



RECRUITMENT PROCESS AND STAFFING SECTOR: A BIG DATA PARADIGM

Computer Science

Vidyalakshmi. V Computer Science Department, Kasturba Gandhi Degree and PG College, Osmania University, Hyderabad

ABSTRACT

Finding the right talent is a key advantage in the recruitment and staffing sector. Internet is considered as wealth of information and data flow is abundant from all sources and directions. In the current situation recruiters find identifying the suitable candidates a challenging task. Presently data flow is enormous and is in either structured or unstructured format. It is not the amount of data that is important but the challenge is to analyse, visualize, query and update the data. The traditional way of keying suitable words or using Boolean search might no longer be efficient and inclusive enough for the recruiters to uncover the right candidate for the job. Big data is helping all growing organizations to find their perfect engineers, developers and executives with advanced sourcing and screening methodologies. This research paper provides an understanding about how recruitment process and staffing in an organization occurs using Big Data approach.

KEYWORDS

Big data, data analytics, algorithms, recruitment process

Introduction:

Recruitment Process & Staffing Sector

Recruitment refers to the process involved in selecting a suitable candidate for a job opportunity. The process can be for a new position in the organization or for an existing position that is to be filled. It is an organization specific process to select a suitable candidate for a position. The recruitment process is generally taken care by the human resource wing of an organization which is also sometimes driven by third party organizations.

Many organizations use traditional methods in hiring the candidates. This traditional method holds good only when the amount of data flow is regulated. But in the current scenario of big data the flow of data is enormous.

However, **data analytics** and Big Data are beginning to shape the evolution of the recruitment process. With availability of vast amounts of data aggregated from multiple sources—especially social media channels where prospective candidates usually leave their digital *'thought prints'*—and the ability to transform all that information into intelligence using powerful algorithms, recruiters now have the opportunity to rely more on facts than on intuition before they issue job offers.

This new realm of “people analytics” or “talent analytics” – which refers to the use of data analytics for making people-decisions—is now contributing to the hiring process including recruitment marketing, filtering prospective candidates, identifying outliers, planning interview questions, and determining who to retain and promote etc.[1]

Need for Study:

Big data is future of recruiting. Recruiting does not only depend on traditional methods or not what one feels about the candidate. It will become unfair to take recruitment decisions based on one's intuition. When a recruitment process is undergone, several things have to be considered, the internal variables, external variables and the right tools to analyse it. It is not only the tool that will help in analysing; it also depends on the right people who can use the tool. This in turn will help the recruiters in finding the healthy candidates suitable for a particular job.

An innovative recruitment team always looks for a best tool for a recruitment process. The team look for recruitment methods starting from advertising or campaigning, collecting the resume, shortlisting the candidate, posting the job in a particular post and on boarding as the final stage in the recruitment. This complete process accomplished with a tool reduces much time and also discrepancies.

Therefore it is necessary to use proper tool to model the process that will lead to correct results for decision making.

Research Objectives:

The objectives of this paper are as follows:

1. To study positive impact of Big Data in recruitment process.

2. To have an insight on the technical methodologies used sourcing and screening phases of the recruitment process of any organization.
3. To have an understanding on commonly used big data tool “Rapid Miner”.

Big Data:

Data has gone from stock to a flow, that is data which was stationary and static has moved to fluid and dynamic.

Data is everywhere. In fact, the amount of digital data that exists is growing at a rapid rate. Currently more than 2.7 zettabytes of data exist in today's digital universe, and that is projected to grow to 180 zettabytes in 2025.

Big data is high-volume and high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making, and process automation. Big Data analytics find insights that help organizations make better business decisions. [2]

The new benefits that big data analytics brings to the table, however, are speed and efficiency. A few years ago, a business organization would have gathered information, run analytics and unearthed information that could be used for future decisions but today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before. [3]

Big Data as business engine of opportunity

Today big data can be considered as a data engine. To boost performance, it's a matter of assembling the right components in a seamless, stable and sustainable way. Those components of big data include:

Data Sources: operational and functional systems, machine logs and sensors, Web and social and many other sources

Data Platforms, Warehouses and Discovery Platforms: that enable the capture and management of data, and then – critically – its conversion into customer insights and, ultimately, action

Big Data Analytics Tools and Apps: the “front end” used by executives, analysts, managers and others to access customer insights, models scenarios and otherwise do their jobs and manage the business. [4]

The data has to be analysed to improve organization's performance by seeking professional help who can make a good sense of all the data available.

How Big Data is used in Recruitment

On an average a company spends a lot of amount on training and developing new employees. One has to improve employee retention

and cut down on training costs. This can be achieved by implementing the concept of Big Data.

Big Data can help hiring the right employees for a particular position in an organization by scrutinizing thousands of data on employee files that is generated in hourly basis. The data collected will help the recruitment team to pay attention to the criteria that delivers result in the hiring process and access its candidates and helps in ranking them in an order suitable for the organization. The ranking helps to establish and predict patterns in detail about a particular candidate's strength and weaknesses which happens to be stumbling block, so which helps the recruiters to dive deep in to interview process before they actually recruit a person and also focus on those candidates who are recruited by providing them with a proper training.

Big data enables HR to forecast future gaps in sourcing that may affect the ability of business to generate revenue. It also enhances HR's ability to be both creative and pro-active in its handling of the complex challenges of talent acquisition and workforce management. [5]

Data accumulated is not for the purpose of hiring more number of candidates (quantity) but for hiring a right candidate (quality). Big data saves time and money.

The recruitment process starts with identifying need for a talent and ends at on-boarding of the selected candidate into the organization and closing the demand.

The recruitment has 5 different phases as outlined below:

1. Sourcing
2. Screening
3. Scheduling
4. Evaluating
5. Closing

While all the above phases are very crucial for success of the recruitment process, the sourcing and screening phase forms the key phase of the process.

Sourcing:

Sourcing is gathering of information about various candidates through different sources like social media, job portal etc. while screening refers to reading through the different profiles received for a particular opening / position.

Social professional networks and the internet give human resource leaders access to an unprecedented number of candidates. The candidates are broadly classified into active job seekers and passive job seekers. While passive job seekers form 65-75% of the market, the active job seekers form only 15-25% who is searching for jobs. So the HR should not only focus on the active seekers but also on the passive seekers.

With the vast sources available for sourcing of resumes / profiles, big data helps here in speeding up the process and to source the right set of profiles which can be further screened to have optimal selection of candidates to be evaluated.

There are various algorithms proven for successful results which the organization can leverage to have reduced turnaround time for this phase of the recruitment process.

Data mining also known popularly as talent mining is the technique widely used by firms to aggregate data from variety of different channels. This technique can be very effectively used to screen the resumes that were collected through various sources. The predictive analysis technique along with data mining / talent mining technique that helps in assessing the duration for which the prospective candidate would remain with the organization. This is very crucial information on assessing the longevity and stability of a candidate in the organization.

Resume Screening:

This phase of the recruitment is very ardent, where the recruiters have to screen many resumes / profiles before even shortlisting the candidates to be called for assessment process. The technology would help the recruitment team manifold times by reducing the time and cost spent on screening many resumes (generally scaling to a size of few

hundreds). The technology also helps by eliminating human bias in resume screening process.

The big data algorithms can slice and dice the given set of resumes and can very efficiently and quickly segregate the resumes to three

categories namely:

1. Least likely to be recruited
2. Most likely to be recruited
3. In-between applicants which would require attention of human recruiter.

Studies have proven that technology based big algorithms have given 500% ROI to the organizations. **Below two are very commonly used big data algorithms used in screening:**

1.Cosine Similarity

The business use case for cosine similarity involves comparing different candidate's profiles. The algorithmic question is whether two candidate profiles are similar or not. Cosine similarity is perhaps the simplest way to determine this.

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and it is less than 1 for any other angle in the interval $[0, 2\pi)$. It is thus a judgment of orientation and not magnitude: two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1, independent of their magnitude. Cosine similarity is particularly used in positive space, where the outcome is neatly bounded in $[0, 1]$. The name derives from the term "direction cosine": in this case, note that unit vectors are maximally "similar" if they're parallel and maximally "dissimilar" if they're orthogonal (perpendicular). This is analogous to the cosine, which is unity (maximum value) when the segments subtend a zero angle and zero (uncorrelated) when the segments are perpendicular.

In information retrieval and text mining, each term is notionally assigned a different dimension and a document is characterised by a vector where the value of each dimension corresponds to the number of times that term appears in the document. Cosine similarity then gives a useful measure of how similar two documents are likely to be in terms of their subject matter. [7]

Through this it becomes easier for the recruiter to scrutinize the resumes and remove the duplicate ones from the corpus.

2.Term Frequency -Inverse Document Frequency

In information retrieval, tf-idf or TFIDF, abbreviation for Term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.[1] It is often used as a weighting factor in searches of information retrieval, text mining, and user modelling. The tf-idf value increases proportionally to the number of times a word appears in the document and is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general. Nowadays, tf-idf is one of the most popular term-weighting schemes; 83% of text-based recommender systems in the domain of digital libraries use tf-idf.[8].

This algorithm does a more efficient screening based on word-based approach. This algorithm has higher processing rate on the count of profiles to be screened.

On completion of the screening phase, we have the shortlisted candidates who go-through the assessment process through written tests / technical discussions. Based on the outcome of the assessment phase, the candidate is either on-boarded or rejected.

With the evolving technologies, the organizations are driving towards automation and lean techniques. There is enormous human effort spent on recruitment which is a continuous process. With the concept of Big data there are certain tools developed to help the HR team to conduct the recruitment process through tools with reduced human effort. One such commonly used tool is Rapid Miner.

Rapid Miner

Rapid Miner is a software platform for data science teams that unites data preparation, machine learning, and predictive model deployment.

It is a powerful data mining tool for rapidly building predictive analytic workflows. This all-in-one tool features hundreds of data preparation and machine learning algorithms to support all your data mining projects.

The process outlines six phases:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modelling
5. Evaluation
6. Deployment

The first two phases, business understanding and data understanding, are both preliminary activities. It is important to first define what one would like to know and what questions one would like to answer and then make sure that the data is centralized, reliable, accurate, and complete.

Once it is defined, data is gathered, it's time to prepare the data – this is where one can start to use datamining tools. Datamining software can assist in data preparation, modelling, evaluation, and deployment. Data preparation includes activities like joining or reducing data sets, handling missing data, etc.

The modelling phase in data mining is where mathematical algorithm is used to find pattern(s) that may be present in the data. This pattern is a model that can be applied to new data. Data mining algorithms, at a high level, fall into two categories – supervised learning algorithms and unsupervised learning algorithms. Supervised learning algorithms require a known output, sometimes called a label or target. Supervised learning algorithms include Naïve Bayes, Decision Tree, Neural Networks, SVMs, Logistic Regression, etc. Unsupervised learning algorithms do not require a predefined set of outputs but rather look for patterns or trends without any label or target. These algorithms include k-Means Clustering, Anomaly Detection, and Association Mining.

Data evaluation is the phase that will tell how good or bad the model is. Cross-validation and testing for false positives are examples of evaluation techniques available in data mining tools. The deployment phase is the point at which results are used. [6]

Conclusion:

Tools like Rapid Miner can be configured with algorithms to perform sourcing and screening. Right algorithm must be chosen for the screening part. For the sourcing part, the data set of existing resumes in the organizations' data base maybe deployed for testing. The idea is to run the algorithms for different kinds of positions to be staffed and the desired output is all the relevant resumes and none of the irrelevant resumes must be returned in the search result. By fine tuning our algorithm to achieve both results - all relevant resumes and no irrelevant resumes for at least 30 sourcing instances, the right algorithm is selected.

The selected algorithm is configured in the tools run for a few months and the variance between the manual version and this algorithmic version is compared to update the algorithm. The most difficult challenge to overcome is to parse resumes that are written in different unstructured formats into a common framework so the demand can be compared against this resume and level of relevance be rated. It is at this stage of extraction do most organizations face challenge and getting algorithms right. This maybe explored in further theoretical research on the area.

References:

1. - paragraph 4,5 <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
2. - Big Data paragraph 3 https://www.sas.com/en_us/insights/analytics/big-data-analytics.html#
3. - History and Evolution –paragraph 3 <http://www.icollinsglobaltechnologies.com/BigDataBusinessEngine.html>
4. - paragraph 1,2,3 <http://www.unleashgroup.io/news/5-ways-big-data-transforms-your-hiring-process/>
5. - paragraph Enhances predictions & Assists HR's creativity <https://rapidminer.com/data-mining-tools-try-rapidminer/>
6. - paragraph Data mining process and tools https://en.wikipedia.org/wiki/Cosine_similarity
7. - paragraph 1,2 <https://en.wikipedia.org/wiki/TF%E2%80%93idf>
8. - paragraph 1