Text Classification and Sentiment Analysis

**Improving Customer Satisfaction through Automated Review Sentiment Analysis**

**Overview**

You are a **product manager** at an e-commerce company like Amazon. Your company receives thousands of customer reviews daily for various products. While ratings are available, they don’t always capture customer sentiment accurately — especially when the text and rating conflict.

**Business Problem:**  
Your goal is to build an automated **Sentiment Analysis System** to classify incoming product reviews as **Positive** or **Negative**. This will help:

* Monitor product performance in real-time
* Identify products with sudden spikes in negative feedback
* Improve customer satisfaction by addressing complaints faster

**Dataset**

* Each row is **one Amazon customer review** for some product.
* The dataset has been **pre-labeled** — so you already know if the review is positive or negative.
* This is a **typical sentiment analysis dataset**, often used for training or testing natural language processing (NLP) models.

**📊 Rows:**

10,000 customer reviews

**📑 Columns:**

1. **label**
   * **Type:** Text (Categorical)
   * **Values:** pos (positive) or neg (negative)
   * **What it means:** The sentiment of the review — whether the customer’s feedback is positive or negative.
2. **review**
   * **Type:** Text
   * **Values:** Actual text of the review written by the customer.
   * **What it means:** What the customer said about the product in their own words — for example: *“Excellent Soundtrack! I love it!”*

**Tasks**

**Data Cleaning:** Remove duplicates, handle missing reviews if any, preprocess text (lowercasing, stopwords removal).

**Exploratory Analysis:** Word clouds, sentiment distribution, most common positive/negative words.

**Model Development:** Use NLP techniques (TF-IDF, Word2Vec, or BERT embeddings) with models like Logistic Regression, SVM, or Neural Networks.

**Validation:** Use train/test split, cross-validation, and metrics like accuracy, F1-score.

**Submission Guidelines**

* Your submission should include a comprehensive report and the complete codebase.
* Your code should be well-documented and include comments explaining the major steps.

**Evaluation Criteria**

* Correct implementation of data preprocessing and feature extraction.
* Accuracy and robustness of the classification model.
* Depth and insightfulness of the sentiment analysis.
* Clarity and thoroughness of the evaluation and discussion sections.
* Overall quality and organization of the report and code.

Good luck, and we look forward to your insightful analysis of the given dataset!