

# STUDENT PROFILE SCORING BASED ON RESUME AND WEBSITE DATA

MURALIDHAR A<sup>1</sup>, J V THOMAS ABRAHAM<sup>2</sup>, SRI SAI SARAN REDDY YETURU<sup>3</sup>

<sup>1,2,3</sup>School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India.

E-mail: [muralidhar.a@vit.ac.in](mailto:muralidhar.a@vit.ac.in)

Received: 28.04.2020

Revised: 01.06.2020

Accepted: 20.06.2020

**Abstract**— In the present scenario, recruiters are using automated resume evaluators for shortlisting the applicants. This existing method is taken a step further by automating the recruiting process to a maximum extent in this proposed system. The proposed system analyses the electronic resume and evaluates them based on a particular job description given. Later the data from their social profiles like GitHub, Code chef, and Hacker rank is gathered using web scraping techniques that are used for further evaluation of the applicant. This proposed system is useful to the recruiters to select the most fitting applicant without any inclinations or human errors, which benefits the organization finally. Since a machine performs the entire process, there are very minimal possibilities of errors or biases, which helps the applicant also. To determine the perfect fit for the job, they are ranked based on the scores secured.

**Index Terms**— Automation, Machine learning, Artificial intelligence, Resume evaluator, Web Scraping, Job application, Website evaluation.

## 1 INTRODUCTION

Automation is one of the most heard terms nowadays, we humans are nearly working towards automating everything [1]. In a similar fashion, why shouldn't we automate the recruiting process? Automation reduces the wastage of human resources and other resources. The result obtained by following this method would be fair.

Many giant corporations receive thousands of job applications for each availability [2]. It would be a robust process for the firm to evaluate and manage these large numbers of applications. This process of evaluation must be performed within minimal time to carry on the further recruitment process smoothly. This evaluation to be done in minimal time requires a bunch of human and additional resources. Furthermore, the evaluation and segregation of applications should be delivered, with no inclination or error. Automating this process resolves all the problems of evaluating and segregating the applications in minimal time and without any intentional bias.

As per the conventional recruiting process, after selecting the appropriate applications, these candidates are invited for an examination or a direct in-person interview [3]. In the process of the in-person interview, the interviewer may be taken away with his/her personal opinions or emotions without acknowledging the genuine skill in the candidate. This process may lead to the refusal of the best candidates, which is not desirable for the candidate as well as the organization as hiring skilled candidates is at most important for a successful firm. Moreover, applicants could be affected by their physical and psychic status. The applicant may display nervousness and underperform at that particular moment. This situation could be overlooked by the interviewer, which directs the rejection of skillful candidates.

In the proposed system, the applicants are assessed both based on their resume and their practical or domain knowledge. The evaluation is being performed using the data scraped from their social websites like GitHub, Code chef, and Hacker rank. During the initial round, the candidates are evaluated based on their resume. Later they are evaluated based on their social website data. Finally, a shortlist is obtained, based on both the scores consolidated.

## 2 LITERATURE SURVEY

There are some automated hiring systems available which either shortlist based on the resumes collected or based on the social media data.

Chou, Yi-Chi, Chun-Yen Chao, and Han-Yen Yu have proposed a hiring system based on text mining in the year 2019. They have gathered the resumes from the recruitment websites and evaluated them. The complete system is built for Chinese resumes. They have taken the education and experience of the candidate into consideration. The report of all the candidates is made for a perfect comparison.[1]

Mohamed, Ashif, has designed the recruiter-resume ranking system in the year 2018. In this system, Information extraction is done from resumes using NLP and stored in the database. When companies give their requirements, they match them with resumes and recommend candidates using a candidate ranking algorithm.[2]

Khan, Tabrez, proposed a resume ranking system. In this system, they have used Natural Language processing techniques for the evaluation of the resume. They have taken the part of GitHub and Linked in profile data to give the score. They have also created visual reports based on each candidate's resume.[3]

Bollinger, Jacob, David Hardtke, and Ben Martin have used social data for resume job matching. In this work, they have scraped the data from various social sites like Facebook, Linked-in. They have matched the data with job requirements and then recommended a candidate. This method helps the human resource team to understand the capabilities of the candidate like strengths and weaknesses but not much help for the assessment of technical aspects.[4]

Apart from all these mentioned models, in this proposed system, various additional attributes have been taken into consideration for the evaluation of the resume. Also, technically based websites chosen to know the candidate's practical knowledge. In the following segment, the proposed methodology is being described.

### 3 METHODOLOGY

The primary purpose of proposing this hiring system is to make sure that no extra resources are used and no bias in the selection process. There are two steps in the evaluation of the candidate in this system, and they are resume evaluation and social websites evaluation. For assessment based on the resume, the requirements mentioned in the job description are matched with resume and score is allotted to each applicant. Then the shortlisted candidates are assessed based on their social websites. For collecting data on social websites, we use web scraping techniques. A score is allotted for the applicants based on their social data. Finally, both these scores are combined, and the final shortlist is prepared.

#### 3.1 System Flow

As demonstrated in Fig.1, the first step is to gather the data, i.e., both the resumes and job descriptions. The collected data is processed and then stored in CSV format, which is easy to reuse. After this data collection process, the resumes are evaluated based on a particular job description, and scores are allocated. The first shortlist is made here, the candidates who do not reach the minimum requirements are rejected. Then the remaining applicant's social data is collected and evaluated. Scores based on social website data is also provided. Both the scores are combined, and the final shortlist is made and stored in separate CSV.

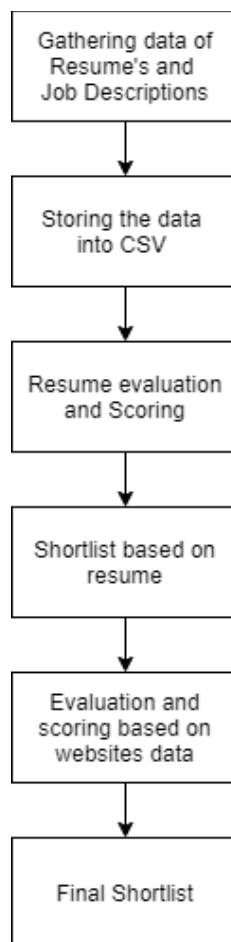


Fig. 1. The brief flow of proposed system.

**3.2 Data Gathering and Preprocessing**

Data collection and pre-processing is a crucial step in designing any system. Here, we have to gather the data of the resume from the applicants and job description data from the hiring team. The resumes and job descriptions are collected in a specific format.

The gathered data has to be pre-processed and stored in a structured format, which makes accessing the data simple. In the pre-processing step, the obtained resumes and job descriptions are converted into CSV format. Table1 and Table2 show the features extracted from collected resume data and job description data, respectively.

**Table1. Features of Resume Data**

<b>S.no</b>	<b>Attributes of Resume Data</b>
1	Name
2	Mail
3	Address
4	Highest Education
5	CGPA
6	12 <sup>th</sup> Score
7	10 <sup>th</sup> Score
8	Programming Skills
9	Domains Interested in (Domains in which Projects are done)
10	Internship Domain
11	Internship Duration
12	Experience
13	GitHub Username
14	Hacker rank Username
15	Code Chef Username
16	Willing to Relocate?

**Table2. Features of Job Description Data**

<b>S.no</b>	<b>Attributes of Job Description Data</b>
1	Company name
2	Role offered
3	10 <sup>th</sup> Cut-off
4	12 <sup>th</sup> Cut-off
5	Package
6	Eligibility Branches
7	Location
8	Skills Required
9	Job Domain
10	Experience Required

Apart from the data collected for the resume and job description, we have to gather the data from the websites. Retrieving data from the websites is the most significant part of the evaluation of the candidate’s profiles. For obtaining this information in an automated way, we use web-scraping techniques. Web scraping is used to extract a large amount of data from the websites and store it in local devices in a structured format [8]. In general, the data displayed on the website can only be seen when we use a web browser. For using this data in large applications, it is not feasible to manually copy paste the data. This manual process takes much time, which is not preferred. To overcome this and to use the data present in the websites, a web scraping technique is used.

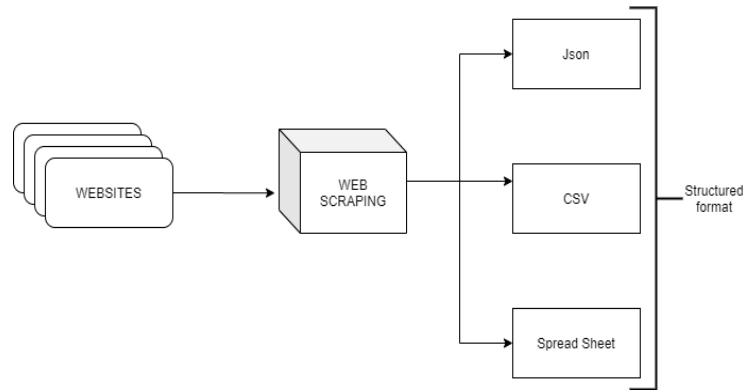


Fig. 2. Web Scraping Overview.

Relevant fields of data are being scraped from each of the websites. Some of the most crucial fields retrieved and used for scoring are mentioned below.

- GitHub -number of followers, the number of repositories.
- Code Chef -Ratings
- Hacker Rank -Score for each programming skill

**4 RESUME EVALUATION**

For the evaluation of the resume, we consider both the resume data and the job description data. By comparing the requirements given in the job description with the qualities mentioned in the resume, the final score is obtained. Firstly, basic cut-off requirements such as branch, minimum CGPA, 10th, and 12th scores are examined. The applications which do not meet the conditions are rejected. After this rejection, the remaining applications are assessed for scoring purposes. We use a weighted scoring technique for assigning a score for each candidate. Weight for each of the fields is as follows.

- programming skills-50%
- domain skills-20%
- internship domain-20%
- internship duration-10%.

Consolidating all these scores out of 100, a final shortlist based on the resume scores is made and stored.

**5 WEBSITE EVALUATION**

After retrieval of the data from the websites using web scraping techniques, we have to apply scoring algorithms on the data. The final scores based on websites are obtained after applying the data in the proposed algorithm. A total of 50% weightage is given to the score obtained based on the websites. The other 50% is obtained from the resume data. The primary purpose of evaluating the websites and giving scores based on them is to know the candidate’s practical knowledge.

Three different scores are obtained for each candidate based on their GitHub, Code Chef, and Hacker rank profiles. After the score is obtained by each applicant for each website, all of them are combined to form the website-based score. Weightage of score based on each website is as follows.

- Hacker rank Programming skills-60%
- Code Chef Rating-20%
- GitHub repositories and followers-20%

The second shortlist based on the website score is again stored in another CSV.

**6 RESULTS**

The approximate time taken by this system for evaluating the different number of resumes is calculated. Time calculated includes the time taken from the conversion of resumes into CSV to providing the resume score for each candidate. Fig.3 depicts the time taken in seconds for the evaluation of the resumes.

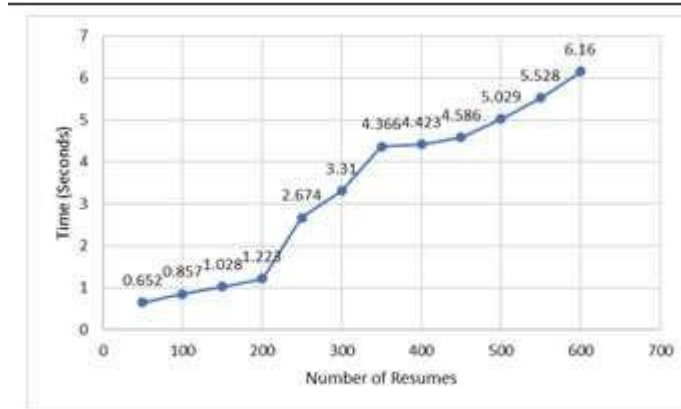


Fig. 3. Time taken for resume evaluation.

In the above figure, the Y-axis represents the time taken in seconds, and X-axis represents the number of resumes being evaluated. The time is measured based on several iterations of execution. The model is being executed several times for the time taken for the website evaluation, and the average time is gathered. The average time taken to scrape the data from all three websites and evaluate them is about 7.9 seconds. Some of the factors influencing the time taken by the system are server-side load, connectivity at the time of scraping, etc.

As both the resume and website-based scores are combined for preparing the final shortlisted candidates. The outcome of the system would be a CSV file that has the shortlisted candidate’s information. The information would be candidate-id, name of the candidate, resume score, website score, and final score. All three scores are mentioned in the final output. The hiring team may use any of these three scores or maybe allotting cut-offs for each score for shortlisting based on the requirements.

**7 CONCLUSION**

Various aspects like programming skills, domain skills, competitive programming skills (from website data), etc., are considered for the scoring, which makes it easier to find the perfect fit for the job. The proposed system is beneficial for both the recruiting team and the candidates. For the recruiting team, it is useful as it requires very minimal requirements compared to the conventional hiring process. As it also prevents any kind of human errors or bias, it is helpful for the candidates.

**8 FUTURE WORK**

Apart from shortlisting candidates based only on technical stuff, we can also use their social profile data to know more regarding them. With the social profile data of the applicant, we can understand their personality traits, psychometric scores, etc. Utilizing both the technical and social score gives a complete overview of the candidate. In the future, to produce such completeness, we would also use social data.

**REFERENCES**

[1] Adam C. Uzialko : “Workplace Automation is Everywhere, and It’s Not Just About Robots”, 2019 [Online], Available: <https://www.businessnewsdaily.com/9835-automation-tech-workforce.html>

[2] Ivana: “70 Recruitment Statistics for Attracting Top Talent in 2020”, 2019 [Online], Available: <https://www.smallbizgenius.net/by-the-numbers/recruitment-statistics/>

[3] jobsoid :” Recruiting Process (From sourcing to onboarding everything in between)”, 2019 [Online], Available: <https://www.jobsoid.com/recruitment-process/>

[4] 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 (ICAIIIC). IEEE, 2019.

[5] Mohamed, Ashif, et al. “Smart Talents Recruiter-Resume Ranking and Recommendation System.” 2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS). IEEE, 2018.

[6] Khan, Tabrez. “Resume Ranking using NLP and Machine Learning.” (2016). Available: [http://www.ijirce.com/upload/2016/april/218\\_Intelligent.pdf](http://www.ijirce.com/upload/2016/april/218_Intelligent.pdf)

[7] Bollinger, Jacob, David Hardtke, and Ben Martin. “Using social data for resume job matching.” Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media. 2012.

[8] Saurkar, A., Pathare, K., & Gode, S. (2018). An Overview On Web Scraping Techniques And Tools. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4), 363-367.