FECSIT, (2025)

Frontiers in Emerging Computer Science and Information Technology (Open Access)

Volume 02, Issue 05, May 2025,
Publish Date: 01-05-2025
PageNo.01-07

Leveraging Web Data Harvesting for Product Recommendation Systems: A Comprehensive Review of Methodologies and Use Cases

Prof. Elizabeth Schneider
Department of Computer Science, University College London, London, United Kingdom

Prof. Thomas J. Carter
School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

ABSTRACT

Product recommendation systems have become essential tools for enhancing user engagement and driving sales across e-commerce platforms. With the proliferation of online data sources, web data harvesting offers powerful capabilities to enrich recommendation models with real-time, diverse, and contextually relevant information. This paper presents a comprehensive review of methodologies and use cases related to leveraging web data harvesting for product recommendation. It examines key techniques, including web scraping, API integration, and semantic enrichment, outlining their roles in collecting product metadata, user reviews, competitor pricing, and emerging trends. Additionally, it explores how harvested data can be integrated into collaborative filtering, content-based, and hybrid recommendation frameworks to improve personalization and accuracy. The review also discusses ethical considerations, legal compliance, data quality challenges, and strategies for scalable implementation. By synthesizing current practices and applications, this work aims to guide researchers and practitioners in developing more effective, data-driven recommendation systems.

KEYWORDS: Web data harvesting, product recommendation systems, web scraping, collaborative filtering, content-based filtering, hybrid recommender systems, e-commerce analytics, data integration, personalization, big data in marketing.

INTRODUCTION

In the contemporary digital economy, online platforms have become pivotal avenues for commerce, information exchange, and social interaction. A key feature driving consumer engagement and sales on these platforms is the product recommendation system [Kang & Wang, 2024; Roy & Dutta, 2022]. These systems, powered by sophisticated algorithms, aim to predict user preferences and suggest items, services, or content that are most likely to be of interest, thereby enhancing user experience, increasing sales, and reducing information overload [Soni et al., 2024]. The effectiveness of a recommendation system hinges critically on the quantity, quality, and relevance of the data it can access and process.

Traditional recommendation systems often rely on internal user interaction data (e.g., purchase history, ratings, clicks) or explicit user preferences. However, a vast ocean of publicly available, unstructured data exists on the web, offering rich and diverse insights into product attributes, consumer opinions, market trends, and competitive landscapes [Barbera et al., 2023; Guyt et al., 2024; de Haan et al., 2024]. Web scraping, the automated extraction of data from websites, provides a powerful mechanism to tap into this external data reservoir [Lotfi et al., 2021; Beveridge & Gallagher, 2021]. By programmatically collecting information that is not readily available through APIs or structured databases, web scraping can significantly augment the data foundation for recommendation systems, enabling more nuanced, comprehensive, and real-time insights.

The integration of web-scraped data into recommendation engines holds the potential to address several limitations of internal-data-only approaches, such as the cold-start problem (recommending to new users or new items) and data sparsity. It also allows for a reduction of information asymmetry in e-commerce, providing a broader view of market offerings [Hadasik, 2024]. Despite its considerable advantages, web scraping for recommendation systems also introduces technical, ethical, and legal challenges, including dealing with dynamic website structures, anti-scraping measures, and respecting terms of service [Tabaku & Ali, 2021].

This article aims to provide a comprehensive review of the techniques and applications of leveraging web data harvesting (web scraping) for developing and enhancing product recommendation systems. We will delve into the methodologies involved in acquiring and processing web data, explore various applications across different domains, and discuss the inherent challenges and future directions in this evolving field.

METHODS

This article adopts a conceptual review methodology to synthesize existing knowledge on web scraping techniques and their applications in product recommendation systems. The approach involves identifying, categorizing, and interpreting relevant studies from the provided set of references to build a coherent understanding of the subject matter. No new empirical data collection or analysis was performed for this review.

Data Source and Selection

The primary data source for this review consists of the list of academic and industry publications provided by the user. These references were chosen for their direct relevance to:

- Web scraping techniques and methodologies [Lotfi et al., 2021; Beveridge & Gallagher, 2021; Fikri et al., 2022; Nurkholis et al., 2023; Park & Shin, 2022].
- The application of web data in general business research and retailing [Guyt et al., 2024; de Haan et al., 2024; Barbera et al., 2023].
- The principles and various types of recommender systems [Roy & Dutta, 2022; Kang & Wang, 2024; Soni et al., 2024; Liu et al., 2024].
- Specific instances where web scraping has been explicitly used to inform or build recommendation systems across different product or service domains [Campos Macias et al., 2022; Pavitha et al., 2022; Reynaldi & Istiono, 2023; Miao et al., 2023; Rostami et al., 2024; Abolghasemi et al., 2024; Rodrigues et al., 2024; Vijayakumar & Jagatheeshkumar, 2024; Akshay et al., 2024].
- Challenges associated with web scraping and data integration [Tabaku & Ali, 2021; Hadasik, 2024; Kudo et al., 2022].

Data Extraction and Synthesis

Information from each relevant reference was extracted and conceptually grouped into themes. The key elements extracted included:

- Definition and Purpose of Web Scraping: Understanding its role in data acquisition.
- Technical Approaches to Web Scraping: Tools, programming languages (e.g., Python), and fundamental steps involved in crawling, parsing, and extracting data [Fikri et al., 2022; Park & Shin, 2022].
- Types of Data Extracted: Product specifications, prices, reviews, user comments, social media data, and other unstructured text [Guyt et al., 2024].
- Data Preprocessing for Recommendation Systems: Methods for cleaning, normalizing, and structuring the scraped data, including the creation of master data [Kudo et al., 2022].
- Integration with Recommendation Algorithms: How the extracted data is used as input for various recommendation models (e.g., content-based filtering, collaborative filtering, hybrid approaches).
- Specific Applications: Documenting diverse case studies and examples where web scraping has been successfully applied to build or enhance different types of recommender systems.
- Identified Challenges: Technical hurdles (e.g., website changes), legal/ethical considerations, and data quality issues (e.g., unstructured data handling [de Haan et al., 2024]).

The extracted insights were then synthesized to identify recurring patterns, categorize different techniques, and highlight the benefits and drawbacks of web scraping in the context of product recommendations. The process involved cross-referencing information across multiple sources to ensure a comprehensive and balanced perspective.

RESULTS

The review of the literature reveals a robust landscape of web scraping techniques and their diverse applications in enhancing product recommendation systems. The findings demonstrate that web data harvesting is a versatile tool capable of enriching recommendation models with external, real-time, and comprehensive information.

1. Web Scraping Techniques and Data Collection

Web scraping fundamentally involves automating the process of accessing web pages, extracting relevant information, and storing it in a structured format.

- Core Process: Typically, this involves sending HTTP requests to web servers, parsing the HTML or XML content of the response, and extracting specific data points using selectors (e.g., CSS selectors, XPath). Programming languages like Python are frequently used for this purpose due to their rich libraries (e.g., BeautifulSoup, Scrapy) [Fikri et al., 2022; Nurkholis et al., 2023]. Novel approaches, such as "scratch programming blocks," are also being developed to simplify web scraping for educational or less technical users [Park & Shin, 2022].
- Crawling vs. Scraping: The process often begins with web crawling, where automated bots navigate links to

discover and retrieve web pages, followed by scraping to extract specific data from those pages [Lotfi et al., 2021]. This allows for the collection of large volumes of data across various sites.

- Data Types: The data extracted can vary widely depending on the recommendation task. For product recommendations, this commonly includes:
    - Product Attributes: Names, descriptions, categories, specifications, images.
    - Pricing Information: Current prices, historical prices, discount details [Meyberg et al., 2024].
    - User-Generated Content (UGC): Product reviews, ratings, comments, and forum discussions [Guyt et al., 2024].
    - External Context: Information from news articles [Fikri et al., 2022], social media (e.g., Twitter data for e-commerce recommenders) [Akshay et al., 2024], or specialized knowledge bases.
- Data Preprocessing: Raw scraped data is often unstructured or semi-structured and requires significant cleaning, normalization, and transformation. This includes text cleaning, entity recognition, and sometimes, the creation of master data through semi-structuring and extraction methods to ensure consistency and usability for downstream recommendation algorithms [Kudo et al., 2022]. Research on unstructured data in business highlights the need for a structured approach to its utilization [de Haan et al., 2024].

2. Applications of Web Scraping in Recommendation Systems

Web scraping has been applied to various types of recommendation systems, enriching them with external data and enabling novel recommendation approaches.

- Product Recommendations (General E-commerce):
    - By scraping product details, prices, and competitor information, recommendation systems can offer more competitive suggestions or identify market gaps. This directly addresses the reduction of information asymmetry for consumers in e-commerce [Hadasik, 2024].
    - Frequent Item Set Mining fusion algorithms, integrated with web data, can provide efficient and accurate personalized product recommendations [Kang & Wang, 2024].
- Content-Based Recommendations:
    - Movie and Anime Recommendations: Web scraping user reviews, plots, genres, and cast information allows for content-based filtering, where recommendations are based on similarities between items a user has liked in the past [Pavitha et al., 2022; Reynaldi & Istiono, 2023]. Sentiment analysis on scraped reviews can further refine these recommendations [Pavitha et al., 2022].
    - E-content Recommendations: For educational platforms, scraping e-content metadata and user feedback (if publicly available) can inform systems that adapt to a user's learning capability, leading to enhanced learning experiences [Vijayakumar & Jagatheeshkumar, 2024].
- Social and Group Recommendations:
    - Group Recommendations: Graph neural approaches leverage relationships and preferences scraped from social networks or review platforms to make recommendations for groups, considering pairwise preferences [Abolghasemi et al., 2024; Rostami et al., 2024].
    - Restaurant and Food Recommendations: Integrating users' long-term and short-term interests with knowledge graphs constructed from scraped data can significantly improve restaurant recommendations [Miao et al., 2023]. Healthy food recommendations can also benefit from deep social community detection approaches on aggregated user data [Rostami et al., 2024].
- Specialized Recommenders:
    - Technology Recommender Systems: Combining web crawling with Natural Language Processing (NLP) on technology-related websites (e.g., academic papers, tech blogs) can build systems that recommend relevant technologies [Campos Macias et al., 2022].
    - Fragrance Recommendation: Utilizing Graph Neural Networks on consumer feedback obtained via web scraping can even craft fragrances based on consumer preferences [Rodrigues et al., 2024].
    - Operational and Maintenance Recommendations: For complex products, web scraping can feed into "equipment portraits" to provide personalized operation and maintenance recommendations within a product-service system framework [Ren et al., 2023].
- General Web Data Applications:
    - Beyond direct product recommendations, web scraping supports broader applications that can indirectly inform recommendation systems, such as constructing multilingual e-learning ontologies [Barwary et al., 2023], identifying stroke outcome measures for medical research [Lee et al., 2022; Sabesan et al., 2023], analyzing

public perceptions from news [Gebretensae, 2024], and powering information systems for strategic watch [Ghoul et al., 2024]. The value of web data scraping has also been demonstrated in applications like analyzing TripAdvisor data [Barbera et al., 2023] and enhancing travel experiences with AI [Londhe et al., 2024].

These examples illustrate the versatility of web scraping in enriching the data landscape for a wide array of recommendation systems, moving beyond internal user data to leverage the vast external intelligence available on the web.

DISCUSSION

The extensive applications outlined in the results section firmly establish web scraping as a powerful and increasingly indispensable technique for enhancing product recommendation systems. Its capacity to collect vast amounts of external, often real-time, data from diverse web sources offers significant advantages over relying solely on internal user interaction logs. This external data can provide richer product attributes, broader market insights, and more nuanced user preferences, effectively tackling challenges like the cold-start problem and data sparsity that plague many traditional recommenders.

Advantages of Web Scraping for Recommendations

1. Rich and Diverse Data Sources: Web scraping allows access to a heterogeneous array of data, including product specifications, competitive pricing, extensive customer reviews, sentiment analyses, and even social media discussions related to products [Guyt et al., 2024]. This richness enables the construction of more comprehensive user profiles and item representations.
2. Real-time Insights: Unlike static datasets, web scraping can continuously collect updated information, allowing recommendation systems to adapt to trending products, changing prices, or evolving consumer sentiments in near real-time. This dynamic capability is crucial in fast-paced e-commerce environments.
3. Reduced Information Asymmetry: By gathering external market data, businesses can gain a more complete understanding of their competitive landscape and consumer demand, while consumers can receive recommendations informed by a wider array of options and external feedback, reducing information gaps [Hadasik, 2024].
4. Enabling Novel Recommendation Approaches: The availability of rich unstructured data facilitates the application of advanced Natural Language Processing (NLP) and machine learning techniques, such as sentiment analysis [Pavitha et al., 2022], knowledge graphs [Miao et al., 2023], and Graph Neural Networks [Abolghasemi et al., 2024; Rodrigues et al., 2024], leading to more sophisticated and personalized recommendations.

Challenges and Limitations

Despite these advantages, the adoption of web scraping for recommendation systems faces several significant challenges:

1. Technical Robustness and Maintenance: Websites frequently change their structure, HTML elements, or introduce anti-scraping measures (e.g., CAPTCHAs, IP blocking, dynamic content loading) [Tabaku & Ali, 2021]. This requires constant monitoring and adaptation of scraping scripts, making the process labor-intensive and prone to breakage.
2. Data Quality and Preprocessing: Web-scraped data is inherently unstructured, noisy, and inconsistent. It often contains irrelevant information, duplicates, or errors. Extracting meaningful features and transforming raw data into a structured format suitable for recommendation algorithms is a complex and resource-intensive task [Kudo et al., 2022; de Haan et al., 2024]. This often requires advanced NLP techniques and robust data cleaning pipelines.
3. Legal and Ethical Considerations: Web scraping exists in a legal grey area. Issues such as copyright infringement, violation of website terms of service, and privacy concerns (especially when scraping personal information or identifiable user data) pose significant risks [Tabaku & Ali, 2021; Hadasik, 2024]. Ethical guidelines and legal precedents vary across jurisdictions, requiring careful navigation.
4. Scalability and Performance: Scraping large volumes of data from numerous websites can be computationally intensive and time-consuming. Ensuring the scalability of scraping infrastructure and the efficiency of data processing pipelines is crucial for real-time recommendation systems.
5. Distinguishing Valuable Data: Not all publicly available web data is equally valuable for recommendations. Identifying relevant signals amidst noise and effectively integrating heterogeneous data sources remains a challenge.

Future Directions

The future of leveraging web data harvesting for product recommendations lies in addressing these challenges through advanced techniques and ethical considerations:

1. Intelligent and Adaptive Scraping: Developing AI-powered scraping agents that can autonomously adapt to website structural changes, bypass anti-scraping measures more intelligently, and identify relevant data

fields dynamically. This could reduce manual maintenance efforts.
2. Advanced NLP and Semantic Understanding: Further integrating state-of-the-art NLP models (e.g., large language models) for deeper semantic understanding of scraped textual data, extracting subtle sentiments, product features, and relationships that might otherwise be missed. This is particularly relevant for diverse text applications like social media discourses [Shahade et al., 2023] or general innovation articles [Ghoul et al., 2024].
3. Cross-platform Data Fusion: Research into sophisticated methods for integrating and fusing data from multiple scraped web sources with internal data and other external datasets (e.g., IoT data) to create richer, more comprehensive user and item profiles. Graph representation learning could play a significant role in integrating platforms [Putrama & Martinek, 2023].
4. Ethical Frameworks and Legal Compliance: Developing clear industry-wide ethical guidelines and best practices for web scraping, as well as working towards robust legal frameworks that balance data accessibility with privacy and intellectual property rights.
5. Real-time Recommendation Architectures: Designing scalable and efficient architectures that can ingest, process, and integrate scraped data in real-time to provide dynamic and highly relevant recommendations to users.
6. Domain-Specific Applications and Decision Support: Expanding the application of web scraping to highly specialized recommendation contexts, such as personalized operation and maintenance recommendations for complex industrial products [Ren et al., 2023] or integrating with ICT-based decision support systems for precision agriculture [Flores Cayuela et al., 2022] and food safety risk assessment [Talari et al., 2022]. The use of web scraping for "web-based decision support systems" is a growing area [Talari et al., 2022].
7. Enhanced User-Centricity: Focusing on how scraped data can contribute to a more profound understanding of user behavior and preferences, leading to highly personalized recommendations that align with user's specific needs, such as for travel experiences [Londhe et al., 2024] or even in complex scenarios like designing earthquake scenarios based on user-centered recommendations [Marti et al., 2023].

In conclusion, web scraping remains a vital technique for enriching product recommendation systems with invaluable external data. While technical and ethical hurdles persist, ongoing advancements in AI, NLP, and data integration promise to unlock even greater potential. By navigating these complexities thoughtfully, the integration of web data harvesting will continue to drive the evolution of more intelligent, personalized, and effective recommendation experiences across diverse online domains.

REFERENCES

[1] Abolghasemi, R., Viedma, E. H., Engelstad, P., Djenouri,Y., & Yazidi, A. (2024). A Graph Neural Approachfor Group Recommendation System Based onPairwise Preferences. Information Fusion, 107, 102343. https://doi.org/10.1016/j.inffus.2024.102343
[2] Akshay, R. K. P., Rinu, R. T. R., Paul, R. T., & Joy, J.(2024). E-Commerce Recommender System onTwitter using Directed Multilayer Network. https://doi.org//10.21203/rs.3.rs-4223941/v1
[3] Barbera, G., Araujo, L., & Fernandes, S. (2023). TheValue of Web Data Scraping: An Application toTripAdvisor. Big Data and Cognitive Computing, 7(3), 121. https://doi.org/10.3390/bdcc7030121
[4] Barwary, M. J., Jacksi, K., & Al-Zebari, A. (2023).Constructing a Multilingual E-Learning Ontologythrough Web Crawling and Scraping. InternationalJournal of Communication Networks andInformation Security (IJCNIS), 15(3), 137-153. https://doi.org/10.17762/ijcnis.v15i3.6241
[5] Beveridge, A., Studies, W., & Gallagher, J. (2021).Project-Oriented Web Scraping in TechnicalCommunication Research. Journal of Business andTechnical Communication, 36(2), 231-250. https://doi.org/10.1177/10506519211064619
[6] Campos Macias, N., Düggelin, W., Ruf, Y., & Hanne, T.(2022). Building a Technology RecommenderSystem Using Web Crawling and Natural LanguageProcessing Technology. Algorithms, 15(8), 272. https://doi.org/10.3390/a15080272
[7] de Haan, E., Padigar, M., El Kihal, S., Kübler, R., &Wieringa, J. E. (2024). Unstructured Data ResearchIn Business: Toward A Structured Approach. Journal of Business Research, 177, 114655. https://doi.org/10.1016/j.jbusres.2024.114655
[8] Fikri, M. R., Handayanto, R. T., & Irwan, D. (2022).Web Scraping Situs Berita Menggunakan BahasaPemograman Python. Journal of Students'Research in Computer Science, 3(1), 123-136. https://doi.org/10.31599/jsrcs.v3i1.1514
[9] Flores Cayuela, C. M., González Perea, R., CamachoPoyato, E., & Montesinos, P. (2022). An Ict-BasedDecision Support System for Precision IrrigationManagement in Outdoor Orange and GreenhouseTomato Crops. Agricultural Water Management, 269, 107686. https://doi.org/10.1016/j.agwat.2022.107686
[10] Gaffey, J., Rajauria, G., McMahon, H., Ravindran, R.,Dominguez, C., Ambye-Jensen, M., Souza, M. F.,Meers, E., Aragonés, M. M., Skunca, D., &Sanders, J. P. M. (2023). Green BiorefinerySystems for The Production Of Climate-SmartSustainable Products from Grasses, Legumes andGreen Crop Residues. Biotechnology Advances,

66,108168.
https://doi.org/10.1016/j.biotechadv.2023.108168

[11] Gebretensae, Y. (2024). Understanding the CulturalCrisis: A Web Scraping Analysis of COVID-19Vaccine Perceptions and Media Patterns. ResearchSquare. https://doi.org/10.21203/rs.3.rs-4297475/v1

[12] Ghoul, D., Patrix, J., Oulmakki, O., & Verny, J. (2024).Information System of Strategic Watch to RankInnovation Article by Machine Learning Models. Procedia Computer Science, 234, 772-779. https://doi.org/10.1016/j.procs.2024.03.063

[13] Guyt, J. Y., Datta, H., & Boegershausen, J. (2024).Unlocking the Potential of Web Data for RetailingResearch. Journal of Retailing, 100(1), 130-147. https://doi.org/10.1016/j.jretai.2024.02.002

[14] Hadasik, B. (2024). Reduction of Information AsymmetryIn E-Commerce: the Web Scraping Approach.

[15] Kang, L., & Wang, Y. (2024). Efficient and AccuratePersonalized Product Recommendations ThroughFrequent Item Set Mining Fusion Algorithm. Heliyon, 10(3), 25044. https://doi.org/10.1016/j.heliyon 2024.e25044

[16] Kudo, T., Yamamoto, T., & Watanabe, T. (2022). Three-Step Master Data Creation Method from Big Data:Scraping, Semi-Structuring, and Extraction. Procedia Computer Science, 207, 360-369. https://doi.org/10.1016/j.procs.2022.09.070

[17] Lee, M. J., Kang, J., Hreha, K., & Pappadis, M. (2022).A Novel Web Scraping Approach to Identify StrokeOutcome Measures: A Feasibility Study. Archivesof Physical Medicine and Rehabilitation, 103(3),30. https://doi.org/10.1016/j.apmr.2022.01.082

[18] Liu, Q., Yu, M., & Bai, M. (2024). A Study on ARecommendation Algorithm Based on SpectralClustering and Gru. IScience, 27(2), 108660. https://doi.org/10.1016/j.isci.2023.108660

[19] Londhe, K., Dharmadhikari, N., Zaveri, P., & Sakoglu,U. (2024). Enhanced Travel Experience usingArtificial Intelligence: A Data-driven Approach. Procedia Computer Science, 235, 1920-1928. https://doi.org/10.1016/j.procs.2024.04.182

[20] Lotfi, C., Srinivasan, S., Ertz, M., & Latrous, I. (2021).Web Scraping Techniques and Applications: ALiterature Review. SCRS Conference Proceedingson Intelligent Systems, 381-394. https://doi.org/10.52458/978-93-91842-08-6-38

[21] Mahmuddah, L. A. A., Wibowo, S. A., & Budiman, G.(2022). Generating Information of Url Based onWeb Scraping Using Yolov3 Face RecognitionTechnology. IJAIT (International Journal ofApplied Information Technology), 5(2), 112-122. https://doi.org/10.25124/ijait.v5i02.3910

[22] Marti, M., Dallo, I., Roth, P., Papadopoulos, A. N., &Zaugg, S. (2023). Illustrating the Impact ofEarthquakes: Evidence-Based and User-CenteredRecommendations on How to Design EarthquakeScenarios and Rapid Impact Assessments. International Journal of Disaster Risk Reduction, 90, 103674. https://doi.org/10.1016/j.ijdrr.2023.103674

[23] Meyberg, C., Rendtel, U., & Leerhoff, H. (2024). FlatRent Price Prediction in Berlin with Web Scraping. AStA Wirtschafts- Und Sozialstatistisches Archiv, 18(2), 245-278. https://doi.org/10.1007/s11943-024-00340-6

[24] Miao, L., Li, X., Yu, D., Ren, Y., Huang, Y., & Cao, S.(2023). Integrating Users' Long-Term and Short-Term Interests with Knowledge Graph to ImproveRestaurant Recommendation. Journal of King SaudUniversity - Computer and Information Sciences, 35(9), 101735. https://doi.org/10.1016/j.jksuci.2023.101735

[25] Nurkholis, A., Fernando, Y., & Ans, F. A. (2023). MetodeVector Space Model Untuk Web Scraping PadaWebsite Freelance. INTI Nusa Mandiri, 18(1), 52-58. https://doi.org/10.33480/inti.v18i1.4266

[26] Park, Y., & Shin, Y. (2022). Novel Scratch ProgrammingBlocks for Web Scraping. Electronics, 11(16),2584. https://doi.org/10.3390/electronics11162584

[27] Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit,A., Patel, A., & Shashidhar, R. (2022). MovieRecommendation and Sentiment Analysis UsingMachine Learning. Global TransitionsProceedings, 3(1), 279-284. https://doi.org/10.1016/j.gltp.2022.03.012

[28] Pawar, S., & Chiplunkar, N. (2022). Dynamic Searchingof Web Services Through Web Scraping.

[29] Putrama, I. M., & Martinek, P. (2023). IntegratingPlatforms through Content-Based GraphRepresentation Learning. International Journal ofInformation Management Data Insights, 3(2),100200. https://doi.org/10.1016/j.jjimei.2023.100200

[30] Rejeb, A., Rejeb, K., Appolloni, A., Treiblmaier, H., &Iranmanesh, M. (2024). Exploring The Impact ofChatgpt on Education: A Web Mining AndMachine Learning Approach. The InternationalJournal of Management Education, 22(1), 100932. https://doi.org/10.1016/j.ijme.2024.100932

[31] Ren, S., Shi, L., Liu, Y., Cai, W., & Zhang, Y. (2023). APersonalised Operation and Maintenance Approachfor Complex Products Based oOn EquipmentPortrait Of Product-Service System. Robotics andComputer-Integrated Manufacturing, 80, 102485. https://doi.org/10.1016/j.rcim.2022.102485

[32] Reynaldi, & Istiono, W. (2023). Content-based Filteringand Web Scraping in Website for RecommendedAnime. Asian Journal of Research in ComputerScience, 15(2), 32-42. https://doi.org/10.9734/ajrcos/2023/v15i2318

[33] Rodrigues, B. C. L., Santana, V. V., Queiroz, L. P.,Rebello, C. M., & B. R. Nogueira, I. (2024).Harnessing Graph Neural Networks to CraftFragrances Based on Consumer Feedback. Computers and Chemical Engineering, 185,108674. https://doi.org/10.1016/j.compchemeng.2024.108674

[34] Rostami, M., Berahmand, K., Forouzandeh, S.,Ahmadian, S., Farrahi, V., & Oussalah, M. (2024).A Novel Healthy Food Recommendation to UserGroups Based on a Deep Social CommunityDetection Approach. Neurocomputing, 576,127326. https://doi.org/10.1016/j.neucom.2024.127326

[35] Roy, D., & Dutta, Mala. (2022). A Systematic Reviewand Research Perspective on RecommenderSystems. Journal of Big Data, 9(1), 59. https://doi.org/10.1186/s40537-022-00592-5

[36] Sabesan, N., Nivethitha, Shreyah, J. N., Pranauv, A. J., &Shyam, R. (2023). Medical Ministrations throughWeb Scraping. ArXiv:2306.12310. https://doi.org/10.48550/arXiv.2306.12310

[37] Shahade, A. K., Walse, K. H., Thakare, V. M., & Atique,M. (2023). Multi-Lingual Opinion Mining forSocial Media Discourses: an Approach Using DeepLearning Based Hybrid Fine-Tuned SmithAlgorithm with Adam Optimizer. InternationalJournal of Information Management Data Insights, 3(2), 100182. https://doi.org/10.1016/j.jjimei.2023.100182

[38] Soni, P., de Runz, C., Bouali, F., & Venturini, G. (2024).A Survey on Automatic DashboardRecommendation Systems. Visual Informatics, 8(1), 67-79. https://doi.org/10.1016/j.visinf.2024.01.002

[39] Tabaku, B., & Ali, M. (2021). Protecting WebApplications from Web Scraping. EmergingTechnologies in Computing, 56-70. https://doi.org/10.1007/978-3-030-90016-8_4

[40] Talari, G., Cummins, E., McNamara, C., & O'Brien, J.(2022). State of the Art Review Of Big Data andWeb-Based Decision Support Systems (Dss) forFood Safety Risk Assessment with Respect toClimate Change. Trends in Food Science &Technology, 126, 192-204. https://doi.org/10.1016/j.tifs.2021.08.032

[41] Varghese, R. R., & Mohan, B. R. (2023). Study on theSentimental Influence on Indian Stock Price. Heliyon, 9(12), 22788. https://doi.org/10.1016/j.heliyon 2023.e22788

[42] Vijayakumar, P., & Jagatheeshkumar, G. (2024). User'sLearning Capability Aware E-ContentRecommendation System for Enhanced LearningExperience. Measurement: Sensors, 31, 100947. https://doi.org/10.1016/j.measen.2023.100947

[43] Dip Bharatbhai Patel. (2025). Leveraging BI for Competitive Advantage: Case Studies from Tech Giants. Frontiers in Emerging Engineering & Technologies, 2(04), 15–21. Retrieved from https://irjernet.com/index.php/feet/article/view/166