Assignment 1

Group Members:

Ahmad Raza 220088
Harsh Agrawal 220425

Indian Institute of Technology Kanpur

CGS616: Human Centered Computing

Instructor: Prof. Pragathi Balasubramani

Download Link for Code and Plots: Click here

Detection of Fake Reviews: An Analytical Approach

March 9, 2025

1 Introduction

In the era of online shopping and digital marketplaces, customer reviews play a crucial role in influencing consumer decisions. However, the presence of fake reviews misleads potential buyers and affects the credibility of e-commerce platforms. This project aims to develop a systematic approach to detecting fake reviews using data analysis and filtering techniques.

2 Objective

The primary objective of this study is to identify and label fake reviews based on specific textual and behavioral criteria. By applying exploratory data analysis (EDA) and filtering techniques, we aim to distinguish authentic user reviews from potentially deceptive ones.

3 Tools and Libraries Used

- **pandas** – Used for data loading, manipulation, and filtering.
- **numpy** – Utilized for numerical operations and handling arrays.
- **nltk** – Applied for Natural Language Processing (NLP) tasks, specifically for stopword handling.
- **textblob** – Used for sentiment analysis of reviews.
- **numpy.vander** – Employed to generate Vandermonde matrices, possibly for polynomial feature transformations.
- **re (Regular Expressions)** – Employed for text cleaning and processing.
- **matplotlib.pyplot** – Utilized for creating visualizations.
- **seaborn** – Used to enhance statistical data visualizations.

4 Methodology

For our analysis, we utilized the **Appliances** dataset from the Amazon Reviews 2023 repository. This dataset originally contained approximately **2 million reviews**. To ensure efficient processing and analysis, we extracted a representative **subset of around 0.8 million reviews** for our study.

4.1 Exploratory Data Analysis (EDA)

- The dataset consists of multiple review attributes, including **title, text, rating, sentiment, timestamp, and user behavior**.
- Data cleaning steps were applied, including handling missing values and filtering relevant textual data.
- Distribution analysis was conducted on **expected sentiment, final review sentiment, and rating** to understand review trends.
- Reviews with **empty ”title” or ”text” columns** were removed.
- Non-alphabetic characters were ignored to ensure only meaningful text was considered.

5 Criteria for Identifying Fake Reviews

Several logical conditions were applied to classify reviews as **fake (1,2,3) or authentic (0)**:

5.1 *Duplicate Review Detection*

- If a review's **text** column appeared in at least **two or more instances**, and its word count was **three or more**, it was labeled as **fake=1**.

5.2 *Sentiment-Based Filtering*

- To determine the authenticity of reviews, we applied a sentiment analysis approach using two tools: **TextBlob** and **VADER**. The filtering process followed these steps:

1. **Dual Sentiment Analysis**
   - We analyzed the sentiment of each review using both **TextBlob** and **VADER**.
   - If both tools returned the same sentiment classification (positive, neutral, or negative), we considered that as the **final sentiment** of the review.
2. **Expected Sentiment Assignment**
   - Based on the rating provided in the review:
     - **Ratings 1 or 2** → Negative sentiment
     - **Rating 3** → Neutral sentiment
     - **Ratings 4 or 5** → Positive sentiment
3. **Fake Review Identification**
   - We compared the **expected sentiment** (from the rating) with the **final sentiment** (determined by TextBlob + VADER).
   - If there was a **conflict** between the two sentiments, the review was labeled as **fake=2**.
   - If the sentiments matched, the review was **initially considered non-fake** (subject to further filt

5.3 *Suspicious User-Based Fake Review Detection*

- To identify potentially fake reviews based on user activity, we analyzed the number of reviews submitted by each user. The following steps were implemented:
  - **Counting User Reviews:** We calculated the total number of reviews submitted by each unique `user_id`.
  - **Defining a Threshold:** Users who submitted more than 20 reviews were considered **suspicious**.
  - **Marking Suspicious Reviews:** Any review submitted by these suspicious users was labeled as **fake=3**.
- This approach assumes that users submitting an unusually high number of reviews may be engaged in spam or fraudulent activities. The threshold of 20 reviews was chosen based on empirical observation and can be adjusted for different datasets.

5.4 *Final Criterion*

- After applying all previous conditions for detecting fake reviews, we implemented a final filtering criterion based on **rating** and **sentiment alignment**.

1. **For Reviews with Ratings 2 and 4**
   - A review was marked as **non-fake** (`fake = 0`) if:
     - The **expected sentiment** (derived from the rating) matched the **final sentiment** (determined using TextBlob and VADER).
     - **OR** the **final sentiment** was *neutral*.
     - **OR** the **expected sentiment** itself was *neutral*.
   - This rule ensures that reviews with these ratings are considered **non-fake** as long as there is no strong contradiction in sentiment classification.
2. **For Reviews with Rating 3**
   - All reviews with a rating of **3** were directly labeled as **non-fake** (`fake = 0`).

- Since a rating of **3** is typically neutral and subjective, we assumed it does not contribute to fake review detection.
- This final filtering step further refines the fake review detection by ensuring consistency in sentiment-based classification while allowing flexibility for neutral cases.

6 Results and Analytics

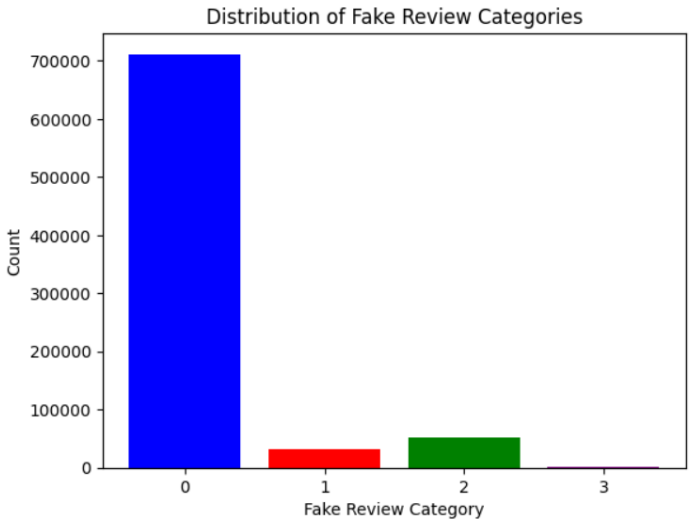

Figure 1: Distribution of Fake Review Categories

- **Figure 1 Interpretation:** The bar chart illustrates the distribution of fake review categories. Most reviews fall under category 0 (non-fake), while only a small fraction belongs to categories 1, 2, and 3, which represent varying levels of fake reviews. This suggests that the majority of reviews appear to be genuine.

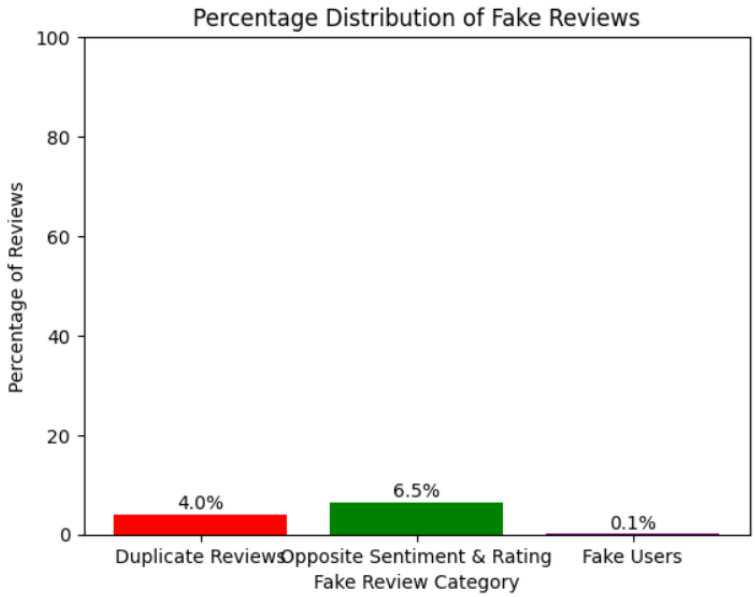

Figure 2: Percentage Distribution of Fake Reviews

- **Figure 2 Interpretation:** The second figure presents the percentage distribution of fake reviews. **Duplicate reviews (4.0%) a**nd **sentiment-rating mismatches (6.5%)** account for most fake reviews, whereas fake users contribute only **0.1%**. This implies that review duplication and sentiment inconsistency are the primary indicators of fake reviews.

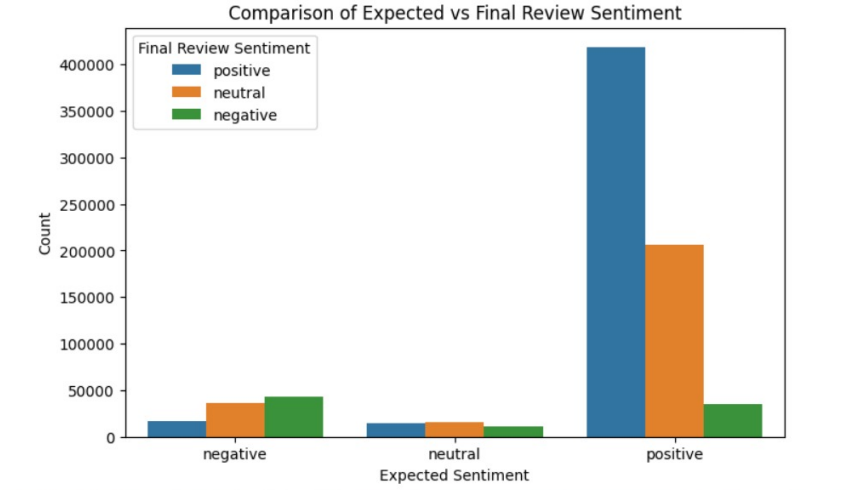

Figure 3: Comparison of Expected vs Final Review Sentiment

- **Figure 3 Interpretation:** The bar chart compares the **expected sentiment** (based on ratings) with the **final sentiment** (determined using sentiment analysis). Most reviews with an expected **positive sentiment** were classified as **positive**, while a significant portion was labeled as **neutral**. For negative and neutral expectations, the distribution of final sentiments is more balanced, indicating inconsistencies in sentiment classification.

7 Relation between Presence of fake reviews and price

To begin our analysis, we plotted a scatter plot to visualize the relationship between price and the percentage of fake reviews (Figure 4).

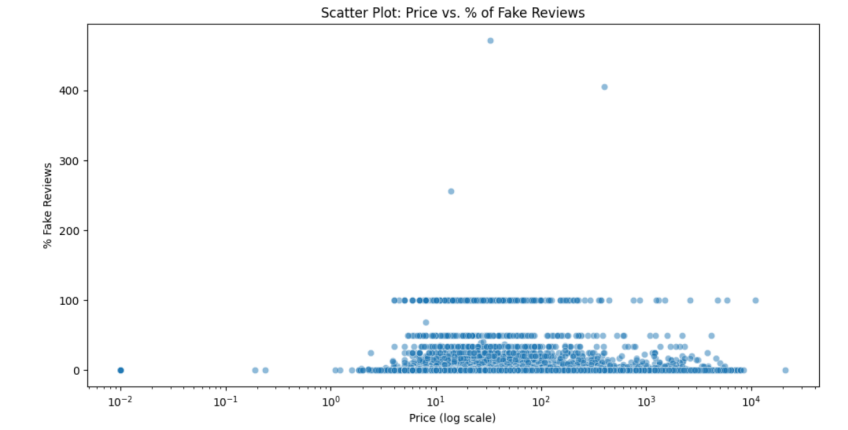

Figure 4: Scatter Plot: Price vs. % of Fake Reviews

The scatter plot does not show any clear pattern, making it difficult to deduce a direct relationship between price and fake reviews.

7.1 Regression Analysis

We applied different regression models to analyze the correlation between price and fake reviews:

Model	R^2 Score	Mean Squared Error (MSE)
Linear Regression	-0.0002	12511.19
Polynomial Regression (Degree 2)	-0.0006	12516.78
Random Forest Regressor	-0.0174	12726.73

Table 1: Regression Model Performance

As seen in Table 1, all models show poor performance, with near-zero or negative R^2 scores. This indicates that price alone is not a strong predictor of the percentage of fake reviews.

7.2 KDE Analysis: Price Distribution for Different Fake Review Groups

To gain further insights, we plotted a Kernel Density Estimate (KDE) graph comparing price distributions for products with high vs. low fake reviews (Figure 5).

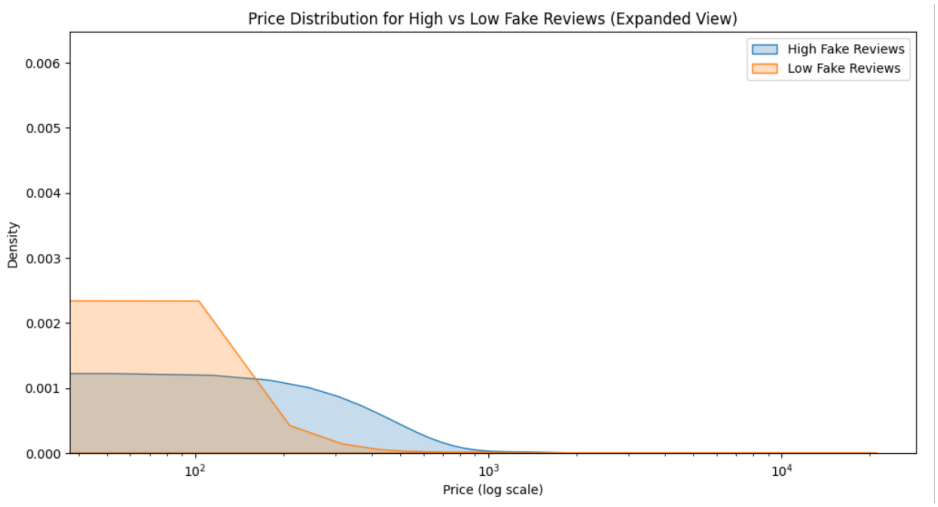

Figure 5: Price Distribution for High vs. Low Fake Reviews (Log Scale)

- **High Fake Reviews** (More than 50% fake reviews) - Blue Curve
- **Low Fake Reviews** (Less than 10% fake reviews) - Orange Curve

7.3 Key Observations

- Most products in both categories (high and low fake reviews) are priced in the lower range.
- The distribution of high fake review products extends further into higher price ranges compared to low fake review products.
- This suggests that expensive products tend to have a higher percentage of fake reviews.

7.4 Conclusion

From our analysis, we find that price alone is not a strong indicator of fake reviews. However, KDE analysis suggests that higher-priced products are more likely to have a higher percentage of fake reviews. This could be due to increased incentives for sellers of expensive items to manipulate reviews.

End of Report

Contributions

- **Ahmad Raza (220088):** Approx. 65% in making report + 2 out of 4 criteria proposed + 35% in writing main code.

- **Harsh Agrawal (220425):** Approx. 35% in making report + 2 out of 4 criteria proposed + 65% in writing main code.

Please refer to the title page (first page of the PDF) for source link to review codes.