Predictive analytics of HR - A machine learning approach

V. Kakulapati , Kalluri Krishna Chaitanya , Kolli Vamsi Guru Chaitanya & Ponugoti Akshay

Published online: 29 Jul 2020.

Submit your article to this journal

Article views: 2

View related articles

Predictive analytics of HR - A machine learning approach

V. Kakulapati *
Kalluri Krishna Chaitanya [†]
Kolli Vamsi Guru Chaitanya [§]
Ponugoti Akshay [‡]
Department of Information Technology
Sreenidhi Institute of Science & Technology
Yamnampet, Ghatkesar
Hyderabad 501301
Telangana
India

Abstract

One of the branches of analytics is HR analytics, which is developing the system HR units in organizations function, principal to sophisticated proficiency, and improved outcomes overall. The usage of analytics by human resources for many years. Though the assortment, processing, and data analysis have been generally manual and specified the nature of HR dynamics, the approach has been constraining HR. The prospect to effort predictive analytics in categorizing the employees furthermost likely to grow promoted. Here we apply machine learning techniques to analyze the employee information for improving his/her position in the organization. Compensation and job performance information from revenue rates and personnel characteristics to payroll and service history, never before have HR executives had such liberated right to use to individual details. In this work, we are applying random forest classification, which facilitates employee classification based on their monthly income and informal way to execute analytics on data. Further, we use clustering techniques based on the performance metrics similarity to analyze employee performance.

*E-mail: `vldms@yahoo.com (corresponding author)`
[†]E-mail: `chaitanyakrishnakalluri@gmail.com`
[§]E-mail: `vamsi.kolli10@gmail.com`
[‡]E-mail: `akahayponugonti12@gmail.com`

1. Introduction

The significant challenge of HR management is to facilitate training for employees on-demand skills and mapping with the organizational goals. Employees are organizations significant assets to overcome competition between different organizations. Human resources people can comprehend as increasingly tenable on the grounds, which provides statistically legitimate information and utilized in generating new policies during the implementation of existing policies.

Human resource analytics is a pursuit loom to contract with assimilates parameters for civilizing employee productivity to develop organizational growth. HR team compatible terms utilized for analyses the employee ability, individual's talent, and balancing the workforce. The main job of HR estimates the requirement of human resources in each part or organization and plan to recruit talented people, placing them in suitable training and development, preparing each employee's progression, employee retention, work commitment, salaries, and other welfare benefits. The crucial role of the organization is human resource analytics that includes a priority-based predictive analysis where they can predict the outcomes of changing strategies or provisions. While recruiting the people, traditional human resource analytics mainly focuses on the items, for example, turnover and hiring cost. Many organizations required a constant and frequent perspective of the employee; accordingly, HR requires executing employee increment.

Consequently, it is converted into essential for HR to extend IT and investment analytical proficiency and capabilities to generate a better return on investment (ROI). In [1], [2],[3] the improvement of expertise when combined with prognostic analytics exponentially increased HR purposes. The insights of human resource predictive analytics produced experiences can't accomplish through the traditional approach as these are responsive and proof-based choice decision systems. Three considerable alterations [4] that have generated a need for prognostic analytics in the human resource are [5]:

  i.   A substantial advance in computing command and its affordability.

  ii. Online human resource data available for analysing

  iii. There is a competition between industries for global talent to secure and track talented peoples.

Every industry or company has its metrics to use to calculate the performance of the employee. Those are time management, attrition intensity, revenue generation, and success rate. Organizations maintain tangible metrics that are decisive to demonstrate top hr professionals and managerial how deliberate HR programs can help influence a company's bottom line. The general parameters are

i.   calculation of turnover rate by monthly

ii. Estimation of revenue per Employee

iii. Capitulate proportion

iv. Employee resources expenditures

v.   the ratio between Human resource to staff ratio

vi. Outlays

vii. Rate of Promotion

viii. Managerial level female percentage

ix. Frequency of a member of staff absent

x.   Member of staff's payment price per each employee

xi. Member of staff's payment price confrontation rate

xii. Employee overtime cost

xiii. employee average age.

The hr management cannot survive for future events that they don't have prescient investigation abilities. The effectiveness of predictive analytics is extensive, and thus the application is wholly related departments of HRM are crucial. These analytics assists companies containing HR-related expenses while improving business functioning in addition to employee commitment and satisfaction. Predictive analytics alter and increasing innovation, which can accomplish 100% exactness in decision making for human resources and is take over conventional investigation in industries.

Many investigations are utilizing machine learning algorithms for predicting employee gross revenue. Those methods are usually problem-specific and difficult to simplify. This is because the employee data is confidential [6], which avoiding in-depth analyses due to employee data is noisy, varying, and contains missing values [7, 8] and performance of model persist evaluation.

2. Related study

Human resource management has various functions, which require methods to get insights into the patterns that assist the organizations in the identification of talented people from the vast employee data. To solving this to identify the skilled persons by the employment of HR analytics of employee data precisely and reasonably connecting with the organizational strategies. These analytics by utilizing different statistical methods, investigation strategy, and techniques to asses employee information and interpreting consequences into suggestive reports [9]. Predictive analysis is utilizing statistical analysis. The management asses the employee perception designs revealed by the decision-makers to the probable prediction of worker behavioral analysis, cost of training, and involvement.

For efficient management, enormous volumes of data to provide organization decision-making issues related to analysis conscience and legitimacy by utilizing the up-to-date technologies. Human resource analysis can involve as input to organizational management. Decision-makers can be alert by using predictive models, and the policies should not deviate the imperative regulations like injustice, incapacities—HR analytics using employee data for decision making. At the same time, they have to take into consideration employee confidentiality, privacy, and reliability.

Predictive analytics of HR management assists the managerial people in executing the HR functions like predicting market trends and needs of skilled persons, the requirement of suitable training, employee salary based on performance, and maintaining the employee data for reward assessment and maintaining the employee privacy. Predictive analytics assist managerial people in decision making about employment, retaining, cost of training, awards, growth of employee career, and administrative efficiency and productivity. The HR predictive analytics estimate the impending consequences and comprehend the implications of theoretical modifications in organizations. HR analytics utilizing predictive analytics to decide for enhancing the performance of the employee and modifying the existing organizational policies [10].

The existing researches focused added on predicting the member of staff revenue utilizing machine learning methods and evaluated the effectiveness of every algorithm. The analysis of HR analytics is collective analysis, such as statistical analysis, modelling, and mining of human resource data, which is achieving significant outcomes [11,12].

Existing investigations built up a model that merges analytical workforce data, values of predictor inconsistencies, and mediating factors such as positive attitude and emotional intelligence to evaluate the performance of the employee. [13]. Because determining the collaboration of cross-functional activities of the employee and emotional intelligence is required for the hr team, this could demonstrate a significant prescient measurement for evaluating employee performance. The hr team struggles to discover into investigation-based competency models that were accurately predicting employee performance. Secure studies are utilizing predictive analytics to assess necessary competencies and correlate with performance factors [14].

Every employee appraisal frequently utilized for estimating his/ her performance analysis. There is no necessity for all performance factors useful for executive employments. An investigation reasoned that specific components could all the more precisely anticipate execution [15]. Many researchers [16-18], whatever the case may exhibit reliable indication for utilizing an employee's present and past activities for career improvement, the estimation of occupation in making or improving abilities for promotion or upgrade their position to more efficiently anticipate performance analysis and distinguish successors. Which is a complex task due to considering most of the managerial jobs of appraisal don't measure this quality in positions? Few tools are utilizing data developed to measure employee performance analysis. The investigators inferred that there is a set of business familiarity, maintaining risks, social influence and adaptability, utilization of input, and reaction to report help to recognize future pioneers [19].

3. Methodology

One of the unsupervised clustering techniques is K-means is vector quantization and its most efficient algorithm in data mining techniques. The objective of this algorithm is partitioning n observations into k clusters n which every consideration has a place in a cluster with the closest mean, serving as a model of the cluster, which brings about a partitioning of the data space. This algorithm minimizes inside-cluster variances called squared Euclidean distances, nevertheless systematic Euclidean similarity distance measure. The mean enhances squared errors, however only the geometric median limits Euclidean. Improved Euclidean results, such as initiated through k-medians and k- medoids.

3.1 *Combination of k-means clustering*

One most significant unsupervised learning is k-mean clustering, which uses a constant enhancement technique. It is also referred to as "naive k-means" due to faster alternatives[6].

The k-mean algorithm proceeds between two steps:

i. Assignment: In this step, assigning every observation to the cluster, where the mean of the observation has the least squared Euclidean distance. Every observation allocated to at least one cluster or more than one.

ii. Update: estimating the centroids of the observations in the latest clusters.

When the assignments no longer change, then the algorithm has converged and no assurance to discover the best possible

4. Implementation result

We take data set from UCI machinery and data set contain above 50 thousand records out of that we took around 1000 records, and 70% of data is considered as train data and remaining as test data. The data set is pre-processing for removing unnecessary elements or noise elements removal. The data set is like

Table 1

The employee data

Employee id	department	region	gender	Recruitment channel	age	previous year rating	Avg Training score
8724	Technology	region_26	m	sourcing	24		77
74430	HR	region_4	f	other	31	3	51
72255	Sales & Marketing	region_13	m	other	31	1	47
38562	Procurement	region_2	f	other	31	2	65
64486	Finance	region_29	m	sourcing	30	4	61
46232	Procurement	region_7	m	sourcing	36	3	68
54542	Finance	region_2	m	other	33	5	57

Contd...

67269	Analytics	region_22	m	sourcing	36	3	85
66174	Technology	region_7	m	other	51	4	75
76303	Technology	region_22	m	sourcing	29	5	76
60245	Sales & Marketing	region_16	m	sourcing	40	5	50
42639	Sales & Marketing	region_17	m	sourcing	40	3	46
30963	Sales & Marketing	region_4	f	other	34	3	52
54055	Analytics	region_24	m	other	37	3	82
42996	Operations	region_11	m	sourcing	30	5	58

Table 2
The employee attrition data

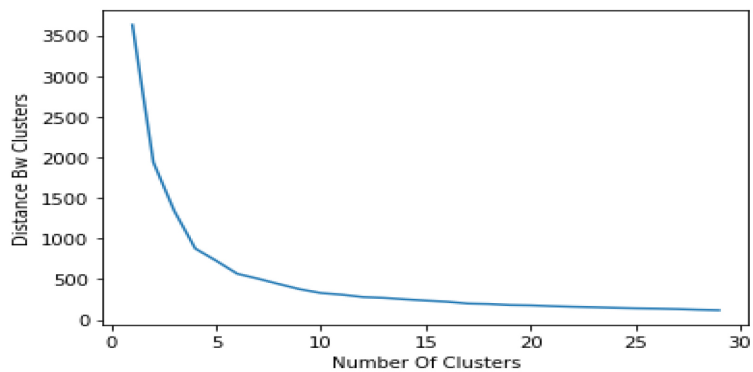|	Age	Attrition	...	YearsSinceLastPromotion	YearsWithCurrManager
0	41	1	...	0	5
1	49	0	...	1	7
2	37	1	...	0	0
3	33	0	...	3	0
4	27	0	...	2	2
5	32	0	...	3	6
6	59	0	...	0	0
7	30	0	...	0	0
8	38	0	...	1	8
9	36	0	...	7	7

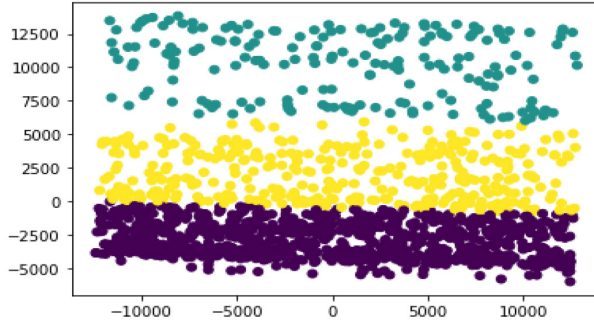**Fig 1**
Understanding Cluster Estimation Graph

Fig 2

The employee salary-wise clustering representation

This graph is acquired to determine how many clusters should take for our K-Means Clustering Algorithm implementation. This graph shows the relationship between how the distance between clusters and the number of clusters is.

From the above graph, we can identify clusters, i.e., the distance between every cluster is distinctive when there are approximately 3-4 clusters taken.

For understanding the graph, we have used RandomForestClassifier to classify based on monthly income and which cluster they fall. We have got to know that People with Monthly Income approx. < 5000 falls under 1st cluster People with Monthly Income approx. between 7000 to 12000 falls under 2nd Cluster People with Monthly Income approx. >12000 falls under the 3rd cluster.
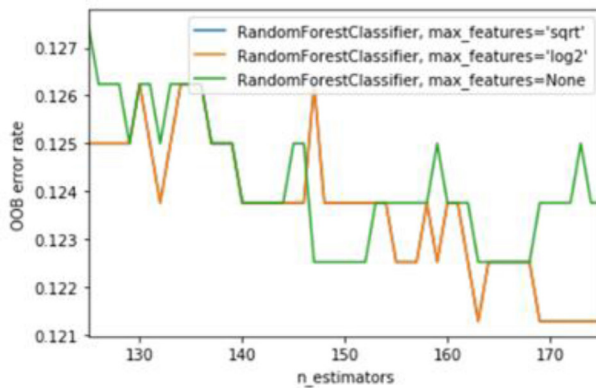

Fig 3

The calculation of error rate for n-estimators

5. Conclusion

We have used the Unsupervised Learning algorithm to automate in which section and which department the HR needs to take employees. We used K-Means clustering to define how many clusters we are going to have to set the attributes on which the HR is going to employ the new candidates or promote the existing candidates according to depending characteristics. In Unsupervised Learning, we can learn and evaluate new patterns, but the accuracy might be less dependent upon data. We classified the clustering model based on Monthly income to determine under which cluster, which can be doing for any attribute depending on the feature set we are taking and the clustering.

6. Future enhancement

In the future, develop employee performance analysis based on the similarity and effectiveness of the factors of risk handling, managerial ability. A framework-based analysis of employee ability and performance analysis can be analyzed. Also, ABE (attribute-based encryption) to protect the combined information is according to workforce' inclinations, which demonstrate considerable benefits in terms of key management to provide privacy even when this data requires to distribute through central data platforms.

References

[1] L. Bassi, "Raging Debates in HR Analytics", People & Strategy, Vol. 34, Issue 2, 2011

[2] M. Molefe, "From Data to Insights: HR Analytics in Organizations," Gordon Institute of Business Science, University of Pretoria, 11 Nov. 2013.

[3] L. Bassi and D. McMurrer, "A Quick Overview of HR Analytics: Why, What, How, and When?" Association for talent development, March 04, 2015

[4] D. Handa and Garima, "Human Resource (HR) Analytics: Emerging Trend In HRM (HRM)", IJRCM, Vol.No. 5, Issue No. 06, June 2014, ISSN 0976-2183

[5] K. Ladimeji, "5 Things that HR Predictive Analytics will Actually Predict." Recruiter (Jan. 23, 2013), sec. I p.1.

[6] Quinn, A., Rycraft, J.R., Schoech, D.: Building a model to predict caseworker and supervisor turnover using a neural network and logistic regression. J. Technol. Hum. Serv. 19, 65–85 (2002)

[7] Chien, C.F., Chen, L.F.: Data mining to improve personnel selection and enhance human capital: a case study in high-technology industry. Expert Syst. Appl. 34, 280–290 (2008)

[8] Punnoose, R., Ajit, P.: Prediction of employee turnover in organizations using machine learning algorithms. Int. J. Adv. Res. Artif. Intell. 5, 22–26 (2016)

[9] Levenson, A. 2011. Using targeted analytics to improve talent decisions. People and Strategy, 34(2), pp 34.

[10] J Fitzenz and J. R. Mattrox II, "Predictive Analytics for Human Resource." Wiley Publication, SAS Institute Inc., Cary, North America, USA, 2014 pp. 2-3

[11] Connors, D. and Mojsilovic, A. Workforce Analytics for the Enterprise: An IBM Approach. Chapter in *Service Science Handbook*, Maglio, P.P., Kieliszewski, C.A. and Spohrer J.C. (ed.s), Springer, 2010.

[12] Naveh, Y., Richter, Y., Altshuler, Y., Gresh, D. L. and Connors, D. P. Workforce optimization: identification and assignment of professional workers using constraint programming. *IBM Journal of Research and Development*, 51, 2007.

[13].Bersin, J., Houston, J., & Kester B. 2014. Talent Analytics in practice: Go from talking to delivering on big data. Deloitte University Press.

[14] Porath, C., & Bateman, T. 2006. Self-Regulation: From Goal Orientation to Job Performance. *Journal of Applied Psychology*, 91:185- 192.

[15] Russell, C. 2001. A Longitudinal Study of Top-Level Executive Performance. *Journal of Applied Psychology*, 86: 560-573.

[16] Morris, S., Daisley, R., Wheeler, M. & Boyer, P. 2014. A Meta-Analysis of the Relationship between Individual Assessments and Job Performance. *Journal of Applied Psychology.* Advance online publication.

[17] Dragoni, L., Tesluk, P. E., Russell, J. E. A., & Oh, I. 2009. Understanding managerial development: Integrating developmental assignments, learning orientation, and access to developmental opportunities in predicting managerial competencies. *Academy of Management Journal*, 52: 731-743.

[18] McCauley, C., Ruderman, M., Ohlott, P. & Morrow, J. 1994. Assessing the development components of managerial jobs. *Journal of Applied Psychology,* 79: 544-560.

[19] Spreitzer, G., McCall Jr., M., & Mahoney, J. 1997. Early identification of international executive potential. *Journal of Applied Psychology,* 82:6-29.