Spring 2024: CS5720 Neural Networks & Deep Learning - ICP-5

Assignment-5
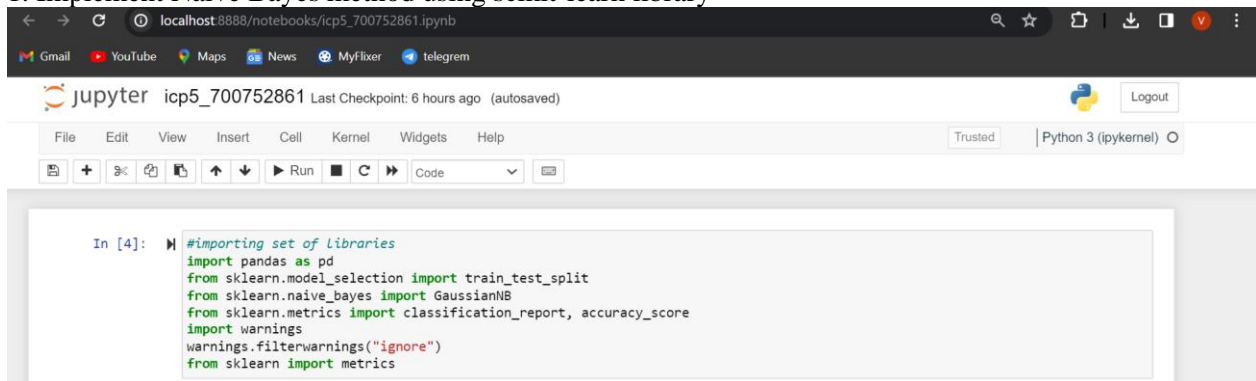
Name: Pushkara Naga Sai Sri Vyshnavi Chakka

STUDENT ID:700752861

GitHub link: https://github.com/PushkaraChakka/Assignment_5_icp5
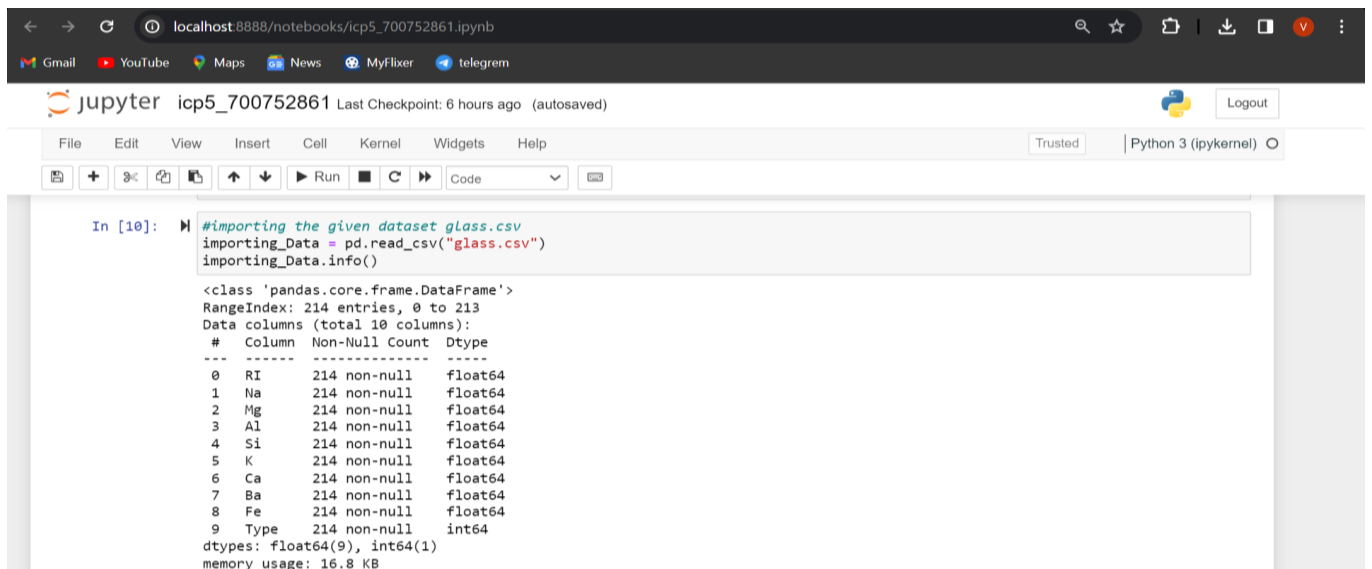
Video link: https://drive.google.com/file/d/1lnyIaHI5aFs3qixSw5Lt5NUP6t_jkxDF/view?usp=sharing

1. Implement Naïve Bayes method using scikit-learn library

Use dataset available with name glass

Use train_test_split to create training and testing part
Evaluate the model on test part using score and
classification_report(y_true, y_pred)


```python
In [12]:  #splitting the dataset which is excluding last columns
          X = importing_Data.iloc[:, :-1]
          y = importing_Data.iloc[:, -1]
          #splitting the dataset into train and test datasets
          X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
          #creating a Gaussian Naive Bayes model
          ge = GaussianNB()
          #fitting train data
          ge.fit(X_train, y_train)
          # predicting the test dataset
          y_pred = ge.predict(X_test)
          # evaluating the model on the test dataset
          print("Accuracy: ", accuracy_score(y_test, y_pred)*100)
          print("Classification Report: \n", classification_report(y_test, y_pred))

          Accuracy:  37.2093023255814
          Classification Report:
                        precision    recall  f1-score   support

                     1       0.19      0.44      0.27         9
                     2       0.33      0.16      0.21        19
                     3       0.33      0.20      0.25         5
                     5       0.00      0.00      0.00         2
                     6       0.67      1.00      0.80         2
                     7       1.00      1.00      1.00         6

              accuracy                           0.37        43
             macro avg       0.42      0.47      0.42        43
          weighted avg       0.40      0.37      0.36        43
```

2. Implement linear SVM method using scikit library
Use the same dataset above


```python
In [13]:  #importing set of libraries
          import pandas as pd
          from sklearn.model_selection import train_test_split
          from sklearn.svm import SVC
          from sklearn.metrics import classification_report, accuracy_score

In [14]:  #loading the glass dataset
          importing_Data = pd.read_csv("glass.csv")
          importing_Data.info()

          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 214 entries, 0 to 213
          Data columns (total 10 columns):
           #   Column  Non-Null Count  Dtype
          ---  ------  --------------  -----
           0   RI      214 non-null    float64
           1   Na      214 non-null    float64
           2   Mg      214 non-null    float64
           3   Al      214 non-null    float64
           4   Si      214 non-null    float64
           5   K       214 non-null    float64
           6   Ca      214 non-null    float64
           7   Ba      214 non-null    float64
           8   Fe      214 non-null    float64
           9   Type    214 non-null    int64
          dtypes: float64(9), int64(1)
          memory usage: 16.8 KB
```
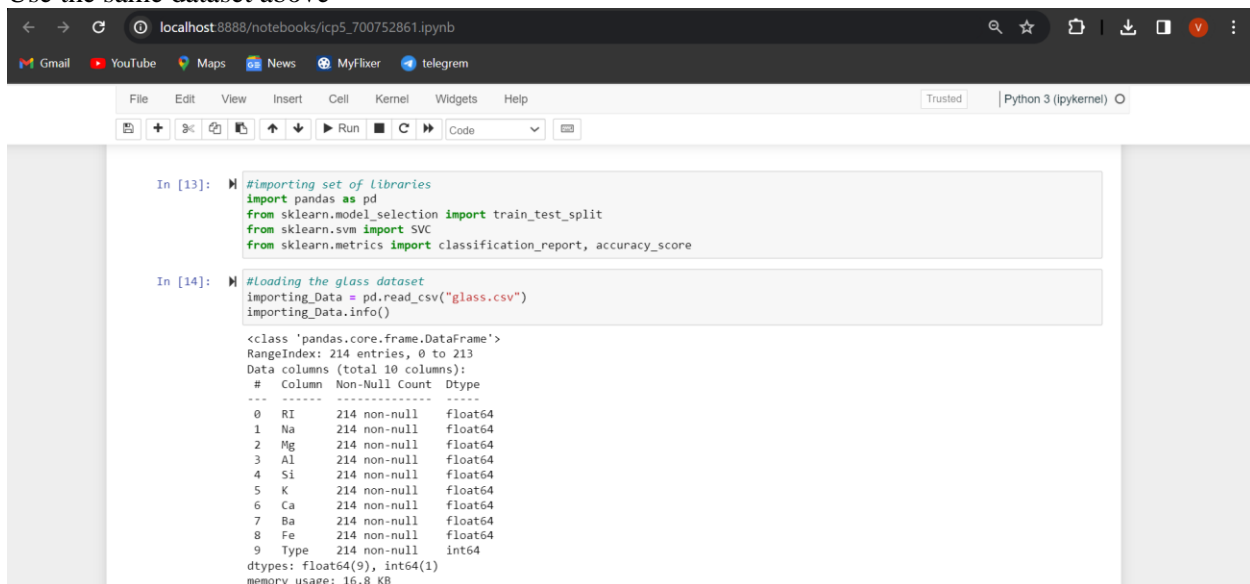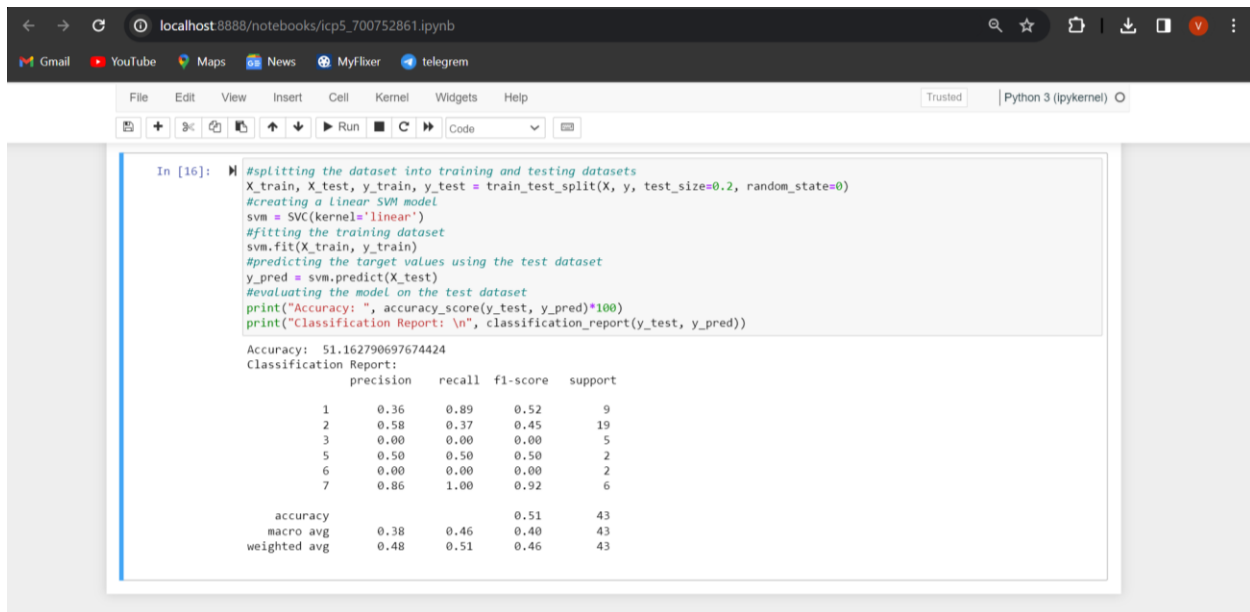
Use train_test_split to create training and testing part
Evaluate the model on test part using score and
classification_report(y_true, y_pred)

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

The biggest difference between the models you're building from a "features" point of view is that Naive Bayes treats them as independent, whereas SVM looks at the interactions between them to a certain degree, as long as you're using a non-linear kernel. So, if you have interactions, and, given your problem, you most likely do, an SVM will be better at capturing those, hence better at the classification task you want.