PBEL PROJECT

SENTIMENT ANALYSIS OF PRODUCT REVIEWS USING NAIVE BAYES

ARTIFICIAL INTELLIGENCE 23 JUNE BATCH-1

INTERN'S NAME:- AVANISH DWIVEDI

TRAINER'S NAME:- MR. ANAND MAURYA

❑ INTRODUCTION

• Sentiment Analysis is a Natural Language Processing (NLP) technique used to determine the emotional tone behind text data.

• It is widely used in analyzing product reviews, social media posts, and customer feedback.

• This project focuses on classifying Amazon product reviews into three categories: Positive, Neutral, and Negative.

• The app is built using Python, trained with machine learning, and deployed using the Streamlit framework.

❑ DATASET OVERVIEW

- Dataset Source: Amazon Product Reviews Dataset (Reviews.csv)
- Total Records: ~500,000+ customer reviews
- Selected Columns:
  - Text (Review content)
  - Score (Rating from 1 to 5)
- Data Cleaning Steps:
  - Removed missing/null values
  - Filtered relevant columns only
  - Balanced the dataset by sampling equal reviews for each class (positive, neutral, negative).

❑ PREPROCESSING

- Converted all reviews to lowercase to maintain consistency
- Removed punctuation marks and special characters
- Removed stopwords (e.g., "is", "the", "and") using NLTK
- Applied word stemming using PorterStemmer (e.g., "playing" → "play")
- Cleaned text saved as new column for model training

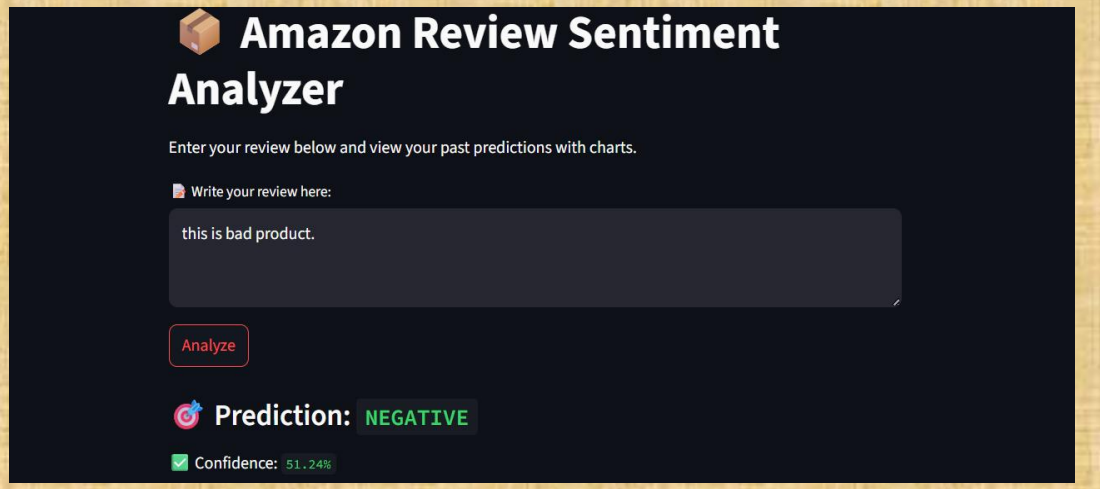
❑ MODEL BUILDING

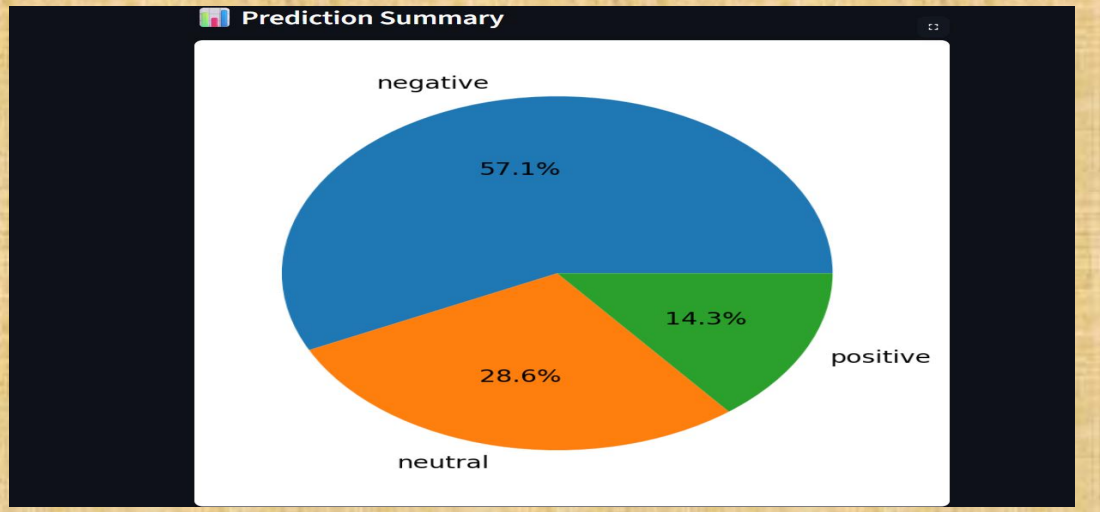
- Text data converted into numerical features using TF-IDF Vectorizer
- Model Used: Multinomial Naive Bayes (best suited for text classification)
- Input Features: Cleaned review text
- Target Labels: Positive, Neutral, Negative
- Libraries Used:
  - Scikit-learn for model training
  - NLTK for text processing
  - Pandas for data handling
- Training performed on a balanced dataset to ensure fair learning

❑ WEB APP USING STREAMLIT

- Built an interactive web app using Streamlit (Python)
- User enters review → app predicts sentiment in real-time
- Displays:
  - Sentiment label + confidence score
  - Pie chart of predictions
  - Recent prediction history
- Extra Features:
  - Save & download history (CSV)
  - Clear history option
- Deployed on Streamlit Cloud

❑ OUTPUT SAMPLE :-





❑ CHALLENGES & SOLUTIONS

• Overfitting Issue:
- Problem: Model performed well on training data but poorly on new inputs
  - Solution: Used TF-IDF + class balancing to improve generalization

• Imbalanced Dataset:
- Problem: More positive reviews than others
  - Solution: Sampled equal data for all 3 classes

• Deployment:
  - Problem: Making model accessible
  - Solution: Used Streamlit + GitHub + Streamlit Cloud

• Text Noise:
  - Problem: Slangs, stopwords, irrelevant words
  - Solution: Applied NLP cleaning and stemming

❑ Conclusion

- Successfully built a sentiment analysis system for product reviews.
- Applied machine learning and NLP to classify reviews into Positive, Neutral, and Negative.
- Created an interactive and user-friendly web app using Streamlit .
- Visualized predictions and stored results with useful features.
- Deployed the app on Streamlit Cloud for public access.

- Future Scope:
  - Use of advanced models like BERT
  - Support for multiple languages

Thank You!