Annals of Emerging Technologies in Computing (AETiC)
Vol. 9, No. 1, 2025

Research Article

Integrating Intelligent Web Scraping Techniques in Internship Management Systems: Enhancing Internship Matching

Hyrmet Mydyti[1] and Andrew Ware[2]

[1]International Business College of Mitrovica (IBC-M), Kosovo
h.mydyti@ibcmitrovica.eu
[2]University of South Wales, UK
andrew.ware@southwales.ac.uk

Received: 14th August 2024; Accepted: 17th December 2024; Published: 1st January 2025

Abstract: The study explores the integration of intelligent web scraping techniques to enhance the internship matching process within internship management systems. The increasing demand for internships necessitates timely and efficient intern matching, a task that conventional manual techniques need help with due to its complexity and time-consuming nature. Intelligent web scraping algorithms and machine learning techniques analyze extensive datasets to match interns with businesses based on competencies, interests, and professional objectives. The integration leverages natural language processing to extract relevant information from internship listings and candidate profiles, enhancing the precision and effectiveness of the matching process. Additionally, clustering and matching algorithms refine recommendations, pairing students with opportunities that fit their competencies and career objectives. However, implementing intelligent web scraping raises ethical concerns, particularly regarding data privacy and algorithmic bias. Ensuring the ethical utilization of these techniques is critical for fair and unbiased internship matching. The research addresses these ethical considerations while proposing a framework for integrating intelligent web scraping into existing systems. The study reviews the literature on web scraping and machine learning in internship management, critically analyzing and synthesizing past research findings to demonstrate the efficacy of these techniques over conventional methods. The study also introduces a theoretical model for effective internship matching, investigating intelligent web scraping and machine learning techniques to optimize the process. Additionally, it examines the benefits, challenges, and limitations of integrating these techniques. The proposed intelligent web scraping approach simplifies internship matching, aligns student strengths with opportunities, enhances onboarding efficiency, and bridges academic learning with practical application.

Keywords: ***Businesses; Internship Management System; Internship Matching; Machine Learning; Natural Language Processing; Web scraping***

1. Introduction

The ongoing growth of technology and digitalization in everyday activities necessitates preparing students for future digital challenges [1]. Artificial intelligence (AI) has transformed various industries and sectors [2]. One domain suitable for AI integration is internship management systems (IMSs). Internships are essential in linking academic learning with real-world work experiences, benefiting both students and businesses [3-5]. Conventional matching methodologies, relying on manual procedures and subjective assessments, frequently impede the effectiveness of internships, leading to mismatches [6-8].

IMSs are a promising process and outcome enhancement area among the domains poised for AI integration. The integration of AI significantly improves internship matching, providing customized, data-driven solutions for both students and businesses [9-12]. AI algorithms assess various elements, such as student competencies, interests, and professional objectives, recommending suitable internship options

Hyrmet Mydyti and Andrew Ware, "Integrating Intelligent Web Scraping Techniques in Internship Management Systems: Enhancing Internship Matching", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, Vol. 9, No. 1, pp. 1-23, 1st January 2025, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2025.01.001, Available: http://aetic.theiaer.org/archive/v9/v9n1/p1.html.

[11,13]. The streamlined matching process enhances achievement and productivity [13]. Web scraping, essential in AI, data science, and cybersecurity, involves the automated extraction of data from websites, benefiting e-commerce, data analytics, and research [14]. Studies indicate a concept shift towards embracing digital solutions, highlighting intelligent web scraping as a beneficial element in this domain [15].

A prime opportunity exists to integrate intelligent web scraping techniques. The increased data availability and machine learning (ML) algorithm advancements enhance internship matching. The process auto-extracts and analyzes vast online data, matching interns with businesses based on competencies, interests, and needs.

The increasing demand for internships across various domains underscores the need for enhanced internship matching processes. Conventional manual techniques, often complex and time-consuming, frequently result in unsuitable connections between students and internship opportunities. This study investigates the benefits of incorporating intelligent web scraping methods into IMSs. The research addresses the question: "How can intelligent web scraping methods improve the efficiency and accuracy of matching students with appropriate internship opportunities?" by exploring how these techniques can enhance matching interns with proper businesses. By harnessing the capabilities of intelligent web scraping techniques, IMSs consider an array of factors, including academic performance, competencies, interests, and career objectives. Subsequently, intelligent web scraping algorithms can propose customized and fitting internship options for each student, ultimately enhancing internships' overall satisfaction and efficiency.

The primary goals of this research are: 1) To investigate intelligent web scraping and ML techniques and algorithms to optimize the internship matching process; 2) To develop a theoretical model for internship matching, focusing on potential improvements over conventional methods; 3) To analyze the benefits, challenges, and limitations of utilizing intelligent web scraping techniques for internship matching. The study investigates the integration of intelligent web scraping and machine learning into IMS to improve matching efficiency. By automating data extraction and utilizing advanced analytics, the study matches student competencies with opportunities while addressing ethical concerns such as data privacy. The research introduces a framework bridging academic knowledge and practical applications, providing a scalable and efficient alternative to conventional methods.

In conclusion, the research explores how intelligent web scraping techniques can elevate the quality of internships. By providing perspectives into the potential of intelligent web scraping techniques in matching students with internships suited to individual profiles, the endeavor aims to enhance student learning and business engagement with interns. Furthermore, the study seeks to demonstrate how intelligent web scraping techniques can bridge the gap between theoretical knowledge acquired by students and practical application in real-world work environments. The approach seeks to equip students with innovative competencies, ensuring preparedness for a constantly transforming future.

2. Technological Progress in Internship Management Systems

2.1. Technology's Role in Internships

Internship management has progressed through technological innovation, providing valuable professional experiences and boosting employability for students and businesses [16-17]. Formerly, internships relied on manual matching through documentation and personal communication [18]. Institutions and students coped with a time-consuming procedure for reports and internship applications [18-19]. By incorporating technology into everyday activities, internship management has seen substantial improvements, promoting a shift toward a broader range of skills needed in the job market [20]. Web and cloud platforms have enhanced conventional methods, facilitating digital solutions and value creation [21]. Early versions of IMSs were online job boards, evolving into web-based portals linking students, businesses, and educational institutions [5]. In the modern era, state-of-the-art technologies enable a dynamic ecosystem of IMS. AI algorithms enhance internship matching [22].

Moreover, virtual and augmented reality provides deep, engaging workplace perspectives [23]. Furthermore, data analytics in this ecosystem enhance student performance insights and internship program effectiveness [24]. The literature review explores this evolution, revealing how technology has streamlined internship management and improved the experience for students, educational institutions, and businesses.

2.2. Advantages of Intelligent Web Scraping: Methodologies and Applications

This section presents a comprehensive academic study centered on methodologies and algorithms that define intelligent web scraping. Web scraping, an essential component of data science and AI, extends beyond data extraction to involve intelligent parsing and organization for real-world applications [14,25]. ML algorithms, crucial for intelligent web scraping techniques, adapt to various layouts and formats, thereby boosting the accuracy and efficiency of data extraction [26-27]. Moreover, natural language processing (NLP) excels in extracting significant insights from text-heavy pages, integrating technologies for both quantitative and qualitative data analysis [28-29]. Intelligent web scraping is key for market and customer sentiment analysis, competitor monitoring, and price tracking in business [14]. In scholarly studies, intelligent web scraping enhances large dataset aggregation. Automated resume evaluation and web scraping revolutionize recruitment by minimizing errors and biases [30-31]. Job classification techniques leverage data scraping from various internship portals and business websites to match candidates with internships, considering their competencies, interests, and experience [32]. Ethical and legal compliance are essential in intelligent web scraping, as emphasized in thoroughly examining its various applications and prospects [14]. The challenging nature of web data necessitates continuous innovation in web scraping methods, driven by the web's rapid progression and the complexities of harvesting data for scientific use [33].

2.3. Advantages of Machine Learning in Internship Matching

ML in internship matching is a novelty in education, indicating AI's growing role in improving student achievement throughout all life cycle stages [34-35]. The study reviews how ML algorithms can be integrated into IMSs to match students with suitable internship options. ML algorithms excel in processing large volumes of data efficiently and accurately, outperforming conventional methods [36-37]. Intelligent algorithms enhance student success by suggesting internships aligned with competencies, interests, and objectives while considering comprehensive educational factors [38]. In IMSs, predictive analytics is critical, serving as a primary digital technology for understanding job market trends [20]. ML algorithms improve effectiveness by predicting student internship suitability through pattern recognition in past performance and learning styles [39].

Consequently, skilled students in one field often thrive in related areas, providing predictive insights for matching students with internships and businesses with suitable candidates. Moreover, incorporating the growing application of ML in self-adaptive systems, these algorithms develop their matching criteria over time through adaptive learning, becoming more accurate by integrating evolving job and education trends [40]. Furthermore, integrating ML in internship matching assesses technical and essential soft competencies and executives' observations that new hires often need further development in social and meta-competencies, which alone are insufficient for hiring [41]. Incorporating ML in critical sectors such as credit lending or criminal justice necessitates addressing challenges such as data privacy, algorithmic bias, and transparent, domain-specific fairness audits to ensure decisions are fact-based and uninfluenced by human biases [42]. In summary, integrating ML with internship matching signifies a notable improvement in educational technology. Applying ML facilitates more customized and adaptive internship matching, significantly improving students' academic and professional journeys.

2.4. Ethical Considerations in AI-Driven Internship Systems

Integrating AI-driven systems has transformed internship placements, improving efficiency and accuracy in matching students with best-suited industry experiences [10,43]. The discussion covers ethical issues like data privacy and algorithmic bias in AI-enabled internship systems with web scraping, highlighting the need for comprehensive study. A balance must be maintained between leveraging AI to enhance system efficiency and safeguarding individual privacy rights, as failure risks trust and misuse [44]. Ethical AI implementation requires adherence to transparency, data minimization, and user consent to safeguard rights. Compliance with regulations like the General Data Protection Regulation (GDPR) is crucial for legal alignment and preserving user data integrity in all processes [44]. In AI-enabled internship matching, data privacy is essential, regardless of AI's comprehensive industry benefits and insights from

varied data sources [37]. Scraping extensive personal data from the web, such as education, social media, and interests, facilitates data mining for customized recommendations and internship program evaluations [37]. Ethical concerns emerge due to diverse data collection practices among researchers [45]. Reliance on ML-based web scraping prompts ethical issues, mainly when methods involve unregulated data collection. Practices involving scraping without explicit consent or beyond legal frameworks underscore ethical challenges [46]. These challenges intensify as ML-based web scraping risks normalizing ethically ambiguous practices. Practical measures should focus on implementing data encryption to protect sensitive intern information, anonymization to safeguard personal data, and data minimization to ensure that only necessary information is collected and retained. Obtaining user consent and adhering to data retention policies that ensure safe storage and proper deletion of data is essential. Additionally, mechanisms must ensure ML systems prevent bias and unfairness in internship matching. Ensuring ethical and credible use of web scraping technologies requires regular security audits and adherence to regulations such as GDPR [47]. While ML enhances internship matching accuracy, mechanisms must prevent bias and unfairness [48]. Algorithmic transparency and accountability remain crucial to balance efficiency with fairness and responsible data practices [46]. Embedding safeguards foster trust and minimize risks of ML-driven web scraping's unethical perception. The security of protecting sensitive data from breaches and misuse must be comprehensively evaluated [45]. This section explores the balance between enhancing AI system efficiency and safeguarding individual privacy rights.

3. Methodology

The research methodology comprises four steps: identifying critical aspects of IMSs, identifying web scraping and ML techniques, developing a theoretical model for matching interns, and identifying the benefits, challenges, and limitations of intelligent web scraping integration. These steps aim to refine internship matching processes (Figure 1).

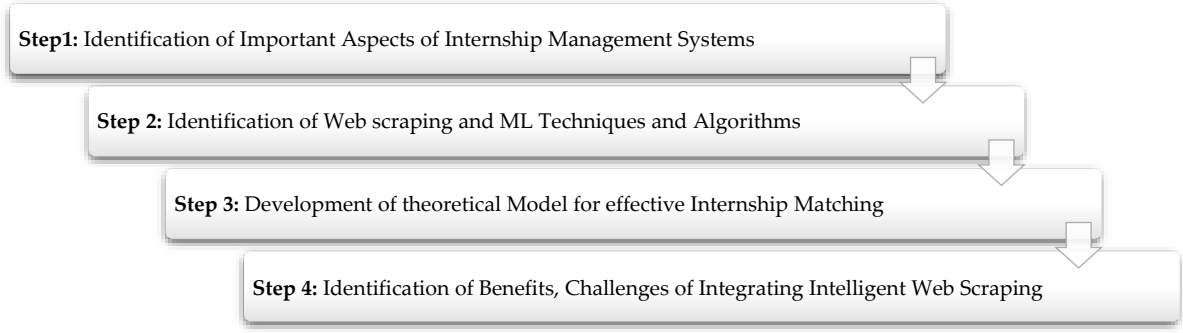

Figure 1. Methodology main steps - Source: Authors' work

3.1. Web Scraping and Machine Learning Techniques in Enhancing Internship Matching Systems

Web scraping and ML techniques have proven critical in internship matching. Table 1 includes a variety of techniques, algorithms, and applications built by researchers to enhance the efficiency of matching systems.

Selected studies explore the integration of web scraping and ML techniques in job-matching systems, significantly contributing to the internship matching field. These research efforts focus on different aspects of system enhancement. A group of studies, including those by Shah *et al.* [49], Sinha *et al.* [48], and Marin and Amel [55], focuses on using web scraping alongside ML to develop recommendation systems. Research emphasizes data preprocessing and entity recognition to enhance the matching process. Another critical element, explored by Vanetik and Kogan [50], Mhamdi *et al.* [52], and Suman *et al.* [57], is the integration of NLP into systems. This integration allows for more effective parsing and understanding of complex job descriptions and applicant profiles. Researchers such as Alsaif *et al.* [58,60], Drozda *et al.* [61], and Verma *et al.* [62] employ an approach that combines web scraping, ML, and NLP to refine algorithms. Data cleaning and advanced preprocessing techniques are explored by Varelas *et al.* [53], Akshay *et al.* [54], Narula *et al.* [65], and Thali *et al.* [51] to ensure the reliability and accuracy of the matching systems. Contributions from Mittal *et al.* [56], Alharbi and Al-Alawi [59], Singh *et al.* [62], and Panchasara *et al.* [64] highlight the application of sophisticated ML models and NLP. This Work addresses specific challenges in matching

candidates with appropriate internships, focusing on targeted data analysis and model applications to meet user needs. Collectively, these 18 studies underscore a trend toward data-driven solutions that enhance the operational efficiency of matching systems.

Table 1. Evaluation of Techniques and Algorithms in Web Scraping and Machine Learning for Matching

#	Author	Techniques Used				Data Preprocessing			Model / Application				Advanced Techniques
		Web scraping	ML	NLP	Techniques Used (Other)	Data Cleaning	Entity Recognition	Data Preprocessing (Other)	Recommendation System	Classification	Clustering	Model / Application (Other)	
1	Shah *et al.* [49]	√	√		√	√		√	√	√	√	√	
2	Sinha *et al.* [48]	√	√		√	√		√	√			√	
3	Vanetik and Kogan [50]		√	√			√					√	√
4	Thali *et al.* [51]	√	√			√			√				√
5	Mhamdi *et al.* [52]	√	√	√		√		√	√	√			√
6	Varelas *et al.* [53]		√	√		√	√	√		√			
7	Akshay *et al.* [54]	√				√			√				
8	Marin and Amel [55]	√	√		√	√		√	√				
9	Mittal *et al.* [56]	√		√	√		√	√		√		√	√
10	Suman *et al.* [57]	√	√	√		√	√		√			√	√
11	Alsaif *et al.* 2022 [58]	√	√	√	√	√	√	√	√			√	
12	Alharbi & Al-Alawi [59]	√	√	√	√	√	√	√				√	
13	Alsaif *et al.* [60]	√	√	√	√	√	√	√	√				√
14	Drozda *et al.* [61]	√	√				√			√		√	√
15	Verma *et al.* [62]	√	√		√			√				√	√
16	Singh *et al.* [63]	√	√	√		√		√	√			√	√
17	Panchasara *et al.* [64]	√	√	√		√		√	√			√	√
18	Narula *et al.* [65]	√	√	√	√	√			√	√		√	√

Based on the evaluation presented, this literature review section establishes a structured foundation for exploring the application of web scraping and machine learning techniques in enhancing internship matching systems. The research categorizes the content into three key parts: intelligent web scraping for extracting listings, machine learning for profile matching, and developing a theoretical model for effective matching. The analysis thoroughly explores these approaches' benefits, challenges, and limitations. Each part builds upon the insights from the reviewed literature, presenting a comprehensive and systematic approach to advancing internship management systems.

4. Internship Matching: Web Scraping and Machine Learning Techniques and Algorithms

The competitive process of matching students with internships frequently fails due to conventional methods [66-67]. Integrating advanced web scraping and ML into IMSs addresses this gap and provides a possible solution [5,15]. Research focuses on technologies and methodologies that enhance the efficiency of internship matching using advanced data extraction and analysis.

Web scraping, a technique for automatically retrieving information from websites, gathers vast amounts of internship data. HTML parsing and API-based extraction facilitate the systematic gathering of structured data, allowing IMSs to maintain an up-to-date repository of opportunities [68-69]. The vast amount of online data necessitates intelligent techniques to enhance extraction. ML and NLP refine data quality [70]. ML identifies patterns, enhancing the categorization and relevance of internships, while NLP understands human language, extracting insights from internship descriptions, business profiles, and student resumes [70-71].

After data extraction, the critical step is cleaning and matching to provide optimal internship recommendations [72]. Algorithms are essential, utilizing clustering to categorize similar internships and matching to align student profiles [94]. In summary, integrating web scraping and ML in IMSs transforms the matching process. Automating data extraction, enhancing data intelligently, and applying advanced

algorithms for cleaning and matching improves efficiency, accuracy, and user experience. The following sections provide comprehensive studies on these technologies.

4.1. Fundamental Web Scraping Technologies

Two core technologies, HTML parsing and API-based extraction, support web scraping by providing different approaches for retrieving valuable data from web sources [14,72].

4.1.1. HTML Parsing

HTML parsing, a core web scraping technique, extracts data directly from the HTML structure of web pages by utilizing the Document Object Model (DOM) hierarchy to identify and retrieve specific elements, such as internship listings, business information, and application deadlines [73-74]. HTML parsing offers flexibility as its main benefit, applying broadly to nearly all HTML web pages. Tools like BeautifulSoup, Scrapy, or Cheerio help developers locate and extract data using CSS selectors or XPath, such as identifying all elements with a "job-listing" class [54,74-75]. However, HTML parsing faces challenges such as frequent website updates and anti-scraping measures like JavaScript or CAPTCHAs, requiring ongoing scraper maintenance to adjust to changes in website structures and ensure efficient data extraction [76].

4.1.2. API-Based Extraction

API-based extraction is a reliable web scraping method, as many websites provide APIs designed for data exchange, enabling developers to access structured data in formats like JSON or XML for specific data retrieval, such as available internships, business profiles, and user feedback [14,68].

API-based extraction delivers robustness and consistency, as APIs are stable and less susceptible to breaking than HTML structures, providing comprehensive data for thorough analysis [73]. Tools like Postman or Insomnia enable interaction and endpoint testing and ensure accurate data retrieval [77]. Although API-based extraction provides benefits, the lack of public APIs, access restrictions, and sometimes not providing specific data limit its effectiveness, requiring a hybrid approach with HTML parsing [68,78]. Ultimately, incorporating HTML parsing and API-based extraction into IMSs improves the capacity to match students with suitable internships by ensuring accurate and comprehensive data collection.

4.2. Intelligent Techniques for Enhanced Data Extraction

Intelligent data extraction techniques like ML and NLP enhance matching students with appropriate internship opportunities [37].

4.2.1. Machine Learning-Based Approaches

ML involves the study of algorithms and statistical models that enable computer systems to carry out specific tasks without explicit programming [66]. ML algorithms analyze vast datasets, using supervised and unsupervised learning techniques to identify patterns and efficiently match student competencies and interests with position criteria [79-80]. Supervised learning, which depends on labeled training data, predicts a student's suitability for a specific internship based on historically successful matches, while unsupervised learning clusters students and internships into groups with similar attributes, enhancing the matching process by identifying the most appropriate matches [80].

4.2.2. Natural Language Processing (NLP)

NLP is a field of research and practical application focused on enabling computers to comprehend and manipulate natural language, whether written or spoken, to carry out valuable tasks [81]. NLP is essential for extracting and analyzing unstructured data from diverse sources, parsing textual data to identify critical information such as necessary competencies, job roles, and candidate qualifications, and creating thorough profiles for students and internships, enabling more accurate matches. For instance, NLP can analyze position descriptions to extract keywords and phrases that characterize the best candidate and subsequently match these with the competencies and experiences provided in student resumes [81-83]. ML and NLP in internship systems automate data extraction and analysis and match students with internships. ML and NLP enhance match accuracy, aligning positions with student competencies and interests and benefiting businesses with well-matched candidates. Intelligent techniques for data extraction provide significant advantages for IMSs, improving outcomes for students and businesses [84].

4.3. Algorithm-Based Approach for Data Cleaning and Matching

4.3.1. Clustering Algorithms

Clustering algorithms are essential in data cleaning and matching, grouping similar data points based on their characteristics [85]. In the context of internship matching, clustering algorithms categorize students based on their competencies and interests, identifying patterns and relationships within the data to enable more accurate and efficient matching [85-86]. The K-Means algorithm is a widely recognized method for clustering, which organizes data into K groups by measuring the mean distance among data points. Each cluster centroid repeatedly reassigns data points to the nearest cluster until the clusters stabilize [87]. The approach is beneficial for identifying distinct groups within the data, such as students with similar competency sets or career objectives [87-88].

The Hierarchical Clustering algorithm constructs a cluster hierarchy by merging or splitting clusters based on similarity. The algorithm initiates by considering each data point as an individual cluster. The process then iteratively merges or splits clusters until achieving the desired level of granularity, which is beneficial for identifying nested or hierarchical relationships within the data [85,89-90].

4.3.2. Matching Algorithms

Matching involves efficiently allocating resources by connecting two entities within a market to attain a stable or optimal state, using preference lists for various matching scenarios [91]. Matching algorithms pair students with suitable internships based on competencies, interests, and objectives, considering academic performance and professional objectives [92]. Collaborative Filtering is a commonly used matching algorithm that recommends internships based on the interests of similar students. The algorithm constructs a matrix of student interests, identifies the students most similar to the target student, and recommends internships that are popular among these similar students [91-92,105]. The Content-Based Filtering algorithm matches internships with students' interests by analyzing internship descriptions, building a matrix of features, and identifying opportunities that align with the student's specific competencies and interests [92]. Combining clustering and matching algorithms, IMSs provide more precise internship recommendations, enhancing student success and increasing employer satisfaction [92-93].

5. Application of Intelligent Web Scraping in Internship Management: Case Studies

The research categorizes the content into three key parts. The approach ensures up-to-date data, categorizes opportunities based on competencies, and predicts the best fits, enhancing internship management efficiency.

5.1. Automated Extraction of Listings from Diverse Sources

Automated extraction of listings from diverse sources utilizing intelligent web scraping techniques is essential for developing a comprehensive management system that provides access to the most relevant and current opportunities [20,94-95]. Web scraping collects information from websites using bots or scripts, mainly utilizing HTML parsing to examine web structures for data elements and API-based extraction for organized data from structured endpoints [14]. Using tools such as BeautifulSoup and Scrapy in Python, HTML parsing allows developers to navigate the DOM of web pages, identify elements like listings, and extract data, but requires frequent updates [96]. API-based extraction is more stable and reliable, accessing structured data via public APIs in formats like JSON or XML [14]. Automating listing extraction ensures up-to-date data, reduces manual effort, and allows staff to focus on strategic tasks while creating a centralized database for easier searches and applications [14]. Moreover, ML algorithms analyze and categorize listings once extracted, identifying patterns to match individuals with opportunities that fit their competencies, interests, and objectives, enhancing system efficiency and effectiveness [97]. Accordingly, intelligent web scraping for automated listings is essential, ensuring current data and leveraging ML to enhance the quality and availability of opportunities [94-97].

Mishra *et al.* [98] demonstrate the significant enhancement of internship management through AI-augmented web scraping. The focus is on automating data collection and analysis from various online platforms, efficiently handling vast data like internship listings, student profiles, and performance metrics. The system matches candidates to opportunities based on competencies, experience, and academic

performance by utilizing NLP and ML techniques. For example, AI-driven systems parse CVs and internship descriptions to generate skill graphs, aiding precise candidate-job matching. The authors illustrate the application in talent acquisition and benchmarking, where the algorithm analyzes job descriptions and candidate work experience to create comprehensive skill graphs. This approach allows recruiters to match candidates with job openings efficiently, enhancing the recruitment process and improving the quality of matches for both students and employers.

Lunn *et al.* [15] explore using web scraping and NLP to manage internship opportunities. Data was extracted from job boards and business websites, creating a comprehensive internship dataset. NLP analyzed job descriptions to identify critical competencies and requirements, highlighting trends and everyday demands. Integrating these technologies streamlined internship management, aligned opportunities with student needs, and improved job market preparedness. The study, focusing on computer science job postings, found that the most frequently offered positions were Software Engineer, Data Scientist, and Software Developer, with Python as the most requested programming language. Most postings required at least a Bachelor's degree, with many preferring or requiring higher degrees. Emphasizing connectivism, the researchers highlighted the importance of accurate information and leveraging networks to share knowledge. They underscored the need to consider industry trends when designing computer science curricula, as salary data from 417 postings showed most salaries above $100,000.

Kumar *et al.* [68] demonstrate the practical application of intelligent web scraping to automate job recommendations. Utilizing tools like Puppeteer and REST APIs, the approach combines Content-Based Filtering (CBF) and Collaborative Filtering (CF) to deliver targeted job recommendations in the engineering sector. Customized web crawlers collect job listings directly from business websites, bypassing the limitations of third-party aggregators. Data preprocessing involves removing HTML tags and stop words, ensuring clean data for analysis. CBF employs TF-IDF and Cosine Similarity to match user profiles with job descriptions, while CF analyzes user behaviors to recommend jobs based on similar interests. This hybrid system enhances recommendation accuracy, creating a personalized, efficient, scalable job recommendation platform. Practical implementation includes a user-friendly web interface, showcasing the potential of intelligent web scraping in optimizing internship and job management processes.

Akshay *et al.* [54] introduce a novel approach to job searching. Leveraging advanced web scraping tools such as Puppeteer and Cheerio, the system extracts job listings from major companies like Google and Apple in real-time, bypassing conventional delays. The user interface features an intuitive registration form and a dynamic homepage that displays personalized job recommendations based on user profiles and interests. The backend, powered by Node.js with Express, ensures efficient data handling and secure storage using MongoDB. The method streamlines job searching by delivering highly relevant opportunities, exemplifying the potential of integrating web scraping and ML technologies to revolutionize job portals and enhance user satisfaction. The innovative system demonstrates how web scraping can provide a more immediate and customized job search experience than conventional methods.

Kumar *et al.* [47] explore intelligent web scraping for job portals in India, addressing job seekers' challenges when finding suitable employment. The research emphasizes web scraping's potential to extract valuable data from job postings, offering insights into in-demand competencies, active industries, and candidate experience levels. Automated data collection from job portals dramatically minimizes the effort and time job seekers need to find relevant opportunities. Python libraries like BeautifulSoup and Selenium ensure data accuracy and relevance. Techniques were employed to maintain updated listings and avoid duplicates. The study highlights the importance of profile screening and personalized job recommendations. ML algorithms enhance job search by notifying job seekers of opportunities that closely match their competencies and interests. An innovative approach demonstrates the practical applications of intelligent web scraping in the employment sector, providing substantial benefits to job seekers and employers.

Dai *et al.* [94] utilized web scraping and clustering techniques to analyze LinkedIn profiles, examining the link between educational background and professional careers. By scraping 5 million public profiles, the researchers used NLP to classify educational backgrounds and cluster professional experiences. Their approach circumvented privacy and access limitations of Online Social Networks (OSNs) by not relying solely on APIs. The study revealed significant correlations between educational degrees and career

trajectories, providing comprehensive insights into LinkedIn users. The research highlights the value of combining scraping and clustering to analyze large-scale social data, enhancing understanding of user behavior and networking dynamics.

Gheorghe *et al.* [33] demonstrate the critical role of intelligent web scraping in data science for efficiently harvesting web data. A case study utilizes Python, Selenium, and Firefox to automate data collection, eliminating manual, error-prone processes. The scraper gathers structured data from dynamic web pages, enabling rapid and accurate analysis, which reveals key insights and patterns. Modern web scraping techniques effectively handle complex web interfaces. Data scientists can swiftly retrieve valuable information, adhering to ethical standards and preventing server overload. The study underscores the significance of adapting web scraping strategies for diverse applications while maintaining data integrity and compliance.

Adekunle *et al.* [73] present an API-based approach for automating data retention from websites, addressing legal and ethical concerns with conventional web scraping. Using Python, Requests, and BeautifulSoup libraries, the method interacts with APIs to extract and store data securely in a local database, minimizing risks of unauthorized data exposure. The approach emphasizes a data retention policy, retaining data only as needed, reducing storage costs, and enhancing privacy. The solution efficiently automates data collection, offering valuable insights into user behavior and market trends. Adaptable for various applications, including market research and customer behavior analysis, the methodology demonstrates versatility and effectiveness in modern data management practices.

Fiesler *et al.* [45] analyzed the terms of service (TOS) of over 100 social media sites to explore restrictions on automated data collection. The analysis revealed that TOS often prohibits data collection, but the language is generally ambiguous, inconsistent, and lacks context. They identified four main restriction categories: prohibitions on automated, manual, and any data collection, along with requirements for permission. Ethical considerations should go beyond TOS, incorporating the data source's context and research purpose. The study highlights the need for a nuanced ethical approach to web scraping, balancing legal constraints with research benefits and societal impacts.

Sirisuriya's [99] research highlights the importance of web scraping, which extracts unstructured data from websites, converting it into structured formats essential for training ML algorithms. It offers several advantages, including data collection from diverse sources, preprocessing through cleaning and formatting, augmenting existing datasets, and supporting competitive analysis. These capabilities significantly enhance the accuracy and robustness of ML models. The study also explores various web scraping techniques and tools, such as BeautifulSoup, Scrapy, and Selenium, and emphasizes web scraping's indispensable role despite the challenges and ethical considerations involved.

5.2. Intelligent Matching of Internships to Candidate Profiles

The intelligent matching process involves analyzing candidate competencies and qualifications using NLP, ML, and data mining [91]. The IMS utilizes NLP to analyze text data from internship descriptions and candidate profiles, extracting relevant keywords and phrases to identify the best matches [100-102]. ML algorithms analyze data to identify patterns that predict the best internship matches, learning from historical data and adapting to new information for increasingly accurate results over time [103]. Data mining identifies correlations between candidate profiles and internships for intelligent matching [104]. Integrating NLP to extract job details and ML for analysis can enhance internship searches, benefiting candidates and recruiters with optimized matching capabilities.

Permana and Pradnyana [10] developed a recommendation system using Artificial Neural Networks (ANN) to match student competencies with suitable internship placements. The system collects data from student questionnaires and tests, which are processed using a recurrent neural network for accurate recommendations. The recommendation process optimizes training accuracy and performance through a feedback architecture focusing on programming, multimedia, and networking competencies. Intelligent matching boosts student motivation and competence, ensuring a better fit between competencies and industry requirements. The recurrent structure enables faster convergence and improved performance, guiding students to select internships that enhance professional development. Addressing the gap between

academic preparation and industry needs reduces graduate unemployment rates and improves labor market readiness.

Yang [13] utilized intelligent web scraping and ANN to match candidate profiles with suitable internships for Physical Education (PE) majors. The model predicted internship success by creating an evaluation index system (EIS) optimized through data preprocessing and feature extraction. Key factors identified included internship preparation, ability, and performance. Extensive experiments validated the model's efficacy, proving superiority to conventional methods. The study highlights the importance of integrating intelligent web scraping and ANN in IMSs to enhance matching precision, improve educational outcomes, and align student competencies with industry requirements, benefiting students and employers.

Maheshwary and Misra [112] present a novel approach using a Siamese convolutional neural network (CNN) to match semi-structured resumes with job descriptions (JDs). The method captures semantic meanings, projecting similar resumes and JDs closer in the semantic space. Using a dataset of 1314 resumes and 3809 JDs, authors demonstrated that their approach outperformed conventional methods like TF-IDF and bag-of-words. The Siamese network's shared parameters ensure consistency and reduce computational load, enhancing scalability and efficiency. Experimental results highlighted the superiority over existing models, emphasizing the potential for intelligent matching in IMSs and optimizing the job search process for both seekers and employers.

Maddumage *et al.* [100] introduced an Intelligent Recruitment System (IRS) with modules like resume parsing, deep learning-based classification, and automatic question generation (AQG). The IRS extracts and ranks resume information for structured processing using NLP and deep learning. The AQG system assesses technical and soft competencies by generating questions from a merged ontology of web and local sources. Using Mamdani's method, a fuzzy inference system (FIS) aids in final decision-making. The comprehensive approach enhances candidate-job matching accuracy and streamlines recruitment, making the process more efficient. The IRS represents a significant advancement in automated recruitment, providing a robust framework for intelligent internship management.

Ujlayan and Sharma [104] enhanced the candidate screening process using latent Dirichlet allocation (LDA) for similarity analysis. The technique automates the extraction and analysis of candidate profiles from vast online databases, improving the recruitment pipeline. The study's case demonstrates that LDA identifies relevant resumes by assessing thematic coherence between candidate attributes and job requirements. The method significantly improved candidate selection efficiency in a recruitment scenario in the IT sector within India's National Capital Region. The approach quickly identified high thematic similarity profiles, reducing manual effort and costs. The approach accelerates hiring and ensures better candidate-job matches, showcasing the potential of intelligent web scraping with NLP and ML in recruitment.

Lunn *et al.* [15] demonstrate the transformative impact of web scraping and NLP on education. Python's Beautiful Soup was used to scrape job postings and gather descriptions and requirements from various websites. NLP techniques analyzed and categorized the data to identify key competencies and qualifications. A matching algorithm intelligently paired candidate profiles with suitable internships based on education, competencies, and experience. The algorithm significantly improved matching accuracy, streamlining the placement process and reducing effort for both candidates and employers. Additionally, the system provided insights into market trends, highlighting in-demand competencies and emerging job sectors.

Saidani *et al.* [25] used Gradient Boosting Models to predict employability based on student and internship contexts, focusing on internship grade, satisfaction, and recruitment offers. Experiments involved XGBoost, CatBoost, and LightGBM classifiers, revealing that internship context, particularly the method and grade, significantly influenced employability predictions. The LightGBM model achieved the highest accuracy, highlighting the critical role of experiential learning in job market readiness. The study emphasizes context-aware modeling to better align academic programs with industry needs, enhancing graduate employability. The case study demonstrates the integration of advanced ML in intelligent IMSs, providing a robust framework for educational institutions to optimize internship programs.

Pasat *et al.* [37] explore SoMeDi, an innovative recruiting platform leveraging sentiment analysis to match candidates' competencies with business needs. By analyzing candidate text inputs, SoMeDi enhances profiling accuracy, aligning professional competencies with job requirements. The architecture includes a

sentiment analysis microservice processing data from social media, yielding personalized internship recommendations. Real-world tests validate SoMeDi's efficiency gains, showcasing AI's role in streamlining recruitment. Integrating this system automates tasks and improves decision-making, highlighting AI's impact on recruitment optimization and outcomes. The study underscores NLP's practical use in recruiting, illustrating AI's transformative potential in enhancing hiring processes for candidates and employers alike.

Lin *et al.* [79] criticized conventional job-matching systems for relying on manual rules and keyword weighting, leading to inefficient searches. Their solution integrates unsupervised feature extraction and deep learning techniques across three modules: feature extraction, classifier training, and ensemble learning. The approach enhances semantic matching between job descriptions and candidate profiles using ML. By deploying the system practically, it analyzed over 47,000 resumes, significantly improving job roles, salaries, education levels, and business size predictions. By employing both Random Forests and Convolutional Neural Networks in ensemble methods, the study demonstrates the transformative potential of intelligent web scraping for optimizing internship and job-matching processes.

Mendez and Bulanadi [105] developed the Job Matcher web app using a V-Model Development Model, integrating user profiles and job requirements for personalized job recommendations. Targeting ICT graduates, the system efficiently matches competencies with partner businesses' needs, reducing mismatches. The evaluation showed high efficiency and user satisfaction, validating efficacy in enhancing employability and cutting recruitment costs. The intelligent web scraping and collaborative filtering approach accelerates job-matching, ensuring a better fit between job seekers and employers. The study highlights how such technologies transform conventional job placements into dynamic systems, emphasizing responsiveness and effectiveness.

Rojas-Galeano *et al.* [101] demonstrated significant advancements in job-resume matching by leveraging advanced algorithms in text mining and NLP. The system efficiently extracts and analyzes unstructured text from internship postings and candidate profiles, involving data collection, preprocessing, and feature extraction. Web scraping tools collect data from diverse online sources, while preprocessing techniques standardize the text. Feature extraction identifies critical attributes such as competencies, experience, and qualifications. ML algorithms match candidate profiles to internship postings based on these features, automating the process and enhancing accuracy by considering the contextual relevance of extracted information. The case study underscores the importance of integrating NLP and ML to streamline recruitment processes, providing a robust framework for intelligent job-matching applicable across various domains.

Suman *et al.* [57] introduced a method combining ML and web scraping to automate data collection from platforms like LinkedIn, ensuring a thorough evaluation of candidate qualifications for specific roles. A fuzzy matching algorithm aligns resumes with job profiles, identifying optimal matches despite data disparities. Integrating diverse data sources and advanced matching techniques boosts efficiency and accuracy in candidate selection. Prototyping validated that the approach saves time and enhances hiring quality by providing unbiased candidate evaluations. Innovation in recruitment technology advances internships and job management, exemplifying practical solutions for modernizing processes.

5.3. Development of Theoretical Model for Effective Internship Matching

The theoretical model proposed for internship matching within an IMS consists of two main sections: internship data extraction and the intelligent matching system. The process aims to streamline and enhance the matching of students' skills and preferences with suitable opportunities, utilizing cutting-edge technologies such as web scraping, NLP, and ML. A step-by-step process outline appears, in Figure 2, below to enhance understanding and visualization.

5.3.1. Internship Data Extraction

Internship data extraction is the foundational phase, leveraging automated intelligent web scraping with tools like BeautifulSoup and Selenium to ensure an up-to-date repository of structured internship opportunities. As shown in Figure 2, this phase involves the following steps:

- **Data Sources:** Reliable platforms, including job boards, business websites, and public APIs, are identified as primary data sources for internship listings.

- **Web Scraping Tools:** Tools like BeautifulSoup (HTML parsing), Selenium (browser automation), and APIs extract structured data, including job titles, descriptions, and requirements, automating browser navigation for retrieving internship details from job boards effectively.
- **Data Storage and Preprocessing:** The extracted data is stored in a database, undergoing deduplication, formatting, and cleaning to ensure accuracy and consistency.

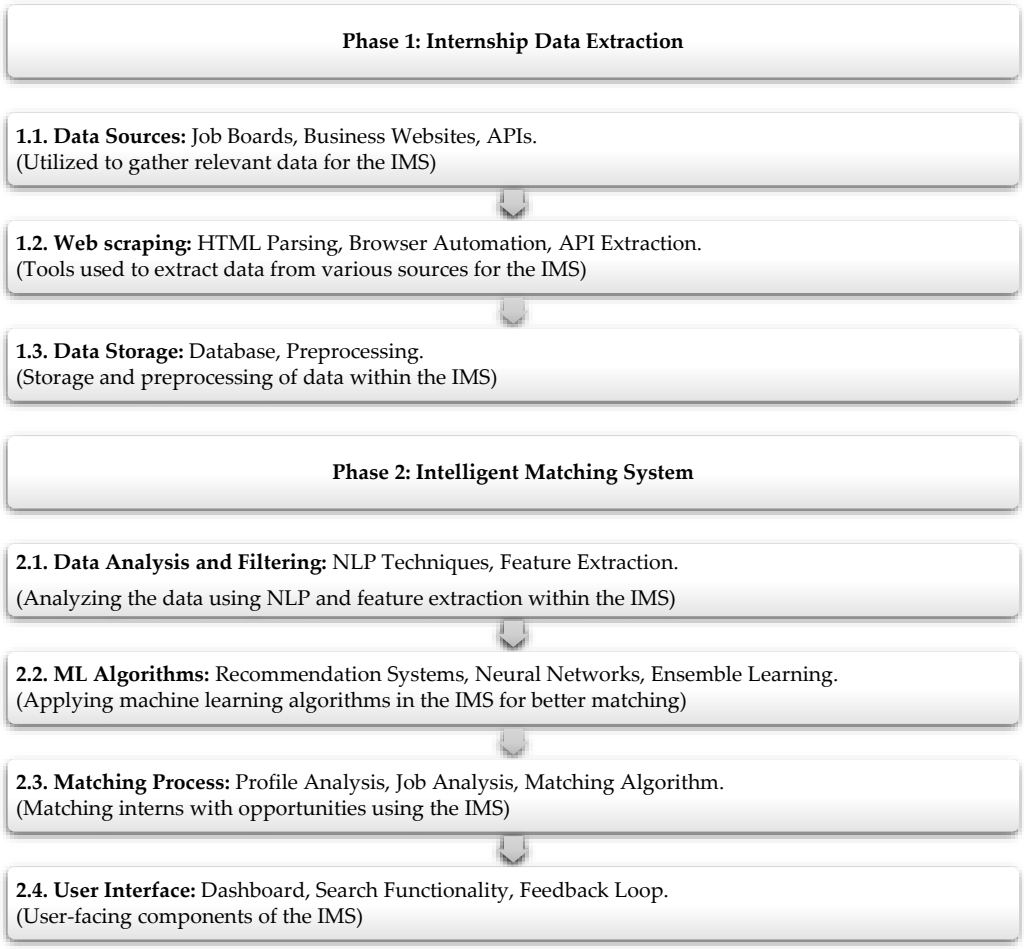

Figure 2. Theoretical Model for Effective Internship Matching within the IMS

5.3.2. Intelligent Matching System

The intelligent matching system transforms the extracted data into actionable insights for effectively matching students with internships. The system, as shown in Figure 2, comprises the following steps:

- **Data Analysis and Filtering**: *a) Feature Extraction*: NLP techniques analyze candidate profiles and internship descriptions to extract critical features, such as competencies, academic backgrounds, and job requirements; *b) Data Filtering:* During preprocessing, the system removes incomplete or irrelevant data entries and evaluates only high-quality data for analysis.
- **Machine Learning Algorithms**: *a) Recommendation Systems*: Analyze historical matches and user feedback to generate personalized internship suggestions.; *b) Neural Networks*: Identify patterns and correlations within the data to predict the most suitable internship options.
- **Matching Process**: a) *Profile and Job Analysis:* Evaluate student competencies, academic achievements, and career objectives against internship requirements and responsibilities; b) *Collaborative and Content-Based Filtering*: Leverage advanced filtering techniques to align student profiles with the most suitable internships.
- **User Interface:** The User Interface Layer delivers personalized internship recommendations via an interactive dashboard, enabling users to search, filter, and provide feedback. Feedback loops enhance the matching process by integrating user preferences and past outcomes.

The layered approach, illustrated in Figure 2, ensures a robust, efficient, and adaptive system for managing internships. It effectively bridges academic competencies with industry requirements while enhancing the user experience.

The benefits of the proposed model are evident through its ability to improve the efficiency and accuracy of internship matching by leveraging advanced technologies. Integrating NLP and ML techniques has demonstrated remarkable progress in matching candidate profiles with internships based on competencies, preferences, and objectives. Insights from existing systems applying similar methodologies highlight increased recommendation precision, higher user satisfaction, and reduced mismatches, thereby strengthening better employer engagement. The theoretical model expands on these validated techniques, ensuring a streamlined and customized matching process that connects the gap between academic learning and professional opportunities. Future implementations could further confirm these benefits through metrics such as precision-recall analysis and user feedback, providing comprehensive evidence to validate the outlined advantages.

6. Benefits, Challenges, and Limitations of Integrating Intelligent Web Scraping

6.1. Benefits of Intelligent Web Scraping

The increasing adoption of intelligent web scraping techniques in IMSs provides significant advantages, including reduced time and resource requirements for matching, enhanced match quality and relevance, and improved experiences for employers and internship seekers [106].

Integrating intelligent web scraping capabilities into management systems minimizes the time and resources needed for matching [10,14]. Tasks such as sourcing internships, aligning student profiles with opportunities, and enabling communication between students and employers, which are challenging with manual approaches, become simplified [92]. These techniques automate the extraction of internship listings from diverse sources, such as business websites and job platforms, maintaining a comprehensive repository of real-time opportunities [20]. ML algorithms further enhance the process by analyzing student profiles and suggesting relevant internships based on interests and academic backgrounds [107]. Automation significantly reduces the workload for administrative staff while ensuring students promptly access opportunities, thereby improving their prospects of obtaining valuable work experiences [108-109].

These techniques boost internship matching efficiency and enhance matches' quality and relevance compared to conventional methods [103]. By utilizing ML and NLP, systems analyze extensive data to objectively match internships with student profiles, ensuring mutually beneficial experiences for both parties [79, 83-84]. Continuous algorithm refinement based on past matches and feedback leads to higher-quality matches and increased stakeholder satisfaction [110-111].

Incorporating these advanced techniques into IMSs enhances the experience for both employers and seekers. The recruitment process is streamlined by providing access to diverse, qualified candidates and automating the matching process, facilitating quick identification and connection with desired competencies, saving time and resources. For internship seekers, these techniques offer greater transparency and accessibility to tailored opportunities, allowing for effortless navigation through diverse internships aligned with their interests and qualifications, thereby maximizing the probability of obtaining an ideal match while ensuring objective and unbiased recommendations [5,113].

Overall, integrating intelligent web scraping methods into IMSs represents a significant advancement in streamlining the internship matching process, promoting efficiency, equity, and satisfaction for employers and internship candidates.

6.2. Challenges and Limitations of Intelligent Web Scraping

Integrating intelligent web scraping techniques into IMSs provides many advantages but also brings challenges, including ethical and legal considerations, data accuracy and reliability, and integration issues with existing systems. Addressing these issues is essential for effective implementation [70,106,114].

One of the main challenges is addressing the ethical and legal considerations of data privacy, consent, and intellectual property rights; organizations must comply with regulations by obtaining explicit consent,

anonymizing data, and establishing clear ethical guidelines to ensure transparency and accountability [99,114].

Moreover, maintaining data accuracy and reliability is a significant obstacle in web scraping. Dynamic web pages frequently update, causing inconsistencies, while variations in website structures challenge algorithms, necessitating robust validation processes, machine learning for anomaly detection, manual verification, and continuous monitoring to ensure data integrity [14,46].

Integrating intelligent web scraping into IMSs presents technical and operational challenges, especially compatibility, scalability, and performance optimization. Businesses must assess systems, refactor outdated ones, and integrate web scraping for smooth data exchange while prioritizing scalability and performance optimization using cloud-based infrastructure and parallel processing techniques. Addressing ethical and legal considerations and ensuring data accuracy and reliability are essential for unlocking web scraping's full benefits [46,115].

6.3. Comparison with Existing Models

Existing IMSs often strive for scalability, accuracy, and adaptability, relying on manual or static methods that limit their efficiency. The proposed model addresses these challenges by automating data collection through intelligent web scraping, improving accuracy with ML and NLP, and enabling effortless integration with various data sources for enhanced flexibility.

Table 2. Comparison with Existing Models

Feature	Existing Models	Proposed Theoretical Model
Data Collection	Manual or API-based scraping	Automated intelligent web scraping
Matching Accuracy	Limited due to static algorithms	Enhanced with ML and NLP
Scalability	Restricted to predefined data	Broad, adaptable to multiple sources
Limitations	Time-consuming, error-prone	Ethical concerns, computational cost

While the proposed model addresses numerous inefficiencies, it introduces challenges such as computational demands and ethical concerns, particularly regarding data privacy. Mitigation strategies, including cloud-based infrastructure and adherence to GDPR, can help resolve these issues. By enhancing adaptability and accuracy, the model provides an innovative approach, more effectively aligning student skills with dynamic internship opportunities than conventional approaches.

6.4. Future Evaluation Metrics and Practical Considerations

This research primarily focuses on providing a theoretical model and conducting a comprehensive review of intelligent web scraping techniques in IMSs. While the integration of practical applications and empirical validation is beyond the scope of this study, these remain critical areas for future exploration. Key performance metrics, such as accuracy, data retrieval speed, and user satisfaction, are critical for assessing the system's practical effectiveness. Metrics such as precision and recall measure accuracy, data retrieval speed evaluates processing efficiency, and surveys or stakeholder interviews assess user satisfaction. Incorporating these metrics into future implementations would provide a benchmark for comparing the proposed model against existing systems. Additionally, justifying algorithm selection - such as the role of NLP and ML - will be vital in future implementations to optimize system design.

Addressing these practical considerations creates a pathway for implementing the theoretical advancements described. Integrating advanced AI, ML, and NLP techniques into internship systems can enhance accuracy, efficiency, and user experience. Focusing on evaluation metrics and innovative design principles ensures that future systems are effective and aligned with ethical standards, bridging academic theory with practical, impactful solutions.

7. Conclusion

The integration of intelligent web scraping techniques within internship management systems has proven transformative, improving the efficiency and effectiveness of the internship matching process. Advanced methodologies in machine learning and natural language processing empower systems to handle large amounts of data, find important details, and provide personalized internship recommendations that align with students' competencies, interests, and career objectives. By automating data collection and

analysis from various sources, intelligent web scraping reduces the time and effort required to connect students with suitable opportunities. The technology ensures better matches between students and internships, benefiting both students and businesses.

However, intelligent web scraping raises critical ethical and legal concerns about data privacy and algorithmic bias. Strong data protection measures, transparent and unbiased algorithms, and regular reviews are essential to ensure responsible data use and maintain fairness and transparency. The changing nature of web data, different website structures, and the need for constant updates present considerable challenges. Data is crucial for optimal system performance, necessitating accuracy and reliability. Addressing these challenges is essential to fully benefit from intelligent web scraping while upholding ethical standards and data integrity.

Future research and development in intelligent internship matching systems will likely concentrate on improvements in artificial intelligence, machine learning, and web scraping technologies. Emerging trends include developing advanced algorithms, such as deep learning and reinforcement learning, to enhance data extraction and matching processes. Integrating data from a broader range of sources, including social media, academic records, and professional networking sites, while enhancing natural language processing capabilities like sentiment analysis, entity recognition, and text summarization will create complete candidate profiles and improve matching accuracy by enabling better understanding and processing of natural language data from diverse sources. Research must also focus on finding and reducing algorithm biases, developing fairness-aware models, and implementing regular reviews to ensure fair and unbiased matching. Additionally, user-friendly design will be crucial, with easy-to-use and interactive platforms for students and employers, personalized dashboards, real-time feedback mechanisms, and adaptive learning systems enhancing the overall user experience.

In conclusion, integrating intelligent web scraping techniques in internship management systems holds significant potential to transform the internship matching process. By automating data extraction, enhancing data analysis through machine learning and natural language processing, and applying advanced algorithms, systems can provide more efficient, accurate, and personalized internship recommendations. These approaches benefit students by aligning them with suitable opportunities and enhancing the overall quality and productivity of internships for businesses. As the demand for internships continues to grow, integrating intelligent web scraping techniques in internship management systems becomes increasingly crucial, providing a comprehensive summary of methodologies, algorithms, and potential benefits while addressing ethical factors and proposing an implementation framework. Stakeholders in education and industry must collaborate to leverage the full potential of intelligent web scraping technologies in internship management systems. Educational institutions should allocate resources to these technologies to improve student job prospects, while businesses should streamline recruitment processes and expand their pool of qualified candidates. Policymakers and regulatory bodies must establish rules for ethical and responsible use. By promoting collaboration, stakeholders can drive effective, efficient, and fair internship matching systems, bridging the gap between academic learning and practical application and ultimately empowering students and businesses.

References

[1] Sussane Kamsker and Peter Slepcevic-Zach, "The Digital Change of Vocational Training and Business Education: What It Takes to Prepare Students for the Future Challenges of the Job Market", *International Journal for Business Education*, ISSN: 2164-2885, p. 6, Vol. 161, No. 1, April 2021, Published by International Association of Business Education (IABE), DOI: 10.30707/IJBE161.1.1648090824.269195, Available: https://eric.ed.gov/?id=EJ1312367.

[2] Anupama Prasanth, Densy John Vadakkan, Priyanka Surendran and Bindhya Thomas, "Role of Artificial Intelligence and Business Decision Making", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Print ISSN: 2156-5570, Online ISSN: 2158-107X, pp. 123-130, Vol. 14, No. 6, June 2023, United Kingdom, Published by Science and Information Organization, DOI: 10.14569/IJACSA.2023.01406103, Available: https://thesai.org/Publications/ViewPaper?Volume=14&Issue=6&Code=IJACSA&SerialNo=103.

[3] Peter N. Kiriri, "An Assessment of the Quality of a Work-Integrated Learning Internship Program in Kenya", *International Journal of Work-Integrated Learning*, Online ISSN: 2538-1032, pp. 257-271, Vol. 20, No. 3, September 2019, Published by International Association for Work-Integrated Learning, New Zealand, Available: https://eric.ed.gov/?id=EJ1232851.

[4] Ratih Hafsarah Maharrani, Abdul Rohman Supriyono, Lutfi Syafirullah and Riyadi Purwanto, "Usability Test for User Centered Design-Based Industrial Internship Management System", *Jurnal Teknologi Informasi dan Pendidikan*, Print ISSN: 2086-4981, Print ISSN: 2086-4981, Online ISSN: 2620-6390, pp. 24-36, Vol. 15, No. 2, September 2022, Indonesia, Published by Universitas Teknologi Indonesia, DOI: 10.24036/jtip.v15i1.555, Available: http://tip.ppj.unp.ac.id/index.php/tip/article/view/555.

[5] Hyrmet Mydyti and Arbana Kadriu, "Using Internship Management System to Improve the Relationship Between Internship Seekers, Employers, and Educational Institutions", *ENTRENOVA - ENTerprise REsearch InNOVAtion*, ISSN: 2706-4735, pp. 97–104, Vol. 6, No. 1, September 2020, Croatia, Published by IRENET - Society for Advancing Innovation and Research in Economy, Available: https://hrcak.srce.hr/ojs/index.php/entrenova/article/view/13437.

[6] Jennifer L. Callahan and C. Edward Watkins Jr., "The Science of Training III: Supervision, Competency, and Internship Training," *Training and Education in Professional Psychology*, Print ISSN: 1931-3918, Online ISSN: 1931-3926, pp. 245-261, Vol. 12, No. 4, November 2018, United States, Published by American Psychological Association, DOI: 10.1037/tep0000208, Available: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ftep0000208.

[7] Arkadiusz Januszewski and Małgorzata Grzeszczak, "Internship of Accounting Students in the Form of E-Learning: Insights from Poland", *Education Sciences*, Online ISSN: 2227-7102, p. 447, Vol. 11, No. 8, August 2021, Switzerland, Published by MDPI, DOI: 10.3390/educsci11080447, Available: https://www.mdpi.com/2227-7102/11/8/447.

[8] Joseph Lobo, "Pre-service teachers' evaluation of the student internship program: Basis for design improvement", *Journal for Educators, Teachers and Trainers (JETT)*, ISSN: 1989-9572, pp. 48-56, Vol. 13, No. 5, December 2022, Granada Spain, Published by Universidad de Granada, DOI: 10.47750/jett.2022.13.05.005, Available: https://jett.labosfor.com/index.php/jett/article/view/1036.

[9] Iryna Nikolaevna Kalinouskaya, "Recruitment Based on Artificial Intelligence Technologies", in *Proceedings of the International Conference "Process Management and Scientific Developments"*, Print ISBN: 978-5-905695-44-5, p. 7, 2nd May 2020, Birmingham, United Kingdom, Published by Scientific Publishing House Infinity, Available: http://naukarus.ru/wp-content/uploads/2020/PMSD%20July%2022%20-%20Part%201.pdf.

[10] A. A. J. Permana and G. A. Pradnyana, "Recommendation systems for internship place using artificial intelligence based on competence", *Journal of Physics: Conference Series*, Print ISSN: 1742-6588, Online ISSN: 1742-6596, p. 012007, Vol. 1165, No. 1, February 2019, United Kingdom, Published by IOP Publishing Ltd., DOI: 10.1088/1742-6596/1165/1/012007, Available: https://iopscience.iop.org/article/10.1088/1742-6596/1165/1/012007.

[11] Zachari Swiecki, Hassan Khosravi, Guanliang Chen, Roberto Martinez-Maldonado, Jason M. Lodge *et al.*, "Assessment in the age of artificial intelligence", *Computers and Education: Artificial Intelligence*, ISSN: 2666920X, p. 100075, Vol. 3, October 2022, Netherlands, Published by Elsevier B.V., DOI: 10.1016/j.caeai.2022.100075, Available: https://www.sciencedirect.com/science/article/pii/S2666920X22000303.

[12] Cuibi Yang, Shuliang Huan and Yong Yang, "A practical teaching mode for colleges supported by artificial intelligence", *International Journal of Emerging Technologies in Learning (iJET)*, Print 1868-8799, Online ISSN: 1863-0383, pp. 195-206, Vol. 15, No. 17, September 2020, Austria, Published by International Association of Online Engineering, DOI: 10.3991/ijet.v15i17.16737, Available: https://online-journals.org/index.php/i-jet/article/view/16737.

[13] Bo Yang, "Internship Effect Prediction for Physical Education Majors Based on Artificial Neural Network", *International Journal of Emerging Technologies in Learning (iJET)*, Print 1868-8799, Online ISSN: 1863-0383, pp. 165-176, Vol. 16, No. 24, December 2021, Austria, Published by International Association of Online Engineering, DOI: 10.3991/ijet.v16i24.27839, Available: https://online-journals.org/index.php/i-jet/article/view/27839.

[14] Moaiad Ahmad Khder, "Web scraping or web crawling: State of art, techniques, approaches and application", *International Journal of Advances in Soft Computing & Its Applications*, Online ISSN: 2074-8523, pp. 144-168, Vol. 13, No. 3, December 2021, Jordan, Published by Al-Zaytoonah University of Jordan, DOI: 10.15849/IJASCA.211128.11, Available: https://ijasca.zuj.edu.jo/PapersUploaded/2021.3.11.pdf.

[15] Stephanie Lunn, Jia Zhu and Monique Ross, "Utilizing web scraping and natural language processing to better inform pedagogical practice", in *Proceedings of the 2020 IEEE Frontiers in Education Conference (FIE)*, Print ISSN: 978-1-7281-8962-8, Online ISSN: 978-1-7281-8961-1, pp. 1-9, Vol. 2020, No. 1-4, 21st-24th October 2020, Uppsala, Sweden, Published by IEEE, DOI: 10.1109/FIE44824.2020.9274270, Available: https://ieeexplore.ieee.org/document/9274270.

[16] Yoser Gadhoum, "A proposed model of a future university in the era of the artificial intelligence transformative society: From why to how", *Creative Education*, Print ISSN: 2151-4755, Online ISSN: 2151-4771, pp. 1098-1119, Vol. 13, No. 3, March 2022, China, Published by Scientific Research Publishing, Inc., DOI: 10.4236/ce.2022.133069, Available: https://www.scirp.org/journal/paperinformation?paperid=116321.

[17] Francisco José García-Peñalvo, Juan Cruz-Benito, David Griffiths and Achilleas Achilleos, "Virtual placements management process supported by technology: Proposal and first results of the semester of code", *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, Online ISSN: 1932-8540, pp. 47-54, Vol. 11, No. 1, February 2016, United States of America, Published by IEEE Education Society, DOI: 10.1109/RITA.2016.2518461, Available: https://ieeexplore.ieee.org/document/7397889.

[18] Yong Chi Fei, Waddah Waheeb and Salasiah Binti Sulaiman, "Student Internships: A Selective Review", *Journal of Applied Technology and Innovation*, Online ISSN: 2600-7304, pp. 1-5, Vol. 4, No. 2, April 2020, Kuala Lumpur, Malaysia, Published by Asia Pacific University of Technology and Innovation, Available: https://jati.apu.edu.my/publication-volumes/.

[19] Fazeera Syuhada Abdullah, Azmi Abu Seman, Norfiza Ibrahim, Noor Aishatun Majid, Nor Mazlina Abdul Wahab et al., "Web-based application of the internship management system", *Journal of Computing Research and Innovation (JCRINN)*, Online ISSN: 2600-8793, pp. 46-51, Vol. 2, No. 3, September 2017, Available: https://jcrinn.com/index.php/jcrinn/article/view/54.

[20] Mirjana Pejic-Bach, Tine Bertoncel, Maja Meško and Živko Krstić, "Text Mining of Industry 4.0 Job Advertisements", *International Journal of Information Management*, ISSN: 02684012, pp. 416-431, Vol. 50, No. C, January 2020, United Kingdom, Published by Elsevier Ltd, DOI: 10.1016/j.ijinfomgt.2019.07.014, Available: https://dl.acm.org/doi/10.1016/j.ijinfomgt.2019.07.014.

[21] Swen Nadkarni and Reinhard Prügl, "Digital Transformation: A Review, Synthesis and Opportunities for Future Research", *Management Review Quarterly*, Print ISSN: 2198-1620, Online ISSN: 2198-1639, pp. 233-341, Vol. 71, No. 2, April 2021, Switzerland, Published by Springer Nature, DOI: 10.1007/s11301-020-00185-7, Available: https://link.springer.com/article/10.1007/s11301-020-00185-7.

[22] Girum Abebe, Marcel Fafchamps, Michael Koelle and Simon Quinn, "Learning management through matching: A field experiment using mechanism design", *National Bureau of Economic Research (NBER) Working Paper Series*, ISSN: 0077-3611, No. w26035, 2019, Published by National Bureau of Economic Research (NBER), Available: https://www.nber.org/system/files/working_papers/w26035/w26035.pdf.

[23] Ekaterina Prasolova-Førland, Mikhail Fominykh and Oscar Ihlen Ekelund, "Empowering young job seekers with virtual reality", in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, ISSN: 2642-5254, pp. 295-302, Vol. 2019, 23rd-27th March 2019, Osaka, Japan, Published by IEEE, DOI: 10.1109/VR.2019.8798179, Available: https://ieeexplore.ieee.org/document/8798179.

[24] Aderibigbe Israel Adekitan and Odunayo Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining", *Heliyon*, ISSN: 2405-8440, p. e01250, Vol. 5, No. 12, December 2019, Netherlands, Published by Elsevier, DOI: 10.1016/j.heliyon.2019.e01250, Available: https://pubmed.ncbi.nlm.nih.gov/30886917/.

[25] Oumaima Saidani, Leila Jamel, Amel Ksibi and Nazik Alturki, "Predicting Student Employability Through the Internship Context Using Gradient Boosting Models", *IEEE Access*, ISSN: 2169-3536, pp. 46472-46489, Vol. 10, August 2022, New York, United States, Published by IEEE, DOI: 10.1109/access.2022.3170421, Available: https://ieeexplore.ieee.org/document/9762923.

[26] Irvin Dongo, Yudith Cardinale, Ana Aguilera, Fabiola Martinez, Yuni Quintero *et al.*, "A qualitative and quantitative comparison between web scraping and API methods for Twitter credibility analysis", *International Journal of Web Information Systems*, Print ISSN: 1744-0084, Online ISSN: 1744-0092, pp. 580-606, Vol. 17, No. 6, December 2021, United Kingdom, Published by Emerald Group Publishing Ltd, DOI: 10.1108/IJWIS-03-2021-0037, Available: https://www.emerald.com/insight/content/doi/10.1108/ijwis-03-2021-0037/full/html.

[27] Kaushal Parikh, Dilip Singh, Dinesh Yadav and Mansing Rathod, "Detection of Web Scraping Using Machine Learning", *Open Access International Journal of Science and Engineering*, ISSN: 2456-3293, pp. 114-118, Vol. 3, No. 2, 2018, India, Published by Open Access International, Available: https://ijcrt.org/papers/IJCRT_183909.pdf.

[28] Meng Cai, "Natural language processing for urban research: A systematic review," *Heliyon*, e06322, Vol. 7, No. 3, March 2021, Netherlands, Published by Elsevier, DOI: 10.1016/j.heliyon.2021.e06322, Available: https://pubmed.ncbi.nlm.nih.gov/33732917/.

[29] Yanshan Wang, Sijia Liu, Naveed Afzal, Majid Rastegar-Mojarad, Liwei Wang *et al.*, "A Comparison of Word Embeddings for the Biomedical Natural Language Processing", *Journal of Biomedical Informatics*, Print ISSN: 1532-0464, Online ISSN: 1532-0480, pp. 12-20, Vol. 87, No. 4, November 2018, San Diego, United States, Published by Academic Press Inc., DOI: 10.1016/j.jbi.2018.09.008, Available: https://pubmed.ncbi.nlm.nih.gov/30217670/.

[30] Aashima Gupta and Mridula Mishra, "Artificial Intelligence for Recruitment and Selection", in *The Adoption and Effect of Artificial Intelligence on Human Resources Management*, Part B, Print ISBN: 978-1-80455-663-4, Online ISBN: 978-1-80455-662-7, pp. 1-11, Vol. 7, February 2023, England, Published by Emerald Publishing Limited, DOI: 10.1108/978-1-80455-662-720230001, Available: https://www.emerald.com/insight/content/doi/10.1108/978-1-80455-662-720230001/full/html.

[31] Muralidhar A., J.V. Thomas Abraham and Sri Sai Saran Reddy Yeturu, "Student Profile Scoring Based on Resume and Website Data", *Journal of Critical Reviews*, Journal of Critical Reviews, Print ISSN: 2394-5125, Online ISSN: 2394-5869, pp. 1646-1650, Vol. 7, No. 15, 2020, India, Published by Innovare Academics Sciences Pvt. Ltd, DOI: 10.31838/jcr.07.15.217, Available: https://jcreview.com/archives/volume-7/issue-15/8629.

[32] Muhammad Haziq Haikal Hisham, Mohd Azri Abdul Aziz and Ahmad Asari Sulaiman, "Job Classification: A Review on Data, Features, and Methods", *Journal of Electrical and Electronic Systems Research*, Online ISSN: 3030-

640X, pp. 91-100, Vol. 21, No. 1, October 2022, Selangor, Malaysia, Published by UiTM Press, DOI: 10.24191/jeesr.v21i1.012, Available: https://jeesr.uitm.edu.my/v1/JEESR/Vol.21/article12.pdf.

[33] Mihai Gheorghe, Florin C. Mihai and Marius Dârdală, "Modern Techniques of Web Scraping for Data Scientists", *Romanian Journal of Human - Computer Interaction*, ISSN: 1843-4460, pp. 63-75, Vol. 11, No. 1, March 2018, Bucuresti, Romania, Published by Matrix Rom, Available: https://api.semanticscholar.org/CorpusID:203647509.

[34] Milena Ilić, Dan Păun, Nevenka Popović Šević, Aleksandra Hadžić and Anca Jianu, "Needs and performance analysis for changes in higher education and implementation of artificial intelligence, machine learning, and extended reality", *Education Sciences*, Online ISSN: 2227-7102, p. 568, Vol. 11, No. 10, October 2021, Switzerland, Published by MDPI, DOI: 10.3390/educsci11100568, Available: https://www.mdpi.com/2227-7102/11/10/568.

[35] Kriti Khare, Brian Stewart and Anshuman Khare, "Artificial Intelligence and the Student Experience: An Institutional Perspective", *IAFOR Journal of Education*, ISSN: 2187-0594, pp. 63-78, Vol. 6, No. 3, December 2018, Japan, Published by The International Academic Forum (IAFOR). DOI: 10.22492/ije.6.3.04, Available: https://iafor.org/journal/iafor-journal-of-education/volume-6-issue-3/article-4/.

[36] Omar Y. Al-Jarrah, Peter D. Yoo, Sameh Muhaidat, George K. Karagiannidis and Khaled Taha, "Efficient Machine Learning for Big Data: A Review", *Big Data Research*, Print ISSN: 2214-5796, pp. 87-93, Vol. 2, No. 3, September 2015, United States, Published by Elsevier Inc., DOI: 10.1016/j.bdr.2015.04.001, Available: https://www.sciencedirect.com/science/article/abs/pii/S2214579615000271.

[37] Adrian Pasat, George Suciu, Andrei Birdici and Iulia Pop, "An internship campaign case study showing results of enhanced recruitment processes using NLP", in *The International Scientific Conference eLearning and Software for Education*, ISSN:2066-026X, pp. 222-231, Vol. 2, 22nd-23rd April 2021, Bucharest, Romania, Published by National Defence University - Carol I Printing House, DOI: 10.12753/2066-026X-21-097, Available: https://www.proquest.com/openview/e1684fc157d884fc09235a5a8be15a65/1?pq-origsite=gscholar&cbl=1876338.

[38] Farouk Ouatik, Mohammed Erritali, Fahd Ouatik, Mostafa Jourhmane, "*Predicting Student Success Using Big Data and Machine Learning Algorithms*", *International Journal of Emerging Technologies in Learning*, Print ISSN: 1863-0383, pp. 236-251, Vol. 17, No. 12, June 2021, Published by International Association of Online Engineering, DOI: 10.3991/ijet.v17i12.30259, Available: https://online-journals.org/index.php/i-jet/article/view/30259.

[39] Juan L. Rastrollo-Guerrero, Juan A. Gómez-Pulido and Arturo Durán-Domínguez, "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review", *Applied Sciences*, ISSN: 2076-3417, p. 1042, Vol. 10, No. 3, 2020, Switzerland, Published by MDPI, DOI: 10.3390/app10031042, Available: https://www.mdpi.com/2076-3417/10/3/1042.

[40] Omid Gheibi, Danny Weyns and Federico Quin, "Applying Machine Learning in Self-Adaptive Systems: A Systematic Literature Review", *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, Print ISSN: 1556-4665, Online ISSN: 1556-4703, pp. 1-37, Vol. 15, No. 3, 2021, United States, Published by ACM, DOI: 10.1145/3469440, Available: https://dl.acm.org/doi/10.1145/3469440.

[41] Ajax Persaud, "Key Competencies for Big Data Analytics Professions: A Multimethod Study", *Information Technology & People*, Print ISSN: 0959-3845, pp. 178-203, Vol. 34, No. 1, January 2021, United Kingdom, Published by Emerald Group Publishing Ltd., DOI: 10.1108/ITP-06-2019-0290, Available: https://www.emerald.com/insight/content/doi/10.1108/itp-06-2019-0290/full/html.

[42] Songül Tolan, "Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges", *arXiv preprint*, arXiv ID: 1901.04730, January 2019, New York, USA, Published by Cornell University - arXiv, DOI: 10.48550/arXiv.1901.04730, Available: https://arxiv.org/pdf/1901.04730.

[43] Norsafwani Ibrahim, Haslizatul Mohamed Hanum, Zainab Abu Bakar and Nur Atiqah Sia Abdullah, "Student-industry matching for internship placement", in *Proceedings of the 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Print ISSN: 978-1-6654-2956-6, Online 978-1-6654-1237-7: 2, pp. 122-126, 15th–16th June 2021, Kuala Lumpur, Malaysia, Published by IEEE, DOI: 10.1109/CAMP51653.2021.9498088, Available: https://ieeexplore.ieee.org/document/9498088.

[44] Yanamala, Anil Kumar Yadav, Srikanth Suryadevara and Venkata Dinesh Reddy Kalli, "Balancing Innovation and Privacy: The Intersection of Data Protection and Artificial Intelligence.", *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence* 15, pp. 1-43, Vol. 13, No. 1, November 2022, India, Available: https://ijmlrcai.com/index.php/Journal/article/view/23.

[45] Casey Fiesler, Nathan Beard and Brian C. Keegan, "No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service", in *Proceedings of the International AAAI Conference on Web and Social Media*, Print ISSN: 2162-3449, Online ISSN: 2334-0770, Vol. 14, No. 1, pp. 187-196, 8th-11th June 2020, Atlanta, Georgia, United States of America, Published by Association for the Advancement of Artificial Intelligence (AAAI), DOI: 10.1609/icwsm.v14i1.7290, Available: https://ojs.aaai.org/index.php/ICWSM/article/view/7290.

[46] Vlad Krotov, Leigh Johnson and Leiser Silva, "Tutorial: Legality and Ethics of Web Scraping", *Communications of the Association for Information Systems*, ISSN: 1529-3181, pp. 555-579, Vol. 47, No. 24, November 2020, United States, Published by Association for Information Systems, DOI: 10.17705/1CAIS.04724, Available: https://aisel.aisnet.org/cais/vol47/iss1/22/.

[47] Ashutosh Kumar, Kinshuk Chauhan and Jaspreet Kaur Grewal, "Web Scraping Job Portals", in *Advancements in Communication and Systems*, A. Kumar Tripathi and V. Shrivastava (Eds.), ISBN: 978-81-955020-7-3, pp. 291-303, April 2024, Delhi, India, Published by Soft Computing Research Society (SCRS), DOI: 10.56155/978-81-955020-7-3-25, Available: https://www.publications.scrs.in/chapter/978-81-955020-7-3/25.

[48] Koustubh Sinha, Priyansh Sharma, Harshit Sharma and Krishna Asawa, "Web scraping and job recommender system", in *Proceedings of the 2023 Second International Conference on Informatics (ICI)*, Print ISBN: 979-8-3503-4384-7, Online ISBN: 979-8-3503-4383-0, pp. 1-6, Vol. 2023, 23rd–25th November 2023, Noida, India, Published by IEEE, DOI: 10.1109/ICI60088.2023.10420941, Available: https://ieeexplore.ieee.org/document/10420941.

[49] Priyanka Shah, Paresh V. Virparya and Hardik B. Pandit, "Model of Recommendation System to Enhance the Employability: Machine Learning Approach", in *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, ISBN: 979-8-4007-0941-8, pp. 1-4, 23rd-24th December 2022, Jaipur, India, Published by ACM, DOI: 10.1145/3590837.3590949, Available: https://dl.acm.org/doi/10.1145/3590837.3590949.

[50] Natalia Vanetik and Genady Kogan, "Job Vacancy Ranking with Sentence Embeddings, Keywords, and Named Entities", *Information*, ISSN: 2078-2489, Vol. 14, No. 8, p. 468, August 2023, Israel, Published by MDPI, DOI: 10.3390/info14080468, Available: https://www.mdpi.com/2078-2489/14/8/468.

[51] Raj Thali, Suyog Mayekar, Shubham More, Sanjana Barhate and Sangeetha Selvan, "Survey on Job Recommendation Systems Using Machine Learning", in *Proceedings of the 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, Print ISBN:979-8-3503-9721-5, Online ISBN: 979-8-3503-9720-8, pp. 453-457, 14th–16th March 2023, Uttarakhand, India, Published by IEEE, DOI: 10.1109/ICIDCA56705.2023.10100122, Available: https://ieeexplore.ieee.org/document/10100122.

[52] D. Mhamdi and M. Azzouazi, M. Y. El Ghoumari, R. Moulouki, & Z. Rachik , "Enhancing Recruitment Process Using Semantic Matching", in *International Conference on Advanced Intelligent Systems for Sustainable Development*, Print ISSN: 978-3-030-90638-2, Online ISSN: 978-3-030-90639-9, pp. 370-378, Vol. 1418, 22nd-27th May 2022, Rabat, Morocco, Published by Springer, Cham, DOI: 10.1007/978-3-030-90639-9_29, Available: https://link.springer.com/chapter/10.1007/978-3-030-90639-9_29.

[53] George Varelas, Dimitris Lagios, Spyros Ntouroukis, Panagiotis Zervas, Kenia Parsons *et al.*, "Employing Natural Language Processing Techniques for Online Job Vacancies Classification", in *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations*, Print ISBN: 978-3-031-08340-2, Online ISBN: 978-3-031-08341-9, pp. 333-344, 17th- 20th June 2022, Crete, Greece, Published by Springer International Publishing, DOI: 10.1007/978-3-031-08341-9_27, Available: https://link.springer.com/chapter/10.1007/978-3-031-08341-9_27.

[54] Akshay Budharapu, Arun Kumar Ravula, Ramyasri Thudum and Arpith Kosuna, "Personalized Job Opportunity Finder Powered by Web Scraping," *MATEC Web of Conferences*, Online ISSN: 2261-236X, p. 01151, Vol. 392, 23rd-25th February 2024, Telangana, India, Published by EDP Sciences, DOI: 10.1051/matecconf/202439201151, Available: https://www.matec-conferences.org/articles/matecconf/pdf/2024/04/matecconf_icmed2024_01151.pdf.

[55] Iuliana Marin amd Hanoosh Amel, "Web platform for job recommendation based on machine learning", in *Proceedings of the 18th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)*, ISSN: 2184-4895, pp. 676-683, Vol. 1, 24th–26th April 2023, Prague, Czech Republic, Published by SciTePress, DOI: 10.5220/0011993600003464, Available: https://www.scitepress.org/Papers/2023/119936/119936.pdf.

[56] Vrinda Mittal, Priyanshu Mehta, Devanjali Relan and Goldie Gabrani, "Methodology for Resume Parsing and Job Domain Prediction", *Journal of Statistics and Management Systems*, Print ISSN: 0972-0510, Online ISSN: 2169-0014, pp. 1265-1274, Vol. 23, No. 7, July 2020, India, Published by Taru Publications, DOI: 10.1080/09720510.2020.1799583, Available: https://www.tandfonline.com/doi/epdf/10.1080/09720510.2020.1799583.

[57] Himanshu Suman, Harsh Tamkiya and Anurag Singh Kushwah, "Candidate Background Verification Using Machine Learning and Fuzzy Matching", *International Journal of Emerging Technologies and Innovative Research*, ISSN: 2349-5162, pp. 1643-1647, Vol. 7, No. 6, June 2020, India, Published by Published by IJ Publication, DOI: 10.1729/Journal.23815, Available: https://www.jetir.org/view?paper=JETIR2006237.

[58] Suleiman Ali Alsaif, Minyar Sassi Hidri, Hassan Ahmed Eleraky, Imen Ferjani and Rimah Amami, "Learning-based Matched Representation System for Job Recommendation", *Computers*, Online ISSN: 2073-431X, p. 161, Vol. 11, No. 11, November 2022, Published by Published by Multidisciplinary Digital Publishing Institute (MDPI), Basel, Switzerland, DOI: 10.3390/computers11110161, Available: https://www.mdpi.com/2073-431X/11/11/161.

[59] Fawzah Alharbi and Adel Ismail Al-Alawi, "Labor Market Prediction Using Machine Learning Methods: A Systematic Literature Review", in *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, Online ISBN: 979-8-3503-7222-9, Print ISBN: 979-8-3503-7223-6, pp. 478-482, 28th–29th January 2024, Manama, Bahrain, Published by IEEE, DOI: 10.1109/ICETSIS61505.2024.10459632, Available: https://ieeexplore.ieee.org/document/10459632.

[60] Suleiman Ali Alsaif, Minyar Sassi Hidri, Imen Ferjani, Hassan Ahmed Eleraky and Adel Hidri, "NLP-based Bi-directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters," *Big Data and Cognitive Computing*, Online ISSN: 2504-2289, p. 147, Vol. 6, No. 4, October 2022, Switzerland, Published by MDPI, DOI: 10.3390/bdcc6040147, Available: https://www.mdpi.com/2504-2289/6/4/147.

[61] Paweł Drozda, Bartosz A. Nowak, Arkadiusz Talun and Leszek Bukowski, "Evaluating Web Crawlers with Machine Learning Algorithms for Accurate Location Extraction from Job Offers", in *Proceedings of the International Conference on Computational Collective Intelligence (ICCCI)*, Print ISBN: 978-3-031-41773-3, Online ISBN: 978-3-031-41774-0, pp. 300-312, Vol. 1864, 27th-29th September 2023, Budapest, Hungary, Published by Springer, Cham, DOI: 10.1007/978-3-031-41774-0_24, Available: https://link.springer.com/chapter/10.1007/978-3-031-41774-0_24.

[62] Amit Verma, Kamal Lamsal and Payal Verma, "An Investigation of Skill Requirements in Artificial Intelligence and Machine Learning Job Advertisements", *Industry and Higher Education*, Print ISSN: 09504222, Online ISSN: 20436858, pp. 63-73, Vol. 36, No. 1, February 2022, United States, Published by SAGE Publications, DOI: 10.1177/0950422221990990, Available: https://journals.sagepub.com/doi/abs/10.1177/0950422221990990.

[63] Swapnil Singh, Deepa Krishnan, Pranit Sehgal, Harshit Sharma, Tarun Surani *et al.*, "Gradient boosting approach for sentiment analysis for job recommendation and candidate profiling", in *2022 IEEE Bombay Section Signature Conference (IBSSC)*, Print ISSN: 2994-2349, Online ISSN: 2994-2357, pp. 1-6, Vol. 1, 8th-10th December 2022, Mumbai, India, Published by IEEE, DOI: 10.1109/IBSSC56953.2022.10037443, Available: https://ieeexplore.ieee.org/document/10037443.

[64] Sahil Panchasara, Rajeev Kumar Gupta and Ashutosh Sharma, "AI Based Job Recommendation System Using BERT", in *Proceedings of the 2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Print ISSN: 2771-134X, Online ISSN: 2771-1358, pp. 1-6, Vol. 1, 18th-19th August 2023, Pune, India, Published by IEEE, DOI: 10.1109/ICCUBEA58933.2023.10392119, Available: https://ieeexplore.ieee.org/document/10392119.

[65] Rachna Narula, Vijay Kumar, Renuka Arora and Rishu Bhatia, "Enhancing Job Recommendations Using NLP and Machine Learning Techniques", *TIJER - International Research Journal*, ISSN: 2349-9249, pp. 826-853, Vol. 10, No. 10, October 2023, Delhi, India, Published by TIJER (IJ Publication) Janvi Wave, Available: https://www.tijer.org/tijer/viewpaperforall.php?paper=TIJERTHE3019.

[66] Mahesh Batta, "Machine Learning Algorithms—a Review", *International Journal of Science and Research (IJSR)*, ISSN: 2319-7064, pp. 381-386, Vol. 9, No. 1, January 2020, India, DOI: 10.21275/ART20203995, Available: https://www.ijsr.net/archive/v9i1/ART20203995.pdf.

[67] Sreeramana Aithal and Varun Shenoy, "ABCD Analysis of Online Campus Placement Model", *IRA-International Journal of Management & Social Sciences*, Print ISSN: 2455-2267, Online ISSN: 2455-2275, pp. 227-244, Vol. 5, No. 2, November 2016, India, Published by IRA Institute of Research Advances, DOI: 10.21013/jmss.v5.n2.p3, Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2875403.

[68] Naresh Kumar, Manish Gupta, Deepak Sharma and Isaac Ofori, "Technical job recommendation system using APIs and web crawling", *Computational Intelligence and Neuroscience*, Print ISSN: 1687-5265, Online ISSN: 1687-5273, Vol. 2022, No. 1, June 2022, United States, Published by Hindawi Publishing Corporation, DOI: 10.1155/2022/7797548, Available: https://onlinelibrary.wiley.com/doi/10.1155/2022/7797548.

[69] Shlok Vishwasrao, Saurabh Birari, Thanmai Gudelli, Hiteshkumar Sharma and Suhas Chavan, "A Review on Training and Placement Management System", *International Research Journal of Engineering and Technology (IRJET)*, Print ISSN: 2395-0056, Online ISSN: 2395-0072, pp. 1-7, Vol. 9, No. 11, November 2022, India, Published by Fast Track Publications. Available: https://www.irjet.net/archives/V9/i11/IRJET-V9I1177.pdf.

[70] David A. Meyerson, Lawrence N. Meyerson, Andrew Bolson and Gregory A. Wilson, "A Legal 'Case' Against the Internship Placement System and a Proposal to Fix the System and the Internship Imbalance", *Training and Education in Professional Psychology*, Print ISSN: 1931-3918, Online ISSN: 1931-3926, pp. 174-184, Vol. 7, No. 3, August 2013, United States, Published by American Psychological Association (APA), DOI: 10.1037/a0033449, Available: https://psycnet.apa.org/record/2013-24408-001.

[71] Gautam Jaiswal, Aryan Uttam, Devesh Dhar Dubey and Pawan Kumar Mall, "Resume Analyser and Job Recommendation System Based on NLP", *in Proceedings of the 2024 2nd International Conference on Disruptive Technologies (ICDT)*, ISBN: 9798350371062, pp. 1584-1587, 15th-16th March 2024, Greater Noida, India, Published by IEEE, DOI: 10.1109/ICDT61202.2024.10489058, Available: https://ieeexplore.ieee.org/document/10489058.

[72] Cibele Santos, Fabrício Góis, Carlos Martins and Felipe da Cunha, "CVO: Curriculum vitae optimization by recommending keywords to undergraduate students", in *Big Data Technologies and Applications*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST), Print ISSN: 1867-8211, Online ISSN: 1867-822X, pp. 279-293, January 2023, Published by Springer Nature Switzerland AG, DOI: 10.1007/978-3-031-33614-0_19, Available: https://link.springer.com/chapter/10.1007/978-3-031-33614-0_19.

[73] Gabriel Adekunle, "Automating Data Retention from a Website Using an Application Programming Interface", *ScienceOpen Preprints*, ISSN: 2199-1006, pp. 220-226, Vol. 2, May 2023, DOI: 10.14293/PR2199.000114.v1, Available: https://www.scienceopen.com/hosted-document?doi=10.14293/PR2199.000114.v1.

[74] Pérez Molano, G., "An Overview of Web Scraping: Technical Aspects and Exercises," *Computer Science*, Vol. 1, No. 1, pp. 1-10, 2023, Published by Polytechnic University of Puerto Rico, San Juan, Puerto Rico, Available: http://hdl.handle.net/20.500.12475/1995.

[75] Vedant Bisht, Renu Choyal, Akshay Singh Negi and Kulvinder Singh, "Utilizing Python for Web Scraping and Incremental Data Extraction", in *Proceedings of the 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Print ISSN: 979-8-3503-4023-5, Online ISSN: 979-8-3503-4022-8, pp. 1450-1455, 11th-13th December 2023, Pudukkottai, India, Published by IEEE, DOI: 10.1109/ICACRS58579.2023.10404702, Available: https://ieeexplore.ieee.org/document/10404702.

[76] Kieron Turk, Sergio Pastrana and Ben Collier, "A tight scrape: Methodological approaches to cybercrime research data collection in adversarial environments", in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, ISBN: 978-1-7281-8598-9, pp. 428-437, September 2020, Published by IEEE, DOI: 10.1109/EuroSPW51379.2020.00064, Available: https://ieeexplore.ieee.org/document/9229744.

[77] Davor Dimoski, Bojana Koteska, Ljupco Pejov and Anastas Mishev, "Testing RESTful APIs – Use Case: RESTful API for Solving Multidimensional Time-Independent Schrödinger Equation", in *Proceedings of the Ninth Workshop on Software Quality Analysis, Monitoring, Improvement, and Applications* (SQAMIA), CEUR Workshop Proceedings, ISSN: 1613-0073, Vol. 3237, October 2022, Published by *National Initiatives for Open Science in Europe (NI4OS Europe)*, Available: https://ceur-ws.org/Vol-3237/paper-kot.pdf.

[78] Justin Littman, Daniel Chudnov, Daniel Kerchner, Christie Peterson, Yecheng Tan *et al.*, "API-based social media collecting as a form of web archiving", *International Journal on Digital Libraries*, Print ISSN: 1432-5012, Online ISSN: 1432-1300, pp. 21-38, Vol. 19, No. 1, March 2018, Germany, Published by Springer Verlag, DOI: 10.1007/s00799-016-0201-7, Available: https://link.springer.com/article/10.1007/s00799-016-0201-7.

[79] Yiou Lin, Hang Lei, Prince Clement Addo and Xiaoyu Li, "Machine learned resume-job matching solution", *arXiv*, Vol. 1607.07657, July 2016, New York, USA, Published by Cornell University - arXiv, DOI: 10.48550/arXiv.1607.07657, Available: https://arxiv.org/abs/1607.07657.

[80] José María Luna-Romera, Fernando Núñez-Hernández, María Martínez-Ballesteros, José C. Riquelme and Carlos Usabiaga Ibáñez, "Analysis of the Evolution of the Spanish Labour Market Through Unsupervised Learning", *IEEE Access*, ISSN: 2169-3536, pp. 121695-121708, Vol. 7, August 2019, Published by IEEE, DOI: 10.1109/ACCESS.2019.2935386, Available: https://ieeexplore.ieee.org/document/8798630.

[81] K. R. Chowdhary, "Natural language processing", in *Fundamentals of Artificial Intelligence*, Print ISBN: 978-81-322-3970-3, Online ISBN: 978-81-322-3972-7, pp. 603-649, April 2020, New Delhi, PublisSpringer, DOI: 10.1007/978-81-322-3972-7_19, Available: https://link.springer.com/chapter/10.1007/978-81-322-3972-7_19.

[82] S. M. Shawal Chowdhury, Mrithika Chowdhury and Arifa Sultana, "Matching job circular with resume using different natural language processing based algorithms", in *International Conference on Machine Intelligence and Emerging Technologies (MIET)*, Print ISSN: 978-3-031-34618-7, Online ISSN: 978-3-031-34619-4, pp. 428-442, Vol. 490, 23rd–25th September 2022, Noakhali, Bangladesh, Published by Springer, DOI: 10.1007/978-3-031-34619-4_34, Available: https://link.springer.com/chapter/10.1007/978-3-031-34619-4_34.

[83] Natalia Vanetik, Alona Kolesnev and Genady Kogan, "NLP-based Screening for IT Job Vacancies: A Case Study", in *Proceedings of the IACT - The 1st International Workshop on Implicit Author Characterization from Texts for Search and Retrieval*, Print ISSN: 1613-0081, Online ISSN: 1613-0073, pp. 13-22, Vol. 3477, July 2023, Taiwan, Published by CEUR Workshop Proceedings, Available: https://ceur-ws.org/Vol-3477/paper2.pdf.

[84] George Suciu, Adrian Pasat, Carmen Nădrag, Cristina Bălăceanu and Alexandru Drosu, "Design of an Internship Recruitment Platform Employing NLP-Based Technologies", in *Proceedings of the ECAI 2018 International Conference*, Print ISBN: 978-1-5386-4902-2, Online ISBN: 978-1-5386-4901-5, pp. 1-6, Vol. 325, 28th-30th July 2018, Iasi, Romania, Published by IEEE, Available: https://www.beiaro.eu/wp-content/uploads/2018/09/.

[85] Absalom E. Ezugwu, Abiodun M. Ikotun, Olaide O. Oyelade, Laith Abualigah, Jeffery O. Agushaka *et al.*, "A Comprehensive Survey of Clustering Algorithms: State-of-the-Art Machine Learning Applications, Taxonomy, Challenges, and Future Research Prospects", *Engineering Applications of Artificial Intelligence*, ISSN 0952-1976, p. 104743, Vol. 110, April 2022, United Kingdom, Published by Elsevier Ltd., DOI: 10.1016/j.engappai.2022.104743, Available: https://www.sciencedirect.com/science/article/abs/pii/S095219762200046X.

[86] Wenxing Hong, Siting Zheng, Huan Wang and Jianchao Shi, "A Job Recommender System Based on User Clustering", *Journal of Computers*, Online ISSN 1796-203X, pp. 1960-1967, Vol. 8, No. 8, August 2013, Finland, DOI: 10.4304/jcp.8.8.1960-1967, Available: https://www.jcomputers.us/vol8/jcp0808-10.pdf.

[87] Kristina Sinaga and Miin-Shen Yang, "Unsupervised K-means clustering algorithm", *IEEE Access*, ISSN: 2169-3536, pp. 80716-80727, Vol. 8, April 2020, United States, Published by IEEE, DOI: 10.1109/ACCESS.2020.2988796, Available: https://ieeexplore.ieee.org/document/9072123.

[88] Ping Ouyang, "The construction of college students' job recommendation model based on improved k-means-CF", *International Journal of Computational Systems Engineering*, Print ISSN: 2046-3391, Online ISSN: 2046-3405, pp. 190-198, Vol. 7, No. 2-4, 2023, Switzerland, Published by InderScience Publishers, DOI: 10.1504/IJCSYSE.2023.132918, Available: https://www.inderscienceonline.com/doi/epdf/10.1504/IJCSYSE.2023.132918.

[89] Vincent Cohen-Addad, Varun Kanade, Frederik Mallmann-Trenn and Claire Mathieu, "Hierarchical Clustering: Objective Functions and Algorithms", *Journal of the ACM (JACM)*, Print ISSN: 0004-5411, Online ISSN: 1557-735X,

pp. 1-42, Vol. 66, No. 4, June 2019, United States, Published by *Association for Computing Machinery (ACM)*, DOI: 10.1145/3321386, Available: https://dl.acm.org/doi/10.1145/3321386.

[90] N. Rajalingam and K. Ranjini, "Hierarchical Clustering Algorithm—A Comparative Study", *International Journal of Computer Applications*, ISSN: 0975-8887, pp. 42-46, Vol. 19, No. 3, April 2011, Published by Foundation of Computer Science (FCS), DOI: 10.5120/2340-3052, Available: https://www.ijcaonline.org/archives/volume19/number3/2340-3052/.

[91] Jing Ren, Feng Xia, Xiangtai Chen, Jiaying Liu, Mingliang Hou *et al.*, "Matching Algorithms: Fundamentals, Applications, and Challenges", *IEEE Transactions on Emerging Topics in Computational Intelligence*, ISSN: 2471-285X, pp. 332-350, Vol. 5, No. 3, August 2021, United States, Published by IEEE, DOI: 10.1109/TETCI.2021.3067655, Available: https://ieeexplore.ieee.org/document/9416305.

[92] Olayemi Olasehinde, Johnson Olanrewaju Victor, Abiona Akeem A. and Daniel Olayiwola Omilegan, "Design and Implementation of a Web-Based Internship Placement Recommendation System: A Case Study of Federal Polytechnic, Ile-Oluji, Nigeria", *International Journal of Scientific & Engineering Research (IJSER)*, Print ISSN: 2229-5518, pp. 398-409, Vol. 13, No. 3, March 2020, United States, Published by IJSER, Available: https://www.ijser.org/journal-volume13-issue3-March-2022-edition.aspx.

[93] Andréia Antunes da Luz, João Luiz Kovaleski, *"Framework Proposal for Management of Knowledge and Technology Transfer in Brazilian Academic Internships", Journal of Technology Management & Innovation*, Print ISSN: 0718-2724, Vol. 13, No. 3, pp. 3-11, September 2018, Brazil, Published by: Journal of Technology Management & Innovation Group, DOI: 10.4067/S0718-27242018000300003, Available: https://www.jotmi.org/index.php/GT/article/view/2576.

[94] Kais Dai, Celia González Nespereira, Ana Fernández Vilas and Rebeca P. Díaz Redondo, "Scraping and Clustering Techniques for the Characterization of LinkedIn Profiles", In *proceedings of the Fourth International Conference on Information Technology Convergence and Services (ITCS)*, 2nd-3rd January 2015, Zurich, Switzerland, Published by arXiv, DOI: 10.48550/arXiv.1505.00989, Available: https://arxiv.org/abs/1505.00989v1.

[95] Yuting Yan, Zhengyan Xie and Hui Chen, "Industry requirements for translators across China before COVID-19: Analyzing 51job listings through web scraping", *Revista Argentina de Clínica Psicológica*, Print ISSN: 1851-7951, Online Print: 0327-6716, p. 768, Vol. 29, No. 4, December 2020, Argentina, Published by Fundacion Aigle, DOI: 10.24205/03276716.2020.881, Available: https://www.revistaclinicapsicologica.com/archivesarticle.php?id=89.

[96] Sarah Chasins, *"Democratizing Web Automation: Programming for Social Scientists and Other Domain Experts"*, Doctoral Dissertation, No. UCB/EECS-2019-139, EECS Department, University of California, October 2019, Berkeley, Published by University of California, Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-139.html.

[97] Cynthia S. Liem, Markus Langer, Andrew Demetriou, Annemarie Hiemstra, Achmadnoer Sukma Wicaksana *et al.*, "Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening", in *Explainable and interpretable models in computer vision and machine learning*, H. Jair Escalante, S. Escalera, I. Guyon, X. Baró, & Y. Güçlütürk (Eds.), Springer, Print ISSN: 978-3-319-98130-7, Online ISSN: 978-3-319-98131-4, pp. 197-253, Vol. 1, November 2018, Published by Springer Nature Switzerland, DOI: 10.1007/978-3-319-98131-4_9, Available: https://link.springer.com/chapter/10.1007/978-3-319-98131-4_9.

[98] Rudresh Mishra, Ricardo Rodriguez and Valentin Portillo, "An AI based talent acquisition and benchmarking for job", *arXiv preprint arXiv:2009.09088*, August 2020, New York, USA, Published by Cornell University - arXiv, DOI: 10.48550/arXiv.2009.09088, Available: https://arxiv.org/pdf/2009.09088.

[99] SCM De S Sirisuriya, "Importance of Web Scraping as a Data Source for Machine Learning Algorithms - Review", in *Proceedings of the 2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*, Print ISBN: 979-8-3503-2364-1, Online ISBN: 979-8-3503-2363-4, pp. 134-139, 25th 26th August 2023, Peradeniya, Sri Lanka, DOI: 10.1109/ICIIS58898.2023.10253502, Available: https://ieeexplore.ieee.org/document/10253502.

[100] Chamila Maddumage, Dulanjaya Senevirathne, Isuru Gayashan, Tharusha Shehan and Sagara Sumathipala, "Intelligent Recruitment System", in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Print ISBN: 978-1-5386-8076-6, Online ISBN: 978-1-5386-8075-9, pp. 1-6, 29th-31st March 2019, Bombay, India, Published by IEEE, DOI: 10.1109/I2CT45611.2019.9033836, Available: https://ieeexplore.ieee.org/document/9033836.

[101] Sergio Rojas-Galeano, Jorge Posada, Esteban Ordoñez, "A Bibliometric Perspective on AI Research for Job-Résumé Matching", *The Scientific World Journal*, Print ISSN: 2356-6140, Online ISSN: 1537-744X, pp. 1-15, Vol. 2022, October 2022, Published by Hindawi Limited, DOI: 10.1155/2022/8002363, Available: https://onlinelibrary.wiley.com/doi/10.1155/2022/8002363.

[102] Thanveer Shaik, Xiaohui Tao, Yan Li, Christopher Dann, Jacquie McDonald *et al.*, "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis", *IEEE Access*, Online ISSN: 2169-3536, pp. 56720-56739, Vol. 10, October 2022, United Published by IEEE, DOI: 10.1109/ACCESS.2022.3177752, Available: https://ieeexplore.ieee.org/document/9781308.

[103] Gehad ElSharkawy, Yehia Helmy and Engy Yehia, "Employability Prediction of Information Technology Graduates Using Machine Learning Algorithms", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Print ISSN: 2158-107X, Online ISSN: 2156-5570, pp. 359-367, Vol. 13, No. 10, October 2022, United

Kingdom, Published by Science and Information Organization, DOI: 10.14569/IJACSA.2022.0131043, Available: https://thesai.org/Publications/ViewPaper?Volume=13&Issue=10&Code=IJACSA&SerialNo=43.

[104] Manisha Sharma and Anshul Ujlayan, "A Study to Enhance Candidate Screening Process Using Similarity Analysis", *International Journal of Business and Data Analytics (IJBDA)*, Print ISSN: 2515-9100, Online ISSN: 2515-9119, pp. 41-53, Vol. 1, No. 1, January 2019, Geneva, Switzerland, Published by Inderscience Publishers, DOI: 10.1504/IJBDA.2019.10020201, Available: http://www.inderscience.com/storage/f113712185692410.pdf.

[105] Jeanky S. Mendez and Jehan D. Bulanadi, "Job Matcher: A Web Application Job Placement Using Collaborative Filtering Recommender System", *International Journal of Research Studies in Education (IJRSE)*, Online ISSN: 2243-7711, pp. 103-120, Vol. 9, No. 2, July 2020, Philippines, Published by Consortia Academia, DOI: 10.5861/ijrse.2020.5810, Available: https://consortiacademia.org/10-5861-ijrse-2020-5810/.

[106] Pushya Chaparala, Amar Jukuntla, V. Sasidhar Reddy, V. Vishnu Vinayak and T. Pavani Sudha, "Extraction and Updating of Alumni Information from Web Profiles Using Web Scraping", in *Proceedings of the 2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*, Print ISBN: 979-8-3503-1495-3, Online ISBN: 979-8-3503-1494-6, pp. 1-7, 15th 16th September 2023, Kottayam, India, Published by IEEE, DOI: 10.1109/iQ-CCHESS56596.2023.10391404, Available: https://ieeexplore.ieee.org/document/10391404.

[107] Evis Trandafili, Alban Allkoçi, Elinda Kajo and Aleksandër Xhuvani, "Discovery and Evaluation of Student's Profiles with Machine Learning", in *Proceedings of the Fifth Balkan Conference in Informatics*, ISBN: 978-1-4503-1240-0, pp. 174-179, 16th-20th September 2012, Novi Sad, Serbia, Published by ACM, DOI: 10.1145/2371316.2371350, Available: https://dl.acm.org/doi/10.1145/2371316.2371350.

[108] Cherry D. Casuat and Enrique D. Festijo, "Predicting Students' Employability Using Machine Learning Approach", in *Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, Print ISBN: 978-1-7281-4083-4, Online ISBN: 978-1-7281-4082-7, pp. 1-5, 20th-21st December 2019, Kuala Lumpur, Malaysia, Published by IEEE, DOI: 10.1109/ICETAS48360.2019.9117338, Available: https://ieeexplore.ieee.org/document/9117338.

[109] Linsey S. Hugo, "Predicting Employment Through Machine Learning", *National Association of Colleges and Employers (NACE) Journal*, ISSN: 1542-2046, Vol. 2019, May 2019, Bethlehem, Pennsylvania, Published by NACE, Available: https://ebiztest.naceweb.org/career-development/trends-and-predictions/predicting-employment-through-machine-learning.

[110] Ni Luh Ratniasih and Ni Wayan Ninik Jayanti, "Sentiment Analysis of Stakeholder Satisfaction Measurement", *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*, Print ISSN: 1978-1520, Online ISSN: 2460-7258, pp. 217-226, Vol. 16, No. 2, July 2022, Yogyakarta, Indonesia, Published by IndoCEISS, DOI: 10.22146/ijccs.72245, Available: https://jurnal.ugm.ac.id/ijccs/article/view/72245.

[111] R. Deepalakshmi, J. Jeyarani, K. Balamurugan, R. Gokul, S. Krishnamoorthi *et al.*, "Efficiency and Effectiveness of Feedback Management in E-Commerce Using Machine Learning Algorithm", *Journal of Information Technology and Digital World*, Online ISSN: 2582-418X, pp. 341-351, Vol. 5, No. 4, November 2023, Chennai, India, Published by: Inventive Research Organization, DOI: 10.36548/jitdw.2023.4.001, Available: https://irojournals.com/itdw/article/view/5/4/1.

[112] Saket Maheshwary and Hemant Misra, "Matching Resumes to Jobs via Deep Siamese Network", in *Companion Proceedings of the The Web Conference 2018*, ISBN: 9781450356404, pp. 87-88, 23rd-27th April 2018, Lyon, France, Published by International World Wide Web Conferences Steering Committee, DOI: 10.1145/3184558.3186942, Available: https://dl.acm.org/doi/10.1145/3184558.3186942.

[113] Sundarambal Balaraman, Chokiyan Karthikeyini, Bommi R. M. and Suresh Subramanian, "AI-based Content Management System", in *Proceedings of the 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, Print ISBN: 979-8-3503-4892-7, Online ISBN: 979-8-3503-4891-0, pp. 1-8, 21st–23rd December 2023, Chennai, India, Published by IEEE, DOI: 10.1109/ICDSAAI59313.2023.10452640, Available: https://ieeexplore.ieee.org/document/10452640.

[114] Alex Luscombe, Kevin Dick and Kevin Walby, "Algorithmic thinking in the public interest: Navigating technical, legal, and ethical hurdles to web scraping in the social sciences", *Quality & Quantity*, Print ISSN: 0033-5177, Online ISSN: 1573-7845, pp. 1023-1044, Vol. 56, No. 3, May 2022, Netherlands, Published by Springer Netherlands, DOI: 10.1007/s11135-021-01164-0, Available: https://link.springer.com/article/10.1007/s11135-021-01164-0.

[115] Adrian-Costin Tundrea, Celestin Draganescu and Cosmin Popa, "Integrating Intelligent Transport Systems in a Risk Management System of Systems", in *Proceedings of the 21st International Conference on Control Systems and Computer Science (CSCS)*, ISBN: 978-1-5386-1839-4, pp. 385-391, 29th -31st May 2017, Bucharest, Romania, Published by IEEE, DOI: 10.1109/CSCS.2017.60, Available: https://ieeexplore.ieee.org/document/7968588.