Sentiment Analysis Report Using spaCy

Stefan W. Vorster 2025/06/08

1. Introduction

This report outlines the process and results of implementing a sentiment analysis model on Amazon product reviews using the spaCy library with the en_core_web_md language model. The goal was to classify each review's sentiment based on its textual content and compare it with a sentiment inferred from its numeric rating.

2. Dataset Description

The dataset contains Amazon product reviews with the following relevant fields:

- reviews.text: The written content of the review.
- reviews.rating: The numerical rating given by the reviewer (typically on a scale from 1 to 5).
- reviews.username and reviews.date: Metadata used for sorting and identifying reviews.

The dataset is suitable for sentiment analysis as it combines both numerical ratings and user-generated review text.

3. Data Preprocessing

To prepare the dataset for sentiment analysis, the following steps were performed:

- Converted the reviews.date column into datetime format
- Rows with missing reviews.username and reviews.date values were dropped.
- Spam was taken into consideration, but none was found. (By counting all unique usernames).

Text Cleaning:

A custom clean_text() function was applied to the reviews.text field to:

- Convert to lowercase
- Remove punctuation and special characters
- Remove extra whitespace

Sentiment Labeling

Two sentiment labels were used:

- Rating-Based Sentiment:
    - *Negative*: Ratings 1–2
    - *Neutral*: Rating 3
    - *Positive*: Ratings 4–5
- Predicted Sentiment (from spaCy):
    - Each cleaned review was processed using the en_core_web_md spaCy model, and a rule-based approach was applied to classify its sentiment as positive, neutral, or negative.

4. Evaluation of Results

Confusion Matrix

The confusion matrix below compares the predicted sentiment (from text) against the sentiment inferred from the rating:

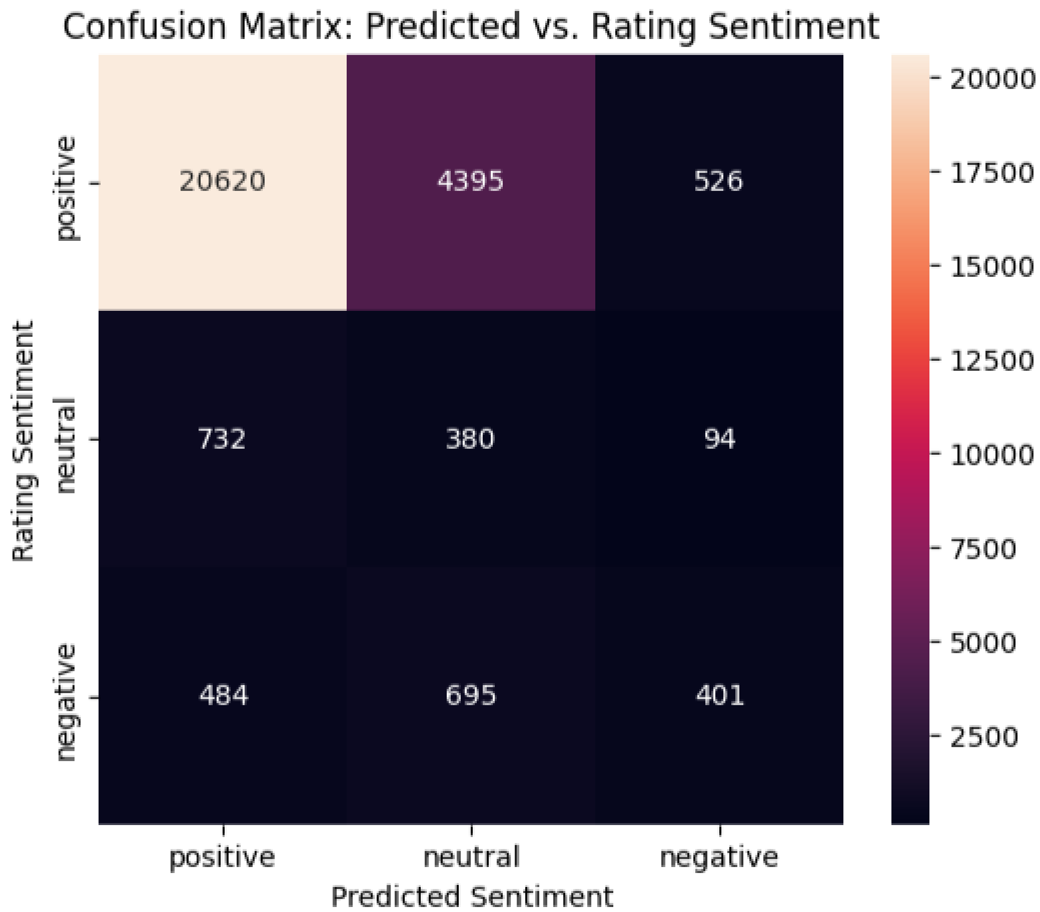

Figure 1: Confusion matrix of Rating Sentiment vs Predicted Sentiment.

Table 1: Precision, Recall, F1-Score and Support values of Sentiments.

Sentiment	Precision	Recall	F1-score	Support
Positive	0.39	0.25	0.31	1580
Neutral	0.07	0.32	0.11	1206
Negative	0.94	0.81	0.87	25541
Macro Avg	0.47	0.46	0.43	28327
Weighted Avg	0.88	0.76	0.81	28327

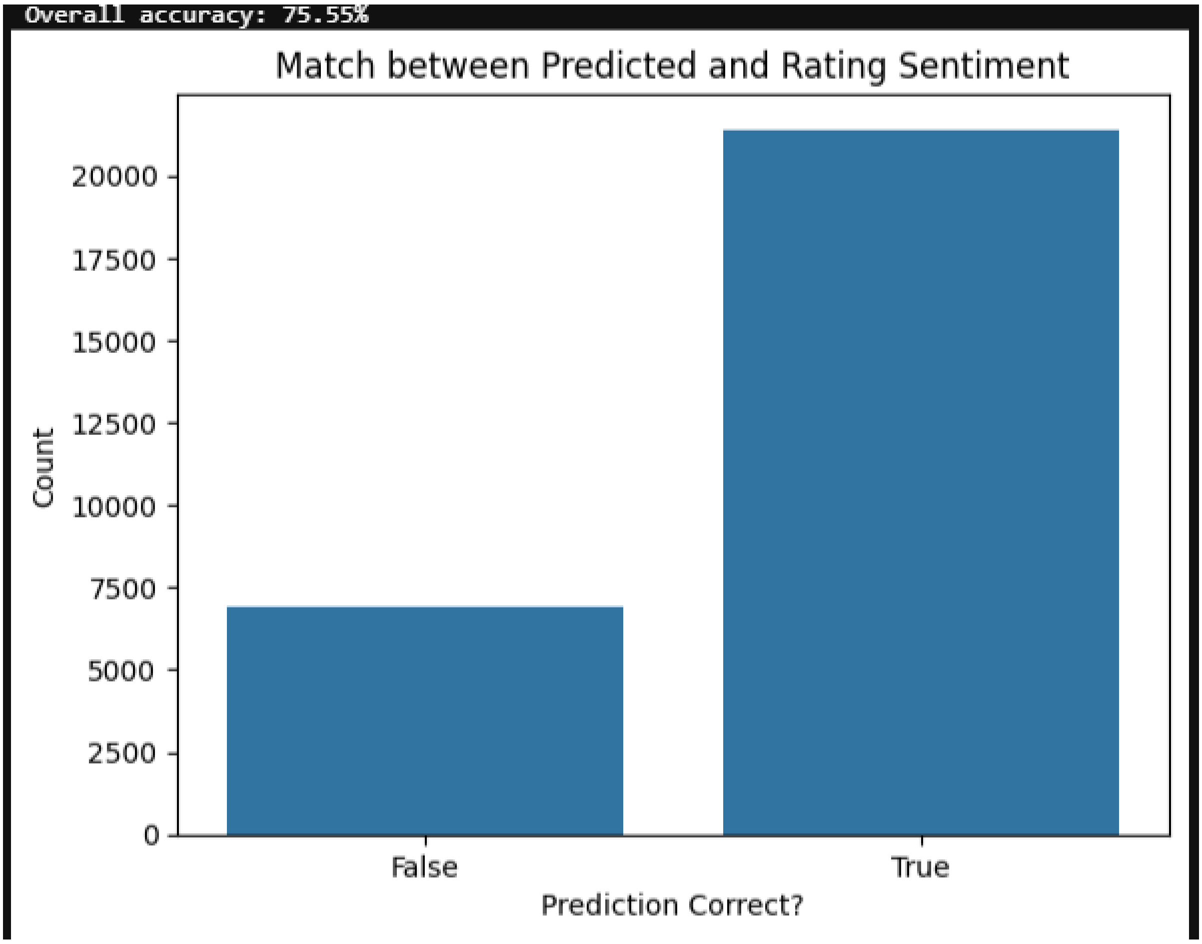

Figure 2: Count of True and False correct predictions.

5. Discussion

The sentiment analysis model implemented using spaCy (en_core_web_md) was evaluated by comparing the predicted sentiment labels from review text with actual sentiment labels derived from review ratings. The overall accuracy achieved was approximately 75.55%, indicating a generally strong performance, especially in identifying negative sentiment—with a precision of 0.94, recall of 0.81, and F1-score of 0.87. However, the model performed poorly on positive and neutral reviews, showing low precision and recall scores in both classes. This performance imbalance suggests that the model tends to over-predict negative sentiment, possibly due to bias in the wording of reviews or an imbalance in the dataset where negative reviews dominate or are more linguistically distinct.

To improve the model's performance, particularly on positive and neutral sentiments, it is recommended to first ensure class balance in the dataset through techniques like oversampling or stratified sampling. Additionally, enhancing the training set with labeled examples of positive and neutral reviews could help the model learn subtler distinctions. Incorporating more advanced models, such as fine-tuned transformer-based architectures (e.g., BERT), could also yield better performance due to their deeper contextual understanding. Lastly, using techniques like cross-validation, error analysis, and adding domain-specific sentiment lexicons might further refine prediction quality.