

# ICode : Crafting Digital Resume for Coding Enthusiasts

Saurav Muke

Department of Computer Engineering  
Pune Institute Of Computer Technology  
Pune, India  
saurav54muke@gmail.com

Samruddhi Ahire

Department of Computer Engineering  
Pune Institute Of Computer Technology  
Pune, India  
samruddhi3915@gmail.com

Dr. Geetanjali Kale

Department of Computer Engineering  
Pune Institute Of Computer Technology  
Pune, India  
gvkale@pict.edu

**Abstract**—Although paper-based resumes may constitute a physical representation of skills, coding experts face several impediments. Limitations such as low accessibility, environmental concerns, and lack of ability to integrate multimedia contents and track participation. However, digital resumes provide a modern forum, which is best suited for the fast-changing tech environment. These tools facilitate an interactive presentation of skills, live coding progress, and interactive projects. Also, the incorporation of web scraping technologies allows programmers to gather and refresh facts from sites such as Codeforces, LeetCode and CodeChef, and makes their profiles up-to-date and informative. Incorporating GitHub and LinkedIn information in the digital resumes boosts them to the advantage of students and recruiters. GitHub highlights coding projects and contributions but LinkedIn creates professional networks and allows for endorsements and recommendations. This new method connects with the old paper resumes and brings an entirely new level of dynamism to the coding world.

**Index Terms**—web scraping, digital resume, visualisation, data analysis.

## I. INTRODUCTION

The current technologically advanced world is marked by innovations and cutting edge developments. This is pushing forward in numerous industries, whereas the paper-based resume remains in uncharted territories. It is very apparent in the coding enthusiasts' community, which is a lively and continually changing environment. However, there are many defects associated with the traditional paper resume, which might not be suitable for the rapidly changing modern job market.

Coding enthusiasts are not only expected to demonstrate their proficiency in traditional paper. This is because the presentation format used must be dynamic and responsive due to the nature of operation of such people. Their professional narrative should contain coding profiles, project showcases, and live updates on the development. These complex attributes are usually not fully covered by paper resumes. Traditional paper resume may as well show the current situation of the applicant, especially if the application was submitted long ago and it is still under the process of screening.

Here comes the "ICode" digital resume platform, which has been specially designed for people having computer science and engineering backgrounds. An idea that goes beyond a

regular resume is embodied in the innovative web-based platform "ICode". It allows real time data from popular sites like Leetcode, Codechef, and codeforces coding profile to be easily incorporated. Also, "ICode" is compatible with widely-known web platforms such as GitHub and LinkedIn. This project allows users to present their coding skills, project successes, and others in a dynamic way. This way, it meets the unique needs of programmers, but also manages to keep pace with today's changing labor market. "ICode" demonstrates the change in the professional presentation landscape and the emergence of sophisticated and dynamic profiles that adequately capture the intricacies of coding world.

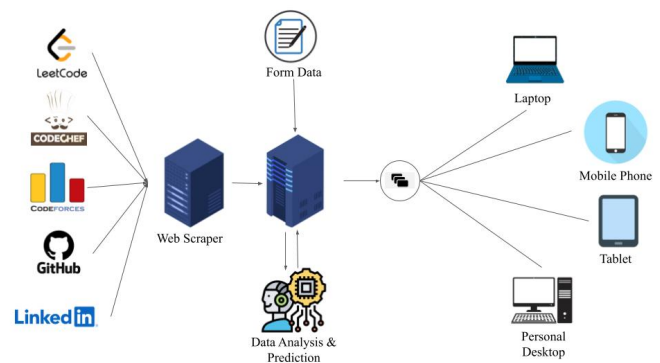


Fig. 1. System Architecture of ICode

The paper has the following structure: The summary of a digital resume for coding enthusiasts is provided in Section 2. The background and relevant literature are presented in Section 3. An overview of web scraping methods and processes for data collection is provided in Section 4. The method of the intended system is contained in Section 5. It concludes in Section 6.

## II. DIGITAL RESUME FOR CODING ENTHUSIASTS

Digital resumes are an advanced, technology-driven avenue for people passionate about software development and coding to showcase their accomplishments, efforts, and experiences towards prospective employers, partners, or the community. Due to the fact of transmitting much information in a playful and visually appealing manner, these digital versions of a person's resume are becoming quite common in the Tech

world. This is an extensive overview of the components of digital resumes for programmers:

- **Engaging Web-Based Style:** Digital resumes are mainly web-based documents and most times they appear as online portfolios or personal websites. In this format, users can interact with multimedia features such as videos or codes demos, click on links, and also view coding projects.
- **Displaying Technical Proficiency:** For coding enthusiasts, digital resumes are a great way to showcase their technical skills. It covers the technologies, tools, frameworks, and programming languages in which they are proficient. These skills can be organized in a more orderly and attractive manner.
- **Highlighting Coding Projects:** One of the most significant benefits of digital resumes is their ability to emphasize coding contributions and projects. Coders can choose to include links to their GitHub repositories, personal websites, apps etc. for more audience to see their work.
- **Competitive Coding Profiles:** Digital resumes may also include links to profiles on competitive coding platforms such as CodeChef, Codeforces, LeetCode, TopCoder, and others. This shows your proficiency in coding and allows viewers to judge your problem solving abilities and ranking.
- **Global Accessibility:** Digital resumes are accessible everywhere. This increases the visibility of coding enthusiasts to potential employers or collaborators worldwide, as they are no longer restricted by geographic boundaries.
- **Easy Updates:** Technology industry is dynamic and professionals must always be updating their projects and skill sets. Digital resumes are easy to update and always provide the viewer with up-to-date information.
- **Personal Branding:** Code enthusiasts' style and personality can be expressed through the layout and content of their digital resumes. This will be an important aspect of their personal branding, particularly in the context of a very competitive job market.

### III. BACKGROUND AND RELATED WORK

Technology improvements have led to a wide range of skills among students, particularly among coding enthusiasts. Because of this, showcasing these skills has become more difficult in the dispersed setting. The interview process still includes traditional paper-based resumes, which is deficient in many ways. One recent development in this industry to streamline the interview process is the digital portfolio or resume.

According to research, recruiters encounter a number of difficulties when seeking for candidates, which is causing a gap. It has been determined that an online resume format is necessary for a smooth hiring process. Research with recruiters has taken into account a number of processes, and a number of studies have been conducted with the aim of identifying the data required for an improved recruitment process [19].

With the use of multiple screening techniques like resumes, background checks, interviews, personality tests, and more, the employee selection process has developed into a sophisticated process. Among these techniques, the expert resume sticks out as being essential for figuring out which candidates move forward in the hiring process. When making decisions about a candidate's qualifications, including their skills, motivation, personality, and suitability for the position, recruiters frequently base their decisions on their resumes. Recruiters essentially use resume data to determine a candidate's employability and make preliminary screening selections [13].

Although India's higher education system has grown significantly, there are still a number of issues with accessibility, employability, and quality. The country's demographic dividend is at risk due to the millions of students pursuing postsecondary education, so growing higher education institutions (HEIs) and addressing the employability gap are essential. This is especially troubling for the engineering industry, where the employability of graduates is comparatively low. According to reports, there is a sizable skill gap in the IT sector, highlighting the necessity for graduates to have knowledge, abilities, skills, and soft skills (KAASOC) in order to succeed in a labor market that is changing quickly due to factors like Industry 4.0. Innovative programs like competitive programming and experiential learning programs are becoming more and more important to improve employability [14].

All the information needed to create a paper-based resume must also be available for digital resume creation. However, there is a wealth of information available to coding enthusiasts that goes far beyond text. Therefore, a standard technique for information retrieval or collection is required. Web scraping becomes effective because of the accessibility of information on the internet [1,4].

One QR-scanning-based digital resume system exists, but it is ineffective for coding enthusiasts because it doesn't cover a lot of information or real-time data [24]. Different information is available on different websites, and its complexity and accessibility vary. Several techniques are available to retrieve the data in order to obtain these information. With regard to specific applications, each technique performs differently [30].

According to a study, user information is crucial, and social and coding-related data are required [21]. It's necessary to gather and scrape the website data. These websites have specific policies regarding access to information, and there are some legal considerations regarding data access [16]. To aid in the recruitment process and provide a better visual representation, all of this data must be presented on a platform.

### IV. DATA COLLECTION USING WEB SCRAPING

#### A. Overview

Web scraping is the way to visit websites automatically, collect their HTML content, and parse it to derive particular information. It is commonly employed for purposes such as data analysis, research, price gathering, and comparison. Scraping a website generally entails making HTTP requests to the Web server, waiting for HTML response and using

HTML parsing tools to discover and extract the required data elements. This includes text, images, hyperlinks, tables and more. Managing dynamic content and pagination forms part of advanced scraping. It is also important to look into ethical and legal issues to ensure that the scraping practices are ethical and legal. If performed correctly, web scraping can give a lot of data and save lots of handmade works that would have been too difficult.

### B. Web scraping techniques

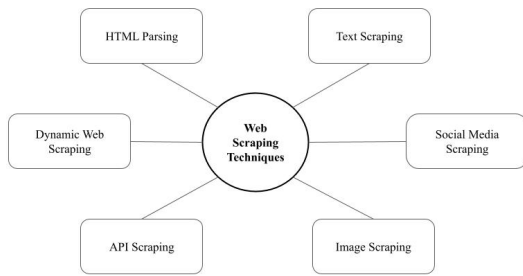


Fig. 2. Web Scraping Techniques

There are different types of web scraping depending on a number of factors and the specific use cases they cater to. The following are some common ways to scrape the web:

- **HTML Parsing:** The simplest sort of web scraping is HTML parsing, another term for static web scraping. It involves getting information from a web page by breaking its HTML code. This is suitable for sites where content loading is not heavily dependent on JavaScript.
- **Dynamic web scraping:** There are certain websites that use JavaScript to dynamically load data. In dynamic web scraping, headless browsers or browser automation libraries like Selenium are used to simulate user interaction, render page, and take out data from dynamically loaded content.
- **API Scraping:** Some websites have APIs that allow developers to retrieve structured data. This is referred to as API scraping, which is the HTTP requests to these APIs for structured data such as JSON or XML.
- **Text Scraping:** This method is often used to retrieve text-based content from websites, like product descriptions, blog posts, or articles. Textual data are extracted and processed based on natural language processing (NLP) techniques.
- **Image Scraping:** The process of gathering images from websites is known as image scraping. It is commonly used in the compilation of image galleries, the gathering of multimedia, and aggregating e-commerce products.
- **Social media scraping:** This form of scraping seeks to retrieve information from social media websites such as Facebook, Instagram, and Twitter. It includes the

collection of user profiles, posts, comments, and other social media engagements.

### C. Working of Web Scraping

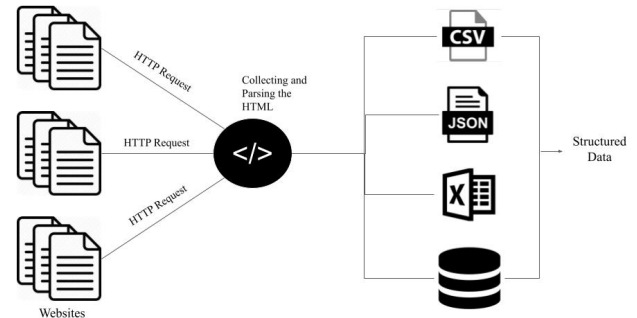


Fig. 3. Working of Web Scraping

- 1) **Sending an HTTP Request:** This process starts by sending an HTTP request to the URL of the target website. This request is directed towards the web server hosting the website. It demands the server to send it back the HTML code of the webpage in effect.
- 2) **Receiving HTML Content:** The web server sends back the page in HTML format to the web server after it has received the request. This HTML contains the page's structure and content, such as the text, images, links, and so on.
- 3) **Parsing the HTML:** A parsing library is employed to pull data from the HTML. Popular libraries such as BeautifulSoup (Python) or Cheerio (JavaScript) can be used for this purpose. The purpose of these libraries is to convert the HTML into a readable and organised format that allows a user to understand the structure of the webpage.
- 4) **Identifying Data Elements:** To scrape data from websites, it is important to locate the precise HTML elements that hold the specific data you want to extract. This is often done using XPath expressions or CSS selectors. These strategies help to pinpoint the specific place where the necessary data is located in respect to the HTML structure.
- 5) **Data Extraction:** It identifies the target elements after which it extracts the data. This may include any kind of material that is posted on the website, including text, pictures, links, and so on.
- 6) **Further Processing or Storage:** After extracting the data, it can also be enhanced within the scraping script to undertake various operations like reworking, analysis, or clearing based on the specified requirements. In place, the information can be stored in a database or file such as CSV and JSON to be used or analyzed at a later time.
- 7) **Handling Dynamic Content:** Modern websites often use JavaScript to load dynamic content. Web scrapers

may have to emulate these interactions in order to access any updated or hidden information contained in this dynamic content.

- 8) **Rate Limiting:** Web scrapers frequently use rate limiting, which regulates the frequency and volume of requests made to the server to ensure responsible scraping, to prevent overloading a website's server.
- 9) **Error Management:** Handling of errors is important in web scraping. This simply means fixing issues of dropped connections, lost data, or website structural adjustments. Robust web scrapers contain an error handling feature, which ensures that the web scraping runs well.
- 10) **Ethical and Legal Considerations:** When web scraping, it's important to adhere to the law as well as ethical guidelines. It includes seeking approvals, compliance with TCs on websites, compliance with copyright and data protection laws. One must also ensure they scrape the web in a responsible and legal manner.

## V. METHODOLOGY

Fig. 4 depicts the module architecture of the purposed system "ICode", which has in three modules: (1) Data Collection , (2) Data Storage and Analysis and (3) User profile generation and monitoring.

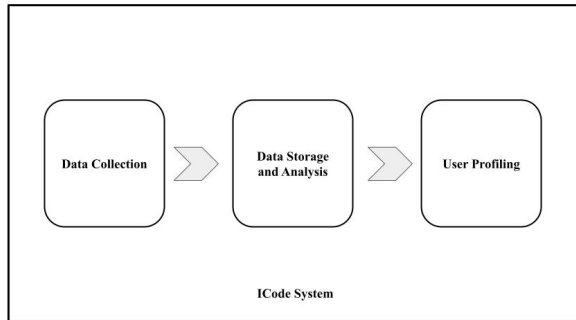


Fig. 4. Module Design of ICode

### A. Data Collection

It is the initial stage or module that is in charge of gathering user data and gathering data from different websites, including linkedin, github, Codechef, Codeforces, and Leetcode. When a user registers, a form is used to obtain their initial basic data. An independent server called the Scaper API gathers data from multiple websites by scraping other websites to provide other information like the development status and coding profile status.

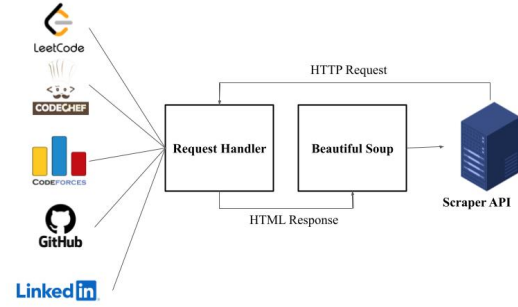


Fig. 5. Working of Scraper API

The Request Handler handles the request sent to the Scraper API, passing control to the appropriate website and retrieving the response as an HTML page. All of the information that needs to be extracted is present on the HTML page. The available web scraping tools, such as BeautifulSoup and Scrapy, are used to extract this data. For this project, beautiful soup has been used to facilitate effective data access. The data is scraped from the HTML page response and parsed using the BeautifulSoup library.

### Algorithm 1 Data parsing using BeautifulSoup

**Require:** URL of the webpage to scrape

**Ensure:** Extracted data from the webpage

```

0: Import Libraries:
0: import requests
0: from bs4 import BeautifulSoup
0: Send an HTTP GET Request:
0: url = 'https://example.com' {Replace with the target URL}
0: response = requests.get(url)
0: Check Response Status:
0: if response.status_code == 200 then
0:   Parse HTML Content:
0:   soup = BeautifulSoup(response.text, 'html.parser')
0:   Locate Relevant Elements:
0:   data = soup.find('div', class_='example - class') {Specify HTML tag and class}
0:   if data is not None then
0:     Extract and Process Data:
0:     extracted_data = data.text
0:     print(extracted_data)
0:   else
0:     print("Data not found.")
0:   end if
0: else
0:   print("Failed to retrieve the webpage. Status code:", response.status_code)
0: end if=0

```

### B. Data Storage and Analysis

The information obtained from several websites and the user-submitted form are diligently stored in a central database. This constitutes the basis of the ICode digital resume platform. The retrieved data is normally structured and can easily be saved in the database from the user's form, for instance, names and academic scores. This highly-organized information serves as the basis of further analysis.

Nonetheless, when handling data extracted from multiple coding profiles on sites including LeetCode, CodeChef, Codeforces, and GitHub, a different problem emerges. The data is inherently unstructured and diverse, which warrants a thorough preprocessing stage before it can be effectively used. This is the preprocessing of the data from these sources through parsing, normalizing and aggregating.

Data retrieved from coding profiles is highly valuable for placement prediction because it offers information about a user's coding skills, problem-solving skills, and overall coding performance. This unstructured data encompasses various dimensions, including:

- **LeetCode and CodeChef Data:** This information includes the solutions that the user has solved, contests in which the user has participated, ranking positions, and problem solutions that have been attempted by the user. The platform should make sense of this information to evaluate the user's coding skill and his/her ability to compete in coding tournaments.
- **Codeforces Data:** Codeforces profiles give information on a user's activity within competitive programming, contests participated, and problem-solving. This data needs to be analyzed for understanding the strengths of the coder.
- **GitHub Statistics:** A user's GitHub repository data like number of repositories, commits, contributions to open-source projects, and coding projects reveals their real world coding experience and collaborative ability.

The "ICode" platform uses this unstructured data undergoes a complete analysis so as to facilitate the prediction of the placement. The analysis encompasses various aspects, including:

- **Coding Proficiency Assessment:** Through their performance on various coding platforms, the platform evaluates the level of coding proficiency of a user. This entails assessing their solved problems, positions in contest and uniformity of participation.
- **Problem-Solving Skills:** The analysis explores user ability to solve problems, including coding competition success, the number of problems faced and ability to handle the complex ones.
- **GitHub Contributions:** To determine a user's real-world coding experience, GitHub looks at contributions like the number of repositories, project collaborations, and the relevance of coding projects to his/her field of study.
- **Academic Scores:** Additionally, academic scores are also included in the analysis, indicating the significance of education when predicting placements.

ICode combines these analyses, creating a unifying code of analysis regarding the user's coding skills, educational background, and practical experience. This all-inclusive profile serves as a foundation for placement predictions, which help both students and recruiters make sound decisions.

### C. User Profile Generation

The "ICode" project culminates in this module which serves as the turning point. This is where all the data that was

painfully collected from various sources and subjected to intensive analysis is systematically arranged and presented in an easily understandable manner.

There, all raw data collected from different coding profiles on platforms like LeetCode, CodeChef, Codeforces, as well as the collected structured data from user-sponsored forms, are combined to create a single, integrated, and complete user profile.

As such, the user profile is constructed in such a way that it is both informative and appealing to the eyes, making it a one-stop place showcasing skills in coding, academic achievements, and practical experience. However, this does not mean that this profile is just a static collection of data about a user professional journey but an interactive profile.

Key features of the user profile within this module include:

- **Data Integration:** This module efficiently combines data from different sources, allowing the data to be presented in a unified manner. A comprehensive picture of the user's academic scores, coding abilities, GitHub contributions, and participation in coding challenges is derived through a harmonious integration of these four.
- **Visualization:** These visualisation elements enhance user experience. Data is presented through charts, graphs, and visual summaries in a way that both an individual looking for a job and recruiters can comprehend the most important points just by taking a quick glance.
- **Real-time Updates:** The user profile is created to be dynamic and is updated in real-time based on changes in the user's coding journey. Similarly, as one participates in coding contests, solves problems, or contributes to GitHub projects, these updates are made on their profile.
- **Analytical Insights:** The module ensures that all the insights from the analysis of the user's data are presented in a clear manner. For instance, it can contain a coding proficiency evaluation, strengths, weaknesses, and how they relate to placement predictions.
- **User Interaction:** Users can navigate through their records, access their achievements and even edit their profiles. Such interactivity allows users to be in control of how they portray themselves to potential employers.
- **Validation:** The user profile is, however, intended not only to be self-presentation but also for validation as well. These profiles can be accessed by the recruiters to have a deep insight of what a candidate is capable of and this helps the employer simplify the recruitment process.

Therefore, this module converts data into a compelling story that allows people to tell their coding story in detail and recruiters to make informed placement decisions. It is a dynamic link between the user's multifaceted journey in coding and the shifting expectations of the job market.

## VI. CONCLUSION

An outdated paper-based resume is deficient in many areas of computer engineering and does not appropriately present the information. The industry revolution also necessitates a shift in the information representation model so that recruiting

teams can more effectively assess applicants. The profile of a coding enthusiast is worth taking into account during the hiring process. that's why it's important to have a digital resume that includes all coding information.

ICode is a digital resume platform that guarantees filling the gaps for coding enthusiasts, by providing a platform that displays all information about the user in real-time and also generates a one-stop solution to showcase coding profile ratings, github stats, and linkedin information.

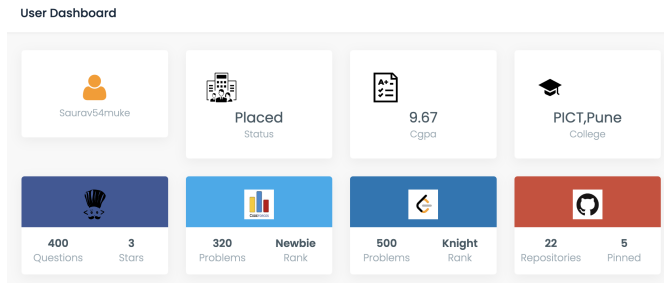


Fig. 6. User dashboard

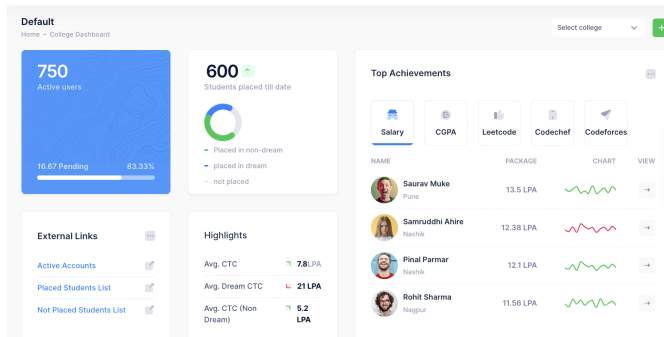


Fig. 7. College dashboard

"ICode" proved to be meeting all of the user's requirements with nearly all of the checkboxes checked, according to reviews done by 300+ college students who were evaluated on a variety of criteria, including accuracy, efficiency, design, ease of use, and portability.

## REFERENCES

- [1] Sirisuriya, De S. "A comparative study on web scraping." (2015).
- [2] Di Mascio, Tania, Luigi Laura, and Marco Temperini. "A framework for personalized competitive programming training." 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET). IEEE, 2018.
- [3] Chou, Yi-Chi, Chun-Yen Chao, and Han-Yen Yu. "A résumé evaluation system based on text mining." 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIC). IEEE, 2019.
- [4] Singrodia, Vidhi, Anirban Mitra, and Subrata Paul. "A review on web scraping and its applications." 2019 international conference on computer communication and informatics (ICCCI). IEEE, 2019.
- [5] Eke, Christopher Ifeanyi, et al. "A survey of user profiling: State-of-the-art, challenges, and solutions." IEEE Access 7 (2019): 144907-144924.
- [6] Anisha, P. R., Nhu Gia Nguyen, and G. Sreelatha. "A Text Mining using Web Scraping for Meaningful Insights." Journal of Physics: Conference Series. Vol. 2089. No. 1. IOP Publishing, 2021.
- [7] Fayzrakhmanov, R. R., Sallinger, E., Spencer, B., Furche, T., Gottlob, G. (2018, April). Browserless web data extraction: challenges and opportunities. In Proceedings of the 2018 World Wide Web Conference (pp. 1095-1104).
- [8] Pal, Ajay Kumar, and Saurabh Pal. "Classification model of prediction for placement of students." International Journal of Modern Education and Computer Science 5, no. 11 (2013): 49.
- [9] Liu, Chung-Chu. "Comparing and Prioritizing Measures of Personal Digital Resume: Job Seekers and Human Resource (HR) Practitioners' Perspectives." Journal of Computers 25.2 (2014).
- [10] Maurya, Laxmi Shanker, Md Shadab Hussain, and Sarita Singh. "Developing classifiers through machine learning algorithms for student placement prediction based on academic performance." Applied Artificial Intelligence 35.6 (2021): 403-420.
- [11] Bhattacharya, Madhumita, and Maggie Hartnett. "E-portfolio assessment in higher education." In 2007 37th annual frontiers in education conference-global engineering: knowledge without borders, opportunities without passports, pp. T1G-19. IEEE, 2007.
- [12] Ribeiro, P. and Guerreiro, P., 2008. Early introduction of competitive programming. Olympiads in Informatics, 2, pp.149-162.
- [13] Wright, Edward W., Theresa A. Domagalski, and Ronald Collins. "Improving employee selection with a revised resume format." Business Communication Quarterly 74, no. 3 (2011): 272-286.
- [14] Nair PR. Increasing employability of Indian engineering graduates through experiential learning programs and competitive programming: Case study. Procedia Computer Science. 2020 Jan 1;172:831-7.
- [15] Kobayashi, Mei, and Koichi Takeda. "Information retrieval on the web." ACM computing surveys (CSUR) 32, no. 2 (2000): 144-173.
- [16] Krotov, Vlad, and Leiser Silva. "Legality and ethics of web scraping." (2018).
- [17] Zamora, A., 2019. Making room for big data: web scraping and an affirmative right to access publicly available information online. J. Bus. Entrepreneurship L., 12, p.203.
- [18] Furtmueller, Elfi, Celeste Wilderom, and Mary Tate. "Managing recruitment and selection in the digital age: e-HRM and resumes." Human Systems Management 30, no. 4 (2011): 243-259.
- [19] Furtmueller, Elfi, Celeste Wilderom, and Roland M. Mueller. "Online resumes: Optimizing design to service recruiters." (2010).
- [20] Dai, Kais, Celia González Nespereira, Ana Fernández Vilas, and Rebeca P. Díaz Redondo. "Scraping and clustering techniques for the characterization of LinkedIn profiles." arXiv preprint arXiv:1505.00989 (2015).
- [21] MURALIDHAR, A., JV THOMAS ABRAHAM, and SRI SAI SARAN REDDY YETURU. "STUDENT PROFILE SCORING BASED ON RESUME AND WEBSITE DATA."
- [22] Essa, A., Ayad, H. (2012, April). Student success system: risk analytics and data visualization using ensembles of predictive models. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 158-161).
- [23] Manjushree, B. S., and G. S. Sharvani. "Survey on Web scraping technology." Wutan Huatan Jisuan Jishu 16, no. 6 (2020): 1-8.

- [24] Kannan, MK Jayanthi, Zerubabel Yonas Nematullah, and Zekrullah Nazeri. "The Digital Resume in an Extended Format."
- [25] Bhatt, Hitarthi, Shraddha Mehta, and Lynette R. D'mello. "Use of ID3 decision tree algorithm for placement prediction." *Int J Comput Sci Inform Technol* 6, no. 5 (2015): 4785-9.
- [26] Johnson, Faustina, and Santosh Kumar Gupta. "Web content mining techniques: a survey." *International journal of computer applications* 47, no. 11 (2012).
- [27] Snell J, Menaldo N. Web scraping in an era of big data 2.0. *Bloomberg Law News*. 2016 Jun 1.
- [28] Khder, Moaiad Ahmad. "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application." *International Journal of Advances in Soft Computing Its Applications* 13.3 (2021).
- [29] Zhao, Bo. "Web scraping." *Encyclopedia of big data* 1 (2017).
- [30] Bale AS, Ghorpade N, Rohith S, Kamalesh S, Rohith R, Rohan BS. Web Scraping Approaches and their Performance on Modern Websites. In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC) 2022 Aug 17 (pp. 956-959). IEEE.