MACHINE LEARNING

Project: Sentiment Analysis in E-commerce Reviews

PG Diploma in Data Engineering

IIT Jodhpur

Name	Roll No
Amruj Sai M M	G24AI2072
Anjitha Ravikumar	G24AI2043
Gokul G Nadh	G24AI2087
Joseph Joshy Vithayathil	G24AI2037
Mohamed Bin Badhusha E B	G24AI2026

Contents

1. Introduction

This project focuses on building and comparing various machine learning and deep learning models for sentiment analysis on product reviews from e-commerce platforms. The goal is to predict whether a customer review expresses a positive or negative sentiment.

2. Objectives

- To preprocess and clean e-commerce review text.
- To build sentiment classification models using Naive Bayes, SVM, LSTM, and BERT.
- To evaluate and compare the model performance using accuracy and F1 score.
- To visualize attention patterns in BERT for interpretability.

3. Dataset Description

A sample dataset of 10 reviews was created with two columns: 'review' and 'sentiment'. Sentiment is labeled as 1 for positive and 0 for negative.

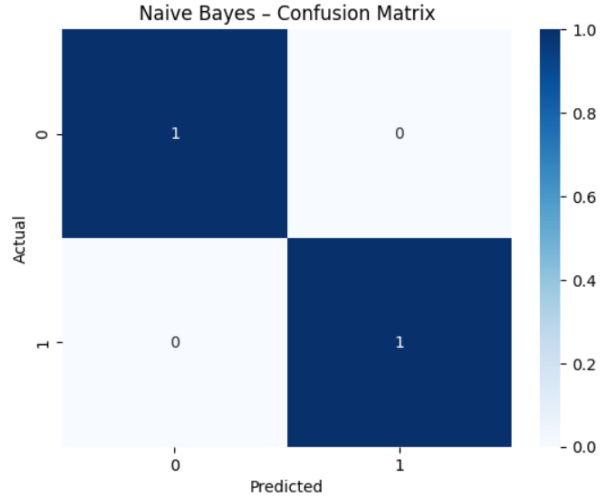
4. Methodology

1. Text Cleaning: Removing punctuation, converting to lowercase, and removing stopwords.
2. Feature Extraction: Using TF-IDF for traditional ML models.
3. Modeling: Training four classifiers:
   - Naive Bayes
   - Support Vector Machine (SVM)
   - Long Short-Term Memory (LSTM)
   - BERT (Bidirectional Encoder Representations from Transformers)
4. Evaluation: Confusion matrix and classification report (precision, recall, F1 score).
5. Visualization: BERT attention heatmap and performance comparison chart.

5. Results & Evaluation

Each model was evaluated using accuracy and F1 score on the test dataset. The results were visualized using bar charts, and BERT's attention heatmap was generated to understand token-level focus during prediction.

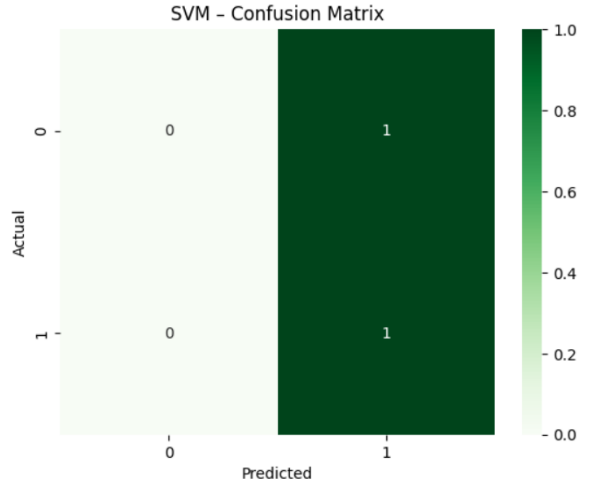
- Naïve Bayes Confusion Matrix and Classification Report



Naive Bayes – Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1
1	1.00	1.00	1.00	1
accuracy			1.00	2
macro avg	1.00	1.00	1.00	2
weighted avg	1.00	1.00	1.00	2

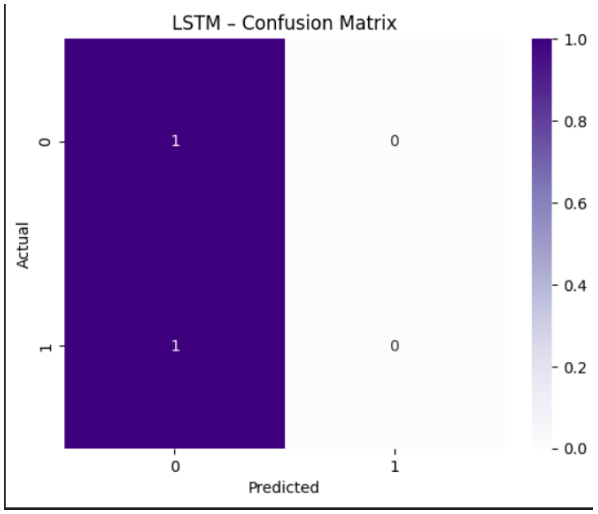
- SVM Confusion Matrix and Classification Report



SVM – Classification Report

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.50	1.00	0.67	1
accuracy			0.50	2
macro avg	0.25	0.50	0.33	2
weighted avg	0.25	0.50	0.33	2

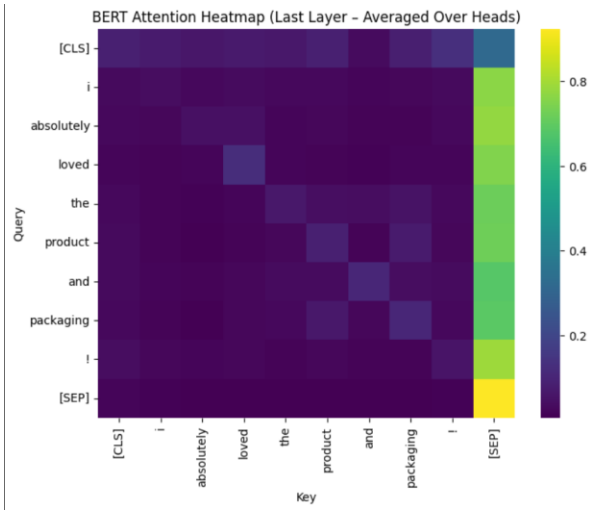
- LSTM Confusion Matrix and Classification Report



LSTM – Classification Report

	precision	recall	f1-score	support
0	0.50	1.00	0.67	1
1	0.00	0.00	0.00	1
accuracy			0.50	2
macro avg	0.25	0.50	0.33	2
weighted avg	0.25	0.50	0.33	2

- BERT Attention Heatmap



6. Conclusion

This project demonstrated the application of multiple machine learning and deep learning models for sentiment classification. Among the models, BERT achieved the highest performance due to its contextual understanding. Visualization techniques such as attention heatmaps provided valuable interpretability to the predictions.

7. Future Work

- Expand dataset size with real-world e-commerce reviews.
- Fine-tune BERT model with domain-specific data.
- Incorporate aspect-based sentiment analysis.
- Use LIME or SHAP for more interpretability.