

# Internship Task-4 Report: Sentiment Analysis

## Project Title: Sentiment Analysis on Text Data

### Objective

The objective of this task is to perform sentiment analysis on textual data such as movie reviews or tweets using Natural Language Processing (NLP) techniques. The goal is to classify each text into positive or negative sentiment.

### Tools and Technologies Used

- Google Colab
- Python
- Libraries: Pandas, NLTK, Scikit-learn, Seaborn, Matplotlib
- Dataset: IMDb 50K Movie Reviews

### Steps Followed

1. Data Loading: IMDb dataset containing 50,000 movie reviews.
2. Preprocessing: Lowercasing, removing HTML tags, punctuation, and stopwords.
3. Splitting: Divided data into training and test sets.
4. Feature Extraction: Used CountVectorizer to convert text into numerical format.
5. Model Training: Trained a Naive Bayes classifier.
6. Evaluation: Calculated accuracy, confusion matrix, and classification report.

### Results and Insights

The model achieved an accuracy of approximately 85% on the test data. It effectively identifies positive and negative reviews. The confusion matrix and classification report showed balanced performance with precision

## Internship Task-4 Report: Sentiment Analysis

and recall above 80%.

### Conclusion

This sentiment analysis model can be extended to other domains such as social media analysis, customer feedback, or product reviews. Natural Language Processing proves to be a powerful tool in extracting insights from unstructured data.