International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Automated CV Classification using Clustering Technique

Prof. Sagar More[1], Bhamare Priyanka[2], Mali Puja[3], Kachave Kalyani[4]

[1]Prof. Sagar More, Dept. of Computer Engineering, GCoE, Dhule(MS) India
[2,3,4]BE Computer

Abstract - *Now a day's many companies have come on ground to crab the talent. Also, digitalization have opened many options for the industry as well as for job seekers. Traditional approach for applying and recruiting has few restrictions. Which leads the job provider and seeker both away from the expectations. To crab the quality from quantity is the challenge for the job provider while to get to the expected place is the challenge for the seeker. This proposed approach with provide benefits to both the sides. It simplifies this the process by calculating a score for resume by using clustering techniques which makes it easy for HR department to get to the eligible candidate*

***Key Words*: clustering, Mining, CV classification**

1. INTRODUCTION

Today's competitions is on high peak. It has become challenging task for the job provider as well as seeker to get the one as they expect. The main challenge is at initial level. For job provider the challenge is the extraction of the required data. In case of job seeker the challenge to get to the place where their talent could be utilized or to the place they deserve. Traditional approach of applying to the job is to look for the job and then apply by sending resume or CV to the place. Manually it is difficult to extract. One method is of filtering but to identify the potential resumes by examining each resume, since these filtered resumes are similar to each other. In this proposed method we have focused in improving the performance of the selection process of resume selection. A method is proposed to identify the special feature and skills accordingly. This solution will provide HR team option to customize each and every job before uploading as per requirement and skill set required.

i. Datamining :

a. Clustering:

By examining one or more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so you can see where the similarities and ranges agree.

Clustering works both ways. There is a certain point which is cluster and identification is used to check either we are correct.

Clustering can also be implemented from opposite perspective.

2. PROBLEM STATEMENT

To develop a system which will allow HR department to publish job opening with customized weighted scheme according to job requirement and provide a portal for candidate to upload their CV/Resume and finally provide the HR team with a classified view of all the CV available for particular job.

2. LITERATURE SURVEY

NingZhong and Yuefeng Li and Sheng-Tang Wu[3] has proposed a technique which used two processes for refining the patterns from the document. Effectiveness was focused by using pattern evolving and pattern deploying methods.

In [4] author has proposed an automated resume extraction and candidate selection system. Which has benefitted the organization hiring structure. Later classification algorithm is used on the data extracted from the above method for identifying the categories of profile. In[6] Ankita Satish Vaidya and et.l has proposed a analyzer system which uses text mining, that extracts the details of the job seeker from its resume. Here, author has used text mining as it is used for extracting the text from unstructured document. Also, the extracted text is converted into data for analysis.

3. PROPOSED SYSTEM

The proposed solution has used data mining techniques. Data mining data mining is about processing data and identifying patterns and trends in that information so that you can decide or judge[1]. Clustering technique of data mining is used for classification and calculation. As Clustering is useful to identify different information because it correlates with other examples so you can see where the similarities and ranges agree[1].

To provide a classified view of CV to HR department

2) To eliminate the manual process of CV classification and shortlisting

3) To reduce time and efforts of HR department in CV selection process

4) Provide HR team option to customize each and every job before uploading as per requirement and skill set required.

The proposed solution to the problem statement is explained below with the algorithm and the system architecture.

4.1 Algorithm

K-Means Clustering:

1) K means is unsupervised machine learning algorithm.

2) K means divides the given dataset into k clusters. Where k means number of clusters.

3) This algorithm will cluster candidate resume score in k clusters.

Steps included in clustering:

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,……..,vc} be the set of centers.

1) Randomly select 'c' cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4) Recalculate the new cluster center by taking means of data points in that cluster

5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step

5. SYSTEM ARCHITECTURE

The proposed system is a client server architecture based system which will use java servlets for client server communication. The front end of system will be developed using the java swing components and backend of system will be developed in java. The system database will consist of relational MYSQL database. As shown in the above figure the system will consist of 3 major modules as follows:

1) HR/Admin Module

2) Candidate Module

3) Web Server

HR/Admin Module:

This module will be run at the HR or admin end. This will be desktop software whose front end will be developed using java swing components. This module will be well protected with username and password and the password will be secured with SHA-1 algorithm. Admin can customize the job as per requirement and assign weight and upload the job. Admin will get a classified view of all the CV submitted by candidates for which k-means clustering algorithm will be used.

Candidate Module:

This module will be run at the candidate end. This will be desktop software whose front end will be developed using java swing components. The candidate will first need to register into the system and then will be presented with an interface where he will answer and select few options which will be part of the candidate CV. The candidate on submission of resume will get to know the score of CV.

Web Server:

The webserver will be developed using java servlets and will be run by using Apache Tomcat application server. Web server will be responsible for all the request and response handling mechanism of system. It will manage the database and provide appropriate data to admin of the system.

6. BENEFITS AND APPLICATIONS

6.1 Easily integrates with your HR software, job portals, and other career sites

It also gives the applicant a more attractive application process, without the need for tedious, manual data entry. Given this fact, terms like "candidate experience," "one-click application," and "employer branding" become more than just empty catch phrases. Integrating CVlizer into the application process, be it on a job portal, a career website, or within your own back office, liberates all involved parties from the daily grind of recruiting, making it more attractive while also increasing your rate of success.

7. RESULT:

7 .1 Home Page for Candidate

7.2 Admin Login

7.3 Candidate Login

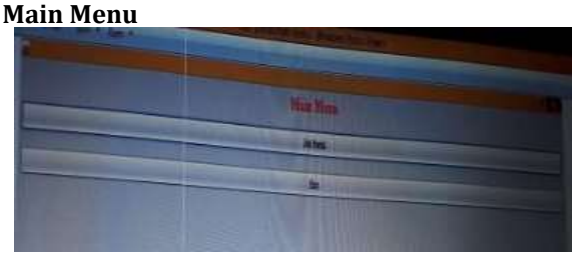
7.4 Main Form

7.5 Form for details (Job)

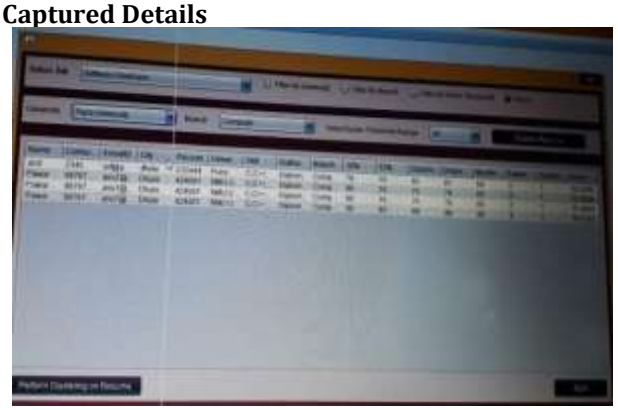
7.6 Form for details (Applicant)

7.7 Candidate Registration form

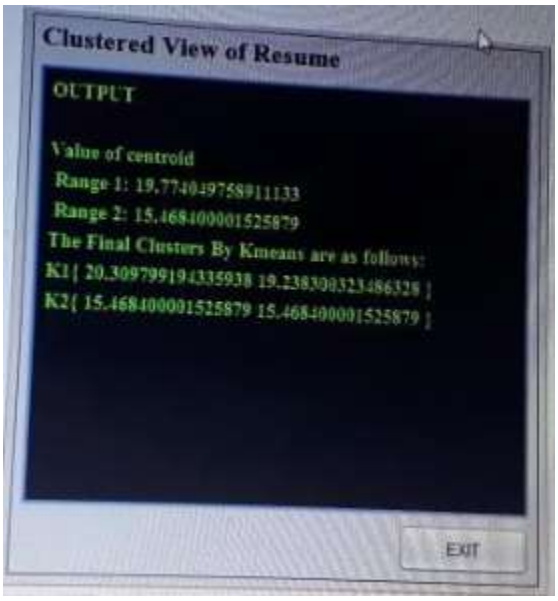
7.8 Main Menu



7.9 Captured Details



7.10 Clustered View



8. CONCLUSION

The problems that the HR team face while shortlisting candidates CV are very time consuming and involves lots of human efforts. This manual approach sometimes even give rise to human errors. So using our proposed system both efforts and time can be reduce in CV shortlisting process along with increase in effiency and accuracy. Considering all the above features our system will definitely prove beneficial to organization and save some precious time and efforts of their HR department

REFERENCES

[1] https://www.ibm.com/developerworks/library/ba-data-mining-techniques/

[2] https://www.hindawi.com/journals/mpe/2018/5761287/

[3] Ning Zhong and Yuefeng Li and Sheng-Tang Wu (2012),"Effective Pattern Discovery For Text Mining", IEEE Transactions On Knowledge And Data Engineering, Volume 24, Issue 1.

[4] Sneha Kumari and Punam Giri (2014)," Automated Resume Extraction and Candidate Selection System", International Journal of Research in Engineering and Technology [IJRET] Volume 3, Issue 1, Page No 206-208.

[5] https://www.edureka.co/blog/clustering-on-bank-data-using-r/

[6] Ankita Satish Vaidya and Pooja Vasant Sawant (2015), "Resume Analyzer an Automated Solution to Recruitment Process", International Journal of Engineering and Technical Research (IJETR), Volume-3, Issue-8.