

# Fakeology – A Fake Review Detection Web Application for Amazon Products

Summer 2021 Capstone Colloquium

Presenter: Zach Shim | Faculty Advisor: Min Chen | Major: CSSE



## Introduction

- Many products have fake reviews.
- Fake reviews can sway consumer purchase decisions.
- Many sellers will manipulate the reviews on their products to deceive people.

## Objectives

- Allow a user to paste links to Amazon product pages.
- Crawl Amazon for relevant review data.
- Feed scraped data into machine learning models for analysis.
- Create a machine learning pipeline that analyzes product data and returns results to the user.

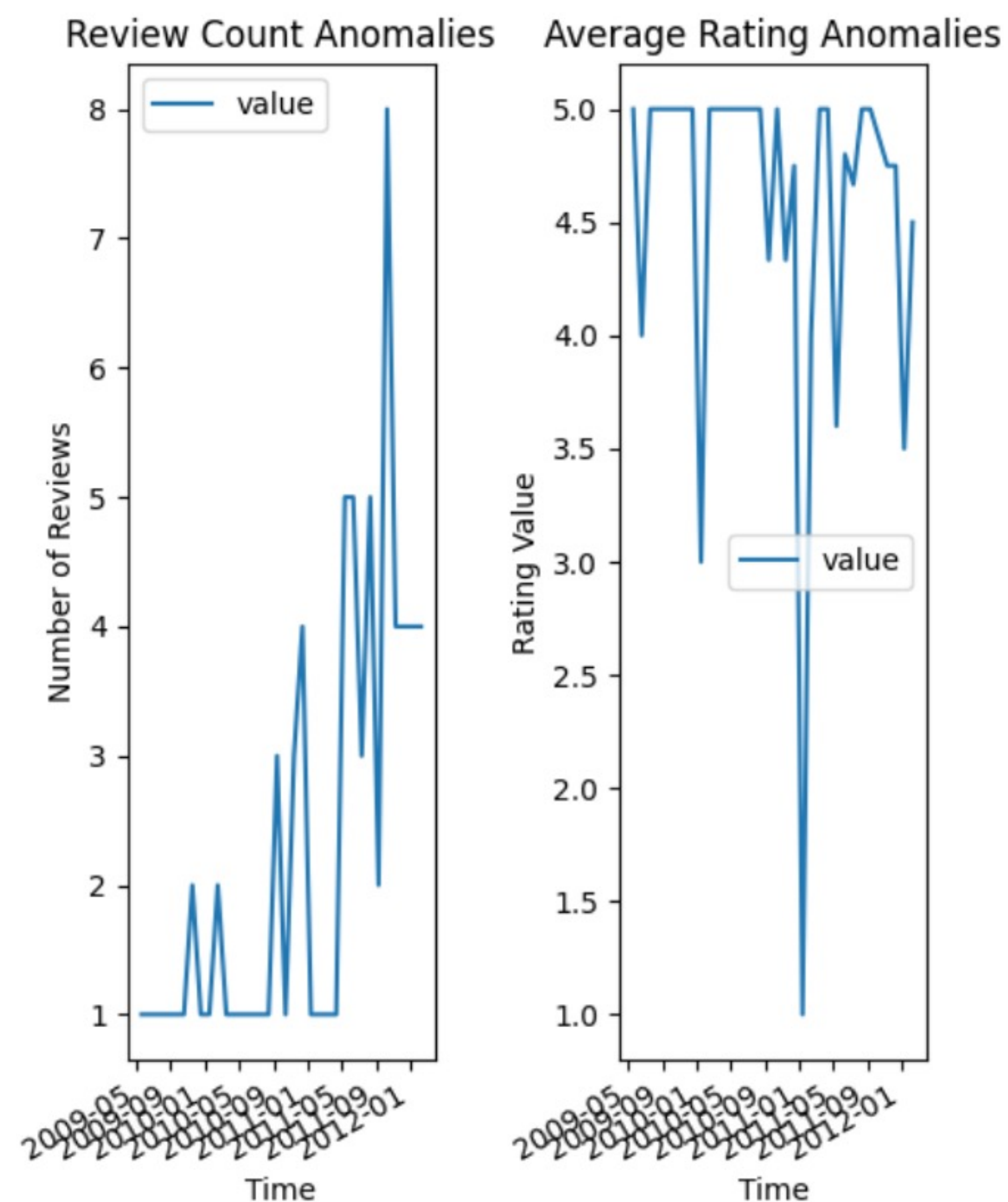
## Methodology

- Natural language processing to preprocess data (stop words, lemmatization, removing punctuation).
- Chain vector transformations to train semantic analysis models.
- Cosine similarity heuristic for an unsupervised similarity index.

## Data Analysis

Review analysis is a three-stage process:

- A. Detection of similarity.
- B. Detection of sentiment.
- C. Detection of anomalies.



## User Interface

- Web-based interfaces that allows users to either see reviews that have already been analyzed or paste product links.
- Amazon web crawler scrapes product review data from a link.
- Ajax requests for chained autofill search bar.

## Screenshots

### Search Analyzed Products

No need to wait for an analysis! Search our repository of analyzed products to see past scores.  
Enter an asin into the search bar to get started

Automotive  Search

- B00002243X
- B00002243Z
- B00008BKX5
- B00008RW9U
- B00008RW9V
- B00009RAX7
- B00009W3G7
- B0000ATZD5

### Analyze an Amazon URL

Paste a link to an Amazon product to analyze it for fake reviews.

Search

\* This field is required.

## Future Work

- Implement new analysis methodologies (e.g., images, videos, emojis, etc.)
- Improve web scraper. (e.g., add proxy cycling service, captcha solver, etc.)
- Distributed clustering for similarity detection.

## Related Coursework

- CSS 342 & 343
- CSS 360 & 370
- CSS 475
- CSS 480

## Tools & Technologies

