UE21CS343BB2 Topics in Deep Learning

DETECTING FAKE REVIEWS ON AMAZON

M. Nikita Mabel   PES2UG21CS263

Maitreyi Kar   PES2UG21CS270

Malavika V   PES2UG21CS271

Malay Shikhar   PES2UG21CS272

Abstract

Detecting fake reviews on e-commerce platforms like Amazon has become a critical challenge due to their significant impact on consumer decision-making. This paper provides an overview of various techniques employed for identifying fake reviews, including sentiment analysis, linguistic analysis, reviewer behavior analysis, and machine learning algorithms. By examining patterns in review content, timing, and reviewer characteristics, researchers and practitioners aim to develop effective methods to differentiate between genuine and deceptive reviews. The success of these techniques holds implications for maintaining the credibility and trustworthiness of online product reviews, thereby enhancing consumer confidence in e-commerce platforms.

1. Dataset

This comprehensive dataset featuring reviews on Amazon products, meticulously collected through web scraping to provide a valuable resource for researchers and practitioners in the field of sentiment analysis and fake review detection. This dataset offers a diverse and extensive collection of reviews spanning various product categories available on the Amazon platform.

Key features:
- Rich diversity
- Large scale
- Authenticity

2. Ensuring that it is not available in the public domain

- This data has been created from scratch using web scraping techniques for each custom-picked products.
- 60 products were selected across various categories of electronic products such as laptops, phones, tablets, device accessories and so on.
- For each product, 4 pages of review data on the amazon website was considered.

3. Efforts taken to create the dataset

- **Requests Library (requests):** This library is employed to send HTTP requests to Amazon's review pages and retrieve the HTML content of those pages.
- **Beautiful Soup (BeautifulSoup):** This library is utilized to parse the HTML content obtained from the requests. BeautifulSoup provides methods to navigate and extract data from HTML documents easily.
- **Parsing HTML:** The script parses the HTML content of Amazon review pages to extract relevant information such as reviewer name, star rating, review title, review date, and review description.
- **CSS Selectors:** CSS selectors are used to target specific HTML elements containing the desired review data.

4. Format

- The dataset is saved in CSV format.
- The scraped data consists of several review items corresponding to the chosen products, where each item consist of the following values: rating (out of 5), title, description (review text)

5. Further research

- **Fake Review Detection:** Developing models to identify fake or deceptive reviews.
- **Sentiment Analysis:** To determine the overall satisfaction or dissatisfaction of customers with specific products.
- **Market Research:** Gaining insights into market trends, customer preferences, and competitive analysis.
- **Product Recommendation Systems:** Using reviews to improve recommendation systems by understanding customer preferences and feedback on similar products.

6. No. of data points

Our scraped dataset from Amazon reviews contains 9000 data points.

7. What ground truth values have been encoded and validity of the same

Given the web-scraped nature of the data, assigning ground truth values (factual and verifiable labels) regarding the sentiment (real/fake) of the reviews becomes unreliable:

- **Source Bias:** Web scraping cannot guarantee the authenticity of the data. There's a high chance of including fake reviews alongside genuine ones.
- **Lack of Verification:** Without Amazon's involvement or user verification methods, it's impossible to definitively label reviews as real or fake.

Therefore, using web-scraped data for tasks like sentiment analysis or identifying fake reviews holds limited validity due to the inherent unreliability of the source.

8. Sample dataset

	Name	Stars	Title	Date	Description
1					
2	**SUMAN**	5.0	5.0 out of 5 stars Best Tablet for Stu	N/A	If u are thinking of buying a tablet for studying under 30k segment, this the tablet to go for. ✏ Although the display is non-laminated,
3	**Sameer S.**	5.0	5.0 out of 5 stars Great product and	N/A	It is a gift that I got for my mother and I am not an IOS user. Her primary use for it is family connection and browsing, so no high end
4	**shoiab alam**	4.0	4.0 out of 5 stars Smooth functionir	N/A	Works good . Been using it for attending online classes and it delivers just right. If the sim slot variant could've been a little affordabl
5	**Shubhanshu Delwal**	5.0	5.0 out of 5 stars Value for money.	N/A	I bought it in sale and I got it in 26000.So after one month I do not at all regret buying it.it is value for money.Good for day to day tasl
6	**Sriram V.**	5.0	5.0 out of 5 stars THE best tablet in	N/A	For people who are casual users and consume mainly content like Hotstar, Netflix etc. with a little productive work like word,excel - r
7	**Narmada khuntia**	5.0	5.0 out of 5 stars Very good	N/A	Very nice product by apple. I am very happy all is good . Thank you amazon
8	**ChaliKahani**	5.0	5.0 out of 5 stars A very well priced	N/A	Using it with an apple pencil and procreate with absolutely no hiccups. Works pretty good for all things considered.The pro models c
9	**Sayed Azaaruddeen**	5.0	5.0 out of 5 stars The Beast - IPAD	N/A	Ipad is very smooth and Fast As expected from AppleHelping me to do my Day to day task and as well Enjoying Games on Large Gi
10	**Sudeep Kumar**	5.0	5.0 out of 5 stars GOOD	N/A	It works fine. It is easy to use. There is no lag. In built app of note making is as good as purchased one. It works fine with Kingone st
11	**phuong**	1.0	1.0 out of 5 stars 1st and last Apple	N/A	The media could not be loaded.

Thank You