Custom Naive Bayes Classifier vs Sklearn GaussianNB

1. Objective

The goal of this assignment is to implement a Naive Bayes classifier from scratch and compare it with scikit-learn's GaussianNB. The classifiers are used to predict weekly stock labels (Green/Red) based on features: mean_return and volatility.

2. Dataset

The dataset contains weekly data from 2020 to 2024 with the following features:

- • mean_return
  • volatility
  • Label (Green/Red)

3. Preprocessing

Labels were encoded as follows:
• Green → 1
• Red → 0

Training set: 2020–2022
Testing set: 2023–2024

4. Custom Naive Bayes Implementation

A Gaussian Naive Bayes classifier was implemented from scratch using the formula for normally distributed features. The model computes class-wise means, variances, and prior probabilities from training data. Predictions are made by computing the log of posterior probabilities for each class.

5. Comparison with Sklearn GaussianNB

Both the custom model and scikit-learn’s GaussianNB were trained and evaluated on the same dataset.

6. Results

Custom Naive Bayes Performance:

- Accuracy: 88.57%
- Confusion Matrix: [[55, 4], [8, 38]]

Sklearn GaussianNB Performance:

- Accuracy: 88.57%
- Confusion Matrix: [[55, 4], [8, 38]]

7. Conclusion

Both the custom and sklearn classifiers performed identically on the test data, indicating that the custom implementation is correct. Naive Bayes proved to be an efficient and simple method for classifying the weekly stock trends based on historical return and volatility.

Attachments

1. Processed_Stock_Data.csv
2. Python code (Naive Bayes Custom vs Sklearn)
3. This Word report