Jonathan Weisskoff

Dec. 9, 2020

Abstract of Capstone

Amazon Fake Reviews Detector

Position: Project Manager, creator of copies exist elsewhere indicator

Myriads of products sold on Amazon are illegitimately reviewed. A fraudulent review can be hard to detect. Even reviews which have a "verified purchase" certificate may be plagiarized. See here for studies about how sellers on Amazon solicit fraudulent reviews.

This project is an experiment in trying to detect fraudulent reviews on Amazon. We provide a website with a form for the user to input an Amazon product URL. Upon submit, the site displays a score indicating how likely the product has fraudulent reviews. To detect the fraudulent reviews, we use indicators of fraud, some of which are not employed by extant commercial fake review detectors so as to strive for a more accurate result.

Indicators used:

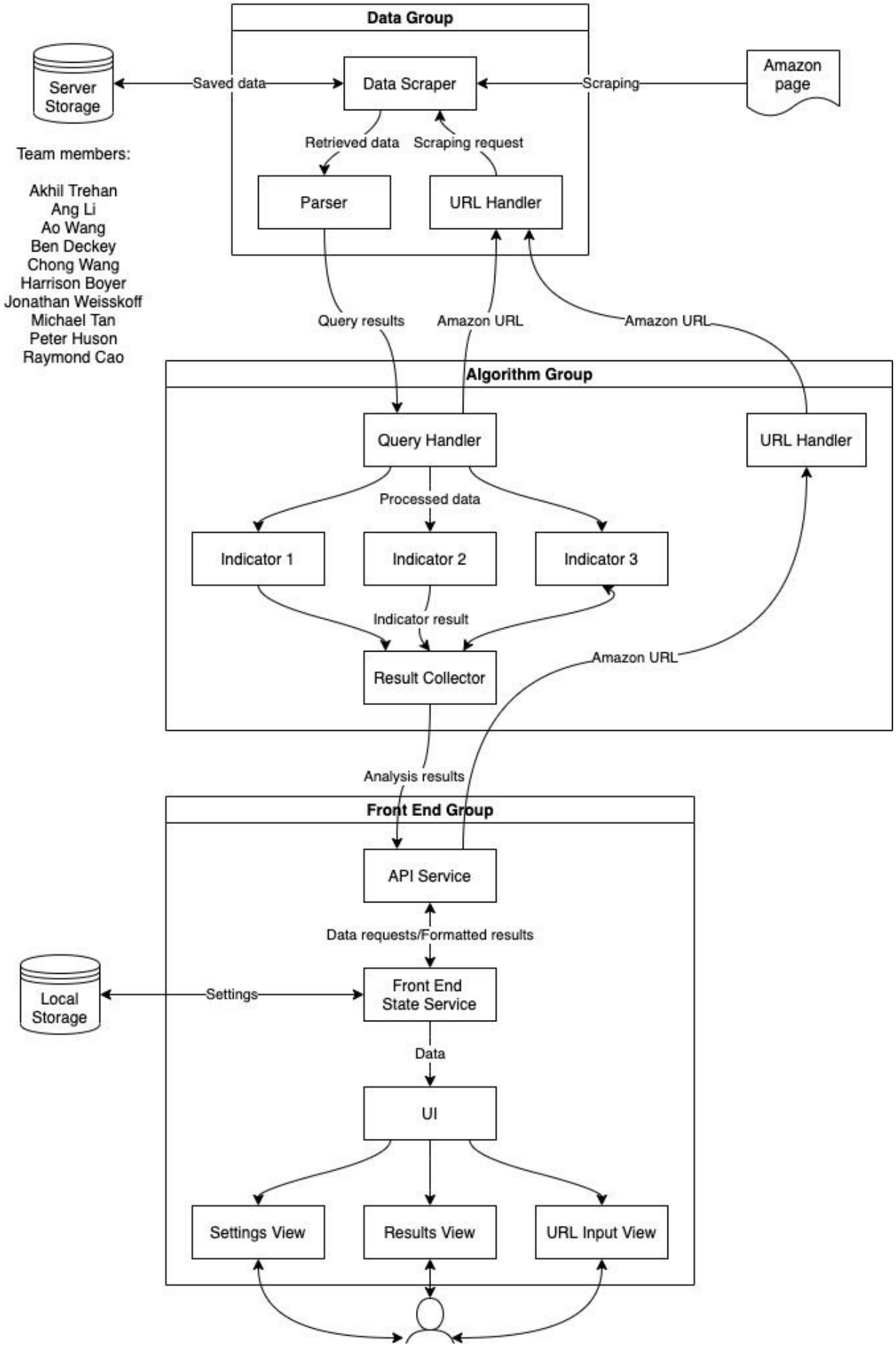
- Copies Exist Elsewhere: Fake reviews will often contain sentences copied from other reviews on Amazon. This indicator detects copied sentences above a given threshold. If enough sentences in enough reviews are copied, then this indicator returns a 1. Otherwise, it returns a 0.
- Tone: From our observations, most fake reviews tend to be positive in order to promote the selling of the product. In this case, checking the extent of positive sentiment of the reviews can be a possible measure: if the tone of a review is extremely positive, chances are that this review is fake. This indicator analyzes the tone of a review and returns a score between 0 and 1, with 1 being super positive and fake.
- Spikes in Positive Reviews: Fraudulent positive reviews are often posted in a short time frame, after and before which the ratings drop. We can catch this pattern searching for positive review spikes in the distribution of review ratings over time. A spike is determined by calculating a running mean and standard

deviation of review scores sorted chronologically. If a score is too far away from this running mean, that review will be flagged. If we encounter a sequence of flagged reviews then this indicator will report a spike in abnormally positive reviews.

- Pairwise similarity: The goal is to detect if any of the reviews on a product have copied from the most helpful ones. The pairwise similarity score uses the frequency-inverse document frequency (TF-IDF) to calculate the similarity between each review with the most helpful ones. To compute the similarity, different words are assigned different importance, which is inversely proportional to their overall frequency in all reviews. This indicator returns a 1 if enough similarity is found, and a zero otherwise.
- Semantics similarity: This indicator tries to detect if there's a pair of reviews that are extremely similar semantically and complements the indicators that detect exact copying. Semantic similarity is a metric defined over a set of documents or terms, where the idea of distance between items is based on the likeness of their meaning or semantic content as opposed to lexicographical similarity. Intuitively, fake reviews left by the same reviewers tend to use different words that are similar semantically.
- Grammar analysis: Often fake reviews contain incorrect grammar. This indicator analyzes the grammar of the review sentences and returns a score between 0 and 1, with 1 being very wrong and fake.

Architecture:

"Amazon Fake Reviews Detector": Software Architecture Design



Screenshots from the website:

Phony Finder

ANALYZE

Results page

(from this product URL: https://www.amazon.com/Machine-Learning-Engineering-Andriy-Burkov/dp/1999579577)

Product URL https://www.amazon.com/Machine-Learning-Engineering-Andriy-Burkov/dp/1999579577

Total Score 0.6657111353507015

Explanation A score 1.0 indicates strong evidence of having fake reviews and a score 0.0 indicates little evidence of having fake reviews.

Indicator	Score	Description
Tone Analysis	1	The score represents how positive and subjective the review is
Positive Review Spike Detection	0	This indicator will output a 1 if an abnormal spike in positive reviews is detected and a 0 otherwise
Pairwise Similarity	0.5547592794589179	The score represents the pairwise similarity between reviews and most helpful reviews.
Semantic Similarity	0.4	This indicator calculates the semantic similarity between individual reviews and tag a review suspicious if it's too similar to another one
Grammer Analysis	0.6583333333333333	The score represents how faulty the grammer of the review is