

1) **What is the task?**

Task: Classify Amazon Book Reviews according to rating.

2) **What ML solution did you choose and, most importantly, why was this an appropriate choice?**

I tried two Naïve Bayes classifiers with their variant

Steps - Load data > Preprocess the data > Split train/dev/test > vectorize the feature and the target variables using TF-IDF vectorizer converting category to numbers > use test data in the inference process.

3) **How did you choose to evaluate success?**

I used the F1 score, accuracy score, Confusion matrix, and classification report to evaluate the models.

4) **What software did you use and why did you choose it?**

Language: Python as all the libraries.

Software: *scikit-learn* has all the machine learning algorithms prewritten, *NumPy*, *pandas*, *matplotlib*,

NLTK and others that are written in python.

5) **what are the results?**

Predicting test data using Multinomial Naive Bayes

```
Accuracy score: 0.83
Precision score: 0.84
Recall score: 0.80
F1 score: 0.83
```

Predicting test data using GaussianNaive Bayes

```
Accuracy score: 0.61
Precision score: 0.68
Recall score: 0.42
F1 score: 0.61
```

Predicting test data using BernoulliNaive Bayes

```
Accuracy score: 0.82
Precision score: 0.84
Recall score: 0.78
F1 score: 0.82
```

Predicting test data using ComplementNaive Bayes

Accuracy score: 0.83
Precision score: 0.84
Recall score: 0.80
F1 score: 0.83

Predicting test data using Logistic Regression

Accuracy score: 0.81
Precision score: 0.80
Recall score: 0.82
F1 score: 0.81

The Support vector machine achieves a better result on the data. Hence it could be used in the production server.