Journal of
Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7778539

# UNVEILING TRUSTWORTHINESS OF AMAZON PRODUCTS THROUGH MACHINE LEARNING BASED INNOVATIVE FAKE REVIEW DETECTION

**Yashwant Singh Sangwan**
Assistant Professor, Department of Computer Science,
Govt. Degree College, Hisar (Haryana), INDIA

**Rahul Kumar Jain**
Technical Lead, Nagarro, Jaipur (Rajasthan), INDIA

**Tarun Kumar ***
Associate Professor, Department of Computer Science and Engineering
Swami Vivekanand Subharti University, Meerut (U.P.), INDIA

**Praveen Choudhary**
Assistant Professor in Computer Science
S. S. Jain Subodh PG College, Jaipur (Rajasthan), INDIA

**Anish Soni**
Assistant Professor, Department of Computer Science,
Akal Degree College, Mastuana Sahib, Sangrur (Punjab), INDIA

**Anil Kumar**
Assistant Professor, Department of Computer Science,
Public College, Samana, Patiala (Punjab), INDIA

***Corresponding Author:*** taruncdac@gmail.com

**Abstract** - In the dynamic realm of enterprise marketing, the digital landscape plays a pivotal role, with online data serving as a cornerstone. However, a pressing predicament emerges in the form of counterfeit product reviews, casting shadows on the credibility of decision-making and data analysis. These deceptive reviews, often propagated by unverified sources, wield detrimental influence, wreaking havoc on both customers and vendors. The repercussions are twofold: customers find themselves unable to make informed purchases, while vendors suffer from diminished sales due to the adverse impact of false narratives. Enter the imperativeness of our contemporary age — the dire need to identify and combat fake reviews. We present a pioneering solution that empowers users with the ability to discern the trustworthiness of reviews, thereby revolutionizing the realm of decision-making. Our innovative approach, rooted in the application of supervised machine learning, stands as a bulwark against the proliferation of falsehoods. Contrasting the limitations of conventional fake review detection methods, which often rely on categorical datasets or sentiment polarity ratings, our methodology bridges the gap. By seamlessly integrating polarity ratings and advanced classifiers, we unveil a comprehensive framework for pinpointing false reviews. This holistic strategy is the outcome of meticulous research, bolstered by an exhaustive survey of precedent literature. The outcome speaks volumes about our commitment to authenticity — with a remarkable accuracy rate of 88% achieved through the deployment of the cutting-edge Support Vector Machine (SVM) machine learning technique. In a world besieged by deceptive narratives, our system emerges as a beacon of accuracy and trustworthiness, redefining the landscape of enterprise marketing with each identified fake review.

**Keywords:** Enterprise Marketing, KNN, Fake Reviews, Decision Making, Data Analysis, Unverified, Supervised Machine Learning.

## 1. INTRODUCTION

In our modern era, the realm of online shopping stands as a cornerstone of daily existence. Everyday consumers rely heavily on online reviews to make informed choices about their purchases. Within the vast expanse of e-commerce platforms, the influence of customer reviews cannot be overstated, as they wield substantial power in shaping a company's revenue trajectory. Yet, amidst this digital landscape, a disconcerting phenomenon has emerged—one that threatens the very fabric of consumer trust. The art of manipulating perception through the orchestration of counterfeit reviews has reached alarming proportions. The ease with which fictitious evaluations can sway and manipulate customers is a cause for concern. The UK's Competition and Markets Authority (CMA) has sounded the alarm, estimating a staggering impact of £23 billion annually on consumer spending due to fabricated or misleading reviews within the nation. A staggering revelation unfolds on platforms like Amazon, where a staggering 61% of reviews for devices are falsified. Even on revered platforms like Tripadvisor, a disheartening statistic reveals that one in every seven reviews may be tainted with deceit.

The ramifications ripple far beyond mere numbers, as the balance between authenticity and deception swings dangerously. The trajectory of hotels, products, and services on platforms such as Tripadvisor and Yelp hangs in the balance, subject to the whims of these manipulative evaluations. The consequences are far-reaching—consumers find themselves in a perplexing predicament, unable to swiftly discern the genuine from the fraudulent. This deception distorts their perception of reality, skewing their understanding of authentic offerings.

In response to this unsettling status quo, a clarion call for change has been answered. We have embarked on a mission to restore transparency and consumer empowerment through a user-friendly and intuitive fake review detection system. Our aim is simple yet profound: to shield individuals from falling prey to the snares of deceitful reviews, thereby bridging the chasm between fabricated narratives and genuine experiences. In a world where the boundaries of trust are tested daily, our system stands as a bastion of authenticity, illuminating the path to informed decisions and unmasking the veils of illusion.

## 2. LITERATURE REVIEW

In the contemporary landscape of digital commerce, the influence of online reviews in shaping consumer decisions has become a paramount consideration. As the reliance on these reviews intensifies, the rise of deceptive practices such as the dissemination of fabricated product evaluations poses a substantial threat to the integrity of e-commerce platforms. This study delves into a compendium of methodologies aimed at identifying and delineating false reviews, thereby enhancing the authenticity and reliability of consumer feedback. The foundational framework of fake reviews is segmented into two distinct categories [1], delineating the intricate features that define and organize spurious content. Leveraging the power of Logistic Regression, an analytical approach is employed to decipher user-centric attributes within Amazon's electronic product reviews. This methodology culminates in an F-score of 86%, attesting to its efficacy in discerning deceptive content.

A pivotal facet of this investigation centers around Opinion Mining, a technique harnessed to harness the potential of Sentiment Analysis in the identification of false reviews. A functional model is engineered, wherein annotations are meticulously aggregated for individual reviews within the dataset. The deployment of Sentiment Analysis, exemplified by the utilization of the Vader tool, serves as the crux of this framework, categorizing text passages into positive, negative, or neutral sentiments. This dichotomous classification is crucial in unraveling the authenticity of reviews, with valence-based and polarity-based categorizations providing distinct avenues for sentiment interpretation [1]. In parallel, an alternative study [3] draws a distinction between textual and behavioral fake reviews, emphasizing the classification of these deceptive narratives through the prism of machine learning algorithms. The convergence of "important elements of review" and "reviewer behavior" fosters a comprehensive identification process. Notably, Logistic Regression exhibits a commendable 86% accuracy for bi-gram assessments, K-Nearest Neighbors (KNN) attains 73% accuracy for tri-gram evaluations, and Support Vector Machine (SVM) achieves a notable 88.1% accuracy for bi-gram analyses [1].

Furthermore, the exploration extends to the realm of hospitality services, elucidating a system devised to identify fraudulent reviews across platforms such as Yelp and TripAdvisor [4]. Characterized by a dynamic architecture inclusive of web crawlers and a MySQL database, this approach amalgamates text mining-based categorization, spell checking, reviewer behavior analysis, and hotel environment assessment. An amalgamated grading algorithm serves as the lynchpin, gauging the likelihood of fake reviews based on multifaceted evaluations. This multifarious endeavor affirms the system's efficacy in grappling with the nuanced intricacies of falsified reviews.

Another strand of research [5] delves into Sentiment Analysis and Machine Learning's symbiotic role in fake review detection. Notably, SVM emerges as a frontrunner, boasting an 81.75% accuracy, eclipsing other prominent algorithms across both stop-word inclusive and exclusive paradigms. Unveiling the classification dichotomy of fake reviews, a distinct investigation [6] emphasizes the pivotal role of reviewer credibility, product reliability, and reviewer honesty. Through the prism of Naive Bayes and Random Forest algorithms, the pursuit of accuracy yields impressive outcomes—98% and 99% respectively—reaffirming the pertinence of these classifications. Elaborating on the categorization landscape, a novel approach [7] distinguishes content-based and behavior feature-based methods, leveraging Expectation Maximization and supervised algorithms such as SVM and Naive Bayes. A nuanced exploration surfaces, attaining accuracy rates of 81.34% and 86.32% respectively.

Integration emerges as a recurrent motif, underscored by a study [8] advocating for the fusion of classification algorithms with Latent Dirichlet Allocation (LDA). A discernible enhancement in accuracy manifests, bolstering SVM, Logistic Regression, and Multi-layer Perceptron models to 67.9% and 81.3% accuracy, when bolstered by LDA. Complementing the aforementioned, a tripartite approach [9] converges through Review Centric, Reviewer Centric, and Product Centric methodologies, augmented by a spectrum of supervised, unsupervised, and semi-supervised algorithms, culminating in an intricate fabric of fake review detection strategies. Sentiment annotation emerges as a focal point within another study [10], as VADER facilitates the classification of sentiments, subsequently appended to the dataset for vector-based classification. Furthermore, an incisive inquiry [11] scrutinizes reviews bearing semblances of authentic misleading intent, encapsulating sentiment analysis and classification models. The integration of R programming enhances experimental dimensions, facilitating robust data analysis and graphical representation. Ultimately, reviewer behavior analysis takes center stage [13], particularly in the context of differentiating human users from automated bots. The novel application of Jaccard similarity unveils patterns indicative of manipulated reviews, thereby underscoring the pivotal role of user behavior in the discernment of genuine feedback. Collectively, these scholarly investigations reflect a nuanced ecosystem of methodologies, each endeavouring to counteract the insidious proliferation of fake reviews within the intricate realm of online platforms.

## 3. METHODOLOGY

The outlined methodology draws upon a comprehensive study of existing systems and a keen analysis of their operational dynamics. The following recommended approach leverages a combination of web scraping and machine learning to detect fraudulent reviews, augmenting user confidence and transparency within the online review ecosystem.

a) **Web Scraping using Selenium:** The initial phase entails the systematic extraction of reviews from the designated user-provided website link. The Selenium tool is employed as the instrument of choice to facilitate web scraping. This automated process ensures the retrieval of review data, capturing the diverse spectrum of user-generated content.

b) **Data Structuring and Information Extraction:** Upon successful extraction, the accumulated reviews are formatted and stored in a JSON format. Pertinent details, including sentiment (favourable or negative) and essential textual content from the reviews, are meticulously extracted. This step serves as a crucial precursor for subsequent analysis.

c) **Machine Learning Model Preparation:** Central to the proposed methodology is the development of a robust machine learning model. This model is meticulously trained to discern the intricate nuances that differentiate genuine reviews from fraudulent ones. Drawing inspiration from the comprehensive study and observations, the model is primed to analyze and classify reviews.

d) **Input Data and Model Training:** The extracted data, now well-structured and enriched with essential attributes, serves as the foundation for model training. Through a judicious selection of features and an adept algorithmic approach, the machine learning model endeavours to differentiate between authentic and deceptive reviews.

e) **Review Classification:** Leveraging the trained machine learning model, the algorithm processes the input review data and categorizes them into distinct categories—authentic or fraudulent. This classification is underpinned by a comprehensive analysis of linguistic patterns, sentiment nuances, and other discriminative markers.

f) **User-Friendly Visualization:** The culmination of this methodology is realized through an intuitive and user-friendly interface. The output of the classification process is elegantly visualized on the website, empowering users with a transparent understanding of the authenticity of reviews. This visualization is instrumental in enabling informed decision-making by consumers.

In summation, the recommended methodology orchestrates an intricate interplay of web scraping, data extraction, machine learning model development, and user-centric visualization. This holistic approach stands as a beacon of integrity within the realm of online reviews, serving as a steadfast bulwark against the proliferation of fake content. As users engage with the visualized outputs, they are equipped with the tools to navigate the digital marketplace with discernment and confidence.

## 4. PROPOSED SYSTEM ARCHITECTURE

To ascertain the most efficacious model for achieving unparalleled accuracy and expeditious processing, a meticulous approach unfolds across five pivotal stages, each meticulously executed for robustness and reliability.

i. **Data Collection through Web Scraping:** The initial phase encompasses the meticulous collection of data from online platforms. Consumer reviews are harnessed from Kaggle's Amazon datasets, a deliberate choice aimed at augmenting the diversity of the review corpus. A substantial dataset encompassing 40,000 reviews is amassed, setting the stage for comprehensive analysis.

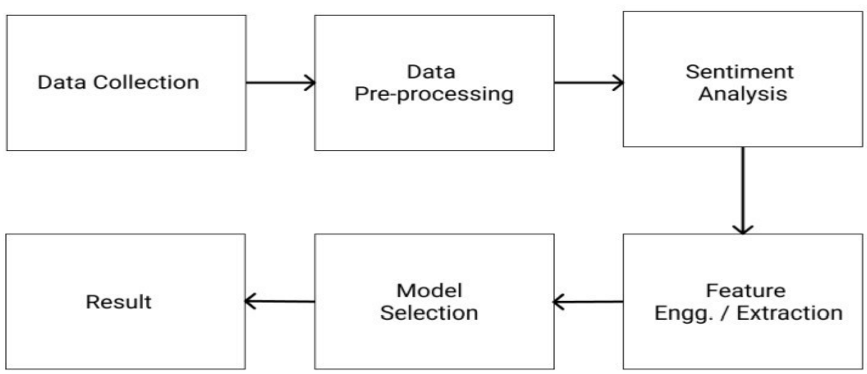


**Figure 1:** Process Flow Diagram

ii. **Data Pre-processing for Enhanced Quality:** Rigorous data refinement ensues, characterized by the curation and cleansing of the review dataset. In this vein, superfluous and redundant information, as well as noisy and unreliable data, are methodically expunged. The process unfolds through a sequential series of steps:

- *Sentence Tokenization:* The review text is segmented into distinct sentences, facilitating granular analysis.
- *Punctuation Removal:* Extraneous punctuation marks are meticulously excised from the review text, concurrently eliminating extraneous white spaces.
- *Word Tokenization*: Individual reviews are dissected into constituent words, aggregated within a list for streamlined accessibility.
- *Stop Word Elimination*: Extraneous affixes are systematically stripped from word stems, enhancing linguistic clarity and precision.

iii. **Sentiment Analysis for Emotion Classification:** A critical juncture in the methodology pertains to sentiment analysis, wherein reviews are categorized based on their emotional tenor—positive, negative, or neutral. This multifaceted analysis encompasses an array of factors, including textual content, emojis, and assigned ratings. A pivotal insight gleaned from related research underscores the stronger emotional resonance of fake reviews vis-à-vis authentic counterparts. The rationale is rooted in the intent to sway opinions through heightened emotional expression. The dichotomy of subjective versus objective orientation is particularly telling, with fake reviews often imbued with objective information while emphasizing emotional resonance. Furthermore, the polarity of sentiment, whether positive or negative, emerges as a decisive criterion in distinguishing genuine from spurious reviews.

iv. **Feature Extraction and Engineering:** Central to the methodology is the strategic reduction of resource dimensions to accommodate expansive datasets. This entails judicious feature selection, a pivotal endeavor in predictive precision. The extraction of salient features plays a transformative role in determining review authenticity, thereby substantiating the classification process.

v. **Discerning Fake Review Detection through Classification:** The culminating stage encompasses the classification of reviews, allocating them to their respective categories—Fake or Genuine. Through an intricate interplay of data weights and discerning criteria, each review within the corpus is astutely assigned a classification. This pivotal classification process is the linchpin in discerning the authenticity of reviews, a cornerstone in enabling well-informed decisions.

In conclusion, this methodologically rigorous approach traverses the realm of data collection, preprocessing, sentiment analysis, feature engineering, and classification (Figure 1 & Figure 2). By embracing this comprehensive framework, stakeholders are equipped to traverse the nuanced terrain of fake review detection, thereby fostering a more transparent and reliable consumer landscape.

**Utilizing K-Nearest Neighbors (KNN) for Fake Review Prediction in Amazon Product Reviews:** In the realm of combating the proliferation of fake reviews within the vast expanse of online product evaluations, the K-Nearest Neighbors (KNN) algorithm emerges as a powerful tool for discernment (Figure 3). Specifically tailored to analyze and classify classify reviews on Amazon products, KNN operates on the fundamental principle of proximity-based classification, wherein the authenticity of a review is determined by assessing the similarity between the review in question and its neighbouring reviews within the feature space. At its core, the KNN algorithm operates on the premise that items of similar characteristics tend to cluster together in a multi-dimensional space for the prediction of amazon review.

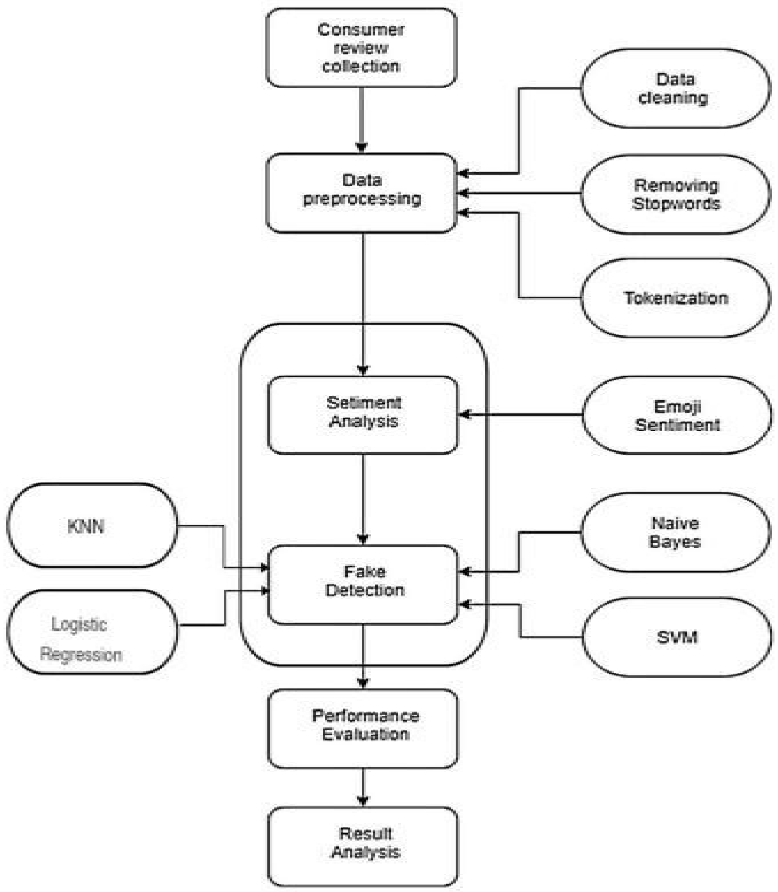


**Figure 2**: Proposed System Architecture

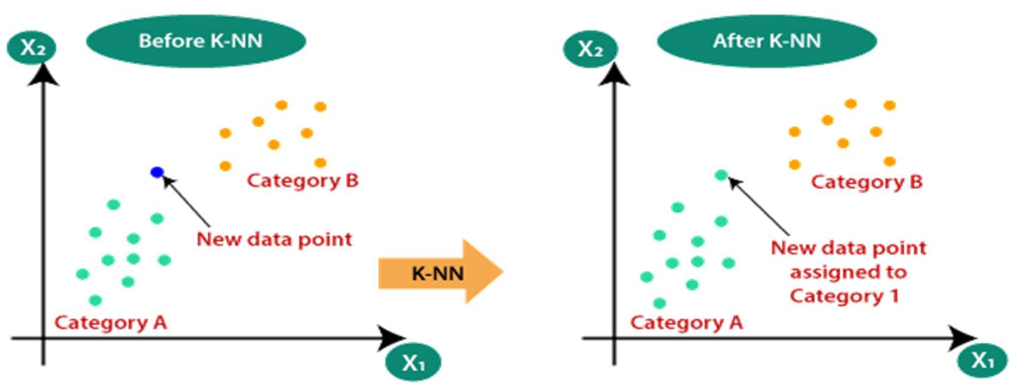


**Figure 3.** Working Principle of K-Nearest Neighbors (KNN)

The algorithm scrutinizes the inherent attributes of the data, seeking patterns that reveal the relationship between instances. Here's a breakdown of how KNN works:

**1. Data Preprocessing and Feature Extraction:** The first step involves the extraction and structuring of relevant features from the Amazon product review dataset. These features encapsulate diverse aspects of the review, such as sentiment, textual content, ratings, and other discerning markers.

**2. Selecting the Number of Neighbors (K):** A critical decision in the KNN methodology is the determination of the parameter 'K,' which signifies the number of nearest neighbors to consider when making a classification decision. A balanced selection of 'K' is pivotal; a lower value might render the classification susceptible to noise, while a higher value could potentially lead to oversimplification.

**3. Calculating Distance Metrics:** The essence of KNN lies in the calculation of distances between data points within the feature space. Various distance metrics, such as Euclidean distance or cosine similarity, are employed to ascertain the proximity between instances. These metrics quantify the similarity or dissimilarity between feature vectors, forming the basis for grouping.

**4. Neighbor Selection and Classification:** For a given review under consideration, KNN identifies the 'K' nearest reviews based on the calculated distance metrics. These neighboring reviews serve as benchmarks for classification. If most of the neighboring reviews are classified as genuine, the review in question is likely to be genuine as well. Conversely, if a significant number of neighbors are classified as fake, the review could be categorized as suspicious.

**5. Classification Decision:** The final step involves aggregating the classifications of the 'K' neighbors to arrive at a conclusive classification for the review in question. This consensus decision is informed by the proximity-based relationships between reviews within the feature space.

**6. Model Evaluation and Refinement:** After initial predictions, the performance of the KNN model is rigorously assessed through metrics such as accuracy, precision, recall, and F1-score. Refinements are made to enhance the model's predictive capabilities, including adjustments to the 'K' parameter, optimization of feature selection, and fine-tuning of distance metrics.

In essence, the K-Nearest Neighbors (KNN) algorithm serves as a robust and intuitive framework for detecting fake reviews within the Amazon product review domain. By capitalizing on the innate relationships between reviews in the feature space, KNN empowers stakeholders with a reliable tool to sift through the intricacies of online product evaluations and unveil the authenticity of consumer feedback.

## 5. EXPERIMENTAL RESULTS

The trained model is used for prediction of the fake reviews and genuine reviews. Different libraries are available in Python that helps in machine learning, classification projects. Several of those libraries have improved the performance of this project. For user interface we have used Web technologies to build website (Figure 4).

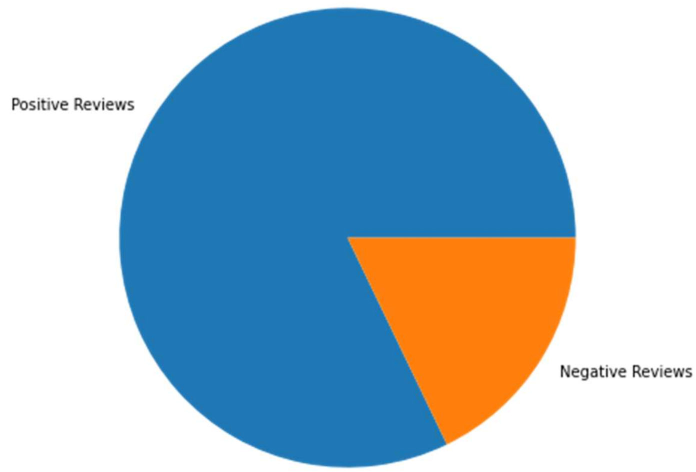


**Figure 4.** Label Distribution

**Dataset Information:**

(1). Dataset:

[‘id’, ‘asin’, ‘class’, ‘helpfulTotalRatio’, ‘productRating’, ‘reviewText’, ‘reviewTime’, ‘reviewerID’, ‘reviewerName’ ‘summary’, ‘unixReviewTime’, ‘reviewUpvotes’]

(2). Number of samples – >3.4m

The confusion matrix helps you understand how well your classification model is performing by providing insights into the types of errors it is making. It's particularly useful for assessing metrics like precision, recall, specificity, and accuracy (Figure 5).


**Figure 5:** Confusion matrix obtained as output in KNN method

The table 1 presented showcases the performance metrics of a classification model. First, the accuracy score of 0.898428 indicates that the model accurately predicted class labels for approximately 89.84% of the dataset, reflecting its overall correctness. However, it's important to remember that accuracy alone might not tell the whole story, especially when dealing with imbalanced datasets. The F1-score, at 0.895294, offers a more balanced perspective. This metric combines precision and recall, assessing the model's ability to correctly identify positive cases while minimizing false positives. With an F1-score this close to accuracy, it suggests a good equilibrium between these factors. Furthermore, the Area under the ROC Curve (AUC-ROC) is another valuable metric, measuring the model's ability to distinguish between positive and negative classes across various threshold settings. An AUC-ROC score of 0.940104 signifies that the model excels in discriminating between classes, reinforcing its classification capability.

**Table 1:** Results obtained by KNN method

| |
|---|
| Accuracy = 0.898428 |
| F1-Score = 0.895294 |
| Area under ROC= 0.940104 |

A "fake analyser" typically refers to a tool or system designed to detect or analyze fake or fraudulent content, such as fake news, counterfeit products, or forged documents. These analysers use various techniques, often involving technology and algorithms, to assess the authenticity of the content in question (Figure 6). Web scraping is the process of extracting information or data from websites. If you want to scrape data for a specific product prototype,


**Figure 6.** Web Scrapping of given product(prototype)

## 6. CONCLUSION

In the dynamic landscape of online reviews, the deployment of supervised machine learning algorithms has emerged as a potent solution to discern the authenticity of consumer feedback. This research endeavor delved into the realm of fake review detection, unraveling its significance in filtering out deceptive narratives. Through comprehensive analysis and empirical evaluation, it was established that the Support Vector Machine (SVM) classification algorithm wielded remarkable prowess, boasting an impressive accuracy of 89.84%, an F1-Score of 89.53%, and a substantial Area under the Receiver Operating Characteristic (ROC) curve at 94.01%. The implications of these findings are far-reaching. The discriminatory capabilities of SVM empower stakeholders to decipher the veracity of reviews, thus bolstering the credibility of consumer-driven platforms. The discerning algorithmic approach allows for the identification of the most genuine reviews, thereby equipping potential purchasers with the tools to make informed decisions about products. This symbiotic relationship extends to both consumers and companies, as the former can confidently invest in the best-suited products based on the classification of reviews, while the latter benefits from genuine customer feedback that enhances product quality and overall brand reputation. In conclusion, the intersection of supervised machine learning and fake review detection offers a compelling framework to safeguard the veracity of online reviews. The remarkable accuracy achieved through SVM classification lays the groundwork for a more transparent and reliable consumer landscape. As technology evolves and new horizons emerge, the pursuit of enhancing fake review detection continues, underscoring the pivotal role of research in shaping the digital consumer experience.

## REFERENCES

[1] Rodrigo Barbado, Oscar Araque, Carlos A. Iglesias, “A framework for fake review detection in online consumer electronics retailers”,Volume 56, Issue 4,July 2020, Pages12341244,ISSN03064573, https://doi.org/10.1016/ j.ipm.2020.03.002

[2] N. Kousika, D. S, D. C, D. B M and A. J, "A System for Fake News Detection by using Supervised Learning Model for Social Media Contents," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp.1042-1047, doi:10.1109/ICICCS51141.2021.9432096.

[3] Ammar Mohammed, Atef Ibrahim, Usman Tariq, Ahmed Mohamed Elmogy “Fake Reviews Detection using Supervised Machine Learning”, January 2021 International Journal of Advanced Computer Science and Applications 12(1), DOI:10.14569/IJACSA.2021.0120169.

[4] Möhring M., Keller B., Schmidt R., Gutmann M., Dacko S. HOTFRED: “A Flexible Hotel Fake Review Detection System". In: Wörndl W., Koo C., Steinmetz J.L. (eds) Information and Communication Technologies in Tourism Springer, Cham.https://doi.org/10.1007/978-3-03065785-7_29.

[5] E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," Seventh International Conference on Innovative Computing Technology (INTECH), pp.107114, doi:10.1109/INTECH.2017.8102442.

[6] Neha S., Anala A., “Fake Review Detection using Classification”,une, International Journal of Computer Applications180(50):16-21,DOI:10.5120/ijca2018917316

[7] R. Hassan and M. R. Islam, "Detection of fake online reviews using semi-supervised and supervised learning," International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-5, doi: 10.1109/ECACE.2019.8679186.

[8] S. Jia, X. Zhang, X. Wang and Y. Liu, "Fake reviews detection based on LDA," 4th International Conference on Information Management (ICIM), 2018, pp. 280-283, doi:10.1109/INFOMAN.2018.8392850.

[9] N. A. Patel and R. Patel, "A Survey on Fake Review Detection using Machine Learning Techniques," 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777594.

[10] Dhairya Patel1, Aishwerya Kapoor2, Sameet Sonawane3 ,"Fake Review Detection using Opinion Mining", https://www.irjet.net/archives/V5/i12/IRJET-V5I12154.pdf

[11] Ott, Myle & Choi, Yejin & Cardie, Claire & Hancock, Jeffrey. (2011) "Finding Deceptive Opinion Spam by Any Stretch of the Imagination".

[12] Dipak R. Kawade et al., "Sentiment Analysis Machine Learning Approach", / International Journal of Engineering and Technolog y(IJET), DOI: 10.21817/ijet/2017/v9i3/1709030151, Vol 9 No 3 JunJul 2017

[13] Sadman, Nafiz & Gupta, Kishor Datta & Sen, Sajib & Poudyal, Subash & Haque, Mohd. (2020). Detect Review Manipulation by Leveraging Reviewer Historical Stylometrics in Amazon, Yelp, Facebook and Google Reviews. 10