

An overview of Existing Approaches for Product Review and Analysis

Chakrapani D S ¹, Shreesha P ², Sneha R Nulageri², Srinidhi G S ² and Srushti S G ²

¹Assistant Professor, Department of Computer Science and Engineering, JNNCE, Shivamogga, Karnataka, India

²UG Students, Department of Computer Science and Engineering, JNNCE, Shivamogga, Karnataka, India

Abstract—In the current digital age, the vast amount of user-generated reviews on e-commerce platforms often overwhelms consumers, making it challenging to make informed purchasing decisions. "ShopInsight - Product Re- views and Analysis Using ML Algorithms" addresses this problem by providing an intelligent, centralized platform that aggregates and analyzes product re- views from multiple online retailers. The rapid growth of e-commerce platforms such as Amazon, Flipkart, Walmart, and Myntra has provided customers with a wide range of product options. However, the availability of the same product across multiple platforms at varying prices, ratings, and reviews makes it difficult for users to identify the best choice. This project proposes the development of a machine learning-based product aggregation and comparison website that integrates product details from multiple e-commerce platforms into a unified interface. The system employs web scraping and API integration to extract key attributes such as product name, price, ratings, reviews, and availability. Natural Language Processing (NLP) techniques are applied to normalize and match similar products across different platforms. A ranking algorithm then sorts the aggregated results, primarily based on ratings, ensuring that users are presented with the most reliable options first. The website provides an efficient, user-friendly interface that reduces decision-making time, improves the online shopping experience, and enables cost-effective purchasing. Future enhancements may include price-drop alerts, sentiment analysis of customer reviews, and personalized product recommendations.

Index Terms—commerce, Product Aggregation, Machine Learning, Web Scraping, Natural Language Processing (NLP), Product Comparison, Recommendation System, Ranking Algorithm, Data Mining, Price Optimization-commerce, Product Aggregation, Machine Learning, Web Scraping, Natural Language Processing (NLP), Product Comparison, Recommendation System, Ranking Algorithm, Data Mining, Price Optimization

I. INTRODUCTION

Today's digital marketplace is so vast that each and every products have multiple choices, consumers often find themselves navigating multiple e-commerce platforms to compare products, prices, and reviews. This project aims to streamline this process by developing a comprehensive website that aggregates product information from various online retailers. The platform will offer users a centralized location to search, compare, and evaluate products across different e-commerce sites, enhancing the shopping experience and enabling more informed purchasing decisions. However, with numerous online platforms available, comparing products across different websites can be time-consuming and challenging. This project aims to develop comprehensive web application that can aggregates product information from various e-commerce platforms, enabling the users to easily compare prices, ratings, and other relevant details in one centralized location. The website will utilize web scraping techniques to gather product data, implement a robust database system for information storage, and employ

machine learning algorithms for accurate product categorization and matching. Key features will include a user-friendly search interface, sorting capabilities based on product ratings, and regular updates to ensure data accuracy. The purpose of this project is to encourage the purchasing decision of a customer to buy a product by comparing the same product in different platforms with different prices, ratings and reviews rather than manually checking the product details in different platforms.

II. LITERATURE SURVEY

Several studies and projects have explored the development of product aggregation and comparison systems to assist online shoppers in making informed purchasing decisions. Existing works emphasize two main strategies for data acquisition: API-based extraction, which provides structured and reliable information but with access restrictions, and web scraping, which offers flexibility but faces challenges such as anti-bot mechanisms and data inconsistency. A key research focus is product matching, where entity resolution techniques like similarity, edit distance, and modern deep learning approaches are applied to identify the same product across multiple platforms.

Additionally, sentiment analysis of customer reviews has been integrated into ranking models to improve reliability beyond numeric ratings. Duplicate detection, price normalization, and category mapping remain critical preprocessing steps for cross-site comparisons. Although mature solutions exist, recent research points to promising opportunities such as leveraging large language models for entity matching, combining sentiment and behavioral signals for ranking allowing the customer to improve their purchasing decisions with the enhanced features.

[1] E-Commerce Customer Behavior Using Machine Learning

Kushank Gupta, Ayush Chaturvedi, Akash Singh, Avinash Rai, Dr. Deepak Asrani, Dr. Komal proposed "E-Commerce Customer Behavior Using Machine Learning" AsraniThe study explores the transformation of e-commerce customer behavior influenced by digital technologies. It highlights how machine learning (ML) techniques can be used to analyze and forecast consumer behavior, especially using data from online reviews. The review focuses on multiple influencing factors such as trust, user experience, social influence, personalization, and post-purchase behavior. It stresses the need for ML in handling big data to accurately predict customer preferences and behavior in the evolving online marketplace.

[2] Products Reviews and Sentimental Analysis System for Ecommerce Website

Pranav Patil, Kiran Nangude, Aditya Rananavare, Rutik Pisal, Pradeep Shinde (Assistant Professor) have done a study on "Products Reviews and Sentimental Analysis System for Ecommerce Website". This paper presents a machine learning-based web application that performs sentiment analysis on product reviews from e-commerce websites. It utilizes Naive Bayes to classify sentiments based on extracted features and helps users understand product sentiments before making purchase decisions. The authors developed a web application that scrapes product reviews from sites, preprocesses the text data, and uses NLP techniques with machine learning algorithms like Naive Bayes and Random Forest to classify sentiments.

[3] Web Scraping Based Site for Product Analysis "Web Scraping Based Site for Product Analysis" was proposed by Ms. C. E. Rajaprabha, B. Noufia Firthous,

C. Suresh Krishna, S. Nisha, R. Srikrishnan. The paper proposes a system that uses ethical web scraping to collect and analyze e-commerce product data. It includes price, descriptions, reviews, and availability. The goal is to help businesses and consumers gain insights and compare products across platforms using a dedicated user interface. This project uses Python tools to scrape and clean product data from e-commerce sites, performs sentiment analysis with VADER, and visualizes insights using charts and word clouds. It's deployed with Flask, featuring automation, security, and scalability.

[4] Ecommerce Assisted by Machine Learning This paper "Ecommerce Assisted by Machine Learning" which is proposed by Nikhil Kandekar, Nehal Pillai, Kshitija Patil, Syed Wilayat, Prof. Mukesh More discusses the integration of machine learning into e-commerce platforms to improve user experience and sales. Key features include product recommendation, dynamic pricing, and fake review detection. These machine learning techniques aim to address limitations of traditional e-commerce such as inaccurate sizing, unverified reviews, and limited customer support, while enhancing personalization and convenience. It presents various machine learning techniques to enhance e-commerce, including product recommendations using filtering methods, fake review detection through sentiment analysis aimed at improving user experience and boosting business performance.

[5] Web Scraping for E-Commerce Websites Gandhe Vineeth Kumar, Hema M S, Aishwarya R, K R Mamatha proposed the "Web Scraping for E-Commerce Websites" focuses on price monitoring and prediction across multiple e-commerce sites using web scraping and machine learning. The proposed tool scrapes prices, alerts users of price drops, and predicts future pricing trends using regression models. This methodology involves scraping product data using BeautifulSoup and Selenium, cleaning and structuring it with pandas, and training multiple regression models. Random Forest was chosen for accurate price prediction, and the system includes email alerts for price drops.

[6] Prediction of Product Rating Based on Polarized Reviews Using Supervised Machine Learning

Raheel Ahmad Khan, Abdul Mannan, Naeem Aslam proposed the "Prediction of Product Rating Based on Polarized Reviews Using Supervised Machine Learning" where the E-commerce platforms allow users to leave textual reviews and star ratings. However, not all reviews include star-level ratings, which are generally preferred for quantitative analysis. This paper proposes an approach to predict numerical ratings from textual reviews using supervised machine learning and deep learning models, leveraging polarization analysis via a lexicon-based method. The approach accounts for the impact of negative or negation words, often overlooked in previous studies. Experiments using Amazon review datasets showed that review polarization can significantly improve rating prediction accuracy.

[7] An approach to improve analysis of product reviews using sentiment analysis

The work done by Geetanjali Tanwar, Swati Tripathi, Divya Shakya, Deepak Gaur "An approach to improve analysis of product reviews using sentiment analysis" Customer opinions are crucial for business success, as reviews and ratings help analyze products and refine marketing strategies. Traditional methods like star ratings and textual feedback are used, with textual feedback being more descriptive and reliable. This paper discusses the implementation of sentiment analysis using the Naïve Bayes model to determine product sentiment ratings, integrates sarcasm detection through a Sequential model, and incorporates aspect-based sentiment analysis (ABSA) to evaluate products on various aspects. This methodology aids sellers and customers in understanding product performance in the market, ultimately clarifying product success.

[8] Amazon Product Reviews: Sentiment Analysis Using Supervised Learning Algorithms

The paper "Amazon Product Reviews: Sentiment Analysis Using Supervised Learning Algorithms" was proposed by Mohibullah Hawlade, Arjan Ghosh, Zaoyad Khan Raad, Wali Ahad Chowdhury, Md. Sazzad Hossain Shehan, Faisal Bin Ashraf explores how sentiment analysis and supervised learning can be applied to Amazon electronic product reviews. Due to the overwhelming volume of customer reviews, sentiment analysis becomes essential in understanding overall product perception. The authors used multiple supervised learning classifiers—Support Vector Machine (SVM), Naive Bayes, Decision Tree, Random Forest, Logistic Regression, and Multi-Layer Perceptron (MLP)—on Amazon product reviews. Three preprocessing techniques—TF-IDF, Bag of Words (BoW), and Word2Vec—were evaluated.

[9] Sentiment Analysis for E-commerce Product Reviews by Deep Learning Model of Bert BiGRU- Softmax

Yi Liu, Jiahuan Lu, Jie Yang, Feng Mao (Affiliation: Management School, Hangzhou Dianzi University, Hangzhou, China) a study on "Sentiment Analysis for E-commerce Product Reviews by Deep Learning Model of Bert BiGRU-Softmax". This study addresses sentiment analysis for e-commerce product reviews, a vital area for businesses to understand consumer attitudes. The paper introduces a hybrid

deep learning model called Bert-BiGRU-Softmax, combining BERT for feature extraction, Bidirectional GRU for semantic encoding, and a Softmax layer with attention mechanisms for sentiment classification. Using a dataset of over 500,000 reviews, the proposed model achieves an accuracy of over 95.5 surpassing traditional models. Trained on large mobile review datasets, it used BERT embeddings and token masking for preprocessing. Compared to models for faster convergence, and strong feature level sentiment detection.

[10] Scraping and Visualization of Product Data from E-commerce Websites

Dr. V. Srividhya, P. Megala have done study on "Scraping and Visualization of Product Data from E-commerce Websites" which focuses on extracting product data from e-commerce websites (web scraping), storing the data in a structured format, analyzing it statistically, and visualizing it. Web scraping is defined as the process of extracting unstructured information from websites and converting it into a structured form such as CSV. The paper implements a three- phase methodology: scraping product data, performing statistical data analysis, and visualizing the results using various charts. This approach supports efficient data extraction and enables businesses and individuals to analyze competitive pricing and product availability across platforms. It is structured into three main phases. First, web scraping was performed using Python's requests and BeautifulSoup4 libraries to extract product names, prices, and ratings from Flipkart and Snapdeal, storing the data in CSV format. Next, statistical analysis using NumPy calculated metrics like mean, standard deviation, max, min, and count to examine pricing trends.

[11] Tool For Review Analysis Of Product

Ruchit Furia, Kaustubh Gaikwad, Kunal Mandalya, Anand Godbole proposed the paper "Tool For Review Analysis Of Product". Today people buy products online rather than manual shopping. These e-commerce websites let their customers write product feedback and review in the form of rating. The review given by these customers helps the company personnel of the product understand where their product stands in the market. At the same time, it helps other fellow customers decide whether this product is suitable for them to buy. Reviews can be fraudulent in a way that they demote the product or advertise different products which might misdirect the customer about the quality of the product. Hence, we have worked towards developing a tool that will classify the reviews as fake or genuine and provide it to the user. The proposed tool will operate on the principle of the Weighted Ensemble Classifier.

Reference	Publication Date	Techniques Used	Dataset	Outcome/ Performance
E-Commerce Customer Behavior Using Machine Learning	2024	Naive Bayes, J48, Logistic Regression, K-Means clustering	Predicts and segments customer behavior	Handles big data, improves personalization and marketing
Products Reviews and Sentimental Analysis System for Ecommerce Website	2024	Naive Bayes, Random Forest, web scraping, NLP	Real-time sentiment analysis web app	Cloud-based UI, supports purchase decisions
Web Scraping Based Site for Product Analysis	2024	Python scraping, VADER, Flask	Product comparison across platforms	Automated, secure, scalable
Ecommerce Assisted by Machine Learning	2022	Filtering, chatbots, fake review detection, 3D sizing	ML features for personalized e-commerce	Comprehensive enhancements
Web Scraping for E-Commerce Websites	2022	Beautiful Soup, Selenium, regression	Price prediction and alerts	Accurate forecasting, alerts
Prediction of Product Rating Based on Polarized Reviews	2022	Lexicon polarity, ML & DL models	Predicts ratings from text reviews	Captures negation, improves accuracy
An Approach to Improve Analysis of Product Reviews	2022	Naive Bayes, sarcasm detection, ABSA	Deeper sentiment insights from reviews	Aspect analysis, sarcasm detection
Amazon Product Reviews: Sentiment Analysis Using Supervised Learning	2021	SVM, Naive Bayes, MLP, BoW, TF-IDF	Sentiment classification on Amazon data	Word2Vec underperformed
Sentiment Analysis by Bert-BiGRU-Softmax	2020	BERT + BiGRU + Softmax w/ attention	Deep learning-based sentiment classification	95.5% accuracy, less interpretable
Tool for Review Analysis of Product	2020	Weighted ensemble (SVM, NB, KNN), Scrapy, Flask	Fake review detection tool	Improves review trust

III. SUMMARY

The product reviews and analysis using ML algorithms emphasizes how ML can handle massive data volumes from e-commerce platforms, particularly on-line reviews. It explores how classification algorithms such as Naïve Bayes, J48, and Logistic Regression can predict customer behavior, while K-Means clustering is used to group similar users. Key influencing factors like trust, personalization, and user experience are discussed, with the overall goal being customer retention and smarter marketing strategies. This aids users in making informed purchase decisions based on collective opinion. It introduces a system that ethically scrapes e-commerce data, analyzes it using VADER sentiment analysis, and presents insights through charts and word clouds. This approach focuses on giving both businesses and consumers comparative product data across different platforms. One study focuses on using classification algorithms like Naïve Bayes, J48 Decision Tree, and Logistic Regression, as well as clustering techniques like K-Means, to predict and segment customer behavior from big data. Several papers explore sentiment analysis, which classifies reviews as positive or negative. Methodologies include using Naive Bayes, Random Forest, and more advanced models like the Bert-BiGRU-Softmax deep learning model, which achieved 95.5% accuracy. Another study uses a lexicon-based polarization analysis to predict numerical star ratings from textual reviews, specifically handling negative words. A tool was developed using a Weighted Ensemble Classifier that combines SVM, Naive Bayes, and KNN to classify reviews as fake or genuine, which helps customers avoid misleading information. Multiple studies detail the use of web scraping tools like Python's BeautifulSoup and Selenium to extract unstructured data from websites and convert it into a structured format like CSV. An additional focus is on review authenticity, where ensemble classifiers (e.g., combining SVM, Naïve Bayes, and KNN) are developed to detect fake vs. genuine reviews, protecting users from deceptive information. The integration of machine learning in product review analysis enables more informed purchasing decisions, improved marketing strategies, and enhanced consumer trust, marking a significant step toward intelligent and ethical e-commerce analytics. This data is then used for statistical analysis and visualized with charts to compare pricing and other trends across different platforms.

REFERENCES

- [1] Kushank Gupta¹, Ayush Chaturvedi², Akash Singh³, Avinash Rai¹, Dr. Deepak Asrani, and Dr. Komal Asrani, "E-Commerce Customer Behavior Using Machine Learning," International Journal of Innovative Research in Computer Science Technology (IJIRST), Volume-12, Special Issue-1, 2024, pp. 324-332.
- [2] Pranav Patil; Kiran Nangude²; Aditya Rananavare³; Rutik Pisal; Pradeep Shinde, "Products Reviews and Sentimental Analysis System for Ecommerce Website," International Journal of Innovative Science and Research Technology, Volume: 9, Issue: 4, 2024.
- [3] Ms. C. E. Rajaprabha¹, B. Noufia Firthous², C. Suresh Krishna³, S. Nisha, R. Srikrishnan, "Web Scraping Based Site for Product Analysis," International Journal for Multidisciplinary Research (IJFMR), Volume: 6, Issue 3, 2024.
- [4] Nikhil Kandekar¹, Nehal Pillai, Kshitija Patil³, Syed Wilayat, Prof. Mukesh More, "Ecommerce Assisted by Machine Learning," International Journal For Research in Applied Science and Engineering Technology, Volume: 10, Issue: III, 2022.
- [5] Gandhe Vineeth Kumar, Hema M S², Aishwarya R³, K R Mamatha, "Web Scraping for E-Commerce Websites," International Journal For Research in Applied Science Engineering Technology, Volume: 10, Issue: VI, 2022.
- [6] Raheel Ahmad Khan¹, Abdul Mannan²³, Naeem Aslam¹, "Prediction of Product Rating based on Polarized Reviews using Supervised Machine Learning," FAST VFAST Transactions on Software Engineering, Volume 10, Number 4, October-December 2022, pp: 01-09.
- [7] Geetanjali Tanwar, Swati Tripathi, Divya Shakyaa, Deepak Gaur, "An approach to improve analysis of product reviews using sentiment analysis," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2022, pp. 1526-1530.
- [8] Mohibullah Hawlader, Wali Ahad Chowdhury, Arjan Ghosh, Md. Sazzad Hossain Shehan, Zaoyad Khan Raad, Faisal Bin Ashraf, "Amazon Product Reviews: Sentiment Analysis Using Supervised Learning Algorithms," 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), 2021.
- [9] Ruchit Furia, Kaustubh Gaikwad, Kunal Mandalya, Anand Godbole, "Tool For Review Analysis Of Product," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2020, pp. 1-5.
- [10] Scraping and Visualization of Product Data from Ecommerce Websites (IJCE 2019)
- [11] Khatter, H. (2022). Web Scraping based Product Comparison Model for E-Commerce Websites.
- [12] Dey, R. (2022). Product recommendation system using machine learning through big data in e-commerce website.
- [13] E. Manohar, P. Jenifer, M. S. Nisha, and B. Benita, "A Collective Data Mining Approach to Predict Customer Behavior," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021.
- [14] V. Shrirame, J. Sabade, H. Soneta and M. Vijayalakshmi, "Consumer Behavior Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020.
- [15] B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. HeGuelton, F. Oble' and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in IEEE Access, vol. 9, pp.