# Sentimental Analysis for Amazon Reviews on Beauty Products

Anwesha Sinha, Armando Miranda, Enoch Yang, Protima Tarafdar
University of Colorado, Boulder
ansi5532@colorado.edu armi8047@colorado.edu
enya0577@colorado.edu prta8577@colorado.edu

## Abstract

*This project covers sentiment analysis of Amazon reviews on products in the beauty category. To analyze the sentiment of the reviews and provide insights into consumer experiences, the team used a Long-Short Term Memory (LSTM) model. The dataset was taken from Kaggle and went through several cleaning processes, including text preparation and the removal of null values. The LSTM model was used to determine market sentiment, track consumer satisfaction, and evaluate product performance and obtained an accuracy of 89%. To improve business outcomes, the project suggests customer management, market analysis, and product creation. Furthermore, the team suggests aspect-specific sentiment analysis and emotional analysis as future steps to improve the analysis's depth, as well as cross-language analysis and social media application to broaden its scope.*

## 1. Introduction

Online reviews from customers have increased as a result of the development of e-commerce platforms. These reviews offer a wealth of knowledge regarding the thoughts and feelings of customers toward goods and services. Sentiment analysis, a branch of natural language processing, has become a well-liked study subject for examining customer reviews and drawing conclusions from massive amounts of text data. Neural networks, particularly Long Short-Term Memory (LSTM) models, have demonstrated promising results in sentiment analysis applications in recent years.

In this study, we use LSTM models to do sentiment analysis on Amazon product reviews. Consumers have access to a large variety of items in the beauty sector, which is a highly competitive market. Determining consumer attitudes toward beauty products is therefore essential for brand management, marketing, and product development. We want to get insights into customer sentiment about beauty items by utilizing the capabilities of LSTM neural networks and to give a thorough analysis of the performance of these models on this particular sort of text data.

The results of this study may have important ramifications for companies in the beauty sector, enabling them to make wise decisions based on client input and raise client happiness.

### 1.1. Problem Statement

Our project involves analyzing Amazon reviews for the beauty product category. We aim to teach our model how to distinguish between positive and negative reviews by performing sentiment analysis. By identifying the most frequently used words in both positive and negative reviews, companies can improve their brand differentiation and work on areas where the product needs improvement.

However, using strings as input requires multiple steps for pre-processing, which we must overcome to ensure our model receives the right data to draw actionable insights. Our data preparation procedure will be covered in more detail below.

It can be frightening for a business to launch new products. This is why these decisions need to be as grounded as possible and be made with data as a center piece. Data willingly provided by customers on their experience is a great way to use readily available sources to drive our business decision making.

In this report, we will be diving deeper into the business problems we are encountering, the methodology followed, and the business insights attained throughout the process.

### 1.2. Business Case

This project's goal is to find out how businesses can use sentiment analysis based on Amazon reviews to learn more about consumer experiences. Customer reviews of Amazon products support the implementation of this project. A company can learn more about consumer perceptions of particular goods and services by reading Amazon product reviews.

This will give businesses the ability to identify the factors behind both favorable and unfavorable customer feedback and put in place efficient strategies to address them appropriately. The assignment assists businesses in using Amazon reviews' sentiment analysis to comprehend customer experiences.

We are trying to achieve a set of goals like tracking the sentiments of the customer vs time, determining which customer segments have the strongest opinions, and gap analysis by identifying the strengths and weaknesses based on positive and negative reviews for strategic placement of the product in the market, determining effective ways of advertising and communication, and prioritizing customer service issues.

## 2. Related Work

Inspired by previous research from the paper Sentimental Analysis for Amazon.com Reviews, Levent G¨uner [1], the research was implemented on Amazon reviews for the beauty category. Three models were performed, in which the best forming model is LSTM whose accuracy is 90% and AUC is 0.96. With LSTM being their best model, they have implemented two other machine learning models which are Multinomial Naïve Bayes and Linear Support Vector Machine. It was also mentioned that this same application can be applied to different categories. With reference to this research paper, our main focus was to implement neural networks for the sentimental analysis of Amazon reviews focusing on beauty products to improve business outcomes, particularly in this industry.

## 3. Data

Amazon Customer Reviews (also known as Product Reviews) is one of Amazon's most recognizable products. Since the first review in 1995, millions of Amazon customers have contributed over a hundred million reviews to express their opinions and describe their experiences with products on the Amazon.com website over a two-decade period. This makes Amazon Customer Reviews a valuable resource to work on Natural Language Processing (NLP), Information Retrieval (IR), and Machine Learning (ML), among other domains.

This dataset was published on a platform called Kaggle. It was originally created specifically to reflect a sample of customer evaluations and opinions, variance in product perception across geographical locations, and promotional purpose or bias in reviews.

The above mentioned Kaggle notebook contained multiple datasets on all sorts of Amazon-sold products' reviews. We think that as a business one of the most uncertain operations is new product launches. Particularly in the beauty industry where products evolve and new trends emerge, making this a very competitive and lucrative industry. This is why, from among all the product categories we decided to utilize the beauty products data and business problem for this project.

### 3.1. Data Description

The dataset pulled from Kaggle on beauty products is metadata that comprises 5113668 unique instances and 15 features. The size of this data is around 22GB. Fifteen features mainly describe the following –

| Feature Name | Description |
|---|---|
| Country Code | Code of the country in two letter formats from where the review was written |
| Customer ID | Unique ID of a customer |
| Review ID | Unique ID of the review |
| Product ID | Unique ID of the product. Product ID could be the same at different geographic locations. |
| Product Parent | A unique number assigned to every product. |
| Product Title | The title of the product |
| Product Category | Defines the product's category or bucket |
| Star Rating | Product is given a rating between 1 to 5 |
| Helpful Votes | Number of helpful votes |
| Total Votes | Number of total votes they received |
| Vine | Review that was written as part of the Vine program |
| Verified Purchase | Indicating that the review is on a verified purchase |
| Review Headline | The title of the review |
| Review Body | Entire review received |
| Review Date | The date the review was written |

Table 1: Data Dictionary

### 3.2. Data Cleaning

Before proceeding to model, data cleaning operations were performed in such a way that the model for sentimental analysis supports it. The operations include as follows –

Those features that contained text were converted into string format so that the NLP model can recognize them as text. These features are:

- Product title
- Review body
- Review headline

Moreover, we have standardized the text included in these columns by lowercasing all its letters, removing special signs and symbols, removing text contractions, removing stop words, among other text standardizing steps to ensure it all followed the same standards for proper analysis.

It was observed that there were a few missing values throughout the different data rows. Given their small quantity relative to the large data size, we decided to get rid of those rows which contained null values. No significant changes to our data integrity were reported.

Finally, we had a column labeled ‘Star Rating’ which contained the star rating corresponding to the product they bought. Such star rating went from one to five (1-5). For a more streamlined labeling and sentiment declaration, we created a new column called ‘sentiment’ which will contain only two values one if the sentiment is positive or zero if the sentiment is negative (1,0). The selected threshold to determine if a review was positive or negative was three. This means that if the review’s star rating was more than three, the review would be labeled as positive, otherwise, negative.

### 3.3. Preliminary Analysis

Getting comfortably acquainted with the data we have at hand is paramount to have a proper base for the subsequent phases which involve a wholistic understanding of the data and how we envision we could use it. With the aforementioned features, we had to identify patterns in our data and what they could mean for our preliminary analysis.

As our main task was to perform sentiment analysis in Amazon beauty product reviews, we wanted to see what the lengths of these reviews were and what their distribution was in terms of length.


Figure 1: Positive Sentiment Word Cloud

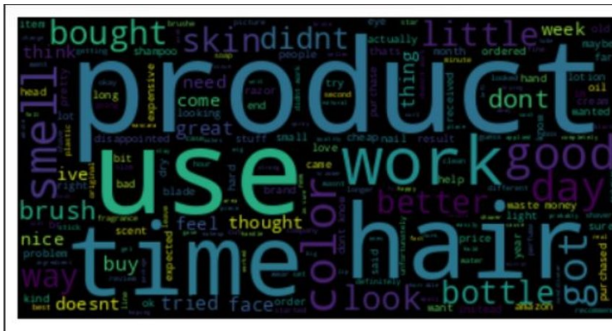


Figure 2: Negative Sentiment Word Cloud

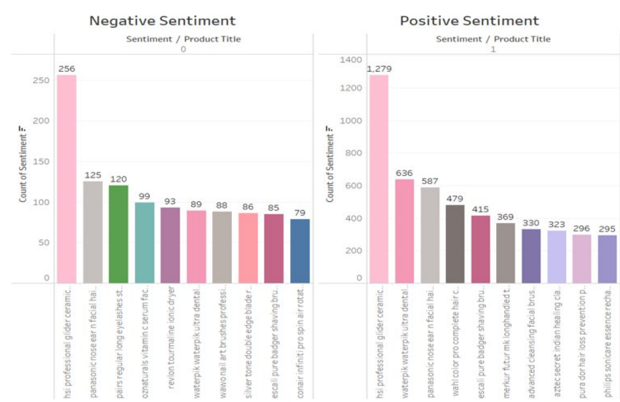


Figure 3: Top 10 Products by Sentiment

By creating a histogram, we realized that most of the reviews were on the sorter spectrum, mainly falling within a four hundred and below number of characters. This left skewness gave us an idea of what length we should specify when working on our LSTM model for tokenizing and ensure the chosen parameter englobed most of the reviews in their entirety.

## 4. Methodology

Recurrent neural networks of the Long Short-Term Memory (LSTM) variety have demonstrated promise in the field of sentiment analysis for natural language processing jobs. LSTM models are used in this project to do sentiment analysis on Amazon reviews of beauty products.

This study aims to investigate the effectiveness of LSTM models in categorizing the sentiment of customer evaluations, to pinpoint the advantages and disadvantages of this methodology, and to assess its potential as a tool for product development and marketing in the cosmetics sector.

To have a more objective view of our results, we have compared its output with the insights gained from the preliminary and exploratory data analysis.

We have prepared the data so that the relevant features would be in the proper shape for it to be fed into the LSTM model. The model was subsequently fitted with the expected data and its parameters have been fine-tuned to yield the best results for both generalization and proper fit. The results of our procedure are to being reported in the subsequent sections.

## 4.1. LSTM Framework

Four primary parts make up the framework that is employed for this project. they are raw text, tokenizer, LSTM, and sigmoid. Natural language processing (NLP) tasks like sentiment analysis and text categorization frequently employ this paradigm.

The input text data that needs to be analyzed is the first component, which is referred to as raw text. This input data may take the shape of tweets, reviews, or news pieces, among others. But first, this text needs to be transformed into a format that the LSTM model can use before proceeding to the analysis.

The tokenizer steps in at this point. The tokenizer is a process that divides the raw text into more manageable chunks, like words or characters. These more manageable pieces, referred to as tokens, can be put into a neural network for additional processing.

The LSTM, or long short-term memory, is the following element. Recurrent neural networks (RNNs) of the LSTM variety are made to process data sequences including text, speech, and video. Because it can handle variable-length inputs and remember long-term dependencies, LSTMs are especially helpful for NLP tasks.

A sigmoid activation function is then applied to the LSTM layer's output. In binary classification problems, the sigmoid function is frequently employed to convert a neural network's output into a probability score between 0 and 1. In this instance, the sigmoid function is employed to determine whether the input text is likely to be positive sentiment or negative sentiment.

Overall, this model architecture is a potent tool for studying text data and drawing conclusions from it. It enables us to create precise sentiment analysis models for a range of applications, including social media monitoring, market research, and analysis of customer comments.

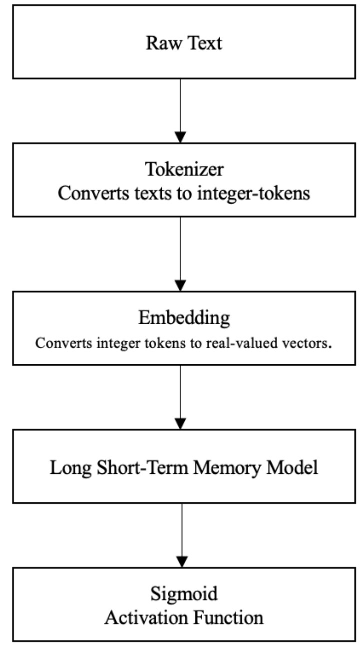


Figure 4: Model Architecture for LSTM

## 4.2. Model Specifications

The feature review body that contains the entire review text, is a raw text as described in the model architecture. This is tokenized using a tokenizer into a numerical format/sequence to create the input data for the LSTM model. Post that, padding is included to keep the sequence's length constant, up to a maximum length of 100. Due to the LSTM model's requirement that inputs have a constant size, this is important.

Next, an embedding layer is added to transform the integer tokens into dense sequence vectors. Each unique token is translated to a vector in a continuous space by the embedding layer. The input length is set to 100, the vector length is set to 32, and the number of different words is set at 5000. This implies that the input sequence will have a length of 100 and that each token will be represented by a 32-dimensional vector.

The 32-size LSTM model is constructed. With the help of the sigmoid activation function, the neural network's output is converted into a probability score between 0 and 1. Binary cross-entropy, which calculates the difference between the predicted and actual labels, is the utilized loss function. Adam is the optimizer, which is a well-liked neural network optimization algorithm.

The model is trained for 5 epochs with a batch size of 64. The number of epochs dictates how many times the complete dataset is used for training, whereas the batch size specifies how many samples are handled at once.

The architecture offered by these model specifications enables precise sentiment analysis. The model can handle variable-length inputs and capture long-term dependencies since it makes use of LSTM neural networks and embedding layers. In addition, selecting the right loss function, optimizer, batch size, and number of epochs is crucial to maximizing the model's performance. These requirements offer a solid foundation for sentiment analysis tasks across a range of applications, such as social media monitoring, market research, and customer feedback analysis.

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.80 | 0.66 | 0.72 | 21,686 |
| 1 | 0.91 | 0.95 | 0.93 | 78,313 |
| Accuracy | | | 0.89 | 99999 |
| Macro Avg | 0.85 | 0.81 | 0.83 | 99999 |
| Weighted Avg | 0.89 | 0.89 | 0.89 | 99999 |

Table 2: Confusion Matrix

## 5. Results

Precision, recall, F1 score, and accuracy are the measures used to gauge the effectiveness of the LSTM model for sentiment analysis. These metrics give information on how well the model can identify positive or negative sentiments in the input text.

Precision is the ratio of correctly predicted positive outcomes to correctly predicted negative outcomes. In this instance, the precision score for the LSTM model was 0.91 for positive feelings and 0.80 for negative sentiments. This indicates that 80% of the model's negative forecasts were accurate while 91% of its positive predictions came true.

Recall is the proportion of true positive cases that were correctly predicted to be true as well as the proportion of true negative cases that were correctly projected to be true. The recall score for the LSTM model was 0.66 for unfavorable sentiments and 0.95 for favorable sentiments. This indicates that the model accurately recognized 66% of the negative cases and 95% of the positive cases, respectively, out of all the negative cases.

A single metric for assessing the performance of the model is provided by the F1 score, which is a weighted average of precision and recall. The F1 score for the LSTM model was 0.72 for unfavorable sentiment and 0.93 for favorable sentiment.

The proportion of all predictions that are accurate is known as accuracy. The LSTM model had an overall accuracy score of 0.89, which meant that 89% of the input text was correctly categorized as having a positive or negative sentiment by the model.

These evaluation metrics show that using the LSTM model to perform sentiment analysis is a good strategy overall. Positive attitudes are particularly well predicted by the model, although negative sentiments may be better predicted.

## 6. Analysis & Recommendations

Review sentiment analysis has significant business implications for companies that rely on customer feedback to inform their operations and decision-making. Three such implications are customer management, market analysis, and product development. Among each of these implications are many recommendations for action.

For customer management, companies can identify areas where their customers are satisfied or dissatisfied. They can then take steps to address customer complaints, improve their services, and ultimately improve customer satisfaction. Additionally, negative reviews can heavily damage a company’s reputation. By identifying and addressing customer complaints, companies gain the ability to prevent negative reviews from propagating, and they can protect their brand image.

In the case of market analysis, positive reviews can be used in marketing and advertising campaigns to highlight a company’s strengths and build brand loyalty. They can also be used as a gauge for the general market sentiment and even competitor products’ sentiments. These sentiments, when identified, can be used to benchmark products against each other and the general market.

Lastly, product development could benefit from sentiment analysis because it can provide insights into what customers like or dislike about a certain product, as well as identify areas for improvement. These valuable opinions could help dictate the direction and purpose of development. To add to that, sentiment analysis is a method for evaluating product performance, whether it be older products or new releases.

Sentiment analysis can have a positive impact on various aspects of a company’s operations. By aiding in customer management, allowing market analysis, and driving product development, it can be seen as a powerful tool that can help companies succeed in a competitive business environment.

## 7. Future Action

After sentiment analysis, there are several directions that we could recommend, depending on the goals and objectives of a company. Many of these directions could be explained by increasing the depth of the model or increasing the breadth of the model.

In increasing the depth of the model, it allows companies to get more specific and higher quality analysis. One way to increase depth is to expand the analysis from negative and positive sentiment to emotional analysis. More powerful emotions such as anger or surprise may reveal significant differences from a binary negative/positive analysis. Another method for increasing depth is to analyze aspect-specific sentiments. The goal of this method is to extract sentiments from specific aspects or features of a product or service.

Increasing the breadth of the model involves increasing the volume of applications. One current limitation is that it the model functions only for reviews written in English. By including more languages such as Chinese or Hindi, the model can expand its domain of application. Another limitation of the model is that it was trained on Amazon beauty products. Applying similar training and analysis to different product markets is one step towards increasing breadth. Another step would be to expand such analysis toward the social media reception of a product.

Whether it be increasing specificity through depth or application through breadth, both pathways offer an immense increase to the potential value of a sentiment analysis model.

## References

[1] Coyne, E., Smit, J., & Güner, L. (2019, March). *Sentiment Analysis for Amazon.com reviews*. ResearchGate. Retrieved May 1, 2023, from https://www.researchgate.net/profile/Levent-Guener/publication/332622380_Sentiment_analysis_for_Amazoncom_reviews/links/5cc08696a6fdcc1d49acb839/Sentiment-analysis-for-Amazoncom-reviews.pdf