

Assignment: Sentiment Analysis using Naive Bayes Classifier

Course Code: CS3151

Topic: Sentiment Analysis using Naive Bayes

CLO: CLO3 – Analyze artificial intelligence techniques for practical problem-solving

Total Marks: 20

Submission Deadline: 28-06-2025

Background:

Naive Bayes is a probabilistic classifier based on Bayes' theorem. It is widely used in natural language processing tasks such as spam detection, topic classification, and sentiment analysis. In this assignment, you will apply the Naive Bayes algorithm to a small corpus for sentiment classification.

Corpus:

Below is a small dataset of movie reviews. Each review is labeled as either positive or negative sentiment:

Review	Sentiment
I loved the movie, it was fantastic!	Positive
The film was boring and too long.	Negative
An excellent and gripping story.	Positive
Poor acting and terrible script.	Negative
Amazing performance by the lead actor.	Positive
Not worth the time, completely dull.	Negative

Instructions and Questions:

Q1. Understanding Naive Bayes (3 marks)

- a. Briefly explain Bayes' theorem and how it applies to text classification.

- b. Why is the Naive Bayes classifier considered 'naive'?

Q2. Preprocessing (4 marks)

- a. Tokenize the reviews and remove stop words.
- b. Create a vocabulary of words from the processed corpus.

Q3. Training (6 marks)

- a. Calculate the prior probabilities for each class (Positive, Negative).
- b. Compute the conditional probabilities (likelihood) of a few selected words for each class using Laplace smoothing.

Q4. Prediction (4 marks)

- a. Given a new review: 'The movie was not interesting and poorly acted.', classify it as Positive or Negative using your model.
- b. Show all intermediate steps including tokenization and probability calculation.

Q5. Reflection (3 marks)

- a. What are the advantages and limitations of using Naive Bayes for sentiment analysis?
- b. Suggest one improvement to increase the accuracy of the classifier.