**Project Report for Stage 2**

**Los Ancianos**

William Hanson

CS 487

New Mexico State University

Las Cruces, NM USA

[billh@nmsu.edu](mailto:billh@nmsu.edu)

Sanford “Jay” Johnston

CS 487

New Mexico State University

Las Cruces, NM USA

[sanford.johnston@gmail.com](mailto:sanford.johnston@gmail.com)

Eloy “Frank” Macha

CS 519

New Mexico State University

Las Cruces, NM USA

[efmacha@nmsu.edu](mailto:efmacha@nmsu.edu)

**1 Introduction and Motivation.**

It’s no secret that the internet has become one of the most important ways that people use to both perform research and purchase goods and services. One of the key tools that potential buyers use is online reviews. An item with an overwhelming number of positive reviews has the perception of being a high-quality product and will attract more potential buyers. On the other hand, an item with mostly negative reviews will be much less likely to attract potential buyers. Similarly, online reviews are one way for a company to receive user feedback on products. As noted in a recent Forbes Technology Council post, 97% of customers use online reviews as a way to find businesses. [1]

Accordingly, companies pay a lot of attention to online reviews. A very positive review may receive a note of thanks from the seller. A very negative review will often receive an apology and an offer to correct the problem.[[1]](#footnote-1) By providing an additional feedback mechanism, reviews can be used to guide product redesign efforts, or even whether to continue selling that product.[2] Therefore, being able to accurately locate and assess both positive and negative reviews is important for business.

However, it is not uncommon to see reviews where the “star” or numerical rating does not match the actual review feedback. A 1-star (poor) rating may have glowing positive feedback, and vice versa. [3] These mis-marked feedback comments make the task of discovering and properly evaluating user feedback more difficult, possibly resulting in faulty business decisions.

**2 Problem Definition**

With many thousands, or even millions of online reviews on sites like Amazon.com, discovering mis-marked feedback via manual review would be untenable for most businesses with significant numbers of products for sale. The problem at hand is to discover user feedback that does not match the numerical or star rating. Put another way, the challenge is to accurately classify online review text as either positive or negative, and then compare that classification with the numerical rating to discover mis-matched or anomalous reviews.

**REFERENCES**

[1] Cory Capoccia. 2018. “Online Reviews are the Best Thing That Ever Happened to Small Business.” *Forbes Technology Council*. April 11, 2018. [https://www.forbes.com/sites/forbestechcouncil/2018/04/11/online-reviews-are-the-best-thing-that-ever-happened-to-small-businesses/](https://www.forbes.com/sites/forbestechcouncil/2018/04/11/online-reviews-are-the-best-thing-that-ever-happened-to-small-businesses/#2eb45c63740a)

[2] Lei Zhang, Xuening Chu, & Deyi Xue. 2018. “Identification of the To-be-improved Product Features Based on Online Reviews for Product Redesign.” *International Journal of Production Research* 57, 8 (Sep 2018), 2464-2479. DOI: <https://doi.org/10.1080/00207543.2018.1521019>

[3] Nishit Shrestha & Fatma Nasoz. 2019. “Deep Learning Sentiment Analysis of Amazon.com Reviews and Ratings.” International Journal of Soft Computing, Artificial Intelligence and Applications 8, 1 (Feb 2019), 1-15. <http://aircconline.com/ijscai/V8N1/8119ijscai01.pdf>

1. One of the authors bought two separate items online recently. Both arrived inoperative. After posting negative reviews, the author was offered both a refund and a replacement item. [↑](#footnote-ref-1)