

COMPARISON REPORT: Multinomial Naive Bayes vs. Bernoulli Naive Bayes on Movie Dataset

1. Introduction:

The objective of this project was to apply two different Naive Bayes classification models on a real-world movie dataset. The dataset contains two key columns:

- ONE-LINE: A short text description or review about the movie.
- LABEL: The sentiment classification of the movie (Positive or Negative).

Our goal was to evaluate the performance of both Multinomial and Bernoulli Naive Bayes models and determine their suitability for this specific dataset.

2. Data Preparation & Cleaning:

Before applying any model, the following preprocessing steps were performed:

- Converted all text to lowercase.
- Removed numbers and punctuation.
- Removed stopwords using NLTK.
- Applied Lemmatization to normalize words.

Vectorization:

- For Multinomial Naive Bayes: CountVectorizer (word frequency).
- For Bernoulli Naive Bayes: CountVectorizer with binary=True (presence/absence of words).

The dataset was split into 70% training and 30% testing sets.

3. Model 1: Multinomial Naive Bayes

Vectorizer Used: CountVectorizer (based on word count frequency).

Results:

- Accuracy: 63.72%
- Detailed Metrics:
 - Negative Class (Label = 0):
 - Precision: 0.61
 - Recall: 0.55
 - F1-Score: 0.57
 - Support: 1075 samples
 - Positive Class (Label = 1):
 - Precision: 0.66
 - Recall: 0.71
 - F1-Score: 0.68
 - Support: 1326 samples

Confusion Matrix:

```
[[586  489]
 [382  944]]
```

Interpretation:

- The model performs better on positive reviews compared to negative ones.
- Struggles with detecting negative reviews (Recall = 0.55).
- Moderate overall performance.

4. Model 2: Bernoulli Naive Bayes

Vectorizer Used: CountVectorizer with binary=True (presence/absence of words).

Results:

- Accuracy: 64.45%
- Detailed Metrics:
 - Negative Class (Label = 0):
 - Precision: 0.62
 - Recall: 0.51
 - F1-Score: 0.56
 - Support: 711 samples
 - Positive Class (Label = 1):
 - Precision: 0.66
 - Recall: 0.75
 - F1-Score: 0.70
 - Support: 890 samples

Confusion Matrix:

```
[[361  350]
 [219  671]]
```

Interpretation:

- Performs similarly to Multinomial but shows slight improvement in detecting positive reviews (Recall = 0.75).
- Has slightly lower performance on detecting negative reviews.

5. Overall Analysis & Suitability:

- Both models demonstrate moderate performance on this dataset.
- Neither model achieves high accuracy or strong detection of both classes.
- There is noticeable difficulty in correctly classifying negative reviews.
- Slight improvement in Bernoulli Naive Bayes, but not significant enough to consider it highly reliable.

6. Conclusion:

- The current dataset, based solely on short ONE-LINE text, appears challenging for Naive Bayes models to achieve high performance.
- Both models may be considered "acceptable" for basic sentiment classification but are NOT highly suitable for accurate, reliable classification tasks on this dataset in its current form.
- To improve performance, further steps are recommended:
 - Experiment with n-grams in vectorization.
 - Improve text preprocessing.

- Consider more advanced models like Logistic Regression or SVM.