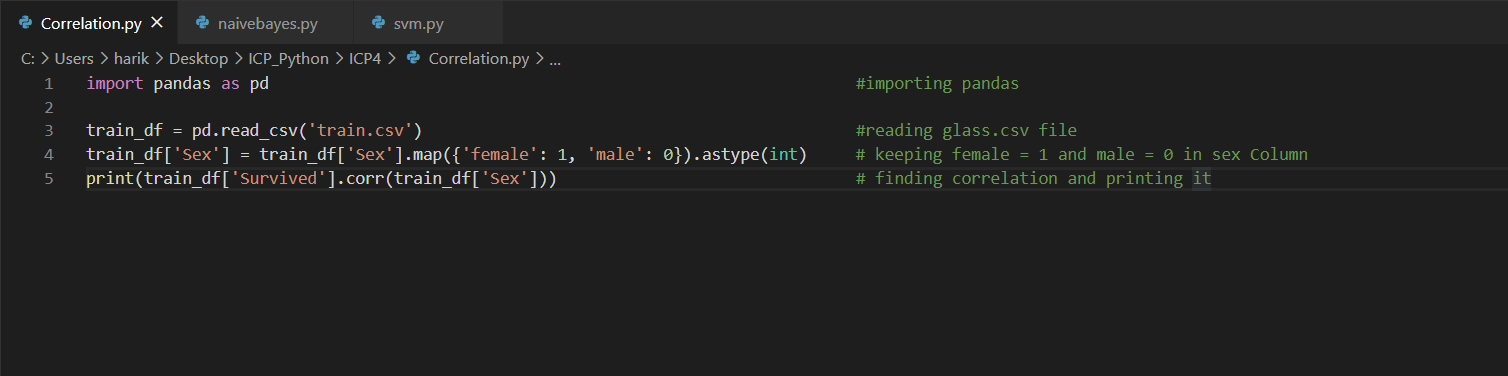
**Answer:**

By including Sex column in the dataset helps us to better analyze the percentage of males and females survived in the Titanic use case and By seeing the value of correlation (0.54 ) , we can include this column in dataset.

**Solution:**

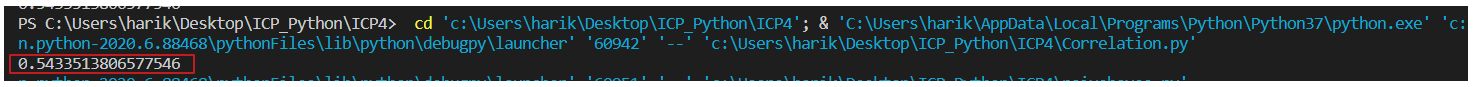
Using Pandas read the use case data and mapped the 'Sex' column categorical values to numerical values then found correlation of Survived column against Sex column using corr() function. Below is the screenshot attached of working code and output.

**Code:**



**Output:**

Below is the screenshot for correlation value between ‘survived’ column and ‘sex’ column

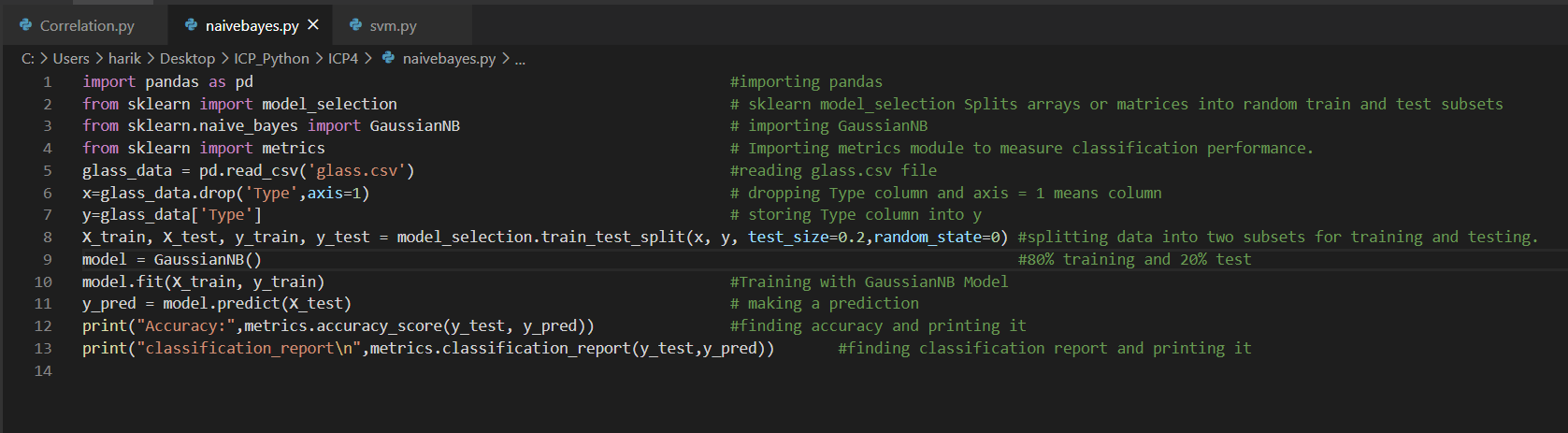


**Question 2. Implement Naïve Bayes method using scikit-learn libraryUse dataset available in**[**https://umkc.box.com/s/ea6wn1cidukan67t02j60nmp1ljln3kdUse**](https://umkc.box.com/s/ea6wn1cidukan67t02j60nmp1ljln3kdUse)**train\_test\_splitto create training and testing partEvaluate the model on testing partusing score and classification\_report(y\_true, y\_pred)**

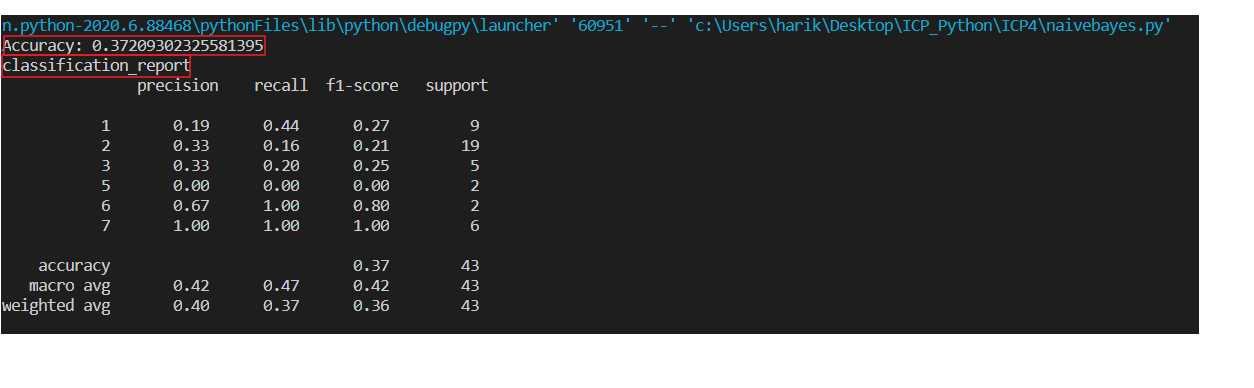
**Solution:**

In this initially imported basic libraries like pandas and scikit learn library then read csv file for data and for training the model dropped a column named 'Type' and assigned to X and used column 'Type' and assigned to y then using train\_test\_split splitted the data to training set(80%) and testing set(20%). Imported gaussianNB method and fitted our train data to it and predicted using test data. At the end accuracy is found using accuracy\_score on test and predicted data and classification report is generated similarly. Below is screenshot attached of working code and output.

**Code:**



**Output:**



**Question 3. Implement linear SVMmethodusing scikit libraryUse the samedataset aboveUse train\_test\_splitto create training and testing partEvaluate the model on testing partusing score and classification\_report(y\_true, y\_pred)Which algorithm you got better accuracy? Can you justify why?**

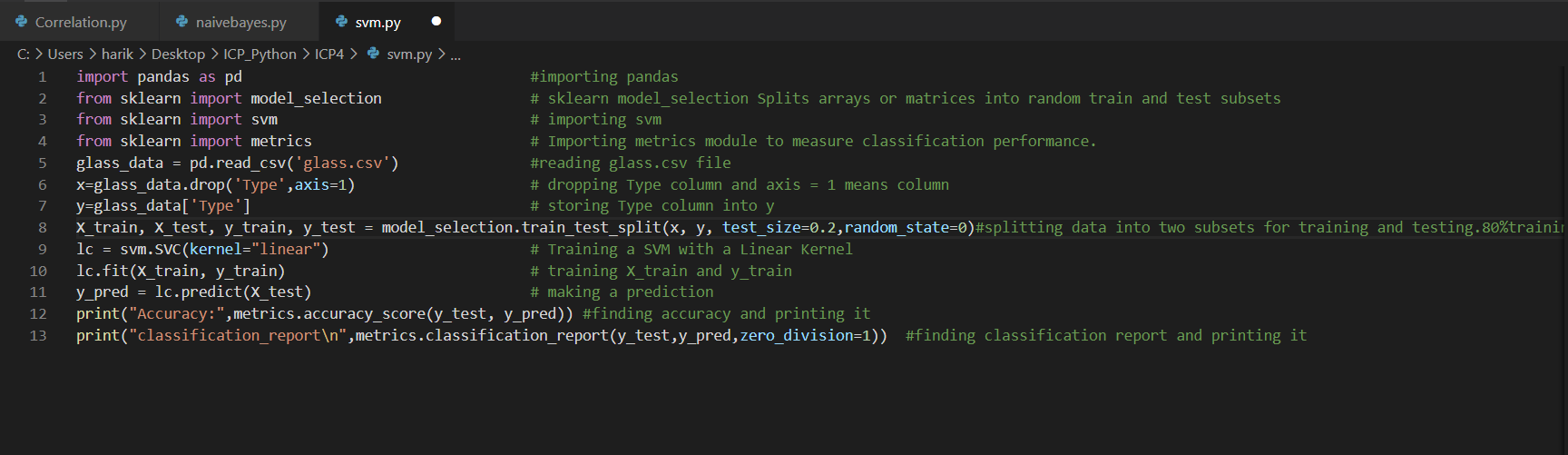
**Answer**

Comparing Naive Bayes and SVM model we achieved better accuracy for SVM than Naive Bayes algorithm. As Naive Bayes algorithm depends on independent features whereas SVM looks for interactions among them to certain degree. So by looking at data set given SVM suits well.

**Solution:**

This task is performed same as task 2 but used SVM algorithm to check the accuracy score. In SVM we used linear model and fitted our training data to it and predicted using testing data.Accuracy score and classification reports are generated at the end.Below is the screenshot attached of working code and output.

**Code:**



**Output:**

