MMDT091
Mentor Ma Nuwai Thet

1. What's the difference between the three Naive Bayes classifiers (BernoulliNB, GaussianNB, and MultinomialNB)?

This week, we have tested and experienced Naive Bayers models. This model is based on probability. In our case, we are predicting whether the transaction is fraud or not. So, I think all three Naive Bayes classifiers, BernoulliNB, would suit our case with different purposes.

This is because GaussianNB may be more useful in predicting continuous data. For MultinomailNB, it will be better for finding the items of the repeated counts.

In our dataset, if we would like to predict the transaction amount, we can use GaussianNB. But if we want to track some words in our dataset, it is better to use MultinomialNB.

2. Find and compare the precision and recall for all three Naive Bayes classifiers.

GAUSSIAN NB Performance:

	precision	recall	f1-score	support
Not Fraud	1.00	0.98	0.99	8541
Fraud	0.38	0.86	0.53	137
accuracy			0.98	8678
macro avg	0.69	0.92	0.76	8678
weighted avg	0.99	0.98	0.98	8678

MULTINOMIAL NB Performance:

	precision	recall	f1-score	support
Not Fraud	0.98	1.00	0.99	8541
Fraud	0.00	0.00	0.00	137
accuracy			0.98	8678
macro avg	0.49	0.50	0.50	8678
weighted avg	0.97	0.98	0.98	8678

```
BERNOULLI NB Performance:
              precision    recall  f1-score   support

   Not Fraud       0.98      1.00      0.99      8541
       Fraud       0.00      0.00      0.00       137

    accuracy                           0.98      8678
   macro avg       0.49      0.50      0.50      8678
weighted avg       0.97      0.98      0.98      8678
```

Model	Precision	Recall
GaussianNB	0.38	0.86
MultinomialNB	0.00	0.00
BernoulliNB	0.00	0.00

These are the precision and recall from three classifiers.

Among all three models, GaussianNB can catch 86% if the fraud and 0.38(62%) are wrongly regarded as fraud.

For multinomials and Bernouli, we can see that they are all predicted as “Not Fraud”.

Also if we check for the f1 score of all classifiers, only the f1 score of GaussianNB(0.53) captures the balance of precision and recall. In this case, we should not only look for accuracy which is around 0.98 for all models, but also for f1 score.

3. Which model would you choose for fraud detection, and why?

In our fraud detection dataset, our best choice will be GaussianNB as it can catch fraud cases. But we need to be careful if it also includes a precision of 0.38. However, among three models, GaussianNB can find the fraud case.