

“Understanding Employee Attrition– An Organizational Change Perspective–using Predictive Analysis Techniques”

Dr.R.V. Dhanalakshmi

Associate Professor, Department of MBA –New Horizon College of Engineering, Bengaluru, Karnataka.

Dr.Richa Tiwari

Associate Professor, School of Legal Studies, Reva University, Bengaluru, Karnataka.

Dr.SheelanMisra

Professor & Head, Department of MBA –New Horizon College of Engineering, Bengaluru, Karnataka.

Dr.R.Punniyamoorthy

Associate Professor, Department of MBA –New Horizon College of Engineering, Bengaluru, Karnataka.

Abstract

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. Accurate predictions enable organizations to take action for retention or succession planning of employees. However, the data for this modeling problem comes from HR Information Systems (HRIS). This paper makes an attempt to study the various Predictive analytics tools which includes Naïve Bayes, Support Vector Machines, Decision – tree & random forests, Logistic regression, Machine Learning & K - nearest neighbours. A theoretical approach is made towards all the above techniques to understand the strength of each technique to understand Employee attrition which leads to Organizational Change.

Key words - Employee Attrition, Predictive analytics Techniques, Organizational Change, Machine Learning

Introduction

The purpose of HRM is measuring employee performance and engagement, studying workforce collaboration patterns, analyzing employee churn and turnover and modelling employee lifetime value. Talent analytics, people analytics, and workforce analytics are the synonyms of HR Analytics. The present situation demands HR analytics to perform workforce optimization and hence it became important for HR to develop IT and finance analytical skills and capabilities to produce better Return on Investment. Three significant changes that have really created a hunger for predictive analytics in HR and these are:

- i. Major boost in computing power and its affordability
- ii. HR big data digitally accessible via cloud storage for processing
- iii. Global talent war to protect and pursue talent, it is about data-derived insights that drive better decisions. includes statistical techniques, machine learning methods, and data mining models that analyse and extract existing and historical facts to make predictions.

Human resource analytics is a pursuit loom to contract with assimilates parameters for civilizing employee productivity to develop organizational growth. 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

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 Modelling

Predictive model combines analytical algorithms and provide results in form of a value or probability scores based on which decisions are taken. Predictive modelling is applied on HR Functions as follows:

- i. Turnover Modelling: Predicts future turnover of business in specific functions, business units, geographies and countries by looking at factors such as commute time, time since last role change, and performance over time. Thus this can scale hiring efforts accordingly, reducing empty desk time and panic hiring, which can lead to lower cost, higher quality hiring.
- ii. Response Modelling: Use old advertising jobs data from previous campaigns to avoid contacting candidates or using channels that don't yield a response and focus on those channels that do work.
- iii. Predictive Retention Modelling: Identify high-risk employees, build profiles, predict vacancies and leadership needs, and understand how risk is distributed throughout the organization.
- iv. Risk Modelling: Develop a profile of candidates with a higher risk of leaving prematurely or performing below standard.
- v. Talent Forecasting: Being able to predict which new hires, based on their profile, and are likely to be high fliers and then moving them in to your high potential programs.

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.

Applying Predictive Analytics in HR

Several organisations have proactively opted predictive analytics for his or her business functions, as an example, finance and risk, client relationship management, selling and sales, and producing. This empowers them to create choices over a scope of activities, as an example, client retention, sales Predictive, insurance rating, and campaign management, provide chain optimization, credit marking, and research[6]. By foreseeing what each client loves, the sales and selling performers will decide on viable selections on product recommendations. Moreover, finance team draws upon predictive modelling and illustration for effective money analysis, which diminishes money risk [7]. Organizations are investigating new potential outcomes for befitting mistreatment the steady increasing intensive volumes of knowledge. Organizations will mine progressive info known with

turnover, investment in training programs, nature of contracts, and alternative basic components of HR administration. This may be used to create applied math models that gauge possibilities and foresee future conduct and patterns crosswise over key HR range. To adopt predictive analytics, HR pioneers need an organized info supply that's adjusted to the business [8]. Whereas HR processes like enlisting and resourcing, men administration, performance and learning, payroll, and time management typically use structured knowledge sources, these are solely integrated with one another. The thanks to utilizing predictive analytic andacknowledging most extreme benefits from the HR info lays in binding the data supply to important business results[9]. The HR's, in addition, wants ability to utilize systematic devices with success supervise talent and enlisting info. HR personnel should be ready adequately to assume an important half in business development by proficiently utilizing examination. HR pioneers have to be compelled to guarantee that examination and its results are adjusted to business objectives [10]

Review of Literature

Employee Attrition

The loss of an organization's employee can be divided into three broad groups, induction crises, natural wastage, and retirement (Bennisonn and Casson, 1984). The lack of career opportunities, challenges and dissatisfaction with the jobscope or conflict with the management have been cited as predictors of high job turnover (Dijkstra, 2008). Wei-Chiang and Ruey-Ming (2007), in their work explored the feasibility of applying the Logit and Probit models, which have been successfully applied to solve nonlinear classification and regression problems, to employee voluntary turnover predictions.

(Adhikari, 2009) distinguished the components influencing worker wearing down in the IT and ITES industry. The paper recognized four elements. Business related issues have the best effect on whittling down. The other three being manager related issue, aptitude of workers and the pay. Pay appeared to have minimal impact on steady loss. turnover aim has effect on steady loss factors, for example, QWL, vocation development, working hours, individual/family reasons, and connection with inner co – specialist, welfare, working condition, and compensation.

Several studies have explored the use of machine learning to predict employee behaviour. Authors used decision trees and Naïve Bayes classifier to predict employee performance. They found that job title was the strongest feature, whereas age did not show any clear effect. The authors explored several data mining algorithms to predict employee churn (or attrition) using a dataset comprising 1575 records and 25 features. Following machine learning algorithms: naïve Bayes, support vector machines, logistic regression, decision trees and random forests were used.

A model for predicting employee attrition, as well as discuss the serious ethical implications of using such a model within organizations. Aset of statistically significant factors that correlate to an employee's decision to quit, and determined to which types of occupations the model may be applied. After applying Principal Component Analysis and classification methods K-Nearest Neighbors and Random Forest, it was Logistic Regression that allowed to simplify the model and predict employee quits with the highest accuracy of our testing methods, achieving a greater than seventy-four percent success rate

Organizational Change & Technology

The recent emergence of the 'datafication' concept reveals that companies are increasingly dependent on data (Lycett 2013). This digital transformation of industry, that is, the digitization of

horizontal and vertical chains (Ardito et al. 2019; Golzer € and Fritzsche 2017), is at the centre of the Industry 4.01 revolution. Big data has been shown a key element of Industry 4.0 and contributes extracting benefits from industrial operations (Golzer and Fritzsche € 2017; Xu, Xu, and Li 2018). Hence, seizing opportunities that arise from big data is considered by several researchers to be the greatest problem faced by companies today (Post and Edmiston 2014).

Employees' emotional attachment to and identification with an organization is concerned with an understanding of organizational goals and values. In the literature, an employee's emotional attachment is associated with organizational loyalty or willingness to dedicate more efforts towards an organization, and organizational identification is concerned with a feeling of pride or a desire to retain a connection with the organization (Chang and Chang, 2007; Chun et al., 2010; Cook and Wall, 1980). Job satisfaction is a key driver underpinning attitudes and behaviors in the workplace (Alegre et al., 2016; Rayton and Yalabik, 2014; Topolosky, 2000) where this relates to how employees think, feel and perceive their jobs (Oliver, 1990; Randall et al., 1999; Spector, 1997). However whilst this shows a dearth of objective factors which relate to organizational change, much of the research in the field of change management does not explicitly examine or relate job satisfaction factors with employee attitudes and behaviors – much less even extend additional indirect company environmental vectors

Predictive analysis techniques

1. Naïve Bayes Theorem

Naive Bayes is a classification strategy that has picked up fame because of its simplicity. The Naive Bayes algorithm utilizes the assumption that every one of the variables are independent to each other, and after that compute"s probabilities, that are utilized for classification. The algorithm works as follows: to get an output function Y given a set of input variables X, the algorithm estimates the values of $P(X|Y)$ and $P(Y)$, and then uses Bayes" rule to compute $P(Y|X)$, which is the required output, for each of the new samples.

2. Support Vector Machines

Support vector machines (SVMs) is a nonprobabilistic supervised machine learning model used for classification and regression. SVMs will train algorithms with assigned classes by separating each class through a decision boundary, also known as a hyperplane [11], [12]. Some problems are considered nonlinear in so far as it is difficult to draw the decision boundary. However, this can be solved by using a kernel function (also known as a kernel trick). This function returns the dot product of the two vectors, where it then maps data points to a new, transformed, high-dimensional space. Moreover, there are several types of kernel function can be used such as linear, Gaussian, and polynomial kernel.

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

3. Decision – tree & random forests

Decision trees are graphical representations of alternative choices that can be made by a business, which enable the decision maker to identify the most suitable option in a particular

circumstance. Decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a topdown recursive divide-and-conquer manner. A greedy strategies is usually used because they are efficient and easy to implement, but they usually lead to sub-optimal models. A bottom-up approach could also be used. Decision tree learning is a method commonly used in data mining.

The goal is to create a model that predicts the value of a target variable based on several input variables. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. In data mining, trees can be described also as the combination of mathematical and computational techniques to aid the description, categorisation and generalisation of a given set of data. Data comes in records of the form: $(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$ (1) The dependent variable, Y, is the target variable that we are trying to understand, classify or generalise. The vector x is composed of the input variables, x_1, x_2, x_3 etc., that are used for that task.

Random Forest

Random forest (RF) is one of the most powerful supervised machine learning algorithms for generating classifications and regressions. RF uses multiple decision trees to train data [15]. Each tree votes for a classification label for a certain dataset, then the RF model chooses which class had the most votes from the decision trees

4. Logistic regression

Logistic regression is a regression model that fits the values to the logistic function. It is useful when the dependent variable is categorical [5]. The general form of the model is $(Y|X, W) = 1/(1+e^{-(w_0+\sum w_i x_i)})$ Logistic regression is often used with regularization techniques to prevent overfitting.

5. Machine Learning

Machine Learning addresses the question of how to build computers that improve automatically through experience. It is one of today's most rapidly growing technical fields, lying at the intersection of Computer Science and Statistics, and at the core of artificial intelligence and data science. Recent progress in machine learning has been driven both by the development of new learning algorithms and theory and by the ongoing explosion in the availability of online data and low – cost computation. The adoption of data – intensive machine – learning methods can be found throughout science, technology and commerce, leading to more evidence – based decision making.

6. K – nearest neighbours

K-nearest neighbours (KNN) is one of the simplest machine learning algorithms and is used for both classification and regression. KNN works by specifying the value of K, which indicates the number of closest training points for a single data point. Each new data point will be classified based on the majority of votes collected from its neighbours.

The KNN algorithm classifies new data dependent on the class of the k closest neighbours. This paper utilizes the estimation of k as 6. The good ways from neighbours can be determined utilizing different distance metrics, for example, Euclidean distance, Manhattan distance, Minkowski distance, and so forth. The class of the new information might be chosen by dominant part vote or by an inverse proposition to the distance computed. KNN is a non-generalizing technique, since the algorithm keeps the majority of its preparation information in memory, perhaps changed into a quick ordering structure, for example, a ball tree or a KD tree. The Manhattan distance is computed using the formula $D = \sum |x_i - y_i|$

Discussions & Conclusion

Organizational change refers to the actions in which a company or business alters a major component of its organization, such as its culture, the underlying technologies or infrastructure it uses to operate, or its internal processes. Employee retention refers to the ability of an organization to retain its employees. Employee retention can be represented by a simple statistic. However, many consider employee retention as relating to the efforts by which employers attempt to retain the employees in their workforce. Talent retention consists of a set of practices and policies adopted by companies to ensure that their above-average professionals remain in it for much longer. With that, it is possible to build a great competitive differential. To ensure retention of Employees from an Organizational change perspective, the Machine Learning approaches have to be adopted for an effective understanding of Employee retention behaviour.

References:

- [1] Applying Advanced Analytics to HR Management Decisions,” James C. Sesil, Pearson Publication, New Jersey, March 2014, pp. 13-25.
- [2] M. Blankenship, “Managing and Measuring Talent Risk,” SHRM Foundation Thought Leaders Conference, Oct 2011.
- [3] L. Smeyers, “7 Benefits of Predictive Retention Modeling (HR analytics).” iNostix (May 6, 2013)
- [4] V. Kakulapati, Kalluri Krishna Chaitanya, Kolli Vamsi Guru Chaitanya & Ponugoti Akshay (2020) Predictive analytics of HR - A machine learning approach, Journal of Statistics and Management Systems, 23:6, 959-969.
- [5] Rohit Punnoose, Pankaj Ajit, “Prediction of Employee Turnover in Organizations using Machine Learning Algorithms, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No. 9, 2016,
- [6] Employee churn prediction V. Vijaya Saradhi†, Girish Keshav Palshikar, Expert Systems with Applications, Elsevier, 38 (2011) 1999–2006.
- [7] Alao D. & Adeyemo A. B. ,Analyzing Employee Attrition Using Decision Tree Algorithms Computing, Information Systems & Development Informatics ,Vol. 4 No. 1 March, 2013.
- [8] Dr. S. Rabiyaatul Basariya & Dr. Ramyarrzgarahmed, “A Study on Attrition – turnover intention of Employees”, International Journal of Civil Engineering and Technology (IJCET), Volume 10, Issue 01, January 2019, pp. 2594–2601.
- [9] Sarah S. Alduayj & Kashif Rajpoot, “Predicting Employee Attrition using Machine Learning”, 2018, 13th International Conference on Innovations in Information Technology.
- [10] Alex Frye et al., (2018), “ Employee Attrition: What Makes an Employee Quit?,” SMU Data Science Review, Volume 1, Number 1, Article 9.

[11] Yves Barlette&PamélaBaillette(2019),” Big data analytics in turbulent contexts: towards organizational change for enhanced agility,” Production Planning & Control,2019.

[12] B. Sri Harsha et al., (2020), Early Prediction of Employee Attrition,” International Journalof Scientific& TechnologyResearch, Volume 9, Issue 03, March 2020