**Assignment-4**

**Naïve Bayes**

(1) Programming: Use Naïve Bayes to predict labels of instances in the testing set (12 games) based on the training set (24 games). Calculate Accuracy, Precision, Recall, and F1 score on the testing result.

*Solution:*

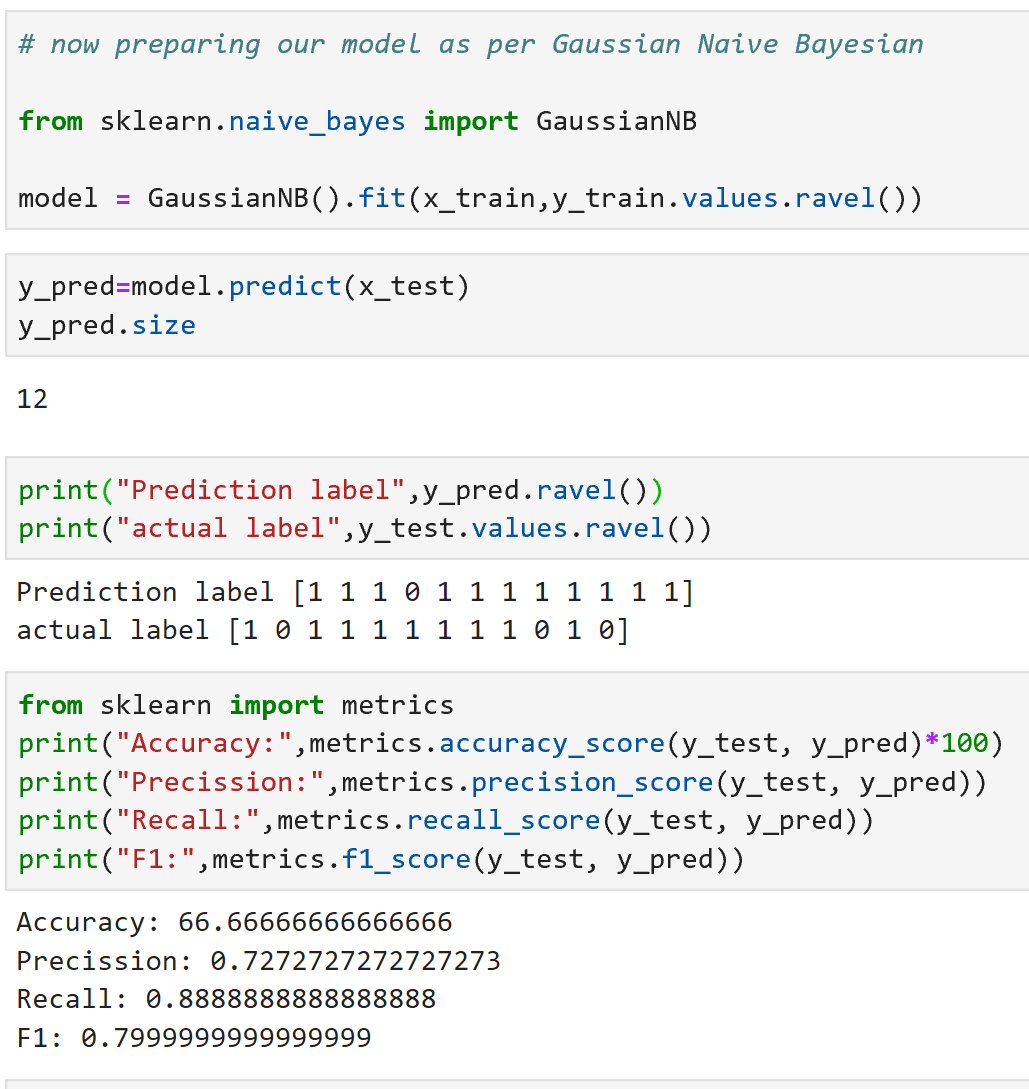
In this problem, I used get\_dummies to encode data to 1 and 0 to add new columns based on the categories. Label Encoding just converts the data to Numerical as 1,2,3 but depending on the data, label encoding introduces a new problem. For example, we have encoded media names into numerical data. This is categorical data and there is no relation, of any kind, between the rows.*The problem here is since there are different numbers in the same column, the model will misunderstand the data to be in some kind of order, 0 < 1 <2.*

Accuracy: 66.67

Precision: 0.72

Recall: 0.89

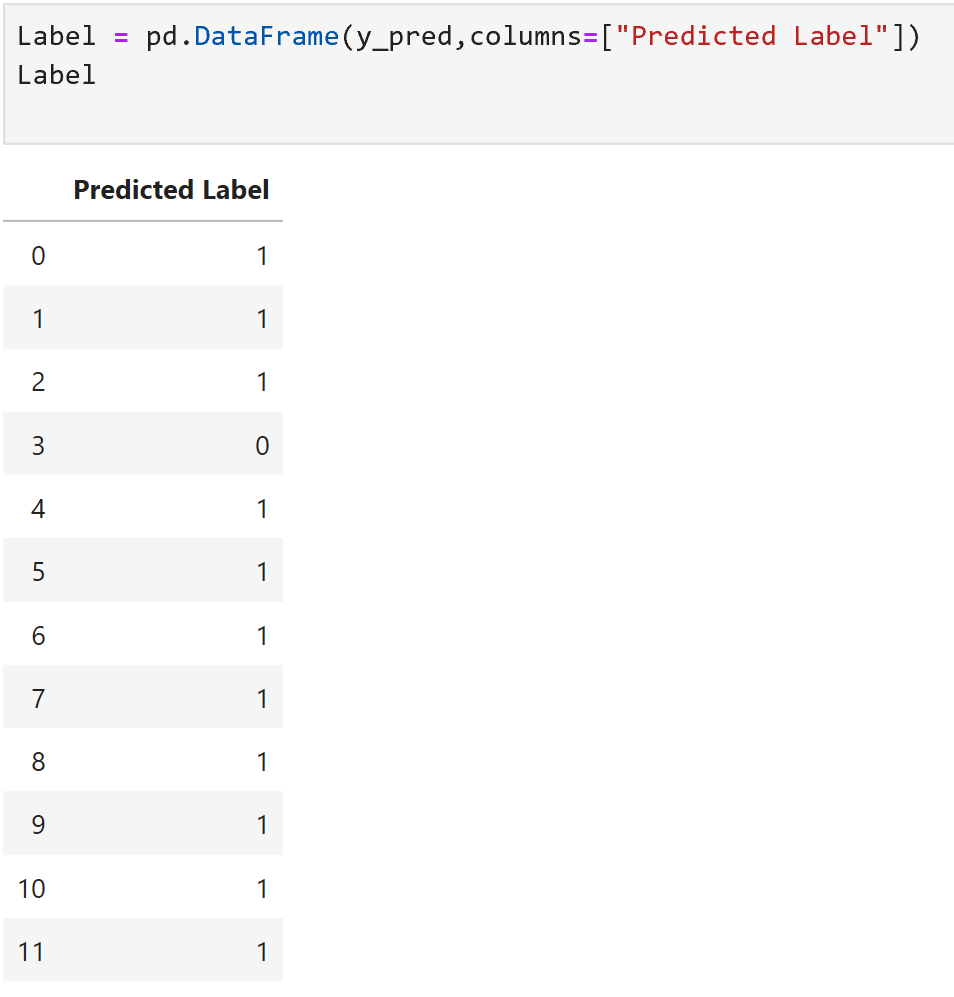
F1: 0.79



(2) Write down prediction labels of the 12 testing games in the PDF.

*Solution:*

Prediction Label:

Win

Win

Win

Lose

Win

Win

Win

Win

Win

Win

Win

Win

(3) Compare Naïve Bayes with ID3 and C4.5, which model is the best, which model performs the worst? Can you explain why?

Solution:

C4.5 is best than ID3 or Naïve bayes.

**Advantages of C4.5 over other Decision Tree systems:**

1. The algorithm inherently employs Single Pass Pruning Process to Mitigate overfitting.
2. It can work with both **Discrete**and **Continuous**Data
3. C4.5 can handle the issue of incomplete data very well.

In our Solution we can see that the Accuracy of C4.5 was 91.66, whereas ID3 accuracy was less and naïve Bayes it is around 66.6%. From this we can infer that C4.5 is comparatively good.