SENTIMENT ANALYSIS

ON AMAZON PRODUCTS
REVIEW DATASET

Problem Statement

E-commerce platforms like Amazon receive an overwhelming volume of customer reviews, making it unfeasible to manually assess the sentiments expressed in these reviews for product evaluation and improvement.

Objective:
To develop a robust sentiment analysis system that utilizes transformer-based models capable of capturing the subtle and complex nature of human language in customer reviews.

Challenge:
Effectively identifying the sentiment whether positive, negative, or neutral embedded in raw, unstructured textual data poses a significant natural language processing challenge.

Proposed Solution

- Use pre-trained transformer models — BERT and RoBERTa — fine-tuned on the Amazon review dataset for sentiment classification.
- Compare the performance of both models in terms of accuracy.

Model Architecture

- Model:
  - This RoBERTa model has a classification head (linear layer on top).
  - It’s a Transformer-based model, pretrained on large corpora.
- Tokenizer: RobertaTokenizer
  - Responsible for converting input text into tokens suitable for RoBERTa.
  - Handles padding, truncation, and special tokens like <s> and </s>.

Data Flow

Preprocessing Steps:

- Lowercasing
- Removing punctuation and numbers
- Removing stop words
- Tokenization using word_tokenize
- Preprocessing is essential to reduce noise before tokenization and modeling.

Dataset Preparation

- Splitting: train_test_split from sklearn
- Encoding: Uses the tokenizer to convert preprocessed text into input_ids and attention_mask.
- Dataloader: Wraps the encoded data in TensorDataset and DataLoader for batching

Training and Evaluation

Training:

- Optimizer: AdamW
- Loss Function: CrossEntropyLoss (default in RobertaForSequenceClassification)

Evaluation

- Evaluation Metrics: accuracy, precision, recall, f1.

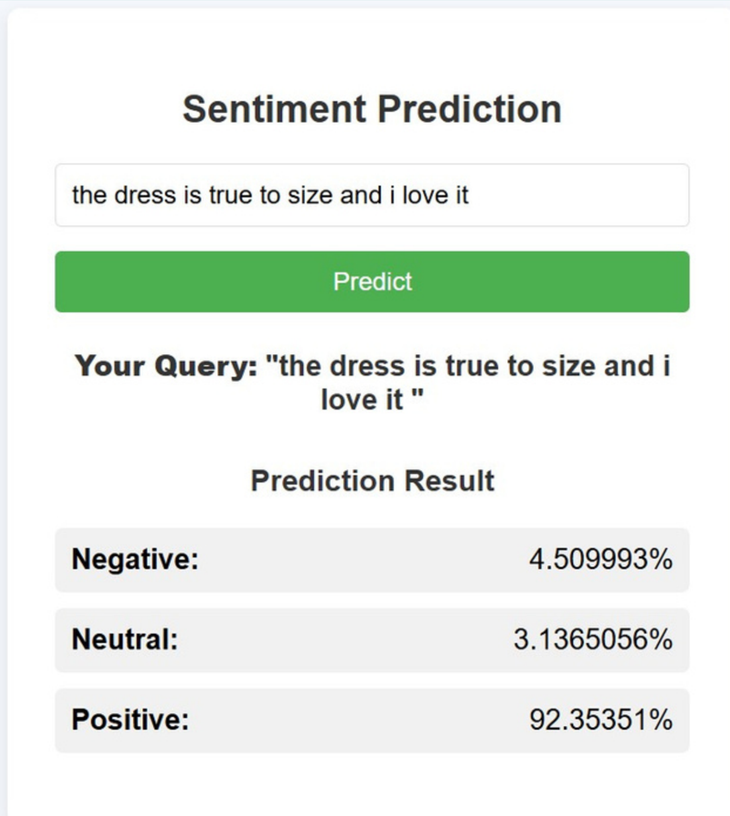
.

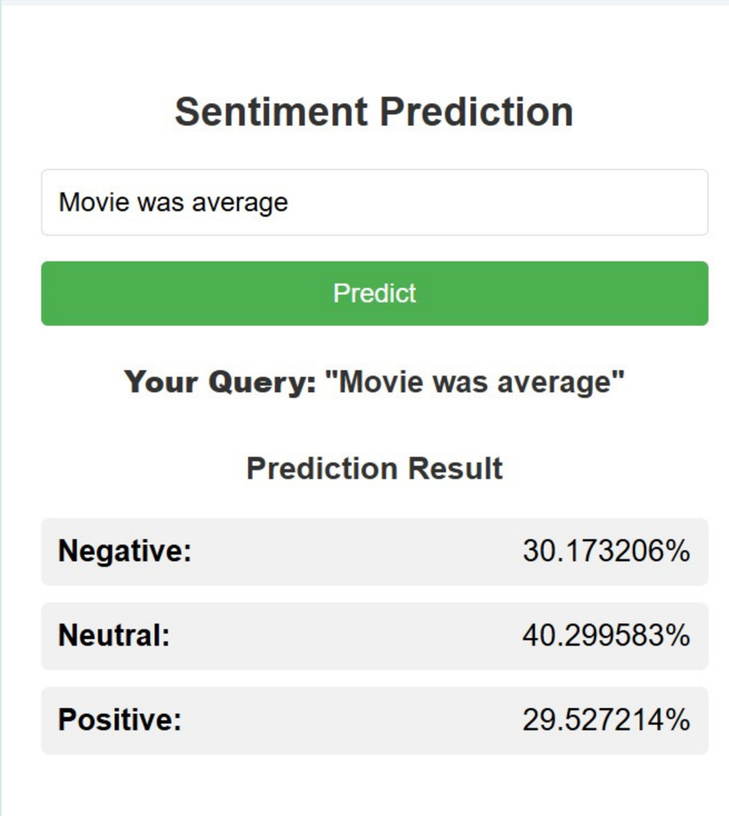
Inference

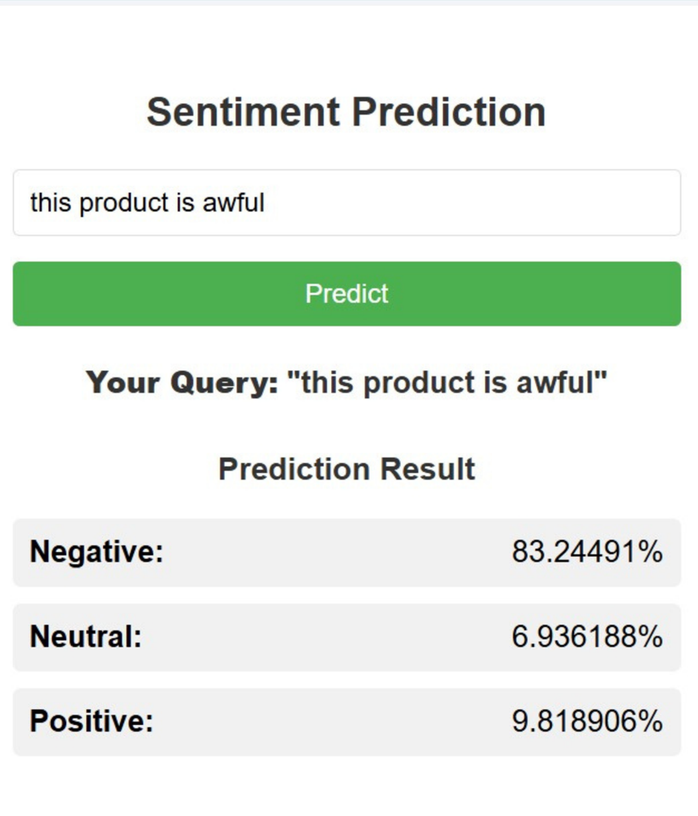
- Takes raw text input
- Applies preprocessing
- Tokenizes and formats for model input
- Returns prediction from the model's whether its positive negative or nureal

.

Examples of the UI predictions







Thank you!