Sentiment Analysis Report – Amazon Customer Reviews

1. Dataset Description

The dataset used for this analysis consists of **Amazon customer reviews** across various products. Each entry contains:

- **Review text** – the written feedback provided by a customer.
- **Rating** – a numerical rating assigned to the product.
- **Date** – the timestamp of when the review was submitted.

This dataset provides valuable information for analyzing customer sentiment and understanding how customers perceive products over time.

---

2. Data Preprocessing

Before performing sentiment analysis, several preprocessing steps were applied to ensure data quality and consistency:

1. **Data Extraction:**
   Extracted the review text column from the dataset.

2. **Handling Missing Values:**
   Removed rows containing missing or null review entries.

3. **Text Cleaning:**

- Converted all text to **lowercase** to maintain uniformity.
- Removed **stop words** (e.g., “the,” “is,” “and”) to focus on meaningful content.
- Eliminated **punctuation, numbers, and non-alphabetic characters** to simplify the text.

4. **Processed Sentences:**
   The cleaned and tokenized text was then stored in a **new column** for further analysis.

---

3. Evaluation of Results

After applying sentiment analysis, the model produced the following outcomes:

- **2 out of 4 reviews** were classified as **positive**.
- **2 out of 4 reviews** were classified as **neutral**.

These results suggest that the sentiment analysis function was **reasonably effective at identifying positive feedback**, though further refinement could improve precision.

---

4. Model Strengths and Limitations

Strengths

- **Efficient and fast** text processing and sentiment scoring.

- **Automated data cleaning** ensures consistent and reliable text input.
- Demonstrates **good accuracy** for general sentiment detection tasks.

Limitations

- **Limited contextual understanding:**
  The model (TextBlob) cannot interpret **sarcasm** or nuanced language.
- **Non-learning model:**
  Since TextBlob is rule-based and not machine learning-driven, it **does not improve with additional data**.

---

5. Conclusion

This project demonstrates the fundamental steps of **text preprocessing** and **sentiment analysis** using a real-world dataset. While TextBlob offers a simple and effective approach for basic sentiment classification, future improvements could involve implementing **machine learning models** (e.g., logistic regression or deep learning with BERT) to enhance contextual understanding and predictive accuracy.

Written by,

Leandro Mendez