Effective Framework for Forecasting Employee Turnover in Organizations

N. Arvinth[1*]

[1*]*Research Associate, National Institute of STEM Research, India. E-mail: nagarajanarvinth@gmail.com, Orcid: https://orcid.org/0009-0000-9798-9828*

Received: 06 June 2024; Revised: 08 July 2024; Accepted: 07 August 2024; Published: 30 September 2024

Abstract

Employee turnover is a crucial issue for businesses, producing significant expenses and problems. Accurate turnover prediction is vital for building talent retention strategies. Conventional models frequently struggle to account for the changing and nonlinear elements that affect employee turnover. In this study, a novel Dynamic Black Hole tuned Logistic Regression (DBH-LR) was proposed to accurately forecast employee turnover in the organization. The employee turnover dataset was gathered from the Kaggle source. Data cleaning is employed to ensure the quality and dependability of the employee dataset. The Black Hole conditions in space, which dynamically modify the locations of potential solutions to converge towards an ideal solution, served as the model for the DBH algorithm. The Logistic Regression (LR) model's parameters are adjusted using DBH to increase the model's forecasting accuracy. The suggested DBH-LR method is tested on the Pythonplatform. The DBH-LR model outperformed existing techniques in terms of recall (97.4%), accuracy (96.3%), F1-score (95.3%), and precision (96.1%). Future studies may examine the use of this methodology in different organizations and worker groups to validate its efficiency.

Keywords: Employee Turnover; Dynamic Black Hole Tuned Logistic Regression (DBH-LR); Organization, Prediction.

I. INTRODUCTION

Employee turnover is a common problem for organizations in several sectors because it links to operational performance, organizational effectiveness, and profitability, as well as staff engagement. By predicting the conditions and circumstances under which the workers may be likely to quit their jobs, the employer can avoid the dangers and costs (Bell, 2022). It is important to predict the turnover correctly because it helps organizations to prepare for their demand of workforce, reduce costs that are incurred on recruiting new employees, and improve productivity. There are several aspects, as well as factors that may be both numerous and intricate, interrelated to employee turnover (Wang & Zhi, 2021). It can encompass such aspects as business practices and policies, including managerial ones, the company's organizational culture, or remuneration systems and levels, as well as personal factors such as work-life balance, career advancement opportunities, and employee satisfaction levels. Another factor greatly determined by changes in the economy and fluctuations particular to a given sector is turnover rates. These characteristics help organizations address underlying problems since they can turn to such factors to recognize signals of turnover (Shafie et al., 2024). Forecasted employee turnover is helpful for strategic planning and makes it possible to develop specific retention tasks. There are ways to identify those employees likely to leave, and organizations can develop individualized retention strategies including; career development programs, individualized incentive structures, and better working conditions. This proactive approach reduces costs that are associated with the turnover of employees, enhances general organizational performance, and enhances the ability of organizations to retain important employees (Salleh et al., 2020). To precisely predict employee turnover in a company, an innovative DBH-LR was presented in this research.

This research is structured into the following sections: related works, methodology, results, and conclusion.

II. RELATED WORKS

To regulate the significant features that influence an employee's decision to relieve an organization and to predict the possibility of a particular person leaving, (Fallucchi et al., 2020) examined how objective elements affect employee turnover. The findings showed the highest recall rate (0.54) since it evaluated a classifier's ability to recognize every positive occurrence and attain a total false negative rate equivalent to 4.5% of all observations. A novel machine learning (ML)-focused strategy was applied by (Jain et al., 2020) to improve various turnover strategies for specific employees. Additionally, an attempt was made to clarify the various factors influencing the worker turnover rate and potential solutions. Based on the findings of their study, it could be determined that random forest (RF) performed better. For the employee turnover and forecasting method (Chaudhary et al., 2022) suggested multi-criteria decision-making (MCDM) and ML-based approach. The CatBoost algorithm provided excellent outcomes for performance evaluations for classified employees and outperformed all employee datasets. By utilizing predictive analytics techniques, (Yahia et al., 2021) allowed human resources (HR) managers to reduce staff turnover by quickly identifying the possibility that a worker would leave. Compared to other studies, they indicated that business travel was the main source of inspiration for workers and that HR policies should take it into account to increase retaining workers. To forecast the probability of staff turnover based on corporate culture and managerial variables (El-Rayes et al., 2020) developed binary classification models using trees. The most effective retention forecasting techniques were found to be those utilizing decision trees and random forests. Additionally, compensation, corporate culture, and the effectiveness of executive leadership were the main factors influencing a worker's choice to leave an organization.

III. METHODOLOGY

The employee turnover data was gathered from the Kaggle platform, the data was then cleaned using the data cleaning process. The DBH-LR approach is proposed for effectively forecasting employee turnover in organizations.

3.1 Data Collection

The employee turnover dataset is gathered from the Kaggle source "https://www.kaggle.com/datasets/marikastewart/employee-turnover". The HR department obtained data on more than ten thousand workers who departed the organization between 2016 and 2020. They utilized data from employee files, exit interviews, and performance reviews. The department, review, promotion, projects, satisfaction, salary, bonus, tenure, left, and avg_hrs_month are the features used in this data set.

3.2 Data Cleaning

Data cleaning is the process of getting data with errors or inconsistencies in data and rectification of these errors. Therefore, it involves missing values, data correction, removing duplicate data, and standardizing values. Through this approach of data cleaning, the analysis is made more accurate and dependable since it is made certain that the data used is accurate, and has no inconsistent data. It is a significant phase to acquire the data suitable for detailed and accurate analysis.

3.3 Dynamic Black Hole tuned Logistic Regression (DBH-LR)

DBH-LR, a new method, is the approach used for the prediction of labor turnover in companies. DBH algorithm is an adaptive optimization technique inspired by black holes. LR is incorporated with this approach employing the DBH algorithm. Through the use of DBH for the tuning of hyperparameters in the LR model, DBH-LR enhances the accuracy of the model forecast. DBH-LR deals with the issue of parameter selection to the given problem of prediction of turnover in the employees, which can be very much influential to the performance of the model. To identify the best configuration to be used in the LR algorithm, the DBH algorithm changes some variables while it searches through the parameters space very efficiently. Combining these methods, DBH-LR strengthens the model's possibility to predict the level of employee turnover thus helping the businesses to use more accurate estimations for managing retention strategies and human capital. The proposed methodology makes sure that an optimal and suitable model for the intricate patterns of employee turnover is created.

- ***LR Algorithm***

LRis a quantitative analysis method employed to estimate the possibility of employee turnover in an organization as inspected by many factors, which is called LR. The research examines the relationship between various factors such as tenure, job satisfaction, and performance with the output variable, turnover. It assists in maintaining aggressive retention efforts by identifying the probability of turnover utilizing a logistical model on the data. The LR model has been widely used in many industries, including anticipating employee turnover. This technique is especially effective for categorizing data into various groups. The target variable in LR is often binary, indicating it can only have values of 1 or 0. This denotes, whether a worker will remain with the organization or leave. Finding the most suitable model that accurately represents the connection between the anticipated variable (employee turnover) and the forecasting factors is the objective of the LR approach. The linear regression algorithm provided in Equation (1) below serves as the basis for the LR procedure.

$$z = g_{\theta}(w) = \theta^{S} w \tag{1}$$

Equation (1) does not effectively predict employee turnover results $\left(z^{(j)}\varepsilon\{0,1\}\right)$. Therefore, the function in Equations (2 and 3) was presented to calculate the possibility that a specific employee, considering their characteristics, will fall into the 1 (leaving) class as opposed to the 0 (not likely to leave) class.

$$O(z=1|w) = g_\theta(w) = \frac{1}{1+\exp\left(-\theta^S w\right)} = \sigma(\theta^S w) \quad (2)$$

$$O(z=0|w) = 1 - O(z=1|w) = 1 - g_\theta(w) \quad (3)$$

The sigmoid function, or Equation (4), allows to keep the value of $\theta^S w$within the [0, 1] interval. This assures that the predicted employee turnover falls within this probability range. Subsequently, it is required for θ to have a value where the probability $O(z=1|w) = g_\theta(w)$ is high when w is in the 1 class and low when w is in the 0 class ($O(z=0|w)$).

$$\sigma(s) = \frac{1}{(1+e^{-s})} \quad (4)$$

The LR approach has been successfully developed and executed.

- ***Dynamic Black Hole Algorithm***

The DBH algorithm can promote enhancing the predictions of employee turnover while considering the anticipations of turnover as well as the settings of the parameters in an optimum manner. To enhance the possibility of identifying and utilizing recognized data signatures of the worker behavior, the DBH algorithm adapts the field of its search, making data predictions more accurate. This strategy ensures that organizations will dynamically enhance their employee management and retention methods to prevent and circumvent possible turnover challenges appropriately. The black hole algorithm will be dynamically enhanced by employing star gravity data. To accomplish this, the kind of gravitational attraction that exists between stars is described and their motion are modified to forecast employee turnover technique while examining the solution space. The swarm contains M stars. Equation (5) defines the location of the j^{th}stars $\left(W_j\right)$.

$$W_j = \left(star_j, \ldots\ldots, star_M, blackhole_c\right) \quad (5)$$

Where $star_i$ denotes the j^{th} star's location and $blackhole_c$denotes the c^{th} black hole's location. The absorption that is operating on star j from star i at a given time k is defined by Equation (6).

$$F_{ji}^{c}(k) = \zeta(k_0)\frac{D_{oj}(s)\times D_{bi}(k)\times\left(star_i(k)-star_j(k)\right)}{\left(C_{ji}(k)+\varepsilon\right)^2\times\left(D_{oj}(k)+D_{bi}(k)\right)} \times \left(\frac{k_0}{k-k_0}\right)^{\propto} \quad (6)$$

Where,

D_{oj}-Power related to star j,

$C_{ji}(k)$-Length between the stars j and$i$,

$\xi(s_0)$-Initial absorption constant,

D_{bi}-Power of star i, and

ε- Small constant.

The overall force in the black hole method is the arbitrarily weighted sum of the different factors, which gives it a random quality, as shown in Equation (7).

$$F_j^c(k) = \sum_{i=1\ i\neq j}^{M} rand_i F_{ji}^c(k) \quad (7)$$

In [0,1], where $rand_i$is located. Therefore, Equation (8) gives the speed of the star j in direction c^{th} at time k.

$$b_j^c(k) = \frac{F_j^c(k)}{D_{jj}(k)} \quad (8)$$

The following is the consideration for the star's next velocity, where D_{jj} is the j^{th}star's power. As a result, velocity and location are determined using Equations (9 and 10).

$$g_j^c(k+1) = rand_j \times g_j^c(k) + g_j^c(k) \quad (9)$$

$$star_j(k+1) = star_j(k) + g_j^c(k+1) \quad (10)$$

$rand_j$ is located in [0, 1]. The search will be made random using this random number.

IV. RESULT

The suggested method was implemented on a Windows 11 laptop equipped with an Intel i5 12th Gen processor, 16 GB of RAM, and a Python 3.10.1 environment. The performance of the suggested method is evaluated with traditional methods such as Gaussian Naïve Bayes (GNB) [5], Decision Tree (DT) [5], and Extra Trees Classifier (ETC) (Raza et al., 2022).

The accuracy of a model indicates its overall predictive correctness by calculating the frequency with which its forecasts correspond to the actual results. Loss measures the variation between anticipated and actual results, and lower loss indicates better model efficiency. While loss indicates forecasting errors, accuracy in a prediction framework indicates how effectively the model forecasts turnover occurrences. The output of accuracy and loss is shown in Figure 1.

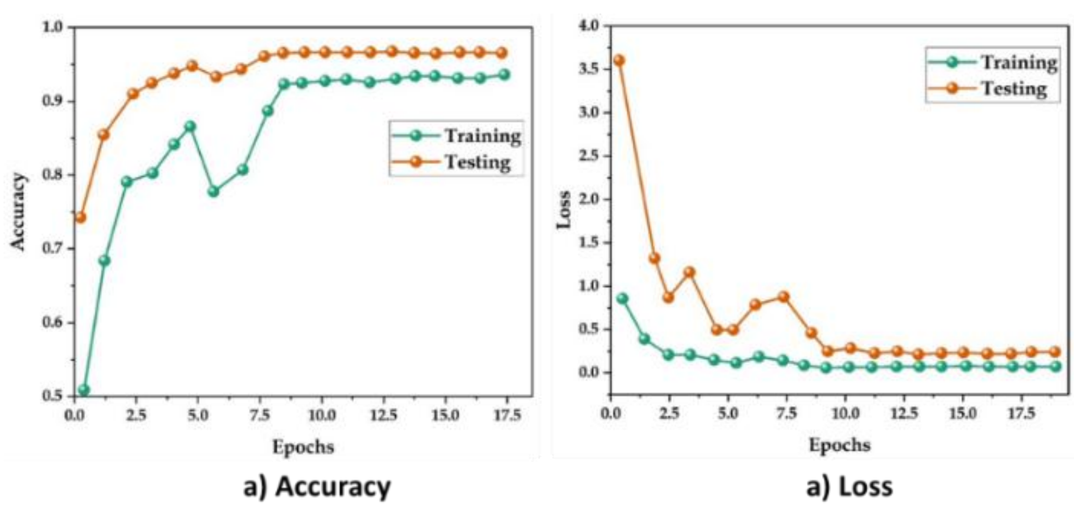


Figure 1: Output of a) Accuracy and b) Loss

The accuracy calculates the percentage of turnover events that were accurately predicted compared to the total number of actual occurrences. It evaluates the model's accuracy in forecasting workers who are most likely to quit, supporting successful retention techniques. The proposed DBH-LR technique has an accuracy of 96.3%, while the conventional approaches: ETC, GNB, and DT have accuracy values of 93%, 78.2%, and 79.2%, as shown in Figure 2 (a). The recall quantifies the percentage of real turnover instances that the model accurately predicted. It demonstrates the model's efficacy in recognizing employees who are likely to leave and demonstrates its capacity to detect true positives. The suggested DBH-LR approach outperforms the traditional, ETC, GNB, and DT approaches with recall rates of 93%, 54.1%, and 36.1%, respectively. Our proposed method has a high recall rate of 97.4%. The output of recall is shown in Figure 2 (b).

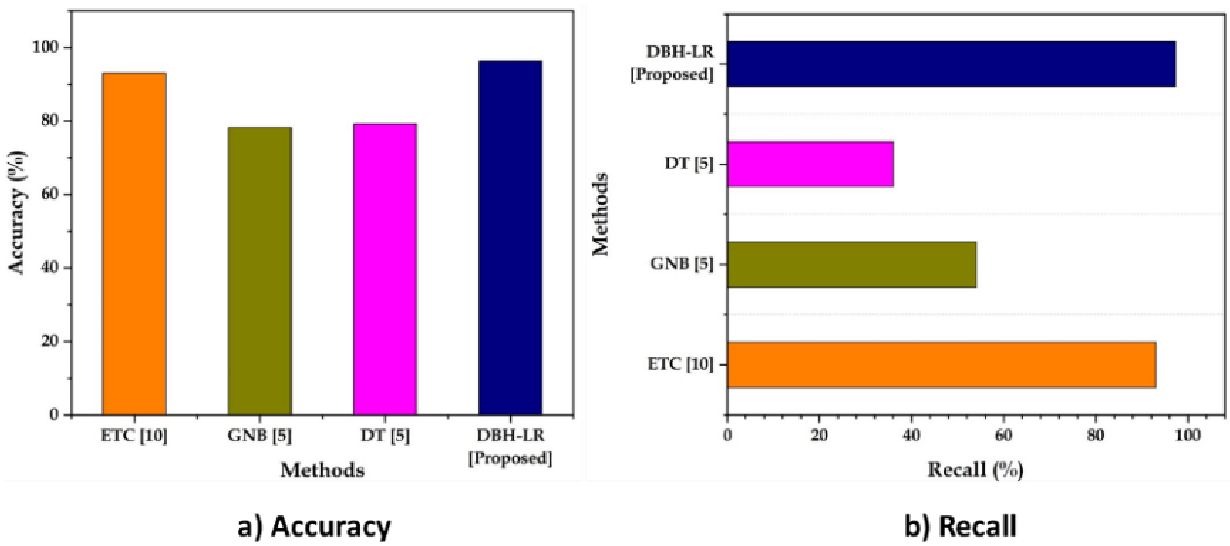
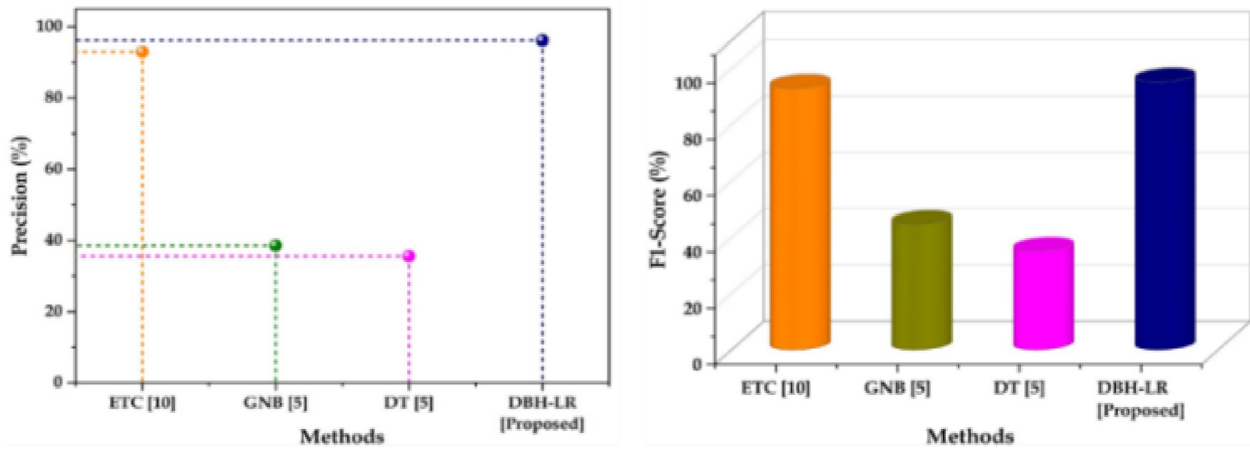


Figure 2: Output of a) Accuracy and b) Recall

The precision quantifies the percentage of accurately forecasted high-turnover instances among all the high-turnover cases that were forecasted. It assesses the model's precision in predicting employees likely to leave, reduces false positives, and improves focused retention tactics. The precisions of the suggested DBH-LR strategy are 96.1% in contrast to the precisions of the traditional approaches, which include the ETC, GNB, and DT, whose precision values are 93%, 38.6%, and 35.6%, as shown in Figure 3 (a).Precision and recall are balanced in the evaluation process using the F1-score measure. It provides a thorough performance evaluation by gauging how well the algorithm predicts employees who are likely to leave while reducing false negatives and positives. Our proposed DBH-LR technique has a 95.3% F1-score and outperforms the conventional ETC, GNB, and DT algorithms, which have recall ratios of 93%, 44.6%, and 35.1%, respectively. The output of F1-score is shown in Figure 3 (b).



a) Precision **b) F1-score**

Figure 3: Result of a) Precision and b) F1-Score

V. CONCLUSION

Employee turnover affects corporate culture, company growth, and productivity, which is an important issue for organizations worldwide. In this study, a novel DBH-LR was proposed to accurately forecast employee turnover in organizations. The suggested approach performed better than previous methods in terms of precision (96.1%), recall (97.4%), accuracy (96.3%), and F1-score (95.3%). It may struggle to anticipate employee turnover when dealing with high-dimensional data, complicated interconnections, and non-linearity. Future research should concentrate on improving the model's capacity for managing high-dimensional data, capturing intricate connections, and dealing with non-linearity by utilizing hybrid techniques or sophisticated algorithms for more precise forecasting of employee turnover.

REFERENCES

[1] Bell, S. M. (2022). *Strategies for Reducing Voluntary Employee Turnover in Small Business*. Walden University.

[2] Wang, X., & Zhi, J. (2021). A machine learning-based analytical framework for employee turnover prediction. *Journal of Management Analytics*, *8*(3), 351-370. https://doi.org/10.1080/23270012.2021.1961318

[3] Shafie, M. R., Khosravi, H., Farhadpour, S., Das, S., & Ahmed, I. (2024). A cluster-based human resources analytics for predicting employee turnover using optimized Artificial Neural Networks and data augmentation. *Decision Analytics Journal*, *11*, 100461. https://doi.org/10.1016/j.dajour.2024.100461

[4] Salleh, A. M. M., Omar, K., Aburumman, O. J., Mat, N. H. N., & Almhairat, M. A. (2020). The impact of career planning and career satisfaction on employee's turnover intention. *Entrepreneurship and Sustainability Issues*, *8*(1), 218. http://doi.org/10.9770/jesi.2020.8.1(14)

[5] Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting employee attrition using machine learning techniques. *Computers*, *9*(4), 86. https://doi.org/10.3390/computers9040086

[6] Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: a machine learning approach. *SN Applied Sciences*, *2*(4), 757. https://doi.org/10.1007/s42452-020-2519-4

[7] Chaudhary, M., Gaur, L., Jhanjhi, N. Z., Masud, M., & Aljahdali, S. (2022). Envisaging Employee Churn Using MCDM and Machine Learning. *Intelligent Automation & Soft Computing*, *33*(2). http://dx.doi.org/10.32604/iasc.2022.023417

[8] Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *Ieee Access*, *9*, 60447-60458. https://doi.org/10.1109/ACCESS.2021.3074559

[9] El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, *28*(6), 1273-1291. https://doi.org/10.1108/IJOA-10-2019-1903

[10] Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. *Applied Sciences*, *12*(13), 6424. https://doi.org/10.3390/app12136424