

Candidate Hiring and Personality Prediction through CV Analysis Using Machine Learning

Ashok Bodhe¹, Sarika Bendbhar², Aanjali More³, Nilam Kendale⁴,

1, 2, 3, 4 Student, Department of Computer Engineering, Siddhant College of Engineering, Pune, Maharashtra, India.

Submitted: 15-05-2022

Revised: 25-05-2022

Accepted: 28-05-2022

ABSTRACT - In the modern era of information technology most organizations have realized that human resource is very important for their success, but the efficiency of the human resource solely depends upon getting right person with right skills for right job. This recruitment application is system in which HR can automate the human resources. Applicant can also register themselves online, view organization requirements and apply for the suitable job. This kind of application plays an important role in simplifying the recruitment process. The system has facilities where candidates can upload the CV's and other academic achievements. Recruitment application make possible for manager to access information that is crucial to managing their staff, which they can see for human resource management, staffing and planning activities. The primary purpose to develop this system is to optimize the recruitment process for an organization. Besides, the equal qualified applicants could be selected by their application based on their qualification and company requirement.

Key Words: Information extraction, Filtering, Ranking, Shortlisting, Sorting.

I. INTRODUCTION

All major industries today are driven by technology. According to current statistics, information available on the internet is about 77% of what we need. This figure is expected to rise exponentially in the near future. Companies are publishing more and more information on the internet about every aspect of their business and their growth. Recruiters receive large numbers of applications through

emails, online job portals, or through services provided by partner staffing companies. Online job portals like monster.com, shine.com, naukri.com, and firstnaukri.com draw in most of the applications. Resumes obtained from such diverse sources are thus difficult to process and store in a unified database format. It becomes very tedious to select the most appropriate ones. Since resumes are structured documents containing information based on the author's thinking and writing skills, they can be created in a multitude of formats (e.g., plain text or structured table, languages). This makes the information extraction (IE) process highly complex. Dynamic filtering techniques are used by the industry to extract relevant resumes. These filtering techniques match hundreds of resumes from the database to a single job posting. Resumes extracted by these filters are generally similar to each other as they satisfy the same search criteria, based mainly on keyword matching. The application fetches resumes which satisfies the requirement of a particular job post.

II. EXISTING SYSTEM

Existing system performs cutthroat process that generally leaves the companies with a good, but not great employee at a significant financial investment. As the industries have grown, their hiring need has rapidly grown. To serve these hiring needs certain consultancy units like employment websites have come into existence. They offer a solution in which the candidate has to upload their information and submit it to the website. Then these websites would search the candidates based on certain keywords. These websites are middle level organizations between the candidate and recruiter.

These websites are not flexible as the candidate has to upload their resume in a particular layout, and these formats changed from system to system. These systems charge a certain amount per resume. According to the survey Monster.com charges Rs.1.44 lac (INR) for 50,000 resumes. Hence such systems are not cost effective.

III. PROPOSED SYSTEM

The motivation behind this system is to develop an application that will assist organizations in the recruitment process. This is a far different approach than employment websites. Our system allows the candidate to enter information about academics, skill set etc. and upload their resumes. The entered information is then analyzed by our system. This makes our search process easy. The analyzing system works on the algorithm that uses ranking, which is a sub domain of Text Mining System reads the information entered by user such as SSC marks, HSC marks, degree aggregate, programming languages known and performs ranking. This acquired information is stored in the database. This stored information can be accessed by HR, HR can simply provide keywords to the system and system will find all the relevant resumes that match with the keywords.

If we take n-1 interviews, we cannot ignore the possibility of nth candidate being better suited than everyone else. The balance of time and quality is an important problem that we face also in our daily lives. In the modern interview process it becomes necessary to handle this situation carefully as the candidate pool is quite large.

Secretary Problem-The establishment of the secretary problem is as follows. Suppose you are a Hiring manager who has a task to hire a secretary for your boss. You take up the challenge but face a problem later on with number of candidates. As candidates who are interested in the job role are quite large in number it might take quite a lot of time to interview everyone. We might employ following two approaches:

- 1) Interview everyone and select the best among them or
- 2) Select the first candidate who meets the requirements.

The problem with first approach is it takes quite a lot of time to interview everyone. While your company might stop at nothing to get only the best employees it might be challenging for small companies or startups to spend that amount of time on hiring.

The second approach then seems promising

and intuitive. It makes sense to select the one candidate who meets the requirement at the earliest. But just as first approach it has a fundamental flaw. The flaw can be emphasized by following questions?

- 1) How do we rank them to appear for the interview?
- 2) How can we be sure that this candidate is best among all the candidates?
- 3) Can we take some more interviews to acquire a better suited candidate?

Above questions are harder to answer than we think. As it is apparent that our first approach which is to take every interview and then select the best performer takes a lot of time, while the second approach which is to select the first candidate who meets the requirements compromises with quality.

IV. SYSTEM ARCHITECTURE

Major components of the proposed system:-

User or Candidate-User or candidate will login to the system in order to go through the process\interview.

Registration: Candidates will be allowed to register through website and can upload their resumes.

Receiving Notification: Candidate will be notified after registration and after getting selected in recruitment process.

HR

Retrieving candidates: HR can view the profile of candidate.

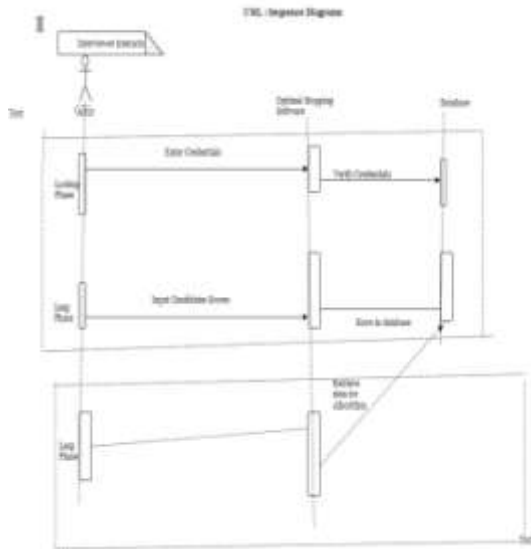
HR can send the notifications to the candidates.

Software

Software extracts the information from CV's using algorithm, Storage as well as retrieval of the information. Software will send notification to candidate.

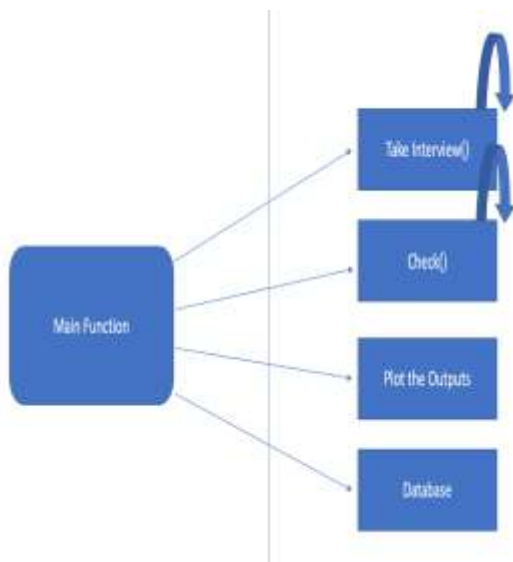
Ranking Algorithm

Each candidate will be scored based on the skillset, experience and academics. Scoring will also be influenced by user profile. The focus in Information



Retrieval research lays on text classification systems which make binary decisions for text document as either relevant or non-relevant with respect to a user's information need. We used precision, recall and F-measure metrics for performance evaluation.

V. BLOCK DIAGRAM



VI. PROBLEM

How do we balance time and quality in the modern interview process? As stated earlier, we will be employing optimal stopping algorithm more specifically the 37% rule along with Machine learning and common intuition to solve this problem.

VII. SOLUTION

Many of the thinkers earlier have tried to automate the process of hiring. Although it seems like we are progressing in terms of technology at a pace at which everything seems possible to automate the simple answer to automated hiring is "We are not there yet." While taking an interview the interview panel looks for their requirements, candidate's experience, interests, energy, confidence, ability to express ideas clearly, etc. apart from their aptitude scores. These analog entities cannot be translated to a single value so that a bot can take an interview and select candidates who meet the requirements. Therefore, in our research we arrived at the conclusion that current interview technique is better - at least for now. In this paper we will optimize this current interview technique which involves taking interview face to face by a panel and based on intuition select a candidate.

VIII. PERFORMANCE MEASURES:

Precision- Precision measures the number of relevant items retrieved as a percentage of the total number of items retrieved.

$$\text{Precision} = \frac{\text{\#relevant items retrieved}}{\text{\#retrieved items}}$$

Recall- Recall measures the number of relevant items retrieved percentage of the number of relevant items in the collection.

$$\text{Recall} = \frac{\text{\#relevant items retrieved}}{\text{\#relevant items}}$$

IX. CONCLUSION AND FUTURE SCOPE

We are representing a highly efficient, cost-effective recruitment system. The system collects and analyzes the data from the websites. We are addressing the problem regarding recruitment of right candidate for right post in economical way. In this, we have implemented a system which helps organization to hire eligible candidates. Because of this system, HR can directly search for required resumes in a fraction. In future scope, the system can be extended by including aptitude tests, language proficiency tests etc.

REFERENCES

- [1]. Brian Christian and Tom Griffiths, "Algorithms to Live By: The Computer Science of Human Decisions", ISBN 9781250118363
- [2]. Zishuo Zhao, "An Asymptotic Analysis on Generalized Secretary Problem", Computer Science and Game Theory (cs.GT), 16 Sep 2019



- [3]. Thomas S. Ferguson, “Who Solved the Secretary Problem?”, *Statistical Science*, Vol. 4, No. 3 (Aug., 1989), pp. 282-289
- [4]. Hill, Theodore P. “Knowing When to Stop: How to Gamble If You Must—the Mathematics of Optimal Stopping.” *American Scientist*, vol. 97, no. 2, 2009, pp. 126–133