A Study on Analytics of Human Resource Management in Big Data

S. Chitra¹, Dr. P. Srivaramangai²
¹Research Scholar, Department of Computer Science, Marudupandiyan College(Affiliated to Bharathidasan University), Thanjavur, Tamilnadu, India
²Associate Professor, Department of Computer Science, Marudupandiyan College(Affiliated to Bharathidasan University), Thanjavur, Tamilnadu, India

ABSTRACT

In the period of Big Data, enterprise management is undergoing marvelous changes. Human resource department, as a vital part of the company, has also been affected by Big Data. The field of human resource development is becoming increasingly common by using data analytics process. This increase in popularity is accompanied by skepticism regarding the ability of human resource professionals to effectively utilize data analytics to reap organizational benefits. Human resource predictive analytics is an application field of data analytics for developing HRM purposes. The purpose of HRM is measuring engagement and employee performance, revising workforce collaboration patterns, researching worker churn and turnover and modeling employee lifetime value. In this paper conferred about the main role of Big Data in human resource management by using predictive analytics.

Keywords: Big Data, Human Resource Management, Analytics, Predictive Analytics

I. INTRODUCTION

We are living in a period when an explosive huge amount of data is being generated everyday. Data from social networking websites, mobile devices, sensors, scientific data & enterprises all are contributing to this enormous explosion in data. This sudden bombardment can be grasped by the reality that we have produced a enormous volume of data in the last two years. Big Data as these large chunks of data is commonly called- has become one of the newest research trends nowadays. In the Research, analysis shows that tapping the potential of this data can assist scientific disciplines, businesses and the public sector contributing to their economic gains in addition to development in every sphere. The necessitate is to develop well-organized systems that can make use of this potential to the maximum, keeping in mind the current challenges related with its analysis, structure, scale, timeliness and privacy. There has been a change in the architecture of data processing systems nowadays, from the centralized architecture to the distributed architecture. Enterprises face the challenge of processing these huge chunks of data, and have set up that none of the existing centralized architectures can powerfully handle this huge volume of data. Therefore, they are utilizing distributed architectures to harness this data.

II. HUMAN RESOURCE MANAGEMENT (HRM)

Resource means that, a source can be used when needed. Each and every resource contains a last point of finish. But, whereas discuss about human resources i.e., creative ideas, aptitudes, innovative techniques, human skills, etc., it may not have a last
point of finish. The resource to compare it depreciates, while human resources appreciate with the passage of time. Amongst all resources needed for an organization Human Resource is the most significant resource. From the periods of kings and politicians have ruled, however only those who have succeeded are the persons, who have diplomatically, used their available human resources. Appropriate organization human resources assure that the proper range and kind of people are available at the right time and place so that organizational needs can be met. Thus, success of several organizations depends on the human resources management. In general terms, Human Resource means that, “The total knowledge, skills, creative abilities, talents, aptitudes, values, attitudes, approaches” in a personnel. Usually, some resources are developed, whereas some are obtained through heredity. When these resources are utilized in a way such that most benefits can be taking out of it, then it is known as “Human Resource Management” [7].

Effective HRM enables employees to contribute effectively and productively to the whole company direction and the achievement of the organization’s goals and objectives. In short, HRM [8] aims at achieving organizational goals meet the employee’s expectations; knowledge development, skills and abilities of employees, improve the quality of working life and manage human resources in an ethical and socially responsible manner. The effective make use of of people is the critical feature in the successful achievement of corporate goals. To this end HR managers have to understand the requirements, aspirations of proactively, face the challenges head on resolve problems amicably in the years ahead.

III. BIG DATA IN HRM

Big data is the collection of huge amount of data which can’t be processed. Data is essential to every field including management. Big data analytics and proper management of Big Data is very much important in various fields of management such as accessing and predicting customer behavior, improved supply chain management, Workforce analysis, Human Resource management, Retention-Attrition, Talent analysis and management, Information Management, Knowledge management, HR planning, Business strategy planning, Management information System, Data Mining etc. Critical decision building systems are becoming more and more data driven.

The Human resource management [9] field includes various stages like planning, organizing, staffing, directing, controlling, recruiting, placement, performance appraisal, compensation and training etc. As a result, with the advancement of time, more and more data are available which is related to HR and obtained from various sources. Today, all the decisions are becoming more data driven than past experience and predictions. In all the fields, phases of HRM Big data analytics became the must to obtain higher, better quality and cost effective outcome.
HR department also faces problem in the minority ways. The Amount of employee information is increasing quickly both in time and with the expansion of company. The previous employee records are also required for various purposes. The skills, behavior and competencies at individual and organization level create the volume even more. All the information cum data is unless HR can store, access and analyze them for the organization speedy. Use of Big data analytics in HR gives improved speed, satisfaction and scale in an organization.

IV. HRM ANALYTICS

HR analytics [10] is a multidisciplinary approach to put together methodology for enhancing the quality of people-related decisions in order to get better individual and organizational performance. There are transposable terms used for HR analytics are talent analytics, people analytics, and workforce analytics. HR analytics plays a role in every characteristic of the HR function, including recruiting, training and development, succession planning, retention, engagement, compensation, and benefits. HR analytics involves “high-end” predictive modelling where what-if scenarios forecast the consequences of changing policies or conditions. Traditional HR analytics focuses on the present, that is, items such as earnings and cost per hire. But the majority organizations lacked a consistent and common view of the workforce and as a result, needed HR analytics to perform workforce optimization and hence it became essential for HR to develop IT and finance analytical skills and capabilities to produce better Return on Investment (ROI) [1], [2], [3]. Further development of technologies when combined with predictive analytics exponentially improved HR purposes in last decade. HRPA generates insights that cannot be achieved all the way through traditional benchmarking as HR analytics is reactive and an evidence-based decision system while HRPA is practical and fact-based decision system. [4] Three significant changes that have really created a hunger for predictive analytics in HR and these are: [5]

- Major boost in computing power and its affordability
- HR big data digitally accessible via cloud storage for processing
- Global talent war to protect and pursue talent streams.

Predictive analytics is dissimilar descriptive analysis which considers external benchmarking data and involves tables, reports, ratios, metrics, dashboards or complex math’s; it is about data-derived insights that drive improved decisions. It includes statistical techniques, machine learning methods, and data mining models that examine and extract existing and historical facts to make predictions. It enables organizations to research the past and look forward to spot trends in key factors related to voluntary termination, absences and other sources of risk. Predictive analytics involves models of organizational systems for prediction of future outcomes and understand the significances of hypothetical changes in organizations. Predictive analytics have led to prescriptive analytics somewhere HR gets decision choice to optimize performance and reshape entire HRM decision making. HR predictive analytics can/may actually help your HR function.

- **Turnover modeling.** In Turnover modeling, it may possibly be able to predict potential turnover in your business within specific functions, countries, business units and geographies by looking at issues like commute time, time since last role modify and performance over time. This means you can level your hiring efforts accordingly, reducing empty desk time and fear on hiring, which can lead to lower cost, better quality hiring.

- **Recruitment advertising measurement.** In response modeling for advertising jobs. You can use your kind of experience from earlier campaigns to avoid contacting candidates or
using channels that don’t yield a reply and focus on those channels so as to do work.

- **Targeted retention.** Targeted retention to find out which employees or groups of employees are at higher risk of churn in the future and focus or target retention activities on those who are additional need and reduce in other areas where it is not required as greatly.

- **Risk Management:** In this management to develop a profile of candidates with a higher risk for leaving prematurely or performing below standard.

- **Talent Forecasting.** Being able to predict which fresh hires of candidates based on their profiles, are likely to be high fliers after moving them into your higher potential programs by using talent forecasting process.

**A. The levels of HR analytics**

In the levels of HR analytics usually starts with the use of reporting and simple metrics, which can then be developed into more complex descriptive analysis models. By Using the simple metrics and reporting is trouble-free, but the degree of intelligence increases when moving towards more complex descriptive analytical models. These different levels of HR analytics are described more detail in Figure 2.

![Figure 2. Degrees of analysis and intelligence](image)

The first levels typically focus on improving organizations’ current processes by collecting and organizing of data. This is frequently easy, but as the amount of data and the number of possible data sources increases the data collecting and reporting also become harder. Frequently some kinds of dashboards are need to do this, which can report different information for example the degrees of performance. It gives you an idea about the current situation, but does not predict the future. When measuring and learning to realize the current way of doing things, value can be added by improving their efficiency. To some extent, more advanced reporting frequently links these processes to the organizational goals, which according to Fitz-end [6] can be measured with “QIPS”. By this, he refers to the “quality”, “innovation”, “productivity” and “service” aspects of business. With these, the typical and common objectives of organization can be measured, which are usually set by the top management. By this method, adding value can be demonstrated more concretely and the results of organization can also be compared to others i.e. bench marking. In arrange to do this, the data from other organizations needs to be collected to do the comparison and in this way the accuracy and utility of the results very much depends on the quality of the data collected. The four upper levels yet again are future oriented, which analyses why something is occurrence and is it expected to continue. These levels also contain different predictive modelling as well as optimization. Commonly, these analytical tools can be classified into different types, which are presented in Figure 3.
Descriptive analysis: In which the aim is to comprehend the past behavior and outcomes and to study and describe the relationships and patterns. It also enables distinguishing the trends from the past and makes comparisons over time or with others. The focus in this level is regularly on cost reduction and process development. Examples of descriptive analysis are dashboards and scorecards, data visualization, OLAP reports, periodic reports and basic data mining.

Predictive analysis: In which tries to predict the future by using past and current data. The focus is on the probabilities and potential force. According to Fitz-enz[6], predicting the future depends on four things, which are (1) understanding the past and current events, (2) trends and drivers behind them, (3) ability to see patterns of consistency but also change and (4) having the right tools to predict the probability of something in the future. As a result of using the data and the descriptions made from the past trends and patterns, the analytics tries to predict the future by giving importance to the patterns. This can be performed by different techniques, for example statistics, modelling and data mining. For Example, predictive analyses are decision trees, CART, genetic algorithms and neural networks.

Prescriptive analysis: In which is regarding decision options and workforce optimization. It analyses complex data, provides information about the achievable decision alternatives and their possible outcomes. One example of prescriptive analysis could be the attrition model, which describes the possibility of employee resignation. Examples of prescriptive analysis are mathematical programming and simulation.

B. Process of HR analytics
In addition to considerate the fundamentals of HR analytics, there requirements are to be an understanding of the actual process related to HR analytics. Fitz-enz[6], who has studied the HR analytics for years now, has formed a framework for HR analytics on the basis of his studies. This framework is known as the Human Capital Management framework for the twenty-first century (HCM:21). The main goal of the Human Capital Management framework is a framework for predicting management and decision-making process that collects, organizes and analyzes data for assessing and predicting the potential usage. The model is illustrated shown as Figure 4 and it consists of four main phases. They are (1) scanning, (2) planning, (3) producing and (4) predicting.

Scanning: In the strategic scan phase, where the external and internal factors and business goals are compared to internal and external labor pools. Three fundamental capitals are mainly taken into focus here are human capital, structural capital and relational capital. All of these capitals in the organization influence each other but are frequently utilized still deal within silos.

Planning: In the strategic planning phase before the scanning done in the first phase offers a foundation, which is the capability planning worksheet. The first step of this phase is the capability planning worksheet which is complete by dividing the workforce into “mission critical” (essential to survive), “unique” (market differentiators),
“important” (operational necessities) and “movable” (outsourcing or eliminating). After this is succession planning, which consists of recognizing employees with most potential, planning development programmers for employees, monitoring the results, and ensuring that plans are aligned with the strategy and key performance indicators (KPIs)?

**Producing:** In the producing phase, consists of process optimization and integrated delivery. In this phase, the different HR services are examined as processes and optimization process aims at improving the efficiency and effectiveness of these different processes for example hiring or retention. Statistical analysis can be utilized in order to identify the most cost-effective choices. Integrated delivery helps to identify the key goals and the company level objectives of different functions or departments.

**Predicting:** In predicting process, includes the strategic, operational and leading indicators. The main goal is to use analytics to identify causal aspects and different correlations in organize to form a comprehensive overview of the current situation. Special focus is directed to predict on operational and leading indicators, which predict the future events.

The last phase of the model is the analytics. According to [6] HR analytics assist to identify connections between inner or outer side of the different factors seen in Figure 4 and described in the previous part. The alignment and integration of the different phases and factors of the HCM:21 model enable more profound analytics, which can be identified with different analytics and algorithms.

HCM:21 framework illustrates the reality of how HR analytics is mainly used as a tool in different planning, managing and decision-making situations. It can be utilized when required, however still the overall focus is on the strategic business context process. Analytical tools HCM: 21 also describe the different situations that HR analytics can be utilized. Because of the different situations, the tools or metrics that are used and also different depending on the strategic business context.

**C. HR Analytics and Metrics**

Metrics are an invaluable tool for HR management to recognize the effectiveness or otherwise of existing hiring and retention programs. The shift towards data driven decision making incorporates both analytics and performance metrics. As the skills shortage squeezes global business leaders, HR must adopt a strategic approach towards talent management. Big data is the starting point.

**People Analytics**

Staff retention is a very real drawback for HR. Research carried out by Hays found that one third of UK based companies experience turnover levels of 21%. Analytics helps HR to recognize and identify the high achiever, as well as employees who are
struggling in their vital roles, evaluating why they leave and why they stay. Data provided by common recruiting, exit interviews and interviewing fresh hires will help businesses to reduce attrition rates. HR analytics is the most important tool to identify the retention rate of the employees. It tells about the company and the reason why the employees leaving and why they staying. The reason could be any, whether they leave for their underperformance, compensation, lack of skills and technical knowledge or anything else. HR analytics could also be helpful to identify skill gaps and areas where employees are struggling. HR can use different tools such as employee satisfaction surveys, team work, team assessments, exit and stay interviews to find out and identify the reason of employee attrition and plan out the strategies to keep hold of them. Nowadays HR analytics management is on the top of the stack and has created the assured advantages of its users. It has become the core part of the company’s strategies and operations and would deliver the outputs in the better way to process.

**Workforce Analytics**

The most efficient workforce planning strikes a delicate balance between what your business requirements and HR’s ability to satisfy those demands. Big data allows employers to more accurately forecast future patterns of employee movement, permitting HR to plan talent and skills needs and regulate its seasonal and long-term recruiting requirements. Workforce analytics is an advanced set of data analysis metrics and tools for wide-ranging workforce performance improvement and measurement. Whereas people analytics focuses on the personnel aspect, workforce analytics focuses on the workforce onto some extent larger scale. Workforce analytics can be measured in a number of ways – whether or not its company specific measurements, an examination of a group of jobs, measurements across a whole industry, and the data applied to analytics is dynamic change. It analyze recruitment, staffing, training and development, personnel, and compensation and benefits, over and above standard ratios that consist of time to fill, cost per hire, accession rate, retention rate, add rate, replacement rate, time to begin and offer acceptance rate. Potential candidates for employment and current employees have to take, workforce analytics can be used to help refine and improve assessments. Organizations can benefit greatly by employing powerful analytics solutions backed to improve performance by using data science and machine learning for managing and developing talent of the employee, helping them create additional objective decisions while reducing cost and time manner. Executives, business leaders and human resource management professionals need the tools to make data-based decisions on whom to hire, simulate headcount plans and get references to address challenges.

Cornerstone workforce analytics offerings eliminate the difficulty for developing organizations who desire to answer important talent management questions, but aren’t quite sure where to begin questioning. Through the variety and volume of talent data growing exponentially, and the velocity at which this is happening, Cornerstone Planning and Cornerstone Insights can help organizations acquire the power of big data from insights to next action.

**Requirement Analytics**

In the Talent analytics provides accurate data on historical hiring patterns to save the risk of a bad hire, slightly than rely on CVs and ‘gut feeling’. Complicated HR software offers online testing, self-assessment of applicants, automated, personalized engagement and an improved candidate experience which extends to offer fresh hires relevant training and development opportunities. The expanse of data available enables HR to identify the talent and skills it requirements Applicant tracking systems are mainly used today’s CRM (customer relationship management), identifying and tracking qualified
candidates through the hiring process stage. Analytics enables HR to build and nurture a talent pool while providing a responsiveness of the behaviour of all those candidates through their engagement with automated e-mails, update with job alert websites and social media. HR analytics make predictive analysis simpler and helps HR to make the best choice based on historical data. The HR easily derives the best candidate to hire from the historical data by using a great HR analytics tool can make the difference. For example if you hired 20 candidates and 10 out of them are hired from a particular background failed at the recruitment, perhaps you will not hire someone again from similar background process. HR analytics prevents you from making mistakes during hire process. Additionally, it also enables the recruiter to learn more about candidates through online resume database, applications, social media profiles, records, interview and training process etc.

Talent Analytics
Talent is defined as each and every one of the people who use their skills and abilities to meet an organization’s business objectives. Talent data is described as insight and information that an organization has available on its talent and skills that is collected from their talent management system (TMS). The TMS includes recruiting, onboarding, learning, improving performance, compensation, workplace collaboration and succession planning. This information can include demographic, mobility/activity, behavioral data associated with the talent from moreover internal or external manner.

Talent data allows organizations to improve employee engagement and productivity status to make better workforce decisions. The Cornerstone Analytics suite is built on top of the world’s largest network of shared talent data across seven talent management dimensions–recruiting, onboarding, training and development, performance management, compensation management, succession planning and workplace collaboration. Cornerstone’s portfolio of analytics offerings applies complicated data science and the most refined machine learning system for talent management to assist organizations harness the power of real-time talent data. The data analytics is four categories of unlocking process such as descriptive, diagnostic, predictive and prescriptive. Cornerstone allows organizations to create smarter, data-driven talent decisions and maximizes business objectives through their people, removing the reliance on guesswork and intuition. The talent analytics provides insight into the engagement levels of your company’s high performing employees and offered data performance, collectively with their common core skills, competencies and typical behaviour. The most accurate data will predict how long a specific skilled group to perform at their peak. HR gains more understanding of the individual motivation of employees, by tracking, evaluating and sharing performance-related data.

V. DATA MINING TECHNIQUES FOR HRM
Data mining is the KDD process concerned with patterns extraction from the data. Nowadays, there are several researchers on solving HRM problems by utilizing Data mining approach. Essentially, the majority of the Data Mining researches in HR problems domain focus on personnel selection task and few apply in other activities like planning, training, talent administration and etc. [12]. Data mining is a collection of techniques for effective automated discovery of patterns in large databases. Those have to be actionable so that they may be used in an enterprise’s decision making process stage. Data mining techniques provides to use several data mining tasks like classification, regression, time series analysis, clustering, summarization, association rules and sequence discovery, to find solutions for a specified problem by using data mining approach. In this paper we discuss below explainnation of some data mining algorithms for HRM.
• **Bayesian Classification:** In the Bayesian classification has been planned that is supported bayes rule of probability.Bayes rule may be a technique to estimation the likelihood of a property given the set of data as proof or input Bayes rule.[3] The approach is named “naïve” as a result of it assumes the independence between the various attribute values. Naïve bayes classification can be viewed as each a predictive and descriptive kind of algorithmic rule. The probabilities are descriptive algorithmic rule and used to predict the class membership for a target tuples. The main advantages of naïve mathematician approach: simple and convenient to use; unlike alternative classification approaches only one scan of the training data is necessary; easily handle data mining value by merely omitting that probability. An advantage of the naïve bayes classifier is needs a small amount of training data to estimate the parameters (means and variances of the variables) needed for classification. The parameters to be projected from sample data are central moments- mean and variance of the variables. As a result of independent variables are assumed, only the variances of the variables for each class ought to be determined and not the whole covariance matrix. Despite of their naïve design and apparently over-simplified assumptions, naïve bayes classifiers have worked quite well in several complex real world situations.

• **Support Vector Machines (SVM):** In general, an SVM uses non-linear mapping of the training set of data with high dimensionality. In other terms, the algorithm searches for an optimal solution by separating hyperplane which performs as a decision boundary between the two classes. An SVM will find out the hyperplane by employing margins (defined by vectors) and vectors (training dataset). Eventhough the training of an SVM takes more time compared to other techniques, the algorithm is assumed to have high accuracy rate owing to its high capability in building non-linear, complex decision boundary. It is also less prone to over fitting [13]. Cao employed an SVM, a straightforward logistic classifier (a combination of algorithms whose core is logistic regression and uses LogitBoost as a simple regression function [14]), naïve Bayes, and a multilayer perception neural network to predict basketball based on the results. Due to the survival of two groups, two outputs were considered as the result. The output result with greater value was preferred as the prediction. He tested and evaluated the practicality of his models by using a scoring process (a process to test a model in prediction of events not yet occurred) [15].

• **K-Nearest neighbor:** In general, these classifiers are mainly based on learning by training samples. Each sample represents a point in an n-dimensional space. All training samples are stored in an n-dimensional pattern space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is described in terms of Euclidean distance, where the Euclidean distance, where the Euclidean distance between two points, X=(x1,x2,……,xn) and Y=(y1,y2,…..,yn) is denoted by d(X, Y). Nearest neighbor classifiers assign or assume equal weight to each attribute. Nearby neighbor classifiers can also be used for prediction, that is, to return a real-valued prediction for a given unknown sample.

• **C4.5 algorithm:** C4.5 is an algorithm used to produce a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is frequently referred to as a statistical classifier. C4.5 is the most well-known inductive learning algorithm, which can be used to construct decision trees in addition to prediction rules. C4.5 is a software extension of
the basic ID3 algorithm designed by Quinlan, but it can also compact with continuous attributes and null attribute values. To compact with the continuous attributes, they should be discretized first. To construct a decision tree from training data, C4.5 tree employs a greedy approach that uses an information theoretic measure (gain ratio) as its guide. Selecting an attribute for the root of the tree divides the training instances into subsets according to the values of the attribute. If the entropy of the class labels in the subsets is less than the entropy of the class labels in the full training set, afterwards information has been gained through splitting on the attribute. C4.5 tree chooses the attribute that achieves the most information to be at the root of the tree. The algorithm is applied recursively to form sub-trees, stop when a given subset contains instances of only one class [9].

- **Bagging Algorithm**: Bagging decision trees, an early ensemble method, builds multiple decision trees by repetitively resampling training data with replacement, and voting the trees for a consensus prediction. On the other hand, there are strong empirical indications that bagging and random subspace methods are much additional robust than boosting in noisy settings [12].

- **Apriori Algorithm**: Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and bigger item sets as long as those item sets appear sufficiently frequently in the database. The frequent item sets determined by Apriori can be used to find out association rules which highlight common trends in the database: this has applications in domains such as market basket analysis. In general, Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further doing well extensions are found [10]. Apriori uses breadth-first search and a Hash tree structure to count candidate item sets professionally. It generates candidate item sets of length k from item sets of length k-1. Then it prunes the candidates which have an uncommon sub pattern. In accordance with the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to make a decision of frequent item sets among the candidates [11].

- **Neural Networks**: Neural Network used gradient descent technique based on biological nervous system having multiple interrelated processing elements. These elements are called as neurons. Rules are extracted from the trained Neural Network to get better interoperability of the learned network. To answer a particular problem NN used neurons which are organized processing elements. Neural Network is used for classification and pattern recognition. An NN changes its structure and adjusts its weight in order to minimize the error. Modification of weight is based on the information that flows internally and externally through network during learning phase. In NN multiclass, issue may be addressed by using multilayer make available for forward technique, in which Neurons have been employed in the output layer rather using one neuron.

**VI. CONCLUSION**

Predictive analytics are a rapidly future trend in Human Resources (HR). The effectiveness of predictive analytics is wider and therefore application in all related areas of HRM is important. The aim of human resources analytics is to provide an organization with insights for successfully managing employees so that business goals can be reached fast and efficiently.
VII. REFERENCES