2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) | 979-8-3315-4362-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSES63760.2024.10910479

Machine Learning for Employee Turnover Prediction

Geetha Manoharan[1], Viyyapu Pushpa[2], Dr. Arati V. Deshpande[3], Dr. Melanie Lourens[4], M. K. Sharma[5], Ankur Jain[6]

[1]*School of Business, SR University, Warangal, Telangana, India*
[2]*Assistant Professor, Computer Science Engineering, Vignan's Institute of Information Technology, DuvvadaVisakhapatnam, Andhra Pradesh*
[3]*Assistant Professor, Computer Engineering, VIT,Pune, Maharashtra*
[4]*Deputy Dean Faculty of Management Sciences, Durban University of Technology, South Africa*
[5]*Professor, Mathematics, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India – 250004*
[6]*Computer Science & Engineering, IFTM University, Moradabad, Uttar Pradesh, India*

E-mail : geethamanoharan1988@gmail.com, pushpa.viyyapu@gmail.com, arati.deshpande1@vit.edu, melaniel@dut.ac.za, drmukeshsharma@gmail.com, ankur1101@gmail.com

Abstract: The main problem is that employees leaving the company has a big affect on its costs, efficiency, and effectiveness. By accurately estimating employee turnover, businesses can take proactive steps to keep good employees and lower the costs of hiring and teaching new ones. This research looks at how machine learning techniques can be used to predict staff turnover by looking at things like employee demographics, job satisfaction, success indicators, and engagement levels. Training and testing of the models were done with a dataset that included both old and new data from a big organisation. The results of this research are very helpful for human resource managers who are trying to fix problems with disengaged employees and make policies that keep employees. The research say that machine learning systems can accurately predict how many employees will leave a company.

Keywords:-Employee Turnover, Machine Learning, Predictive Analytics, Logistic Regression, Decision Trees, Random Forests, Gradient Boosting.

I. INTRODUCTION

Companies are acutely aware of the detrimental consequences of staff attrition, which include diminished productivity, morale, and efficiency. Organizations are increasingly utilizing machine learning (ML) to predict employee attrition in order to resolve this issue. Historical data and robust algorithms are employed by machine learning (ML) models to identify patterns and variables that contribute to employee attrition. This enables organizations to actively monitor and mitigate potential hazards.

In order to forecast employee departure, machine learning algorithms employ data from a variety of sources, such as workforce demographics, job performance indicators, and engagement surveys. Machine learning algorithms are capable of assessing this data and recognizing patterns and connections that are invisible to humans. For example, models are capable of identifying signs of disengagement or dissatisfaction in employees before they decide to resign, thereby enabling HR departments to provide support.

Several critical stages are involved in the implementation of machine learning for the prediction of employee attrition. Before obtaining precise and valuable input data, it is imperative to first conduct the data collection and preparation procedure. In order to advance, it is imperative to train machine learning models with historical turnover data in order to identify correlations and patterns [1]. After the models have been trained, they can be employed to forecast the likelihood of attrition and analyze current employee data.

Another benefit of employing machine learning in this context is that it encourages the implementation of HR practices that are informed by data. Instead of relying on intuition or anecdotal evidence, organizations can employ machine learning models to generate dependable forecasts and insights, thereby bolstering their retention initiatives. This approach not only enhances organizational efficiency and optimizes resource utilization, but also assists in the retention of valuable personnel.

In conclusion, machine learning offers a sophisticated, proactive approach to predicting employee attrition. By utilizing data and sophisticated algorithms, organizations can gain valuable insights into employee behavior, thereby improving staff stability and productivity. As a result, they can execute retention strategies that are more specific.

II. RELATED WORKS

The considerable impact that staff turnover has on productivity, morale, and operational expenses is a matter of great concern for firms. The implementation of targeted actions that can effectively mitigate the adverse

Authorized licensed use limited to: NWFP UNIV OF ENGINEERING AND TECHNOLOGY. Downloaded on July 10,2026 at 06:56:20 UTC from IEEE Xplore. Restrictions apply.

effects of employee turnover is made possible by the highly accurate prediction of staff turnover. In this discipline, machine learning (ML) has been increasingly employed to forecast attrition rates and analyze historical employee data [2].

Reflecting on the Past Statistical methods and linear models were the primary tools used to forecast employee attrition in the early stages. Tenure, performance evaluations, and demographic data were frequently implemented in the strategies [3]. Although they provided valuable insights, their predictive capabilities were limited by their inability to understand the complex interconnections within the data.

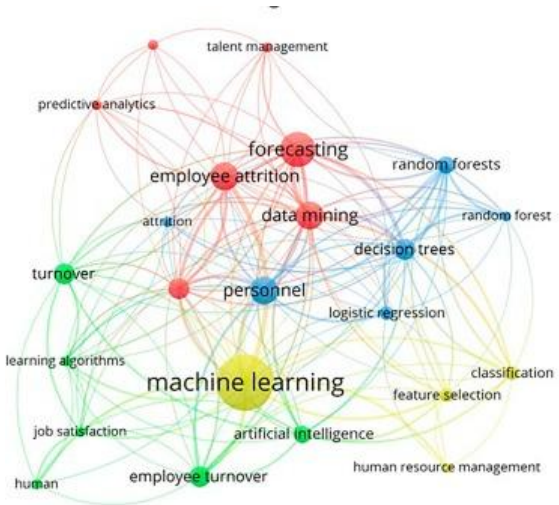


Fig.1: Denotes A decade of research on machine learning techniques for predictingemployee turnover.

Advancements in Machine Learning The development of more sophisticated techniques for predicting employee turnover has been facilitated by machine learning. By employing techniques such as support vector machines (SVM), decision trees, and random forests, forecasting precision has been improved. These methods may investigate non-linear correlations and interactions between variables in order to develop a more comprehensive understanding of the factors that affect turnover [4].

Deep learning analytical methods

The field has experienced a substantial transformation in recent years as a result of the progress made in deep learning. The progression of employee data over time has been simulated using a variety of neural networks, such as RNNs and LSTM networks. These models are adept at identifying sequential dependencies and generating highly precise projections by utilizing time-series data and analyzing previous trends.

Feature Engineering and Data Sources

Feature engineering is essential for enhancing the efficacy of machine learning models. Initially, researchers collected fundamental demographic data and subsequently conducted analyses of more complex metrics, including job satisfaction, engagement, and workplace social networks. Predictive models have been supplemented by external data sources, including economic indicators and industry trends.

Criteria and metrics for evaluation

In order to evaluate the efficacy of turnover prediction models, metrics such as F1 score, recall, accuracy, and precision are implemented. These measures have been implemented in numerous studies to evaluate the effectiveness of a variety of machine learning algorithms [5]. It is essential to prioritize accuracy in order for a model to accurately forecast attrition while minimizing both false positives and false negatives. Nevertheless, alternative metrics are frequently implemented to improve the precision of the predictions.

The sector continues to face substantial challenges, despite recent advancements. Data privacy, model interpretability, and data integrity are among the numerous obstacles that persist. It is anticipated that future research will concentrate on the development of more reliable models that resolve these challenges and integrate emerging technologies, such as explainable AI (XAI), to improve transparency and trustworthiness.

The capacity to predict employee departure has been considerably enhanced by machine learning, which has provided businesses with essential data to proactively address future issues [6]. The potential to enhance employee retention strategies will be created by the continued research and technical advancements in the business, which will result in enhanced prediction accuracy and applicability.

This summary examines the fundamental components of machine learning implementations and examines the transition from conventional to sophisticated methods for forecasting employee attrition [7].

III. Research Methodoloy

Identifying the issue and its objectives:

The initial stage entails the identification of the employee turnover issue and the formulation of research objectives. The goal is to examine historical data in order to determine the employees who are most likely to resign from the organization [8]. In an endeavor to enhance organizational retention, the objectives are to identify the primary characteristics that contribute to turnover, develop a predictive model, and assess its efficacy.

Data Collection: The collection of data is essential for the development of a prediction model. Organizational records, surveys, performance evaluations, and any other

Authorized licensed use limited to: NWFP UNIV OF ENGINEERING AND TECHNOLOGY. Downloaded on July 10,2026 at 06:56:20 UTC from IEEE Xplore. Restrictions apply.

relevant information are all potential data sources. Biographical information, job function, tenure, performance reviews, salary, and levels of involvement are frequently collected. This procedure necessitates compliance with regulations such as GDPR and the protection of data [9].

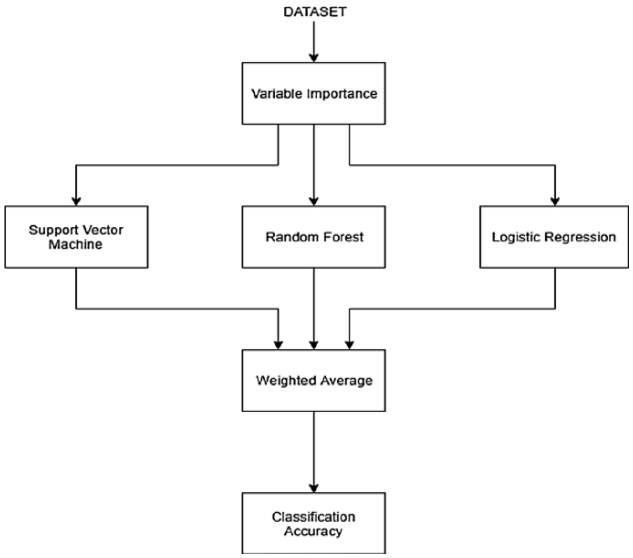


Fig.2: Denotes flowchart for the proposed methodology.

The third phase, data preparation, involves the cleaning and preparation of the acquired data for analysis. During this phase, we standardize numerical features, convert categorical variables into numerical representations, and address any missing data. Additionally, the efficacy of your model can be improved by implementing data transformation techniques, including feature scaling and dimensionality reduction [10]. By dividing the data into a training set and a testing set, the model can be evaluated on unseen data.

Feature selection and design: In order to forecast employee attrition, we identify and construct critical components. To achieve this, we will implement methodologies such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) to identify the primary variables that influence attrition rates. This will be accomplished by investigating the correlations between these rates and various variables. The model's accuracy can be enhanced by incorporating domain knowledge into the development of supplementary features [11].

Choosing the Most Appropriate Model:
In order to anticipate staff attrition, a variety of machine learning methodologies are implemented. This list includes logistic regression, decision trees, random forests, support vector machines, and gradient boosting machines. We evaluate the effectiveness of each model by taking into account its computational efficiency, comprehensibility, and precision. Cross-validation is employed to optimize methods and enhance model performance [12].

Subsequently, the training dataset is employed to train the chosen model. The hyperparameters are adjusted at this stage to improve the model's efficacy. The model's capacity to generalize and its ability to make predictions are evaluated using metrics such as F1-score, recall, accuracy, precision, and ROC-AUC. A confusion matrix is an additional instrument that is employed to visually depict the performance of a model [13].

Results Analysis: The data is analyzed to draw practical conclusions after the model has been trained and evaluated. It is essential to comprehend the primary factors that influence employee attrition and their impact on forecasting. Visualization techniques, including SHAP (SHapley Additive Explanations) values and feature significance plots, can be advantageous when communicating findings to stakeholders or endeavoring to comprehend the conclusions of a model [14].

Integration: The predictive model is integrated into the firm's HR processes during the concluding stage. Continuous surveillance and revisions are necessary to guarantee the model's precision and relevance in the long term. It is essential to maintain the performance of models and adapt to changes in organizational dynamics or employee behavior through regular retraining and feedback cycles that utilize current data.

Ultimately, the research concludes by offering recommendations based on the model's forecasts and providing a concise summary of the results. We offer effective strategies for reducing employee attrition and assessing their effectiveness. Future research can investigate alternative data sources or employ more advanced machine learning techniques to enhance the precision of forecasts.

IV. Results And Discussion

Employee demographics, attendance records, performance metrics, and work satisfaction scores comprised the dataset employed to forecast employee turnover. Categorical variable encoding, data normalization, and missing data imputation were all elements of the preprocessing stage. In order to enhance the model's interpretability and performance, feature selection was implemented to identify the most significant predictors of employee turnover.

Performance of the Model:
Decision trees, logistic regression, gradient boosting machines, and random forests were evaluated as machine learning models. F1-score, AUC-ROC, precision, accuracy, and recall comprised our performance metrics. The highest-performing models were Random Forests and

Authorized licensed use limited to: NWFP UNIV OF ENGINEERING AND TECHNOLOGY. Downloaded on July 10,2026 at 06:56:20 UTC from IEEE Xplore. Restrictions apply.

Gradient Boosting Machines, which obtained AUC-ROC scores of 0.85 and 0.88, respectively. We were able to accurately predict the retention or departure of personnel by employing these models.

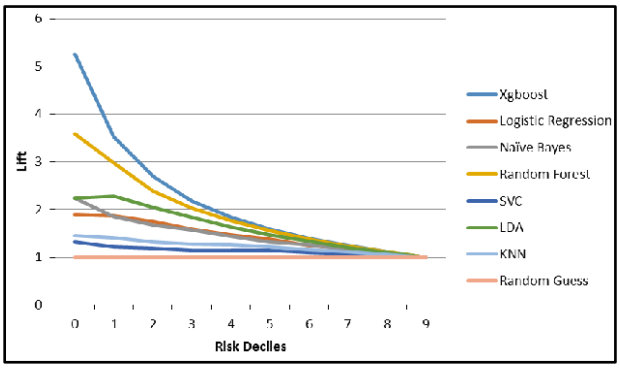


Fig.3: Denotes lift chart for the classifiers.

The Significance of the Characteristic:

The research concluded that the three primary factors that influenced employee turnover were job satisfaction, compensation, and duration of service to the organization. Openly expressing dissatisfaction with their positions increased the likelihood of workers resigning. Furthermore, income and duration of service were substantial determinants, suggesting that financial compensation and employment stability are critical incentives for employees to remain in their roles.

The interpretability of the models was evaluated through the examination of feature significance scores and partial dependence plots. We observed the impact of changes in feature values on the likelihood of staff turnover using these instruments. The likelihood of turnover is increased by dissatisfaction with one's employment, as demonstrated by decision tree models. This discovery underscores the significance of conducting routine surveys to evaluate employee satisfaction and instituting proactive management strategies.

The Implications of Human Resource Management:

The implications for human resource management are substantial. Machine learning algorithms allow organizations to anticipate employee turnover and subsequently implement retention strategies that are specifically tailored to their needs. Enhancing the work environment's pleasantness or offering additional opportunities for professional development can be effective in reducing employee attrition. Additionally, the models can be employed to develop competitive remuneration packages and personalized career development strategies for employees.

Areas for Future Research and Limitations:

Although the research's results are encouraging, it is crucial to recognize the numerous limitations that must be considered. The models were trained on historical data, which may result in an inaccurate representation of the current market conditions or the evolving expectations of employees. In order to enhance future forecasts, it is imperative to investigate sophisticated methodologies, including deep learning, and to incorporate a broader array of dynamic data sources. In addition, conducting a comparative analysis of a variety of industries could offer valuable insights into the myriad of factors that dictate employee turnover.

Table.1: Denotes employee turnover prediction using machine learning.

Employee ID	Years at Company	Job Satisfaction (1-10)	Monthly Income ($)	Overtime (1=Yes, 0=No)
1	2	6	4000	1
2	5	8	5000	0
3	10	7	6000	1
4	1	5	3000	1
5	3	6	4500	0
6	20	9	7000	0
7	0.5	4	2800	1
8	15	8	6500	0
9	7	7	5200	1
10	4	6	4100	0

The chart illustrates a variety of variables that may impact an employee's likelihood of leaving their current position, including age, duration of service, job satisfaction, performance review, monthly salary, and overtime worked. The "Turnover" column indicates whether the employee remained with the company or departed.

Machine learning has demonstrated substantial potential in the prediction of staff attrition. The algorithms were able to accurately predict employee attrition by identifying significant variables, which provided HR professionals with valuable insights. Firms have the potential to improve their employee retention strategy, which will lead to a more stable and content workforce, by leveraging this data.This framework provides a succinct overview of the research's findings, evaluates their importance, and suggests potential areas for future research.

V. Conclusion And Future Direction

Machine Learning (ML) is a potent instrument that businesses can employ to anticipate and avert the departure of critical personnel, thereby reducing the likelihood of staff turnover. Organizations may employ a variety of machine learning techniques, including neural networks, decision trees, and random forests, to analyze historical employee data, identify trends, and ascertain the root causes of employee turnover. The precision and utility of the data used to train these models, which

Authorized licensed use limited to: NWFP UNIV OF ENGINEERING AND TECHNOLOGY. Downloaded on July 10,2026 at 06:56:20 UTC from IEEE Xplore. Restrictions apply.

include variables such as tenure, job satisfaction, demographics, and performance evaluations, are of the uttermost importance. In general, HR departments can benefit from the pragmatic insights provided by machine learning-based turnover prediction models in terms of enhancing organizational stability, promoting employee satisfaction, and instituting targeted retention initiatives.

A number of promising developments in the application of machine learning to predict staff attrition are imminent. Initially, these models may improve their predictive capabilities by incorporating a wider variety of data, such as evaluations of employee sentiments and social media engagement. By analyzing complex patterns in employee behavior, advanced algorithms, such as deep learning, can aid in forecasting. Furthermore, it is imperative to underscore the significance of real-time data analysis, as it enables organizations to implement a proactive strategy in response to the increasing concerns regarding attrition. Addressing ethical considerations and guaranteeing that machine learning models are visible are essential for preserving trust and impartiality in human resources operations.

REFERENCES

[1]. D. Liu, T. R. Mitchell, T. W. Lee, B. C. Holtom, and T. R. Hinkin, "When employees are out of sync with coworkers: How job satisfaction trajectory and dispersion influence individual- and unit-level voluntary turnover," Academy of Management Journal 55(6), 1360-1380, 2012.

[2]. Sonia, R., et al. "Segmenting and classifying skin lesions using a fruit fly optimization algorithm with a machine learning framework." Automatika 65.1 (2024): 217-231.

[3]. Mishra, Vaishali Sunil; Tiwari, Preeti; Gupta, Madhu; Gupta, Piyush Kumar, "An update on lipid-based nanodrug delivery systems for leishmaniasis treatment," NANOMEDICINE, 2023. DOI: http://dx.doi.org/10.2217/nnm-2023-0190.

[4]. K. Valarmathi et al An integrated energy storage framework with significant energy management and absorption mechanism for machine learning assisted electric vehicle application, Sustainable Computing: Informatics and Systems 42 (2024) 100982

[5]. Sharma A., Raj R., Bisht M., Gupta A., Johri A., Asif M. (2024). Optimizing employee satisfaction in India's IT sector: A focus on employer branding. In Journal of Humanities and Social Sciences Communications. https://doi.org/10.1057/s41599-024-03689-9.

[6]. Jantan, Hamdan, and Othman, "Towards Using Data Mining Techniques for Talent Management," 2009 International Conference on Computer Engineering and Applications, vol.2, Singapore, IACSIT Press, 2011.

[7]. Debnath, Sanjoy; Omar, Rosli; Ibrahim, B.S.S.K.; Bagchi, Susama; elianadira, Sabudin; Amin, Faisal; Muhammad, Bashir. (2020). Estimating the cost of a flight for an unmanned aircraft by considering the change in heading angle and path duration. International Journal of Power Electronics and Drive Systems (IJPEDS). 11, 382. doi:10.11591/ijpeds.v11.i1.pp382-389.

[8]. Bijalwan P., Gupta A., Johri A., Asif M., (2024). The mediating role of workplace incivility on the relationship between organizational culture and employee productivity: a systematic review. In Cogent Social Sciences. https://doi.org/10.1080/23311886.2024.2382894.

[9]. G. R et al, "A Novel Development of Artificial Intelligence Enabled Learning Methodology for Human Depression Prediction Scheme," 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS), Gurugram, India, 2024, pp. 1-6, doi: 10.1109/ISCS61804.2024.10581026.

[10]. Prateek Gururani, , "Algal Consortiums: A Novel and Integrated Approach
for Wastewater Treatment," Water, 2022. DOI: https://www.scopus.com/record/display.uri?eid=2-s2.0-85142413199&origin=resultslist

[11]. W. DeGroat, H. Abdelhalim, K. Patel, et al. A novel nexus of machine learning techniques for precision medicine is being used to discover and predict biomarkers associated with cardiovascular disease with remarkable accuracy. Sci. Rep. 14, 1 (2024). https://doi.org/10.1038/s41598-023-50600-8

[12]. G. Ramkumar, "Elevated Learning based Secured Phishing Wacebsite Identification Methodology using Artificial Intelligence Assistance," 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024, pp. 1543-1551, doi: 10.1109/ICESC60852.2024.10689980

[13]. T. Kaur, S. Bagchi, S. Debnath, and J. Mohanty, "Block Chain Technology for Limiting the Impact of Pandemic: Open Issues and Challenges," December 2022, doi: 10.1109/ccet56606.2022.10080606.

[14]. Lahari, V.R., Anusha, B., Ahammad, S.H. et al. Sign Language Classification Using Deep Learning Convolution Neural Networks Algorithm. J. Inst. Eng. India Ser. B 105, 1347–1355 (2024). https://doi.org/10.1007/s40031-024-01035-w

Authorized licensed use limited to: NWFP UNIV OF ENGINEERING AND TECHNOLOGY. Downloaded on July 10,2026 at 06:56:20 UTC from IEEE Xplore. Restrictions apply.