

Fake Review Detector and Price Analysis

Using AI, Web Scraping, and large language models

Oves Nadaf, Shakila Siddavatam

1 Master's Student, Department of Computer Science, Abeda Inamdar Senior College, India

2 Head Department, Department of computer science, Abeda Inamdar Senior Collage, India

Abstract

The process of using online product reviews and their prices during purchasing has become widespread over the years. Nevertheless, the accuracy of such data cannot be guaranteed due to fake reviews and concealed prices. Therefore, this research proposes an approach that may help solve the stated problems.

Indeed, in the current world, people tend to depend more on online reviews and prices when making purchases online. But the emerging problem is that fake reviews and misleading pricing have become common practices nowadays. For this reason, a new method is proposed which consists of the integration of web scraping technology and AI models including transformers and large language models.

The system operates through two main stages. The first stage involves extracting real-time information from e-commerce platforms. The second stage consists of employing the refined AI algorithm to detect suspiciously written review texts. Simultaneously, prices will be monitored to provide the most cost-effective offers and detect any irregular price spikes. To simplify the operation of the algorithm and make its output understandable by any user, our system is supplemented with an LLM that gives recommendations on buying products.

Introduction

The dawn of a new age of shopping is upon us. Comparison became effortless due to e-commerce sites, yet comparison became effortless with its attendant issue. Consumers today are absolutely dependent on "trust signals" that appear in the guise of stars and discount prices all around them. These trust signals might be useful if they are truthful; however, consumers might get misled otherwise.

The most dangerous issue is the emergence of "opinion spam," which involves creating fraudulent reviews either to enhance the reputation of a product or damage the standing of a rival. As AI continues to improve its ability to write reviews, it has become virtually impossible for an average person to distinguish the fraud from reality. Pricing also seems to be a trick in disguise, as "dynamic pricing" along with fake "limited time" discounts make things tricky.

However, most existing studies consider them two distinct issues. On the one hand, there are review detection algorithms and, on the other hand, price analysis tools. In practice, these phenomena are frequently interrelated. An unexpected spike in the number of positive comments about an item along with its unpredictable price changes may be considered the strongest evidence of fraud.

That is why we created such a combined approach to solving this problem. The purpose of our study was to develop a mechanism for evaluating the trustworthiness of products based on the combined application of natural language processing and price analysis techniques. In this way, our method offers a more accurate "trust score."

Methodology

Our system was designed with modularity and practicality in mind. We did not want to create an idealized model but a tool that would operate in practice, within the context of a running website. Below is the outline of our design approach.

Data Gathering

The first stage always involves gathering data. We created a web crawler that operates like a conventional browser, accessing specific product pages and extracting two pieces of information – the body of the review itself and the price of the product. By identifying ourselves as a user agent, we allow the website to recognize us as a regular person, enabling us to gather accurate and up-to-date data.

Data Preprocessing

Gathering raw data from the internet can yield messy results. To ensure that our AI algorithms are processing meaningful data, we clean the gathered material by removing the excess HTML coding, adjusting spacing, and normalizing the currency of monetary values.

Fake Review Detection

For the reviews, we used a transformer-based model. Think of this as an AI that understands the “vibe” and context of a sentence, not just individual words. We finetuned it to recognize the specific linguistic patterns that fake reviewers often use—like being overly generic or using forced enthusiasm.

Price Analysis Pipeline

While the review AI is working, our price pipeline is looking at the numbers. It compares prices from different sources to find the average and the absolute lowest price. This helps us see if a “discount” is actually a discount or just a marketing trick.

Large Language Model Integration

Numbers and “genuine/deceptive” labels are great, but they can be hard to interpret. We brought in a Large Language Model (LLM) to act as a translator. It takes all the technical data and writes a short, human-friendly summary. It might say something like, “The reviews look mostly honest, but we found this product \$10 cheaper on another site.”

System Integration and API Design

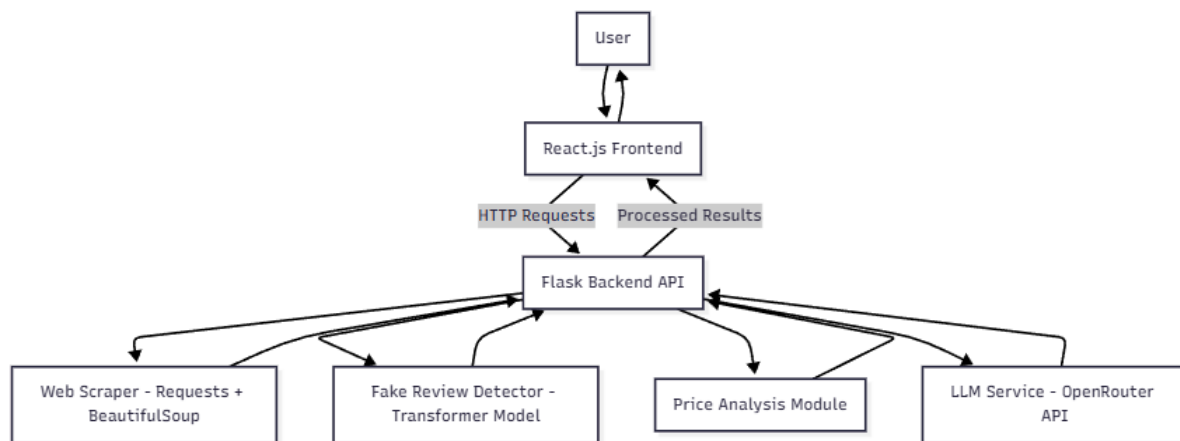
The whole system runs as a backend service using a RESTful architecture. This means it can easily plug into a website or a mobile app. We also made sure to handle things like CORS so the frontend and backend can talk to each other without any hiccups.

Evaluation Approach

Performance of the system is evaluated qualitatively through analysis of classification outputs, confidence scores, and generated summaries. The effectiveness of fake review detection is assessed based on the model's ability to distinguish deceptive linguistic patterns, while price analysis is evaluated through the consistency and clarity of LLM-generated insights. This evaluation approach emphasizes practical usability and real-world relevance.

Design and Implementation

We built this system with a “service-first” mindset. The backend does all the heavy lifting—the scraping, the AI analysis, and the reasoning—while the frontend stays light and easy to use



System Workflow

Step 1. The user copies a URL of a product on our site. Step 2. Our backend works right away and starts scraping all the fresh reviews and prices. Step 3. Our transformer model analyzes the reviews looking for any manipulations there. Step 4. Then the price analyzer runs to compare information from different platforms. Step 5. Our LLM provides a brief summary of what was found during the process. Step 6. As a result, our user gets a detailed report that will help them know what to believe and where to buy a

better deal. Tech Stack: We decided to use Python and Flask as backend technologies due to their speed and great compatibility with AI frameworks. Hugging Face transformers were utilized in the AI field. The core of our “brain” is based on an LLM through the OpenRouter API. React.js was chosen for the frontend.

Technology Stack

Technological Implementation

The proposed technological implementation is based on a selected technology stack designed to provide high performance, flexibility, and easy integration. The backend is implemented using Python and the Flask framework. It creates a lightweight platform for developing REST APIs needed for communication between the frontend application and server services.

Language processing and review detection are based on transformers available in the Hugging Face library. Transformers are used since they allow capturing not only syntactic but also semantic nuances in the text, which are crucial when detecting fake reviews. Web scraping is based on Requests together with BeautifulSoup that parse and retrieve unstructured data from web pages.

Intelligent pricing reasoning and natural-language summarization require a language model integrated into the system architecture. In our case, the system is able to communicate with the OpenRouter API containing the required model that allows for flexible language processing. Finally, we utilize React.js for the frontend, which allows us to build an asynchronous application communicating with the backend via HTTP.

User Interface Overview

Simple and easy-to-use are key goals that characterize the design of the user interface. The system communicates with the user via a clean web interface with input boxes where a user enters product data and performs analysis via specified actions. Once a user submits his or her data, visual cues such as the loading bar appear, signaling that the process has commenced.

The output interface displays the output of the processes in an organized and clear way. Results of fake review detection come with the corresponding confidence scores, allowing users to determine the accuracy of the output. Meanwhile, price analysis is presented in the form of highlighting the lowest price from available options. In addition, LLM-produced summaries of pricing trends and suggestions appear in plain language.

User Interface Screenshots



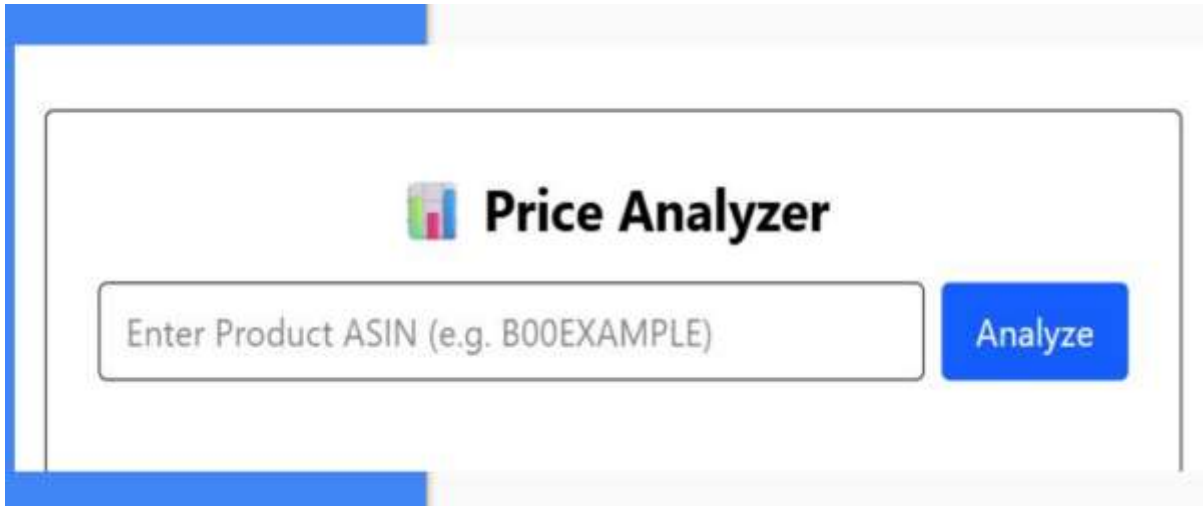
The login page features a light blue background. In the center is a white rounded rectangle with a subtle shadow. At the top of this rectangle is the word "Login" in bold black text. Below it are two input fields: the first is labeled "Email" and contains the placeholder text "Enter your email"; the second is labeled "Password" and contains the placeholder text "Enter your password". At the bottom of the white rectangle is a solid blue button with the word "Login" in white text.

Figure 1.login page



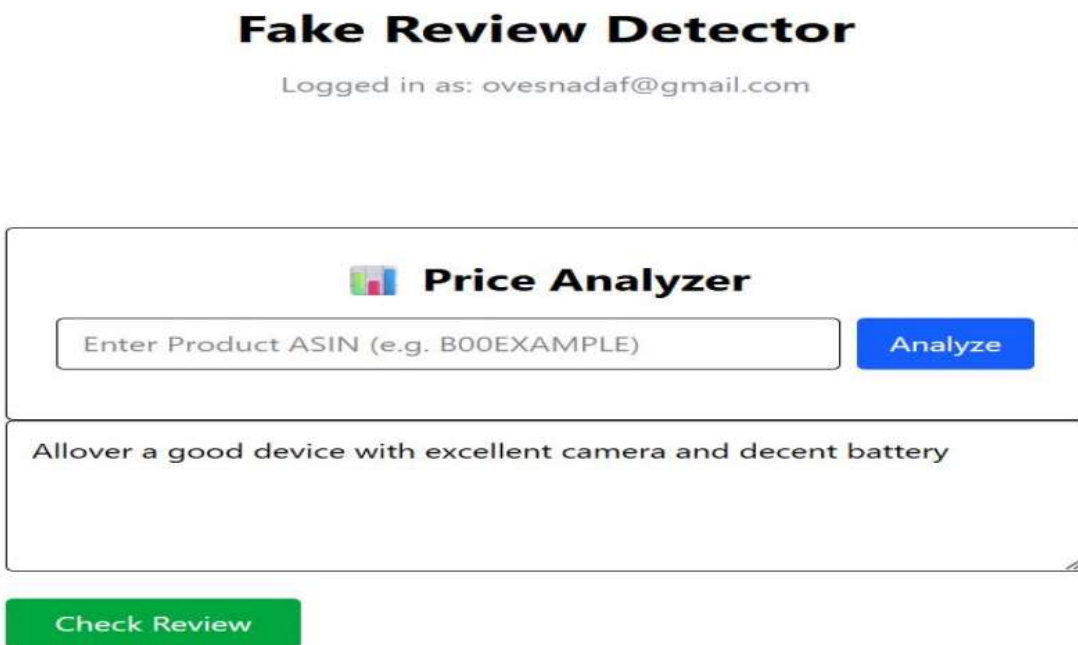
The review component consists of a large, empty rectangular text area with a thin gray border. Inside the top-left corner of this area is the placeholder text "Enter a product review...". Below the text area is a green rounded rectangular button with the text "Check Review" in white.

Figure 2 review component



The interface for the Price Analyzer tool. It features a blue header bar. Below it, a white box contains the title "Price Analyzer" with a small bar chart icon to its left. Underneath the title is a text input field with the placeholder text "Enter Product ASIN (e.g. B00EXAMPLE)". To the right of the input field is a blue button labeled "Analyze".

Figure.3 price analyzer



The interface for the Fake Review Detector tool. It has a white background. At the top, the title "Fake Review Detector" is displayed in bold black text, followed by the text "Logged in as: ovesnadaf@gmail.com". Below this is a section titled "Price Analyzer" with a bar chart icon. This section contains a text input field with the placeholder "Enter Product ASIN (e.g. B00EXAMPLE)" and a blue "Analyze" button. Below the "Price Analyzer" section is a text area containing the review text: "Allover a good device with excellent camera and decent battery". At the bottom of the interface is a green button labeled "Check Review".

Figure 4 main page

Summary of Design Implications

This paper highlights the key design decisions in creating the system. The focus is on making the algorithm interpretable, modular and applicable. The use of automated data processing, transformer-based classification and LLM-reasoning shows how such complex processes can be implemented to overcome issues of trust and transparency in e-commerce.

Discussion

The findings of this study are straightforward. Combining review and price analysis yields much better results than focusing on one of the factors alone. We have identified an interesting trend of how "suspicious" reviews correlate with highly volatile pricing. In other words, attempts to fake reputation and customer satisfaction often coincide with changes in product pricing strategy.

Transformer models proved to be quite proficient in identifying even minor signs that cannot be perceived by older algorithms. Such models can distinguish legitimate customers from bots. Nonetheless, no artificial intelligence is perfect. Therefore, the price analysis acts as a sort of "sanity check." If the model determines that reviews look legitimate but price is very volatile, it may still identify the product as unreliable.

Limitations of the System

There are limitations, too. The execution of these artificial intelligence algorithms is computationally intensive. Web scraping can also become problematic when a website redesigns its interface. It is imperative to factor in "cross-platform fraud" in which the merchant may post fake reviews on five separate platforms to prevent any suspicion.

Conclusion

In conclusion, it should be noted that the proposed system does not solve the issue of cross-platform fraud coordination, in which deceptive entities utilize various marketplaces to deploy their strategies for review distribution and price manipulation. Without federated learning or data sharing infrastructure, the problem of detecting such activity would remain restricted to a particular platform ecosystem.

Future Scope

In the future, we want to add more layers to the analysis:

Images and Profiles: Looking at product photos and the history of the person writing the review could make the system even more accurate

Causal Analysis: We want to figure out if fake reviews actually cause price changes or if they just happen at the same time

Privacy and Scaling: We're looking into "federated learning," which would allow different platforms to share fraud data without actually sharing private user information.

Conclusion

An innovative artificial intelligence approach that integrates analysis of deception in product reviews and the dynamics of pricing trends for improved trust evaluation in online business environments is outlined in this research paper. Combining analysis of linguistic veracity and economic signaling in a single integrated process, the approach goes beyond existing one-dimensional methods in showing how intelligence across multiple domains can produce more accurate trust signals. In the proposed framework, deception in product reviews and price manipulation are no longer separate problems but related aspects of marketplace deception.

Implications for Further Research

This work opens multiple avenues for future research. First, the proposed trust formulation can be extended through **causal and longitudinal modeling** to study how deceptive reviews influence price volatility over time, rather than treating the two signals as independent inputs. Second, **multimodal learning**—incorporating product images, seller metadata, and reviewer social graphs—offers a promising direction for strengthening robustness against increasingly sophisticated adversarial behavior. Third, future studies should explore **privacy-preserving and federated learning approaches** that allow cross-platform collaboration without centralized data sharing. Finally, the trust scoring mechanism itself can be formalized through probabilistic or game-theoretic frameworks to better capture uncertainty, strategic manipulation, and evolving market dynamics.

Implications for Policy and Regulation

From a policy perspective, this research provides a technical foundation for **algorithmic transparency and consumer protection regulations** in digital marketplaces. Regulatory bodies can leverage integrated trust signals to audit platform compliance with fair pricing and review authenticity standards. The system also supports the development of **standardized trust metrics** that can be mandated across platforms, reducing information asymmetry and discouraging deceptive commercial practices. Importantly, the explainability components proposed in this framework align with emerging regulatory requirements for accountable and interpretable AI systems, enabling platforms to justify automated moderation decisions to users and oversight authorities.

Implications for Practical Intervention

In terms of intervention, the proposed framework enables proactive rather than reactive trust enforcement. E-commerce platforms can deploy the system to **flag high-risk products in real time**, trigger human moderation workflows, or temporarily suppress misleading recommendations. For consumers, integrated trust indicators can power **buyer-facing alerts**, warning users of abnormal price behavior or suspicious review patterns before purchase. At the marketplace level, the system can inform **adaptive pricing oversight**, detect coordinated review manipulation campaigns, and guide targeted enforcement actions against repeat offenders. These interventions collectively shift trust management from post-hoc damage control to continuous, intelligence-driven prevention.

Closing Remarks

The following research outlines an advanced artificial intelligence architecture that integrates deception detection of reviews and pricing dynamics in order to improve the reliability of the trust evaluation process within e-commerce networks. Through the integration of linguistic genuineness detection with economic signaling analysis, the described model goes far beyond standard single modality techniques and illustrates the potential impact of multidisciplinary intelligent solutions on building more trustworthy and relevant trust measures. Deceptive reviews and price manipulation are revealed as interrelated aspects of market manipulation activities that distort the decision making process of consumers.

References

1. Salminen, J., et al. *Creating and Detecting Fake Reviews of Online Products*. This research explores the challenges of fake review detection on e-commerce datasets and experiments with language models (e.g., ULMFiT, GPT-2) for generating and classifying fake reviews.
2. Jalal, M. A., & Hussein, M. H. *Fake Reviews Detection in E-Commerce Using Machine Learning Techniques: A Comparative Survey*. This paper surveys various machine learning approaches for classifying fake versus genuine reviews and discusses their strengths and weaknesses.
3. Liu, J. *A Study on Fake Review Detection Based on RoBERTa*. This recent study uses the RoBERTa transformer model combined with behavioral features to enhance false review classification performance.
4. Hajek, P., Hikkerova, L., & Sahut, J.-M. *Fake Review Detection in E-Commerce Platforms Using Aspect-Based Sentiment Analysis*. This research investigates how aspect-level sentiment analysis improves fake review detection across different product categories and highlights practical implications for online platforms.