BIG DATA BASED ENHANCED SKILL-BASED RECRUITMENT DECISION SUPPORT SYSTEM

Ms. Preethi.R[1], Ms. Indhumathi.V[2], Ms. Raajashri.J[3], Mrs. B.Monica Jenefer[4]

Abstract — The objective of this paper is to enhance the skill-based recruitment and help the recruiters in choosing the most deserving candidate for a particular job. It helps people grab job opportunities based on their talents and specialization. This also delivers a helping hand for the recruiting companies to broaden their ways in selecting the most worthy candidate to be employed in their respective firms. The dataset contains all the personal and academic details of the graduates. In the Hadoop platform the dataset could be considerably filtered and reduced based on the companies' requirements .Various resumes that have been sent individually to specific organizations that have openings can be manipulated based on the recruiters requirements. Recruiters can have an insight beyond the curriculum vitae by looking at their activities in social world. We can have parameters that will evaluate candidates based on not just their academic performance but also on his/her skills which is assessed by on-hand work experiences such as internships, hackathons, contests, workshops etc. The organization can now call the candidates from the reduced dataset for the interview process. The dataset would go through a "map and reduce" process for collecting the list of graduates with a good number of achievements and experience in a particular domain and sort them according to requirements. This way it serves as an aid to both the recruiters to choose the best, and to the fresher's to grab the best job opportunity according to their area of specialization or interest.

Keywords-Big data, MapReduce

1. INTRODUCTION

Immense and intricate data is known as big data. Regular data manipulation procedures cannot handle these vast data. The major challenge in big data is data investigation and recovery of   useful information out of it. Volume, Variety, Velocity, Veracity and Value are the main dimensions of big data.

Volume: mass of data. Velocity: measure of data builds up. Variety: different sources of data..

Many challenges are faced by small scale companies and startups in hiring candidates based on their talents. Data about candidate's achievements and talents can be gathered easily the problem lies in interpreting them. Hadoop is a popular tool used to gather the required data.

Big data can be used to address this problem. It can be helpful in the whole recruitment process. There are few companies that collect data about the people's behavior in the social media. Data can be collected from these companies and a personal digital footprint can be interpreted from these data points.

This type of analysis of talents and skills based recruiting is very useful for recruiters. The burgeon amount of available details about unemployed professionals was matched with that of the opening in the organization and the people are recruited accordingly. This approach has made unemployed candidates build a more arduous curriculum vitae and instead of a template one.

Hiring candidates based on their skills instead of hiring them based on their mark/grade might prove useful for improving the company's level in all ways. Analytics of data by this way helps the HR in knowing the reasons of employee's success and retention.

2. RELATED WORKS

The Recruitment System can be amalgamated into Organization's website and thus can publicize for niche of positions and professionals who meet the criteria can apply online.

Features:

- Proclaim a list of positions available within the company.
- Furnish information on the position publicized by the company.
- Manage opportunities available in the workplace
- Administer the opening process from application to acceptance or rejection
- Maintain a database of all candidate application

Depending on a user's access level rights, they will be able to:

- Insert an upcoming opening in the office

[1] UG Student, Department of Computer Science & Engineering, Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India
[2] UG Student, Department of Computer Science & Engineering, Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India
[3] UG Student, Department of Computer Science & Engineering, Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India
[4] Associate Professor, Department of Computer Science & Engineering, Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India

- Set up assessments for potential candidates

The Human Resource executive will be able to administer the applications they have received for various job vacancies. The user will have a list that sets out:
- the name and surname of the applicant
- the vacancy an applicant has applied for
- the start date of the vacancy
- the status of the applicant in the application process assessment
- Shortlisted, interviews, and the expiry date of application.

Traditional recruitment process includes comparing mark/grade of the candidates who had applied for the job.This may lead to a situation where a candidate may be highly talented and has good skills but no good mark/grade.Another candidate with good mark/grade and average skills.The later candidate would get selected.

CANDIDATE'S DETAILS	APTITUDE SCORE	SPEAKING SKILLS	CREATVITY SCOPE	PROGRAMMING SKILLSET	CGPA	EXPERIENCE (in years)
Candidate 1	89%	90%	89%	75%	8.5	Fresher
Candidate 2	89%	95%	91%	90%	8.2	3
Candidate 3	97%	90%	95%	85%	9.3	Fresher
Candidate 4	95%	93%	89%	92%	8.1	6

3. PROPOSED SYSTEM

All the requirements of a candidate depending on the job opening in a specific company can be uploaded in Hadoop distributed file system by that company. Instead of manually searching for a particular skill set in people according to the requirements, all the data about the unemployed candidate can be analyzed and recruited in one platform. The recruiting company has to have a tie up with useful educational sites like 'geeks for geeks, hacker rank', etc. The candidate's behavior in social world tells about the skillset of that person. The types of codes uploaded in GitHub provide an insight about the areas where the person is strong in. The type of questions answered in Quora and Stackoverflow, etc. and the uniqueness of each answer provides an idea about soft skills as well as technical knowledge related to that person. Based on the rank the person had scored in coding websites like hacker rank and geeks for geeks, etc. the talents can be analyzed. The ideas and codes the person had uploaded in these websites can be checked for plagiarism using plagiarism checker. All the data related to the skills of the person after checking for plagiarism can be stored in Hadoop distributed file system by this application without manual intervention. All these data stored in Hadoop distributed file system is analyzed using Map-Reduce. This idea reduces the time of Human Resource executive in the recruiting process. The recruiter might not read the resume of the candidate at all, the recruiter might just enter the current requirements and the candidate's details in the Hadoop distributed file system, this application would compare the recruiter's requirement's details and the candidate's details based on the skills and retrieve the appropriate results accordingly.

3.1.Hadoop Distributed File System

Hadoop File System was developed using distributed file system design. It is run on commodity hardware. Unlike other distributed systems, HDFS is highly fault tolerant and designed using low-cost hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing. The important features of the HDFS are its suitability for the distributed storage and processing. Hadoop provides a command interface to interact with HDFS. The built-in servers of name node and data node help users to easily check the status of cluster. HDFS also streams access to file system data. HDFS provides high security by using file permissions and authentication.
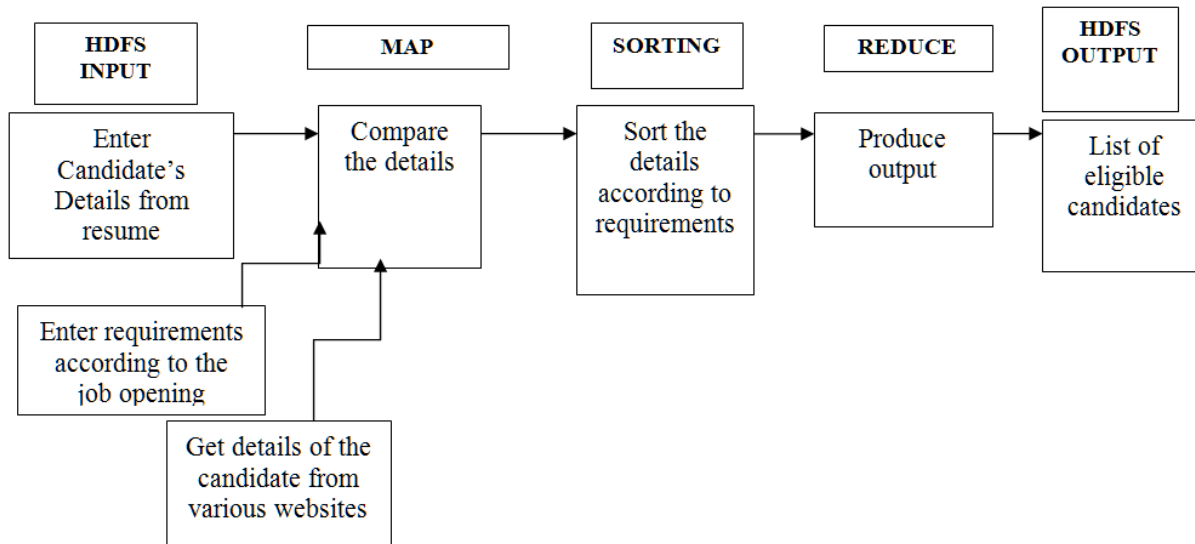
3.2. Map Reduce

The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data from the technical websites and converts it into another set of data that describes the skills of an individual. Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. Compares the requirement's dataset and candidate's details dataset. Produces the matching output required by the company. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

3.3.Weighted Product Model

The weighted product model (WPM) is a popular multi-criteria decision analysis (MCDA) / multi-criteria decision making (MCDM) method. Given is a finite set of decision alternatives described in terms of a number of decision criteria. Each decision alternative is compared with the others by multiplying a number of ratios, one for each decision criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion. If the ratio $P(A/B)$ is greater than or

equal to the value 1, then it indicates that alternative A is more desirable than alternative B. The best alternative is the one that is better than or at least equal to all other alternatives. This model is used in this idea to choose the best result.

3.4. Architecture

```
┌──────────┐      ┌─────────┐     ┌──────────┐    ┌──────────┐    ┌──────────┐
│ HDFS     │      │  MAP    │     │ SORTING  │    │ REDUCE   │    │ HDFS     │
│ INPUT    │      │         │     │          │    │          │    │ OUTPUT   │
└──────────┘      └─────────┘     └──────────┘    └──────────┘    └──────────┘
```

Enter Candidate's Details from resume → Compare the details → Sort the details according to requirements → Produce output → List of eligible candidates

Enter requirements according to the job opening

Get details of the candidate from various websites

3.5.Algorithm
Algorithm Best Candidate

Input: CSV files containing datasets of the candidate.
Output: The most deserving employee for the particular job opening.
for each candidate in dataset,do
if(Candidate's details==invalid),then /*plagiarism of ideas mentioned by the candidate in the social world is checked*/
return null
else if(details==valid),then /*plagiarism of ideas mentioned by the candidate in the social world is   checked*/
if(requirements==skills),then //recruiter's requirements and skills are compared
produce one output record
else return null
The working of this model consists of varied steps that implement the "Map & Reduce" algorithm. Steps concerned in building the model are as follows:

1) Collect the candidate details from the resume and store it in hdfs.
2) Enter requirements according to the job opening in hdfs.
3) Get details of the candidate from various websites.
4) Compare all the datasets.
5) Sort the candidates according to matching of requirements.
6) The list of eligible candidates are given as the output based on how much it matches with   company's requirements

3.6.Implementation
The implementation process involves several steps. Initially the requirement details depending upon the job opening are entered in Hadoop distributed file system. This is generally done by the Human Resource Executive. The performance of the person who has applied for job is retrieved and analyzed from useful websites and they are loaded in another HDFS. Now the requirements of the company and person's skillset is compared and analyzed using Map and Reduce technique. In this the compared data is split and they are sorted according to the needs of the recruiter. Here all the duplicate data is eradicated and the results are displayed.

4. RESULTS AND DISCUSSION
The recruitment of a deserving candidate has been done using the big data analytical tools. Big data analytics, Weighted product model are a couple of concepts which has been used in this method. This reduces the human effort in a big way as it stresses on the machine by itself importing, analyzing and producing the results about the candidates on its own without human intervention. The details of the candidate is stored in the Hdfs and analysed.The output is compared with requirements and the list of eligible candidates for the job opening is returned.

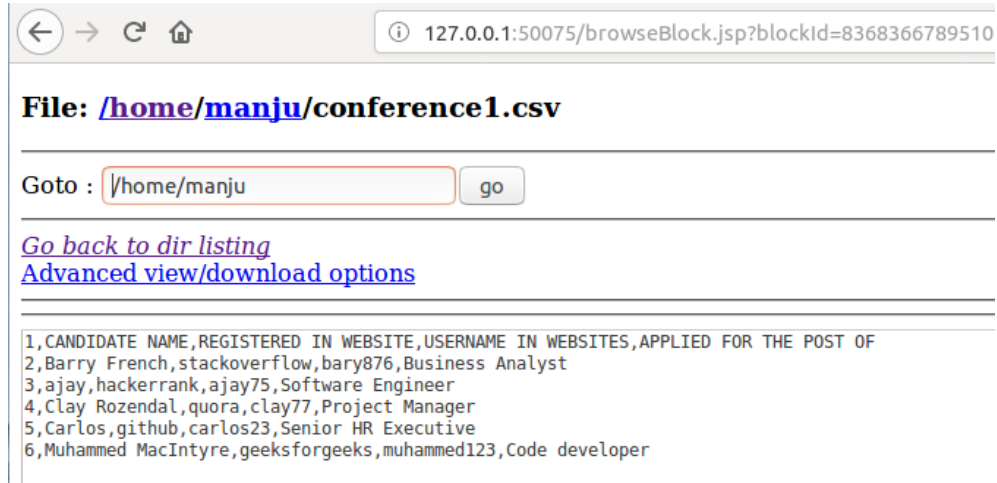TABLE I (Details of the candidate entered by recruiter)

```
←  →  C  ⌂                    ⓘ  127.0.0.1:50075/browseBlock.jsp?blockId=8368366789510

File: /home/manju/conference1.csv

Goto : /home/manju          go

Go back to dir listing
Advanced view/download options

1,CANDIDATE NAME,REGISTERED IN WEBSITE,USERNAME IN WEBSITES,APPLIED FOR THE POST OF
2,Barry French,stackoverflow,bary876,Business Analyst
3,ajay,hackerrank,ajay75,Software Engineer
4,Clay Rozendal,quora,clay77,Project Manager
5,Carlos,github,carlos23,Senior HR Executive
6,Muhammed MacIntyre,geeksforgeeks,muhammed123,Code developer
```

TABLE II (Details of the candidate entered by proposed system)

```
←  →  C  ⌂                    ⓘ  127.0.0.1:50075/browseBlock.jsp?blockId=83683667

File: /home/manju/conference1.csv

Goto : /home/manju          go

Go back to dir listing
Advanced view/download options

1,CANDIDATE NAME,REGISTERED IN WEBSITE,USERNAME IN WEBSITES,APPLIED FOR THE POST OF
2,Barry French,stackoverflow,bary876,Business Analyst
3,ajay,hackerrank,ajay75,Software Engineer
4,Clay Rozendal,quora,clay77,Project Manager
5,Carlos,github,carlos23,Senior HR Executive
6,Muhammed MacIntyre,geeksforgeeks,muhammed123,Code developer
```

TABLE III (Weighted product model inference)

```
←  →  C  ⌂                    ⓘ  127.0.0.1:50075/b

File: /home/manju/conference3.csv

Goto : /home/manju          go

Go back to dir listing
Advanced view/download options

1,,ANALYTICAL SKILLS,LOGICAL THINKING,EFFICIENCY
2,METRIC,0.26,0.5,0.75
3,Barry French,24,78,96
4,ajay,43,26,53
5,Clay Rozendal,32,43,84
6,Carlos,19,77,12
,Muhammed MacIntyre,12,52,39
```
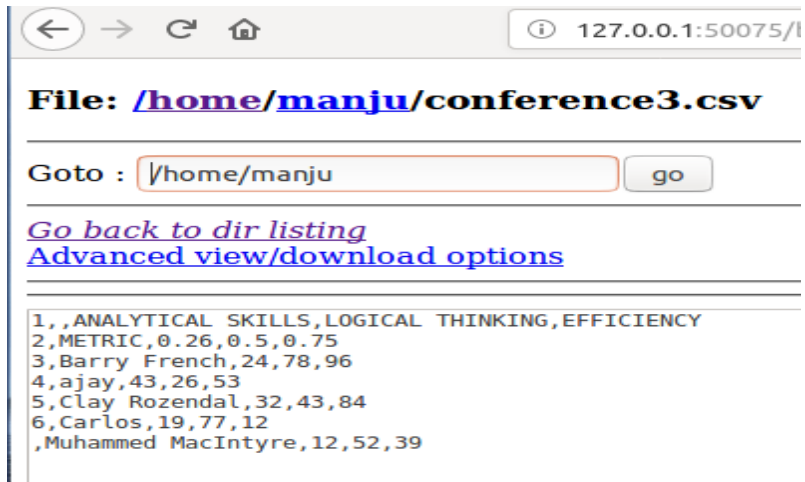
P(Candidate1/Candidate2)=$(24/43)^{0.26}$ *$(78/26)^{0.5}$*$(96/53)^{0.75}$ =0.703<1 which means Candidate 2 is the best.

CANDIDATE ELIGIBILITY ANALYSIS

x axis = candidates
y axis = candidate's score
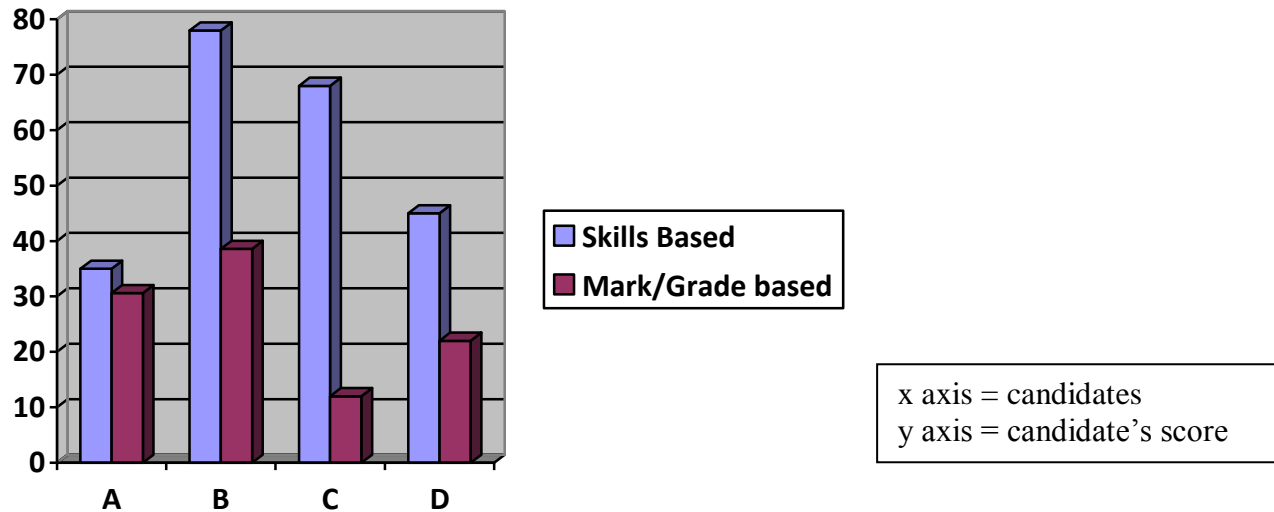
A : Candidate 1
B : Candidate 2
C : Candidate 3
D : Candidate 4

5. CONCLUSION

Enhanced Skill based recruitment support system results in time and effort save for organizations. This system helps the job seekers with high talents to be placed in an appropriate company. Organizations select candidate based on only their marks in their graduate or post graduate studies. It is a major issue to select a perfect candidate for a particular opening of a job because more number of candidates apply to single vacancy. To overcome this, we analyze the enhanced skills of the candidate in addition to the marks/ grades obtained by the candidate instead of giving only importance to mark / grade.The weighted product model allows to choose the best candidate by comparing each skills of all the candidates who have applied.The recruiter enters the requirements according to the job opening in hdfs.The skills details are entered in Hdfs by this proposed system.Both the details are compared and weighted product model is calculated and a best candidate is selected. The results of this study indicate that big data analytics can be used to build an efficient and easier recruitment process. In the future, the next step would ideally be to make a complete and universal model that the organizations can use to help them reduce their human resource time.

6. REFERENCES

[1]   [1] Amir Ikram, Qin Su, Muhammad Fiaz,"Big Data in enterprise management: Transformation of   traditional recruitment strategy".
[2]   [2] Shaokang Dong, Zijian Lei.Pan Zhou,"Job and candidate recommendation with big data support: A contextual online learning approach".
[3]   [3] Pooja Tripathi,Ruchi Agarwal,Tanushi Vashishtha,"Review of job recommender system using big data analytics".
[4]   [4] Dyna Marisa Khairina ; Muhammad Reski Asrian ; Heliza Rahmania Hatta "Decision support system for new employee recruitment using weighted product method", 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2016
[5]   [5] https://www.tutorialspoint.com/hadoop/hadoop_hdfs_overview.htm
[6]   [6] https://arxiv.org/pdf/1801.00377.pdf