Fine-Grained Sentiment Analysis of the Amazon Reviews Dataset

Andrew Bauman, Allison Nowakowski

Marist College

Fall 2019

# Abstract

For our final project, we seek to explore Amazon Customer Review Datasets to see if we can create a model to create a five-star product review system truly reflective of a customer’s feelings related to a given product. Aligning our process to the CRISP-DM model, we seek to employ machine learning algorithms and neural networks to understand, prepare, model, and evaluate the data to reach a conclusion.

# Introduction

The five-star rating system, a well-known review methodology, has been adopted and implemented throughout many industries including e-commerce, hospitality, and healthcare to help accurately capture the feelings a customer associates with a product or service. While it stands to be appear a simple and efficient way of doing so, it is quite simple to misrepresent or skew how said product or service is viewed by potential consumers. Amazon, headquartered in Seattle, Washington and perhaps the world’s largest online retailer6, employs the use of the five-star rating system in an interesting way likely with the hope of mitigating or reducing the misrepresentation the rating system may have. According to its own website, it is explained that:

Amazon calculates the star ratings of a product using machine-learned models instead of a simple average. These models take into account numerous factors including how recent the rating or review is, verified purchase status, and use multiple criteria that establish authenticity of the feedback. The system continues to learn and improve over time.

We do not consider customer ratings without an Amazon Verified status in products' overall star rating until a customer adds more details in the form of a text, image, or video.3

While this statement does not confirm or deny that natural language processes are used in determining a final star value, it certainly alludes to it. We feel that such a process is an interesting method of summarizing presenting past customer experiences to potential ones and seek to replicate the model and test its accuracy, comparing it with other state-of-the-art methods.

# Data Mining Task

Aligning our process to the CRISP-DM model, it is our goal to properly understand the data and the context surrounding it to most effectively prepare, model, and evaluate it with the overall goal of creating and presenting information. For this project, we will employ natural language processing and machine learning methodologies, including neural networks, to evaluate whether or not Amazon’s five-star rating systems are truly representative of a customer’s experience and opinions related to the products they have consumed.

# Data Set

In order to complete our task, we will employ the use of the *Amazon Customer Reviews* Dataset made public by Amazon itself. Containing over a hundred million reviews in multiple languages from varying Amazon marketplaces, written by millions of Amazon customers since the first review in 1995, this product review dataset “was constructed to represent a sample of customer evaluations and opinions, variation in the perception of a product across geographical regions, and promotional intent or bias in reviews”4. The dataset is quite robust and contains the following attributes, in a tab (‘\t’) separated text file (copied from the original documentation):

DATA COLUMNS:

marketplace - 2 letter country code of the marketplace where the review was written.

customer\_id - Random identifier that can be used to aggregate reviews written by a single

author.

review\_id - The unique ID of the review.

product\_id - The unique Product ID the review pertains to. In the multilingual dataset the

reviews for the same product in different countries can be grouped by the

same product\_id.

product\_parent - Random identifier that can be used to aggregate reviews for the same product.

product\_title - Title of the product.

product\_category - Broad product category that can be used to group reviews

(also used to group the dataset into coherent parts).

star\_rating - The 1-5 star rating of the review.

helpful\_votes - Number of helpful votes.

total\_votes - Number of total votes the review received.

vine - Review was written as part of the Vine program.

verified\_purchase - The review is on a verified purchase.

review\_headline - The title of the review.

review\_body - The review text.

review\_date - The date the review was written.

DATA FORMAT

Tab ('\t') separated text file, without quote or escape characters.

First line in each file is header; 1 line corresponds to 1 record.4

In our project, we plan to make use of *review\_id, product category, star\_rating, helpful\_votes, total\_votes, verified\_purchase, review\_headline, and review\_body.*

# 

# Methods and Models

In order to successfully work with the dataset and create useful information, we will use Python libraries including Numpy, Pandas, and SciKit-Learn to properly transform the data and process it in order to create the model. To begin, we plan to work with reviews exclusively made by customers with *Verified Purchases* and separate them by their star-rating groups. From there, we aim to tokenize, lemmatize, and stem the reviews in order to perform a sentiment analysis on them, planning on classifying them as being one of the five-star values. Before feeding the model, we plan to separate the reviews into training and test datasets in order to properly train and examine the model’s accuracy.

# Assessment

After our model was created and the sentiment analysis had successfully been performed, we were able to achieve an accuracy of 55% with Logistic Regression. We find this accuracy to be satisfactory, especially when compared to newer, state-of-the-art methods including BERT and DPCMM Models which were only able to achieve 65.83% and 65.19% accuracy respectively.7

# Presentation and Visualization

We visualize the data, as well as the model, in the provided PowerPoint presentation through the use of histograms and screenshots.

# Bibliography

1. “Amazon Customer Reviews Dataset,” n.d.
2. “About Customer Reviews.” Amazon. Goettsche Partners, 2010. <https://www.amazon.com/gp/help/customer/display.html?nodeId=201967050>.
3. “About Product Star Ratings.” Amazon. Goettsche Partners, 2010. <https://www.amazon.com/gp/help/customer/display.html/ref=?nodeId=G26JKQMMBXWFT6D7>
4. “Amazon Customer Reviews Dataset: README.” Amazon Customer Reviews Dataset. Accessed November 6, 2019. <https://s3.amazonaws.com/amazon-reviews-pds/readme.html>.
5. Papreckis, Paulius. “5 Star Rating Systems Are Finished: What's the Way Forward?” Medium. Prototypr, December 17, 2018. <https://blog.prototypr.io/5-star-rating-systems-are-finished-whats-the-way-forward-b02591439185>.
6. Rouse, Margaret, Margaret Rouse, and Margaret Rouse. “What Is Amazon?” WhatIs.com. Accessed November 6, 2019. <https://whatis.techtarget.com/definition/Amazon>.
7. “Sentiment Analysis on Amazon Review Full” Paperswithcode.com. Accessed November 29, 2019. <https://paperswithcode.com/sota/sentiment-analysis-on-amazon-review-full>.