



Predicting Employee Attrition in Healthcare Using AI Models, Deep Learning and Natural Language Processing

Kiran Veernapu

USA

ABSTRACT

Employee attrition, or turnover, is a significant concern in many industries, particularly in healthcare, where retaining skilled professionals is critical for maintaining high-quality patient care. Healthcare organizations face unique challenges, including high stress, burnout, and work-life balance issues, which can contribute to higher employee turnover rates. Predicting employee attrition can help organizations implement targeted interventions to retain employees, reduce turnover costs, and ensure a stable workforce. This paper explores the use of Artificial Intelligence (AI) and Machine Learning (ML) models to predict employee attrition in the healthcare sector, focusing on applying various AI models, their effectiveness, and the key factors influencing attrition.

ARTICLE HISTORY

Received December 08, 2022
Accepted December 15, 2022
Published December 22, 2022

KEYWORDS

Employee Turnover, Retention, Predicting Employee Turnover, AI, ML, Deep Learning, Voice to Text Analytics, Exit Interviews

Keywords: Employee Turnover, Retention, Predicting Employee Turnover, AI, ML, Deep Learning, Voice to Text Analytics, Exit Interviews

Introduction

Employee attrition is an ongoing challenge in healthcare settings, especially given the high-stress environment, emotional toll, and demanding hours that characterize many healthcare professions. Turnover not only incurs direct costs related to recruitment and training but can also negatively impact patient care and organizational efficiency [1]. The direct cost of turnover includes advertising costs, recruiting selection, ads, background checks, training, and orientation. The indirect costs include the cost of paying overtime to staff replacement labor like temporary workers, the training period for the interim staff, lower productivity on the new hires, and the cost of business opportunity lost due to the non-availability of a skilled healthcare professional. As the employee's expertise and education levels increase, the employee becomes more profitable to the organization. Predicting attrition is essential for early intervention, enabling healthcare administrators to design retention strategies.

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have opened new avenues for predicting and mitigating attrition risks. AI-driven models are capable of identifying patterns in employee behavior and work conditions that contribute to attrition. By leveraging these technologies, healthcare organizations can gain actionable insights that can help optimize workforce management and reduce turnover [2].

Employee Attrition in Healthcare: Challenges and Factors

Understanding the factors contributing to employee attrition in healthcare is crucial for designing effective predictive models. According to Albougami et al, the common reasons for nurses resigning from their jobs are job satisfaction and quality of life. However, healthcare systems must consider all other possible reasons for healthcare professionals to leave their jobs [3]. Some

of the key factors include

- **Job Satisfaction:** Low job satisfaction, whether from poor management, lack of career development opportunities, or inadequate compensation, can increase the likelihood of attrition.
- **Workplace Stress and Burnout:** Healthcare workers, particularly nurses and doctors, often experience high levels of stress due to long working hours, patient demands, and emotional exhaustion.

Albougami et al conducted a survey to understand the reason why healthcare professionals leave their jobs in Saudi Arabia, one of the reasons they figured out from the survey is stress level at the workplace. With a set of questions that leads to the evaluation of what is considered to be stressful in daily work life. It was evident that the survey demonstrated that healthcare employees express high-stress levels [3].

Burnout is a physical and psychological response to constant levels of work-related stress [4]. Out of all professions surgical nurses and other nurses were found to be experiencing higher levels of burnout. The result of burnout in nurses' results in non-emphatical towards patients when they need care and empathy. Albougami et al describe the analysis of 13 surveys conducted to study the reasons for burnout are unsatisfactory work conditions, high workload, and inadequate time for completing a task. Burnout has become one of the most significant contributors to employee turnover in healthcare.

- **Work-Life Balance:** Healthcare professionals often struggle to maintain a healthy work-life balance, which may lead to dissatisfaction and eventual turnover. Work-life balance can denote a person's provision to be able to plan professional activities and not interfere with personal life activities, as well as the amount of time a person spends on work-related activities at home outside work hours.

Contact: Kiran Veernapu, USA.

Akhila Rao & Shailashri did an elaborative illustration of various factors and reasons for work-life balance and the relationship with a nurse leaving a healthcare job [5]. The research findings show that there is a relationship between age and work-life balance, weekend work, long working hours, less leisure time, little time for family, poor remuneration, etc. have a direct impact on job and work-life balance.

• **Organizational Culture and Support:** The work environment and support from colleagues and management can play a significant role in whether an employee chooses to stay or leave.

Mesfin et al conducted a study with a group of healthcare workers in Ethiopia, with dependent and independent variables of job and the elements of organizational behavior. The dependent variables are the physical working environment, relation to the supervisor, communication, and work experience [6]. The independent variables are cultural score, innovation, market culture, and hierarchy culture. The results of the study show that an innovative organizational culture is more preferred that welcomes new ideas, and new energy rather than a hierarchical culture.

• **Career Advancement Opportunities:** Employees are more likely to stay if they see clear career advancement opportunities within the organization. Conducted a study on the factors that can influence career advancement in the jobs of nurses in healthcare. The survey was conducted with variables such as Age, gender, educational qualification, years of experience, department of work, marital status, etc. The statistical analysis shows that the nurse population was looking for training programs, and support for certifications which help advancement in their career [7].

Understanding these factors is critical to developing predictive AI models that can identify at-risk employees based on historical data and contextual attributes.

Methodology Review

The methodologies used to predict employee turnover typically involve focusing on data collection, feature selection, model selection, and evaluation metrics.

Data Collection Attributes

To build effective predictive models, it is essential to gather comprehensive data on employee characteristics and organizational factors. The data should include

- **Demographic Information:** Age, gender, experience level, educational background.
- **Workplace Factors:** Department, tenure, job role, salary, and work hours.
- **Employee Feedback:** Job satisfaction, workplace culture, perceived stress levels, and burnout scores (often obtained from surveys).
- **Performance Data:** Metrics such as patient outcomes, peer reviews, and performance appraisals.
- **Attrition History:** Historical data on employee turnover, including the timing and reasons for departure.

Feature Selection

Selecting the most relevant features is key to model performance. Techniques such as correlation analysis, feature importance ranking, and Principal Component Analysis (PCA) are typically employed to identify the key predictors of attrition. Some possible features might include

- Tenure and length of service
- Job satisfaction scores
- Frequency of overtime or night shifts
- Age and work-life balance scores
- Availability of career advancement opportunities

AI and Machine Learning Models

Several AI and ML models can be used to predict employee attrition in healthcare settings. Common models include

• **Logistic Regression:** Logistic regression is a classification approach that use linear discriminants. The main objective is to find out the likelihood of an input value contributes to a corresponding to a specific class [9]. The model draws a linear border a linear line separating the inputs into two groups based on the probability value. This model is used in lot of fields including neuroscience.

G. Pratibha and N. P. Hegde, proposed a new formula to address a minor prediction error with the regular equation for logistic regression. In a regular equation the employee left the organization with a positive class and the employees still working for the company are assigned with a negative class [9].

• **Random Forests:** A powerful ensemble learning model that aggregates predictions from multiple decision trees. It is effective for capturing complex, nonlinear relationships in data and is particularly useful for handling large datasets with many variables.

• **Support Vector Machines (SVM):** SVM models are effective for classification tasks, especially when there is a clear margin of separation between different classes (e.g., employees who stay vs. employees who leave). SVM is a supervised learning method that analyzes data recognizes patterns for classification, SVM works to find a hyperplane from each data point which is maximum decision surface. M. A. Widyanda and I. Palupi discussed about different variations of SVM algorithms to bring in more clarity in classification of data [10].

• **Gradient Boosting Machines (GBM):** GBMs are highly effective for predictive modeling as they combine the results of multiple weak models to improve accuracy.

• **Neural Networks:** Deep learning models, including neural networks, can be used to uncover intricate patterns in large datasets, though they may require significant computational resources.

Long short-term memory (LSTM) is a special recurrent neural network (RNN) algorithm which is an improvement of RNN to solve the problems of gradient disappearance and gradient explosion in the process of long sequence training [11].

• **Voice Analytics from Exit Interviews:** The advancement in Artificial Intelligence (AI) and Natural Language Processing (NLP) has advanced to know humans search for information, provide feedback through text and voice. Many healthcare organizations suffer due to nurse staffing issues or nurses leaving the organization due to several reasons. There is no systematic feedback mechanism to understand the reason behind them leaving the job. The formal exit interviews with HR are not recorded for privacy. However, the promoting the AI and NLP based intelligence and extracting the important pointers for the reasons for leaving the job help the HR team streamline and optimize the HR program to keep the skilled people on the job. Voce to text analytics and voice analyzer are the advanced

AI models to help with employee turnover through their smart analytics [12].

Khanbhai et al explained the importance of ML and NLP in health informatics to understand the analyze the voice messages and predict the outcome with patients as well as employee. These models can analyze the emotional outcome of a person. [13].

Evaluation Metrics

Machine learning algorithms are advancing in every domain, they are much more adoption and utilization of AI in healthcare, and HR teams can benefit implementing the AI technology in determining the reason for employee turnover. The scientific analysis process of evaluating the employee turnover is to collect the historical data of the employees leaving the organization, clean the data, prepare the data and identify the metrics that impact the turnover of the employee [14].

Fallucchi et al described the process adopted to use AI models in predicting the employee attrition. Their descriptive analysis included the importance of the evaluation metrics that could impact the employee leaving the organization, the analysis was the comparison of the evaluation metric or characteristic of the reasons to the target variable “attrition”. The modeling used different machine learning algorithms to evaluate the reason the performance of the algorithms. The results indicated that key indicators that prompt the employee leaving and some key indicators of the probability of the employee can leave.

Below are few metrics are used

- **Accuracy:** The proportion of correct predictions out of total predictions.
- **Precision and Recall:** Precision measures the proportion of true positive predictions among all positive predictions, while recall measures the proportion of actual positive cases correctly identified.
- **F1 Score:** A harmonic mean of precision and recall, providing a balance between the two.
- **Area Under the ROC Curve (AUC-ROC):** A graphical representation of the trade-off between sensitivity and specificity, helping to evaluate the model's ability to distinguish between classes.

Results and Discussion

Jain, P.K., Jain, M. & Pamula, R did a comparative analysis of the evaluation metrics to see how each variable is impacting the other variable in the overall analysis. Using SVM classification model the results seem to be biased to classify predicting the major class which is the reason why employee not leaving the organization [15]. They used other models to compare the with same sample data they found other models demonstrate a better performance results 99% close to the true cases of employees who left the organization.

S. Yadav, A. Jain and D. Singh used a collection of data to evaluate the models. They used logistic regression, SVM, Random Forest, Decision Tree, and AdaBoost models. Except Logistic Regression they observed other models achieved more than 93% of accuracy and precession [16]. There are others who evaluated the employee attrition using the similar methods and have seen different results [17,18]. The key is to have the model's sufficient realistic data to train the machine learning algorithms. Post model training the predictions show more than 90% of accuracy.

With all these evaluations the research shows that employees who reported high levels of workplace stress and low job satisfaction were significantly more likely to leave. Employees with fewer career advancement opportunities or those in departments with higher turnover rates were also at increased risk of attrition. Age and tenure had a moderate correlation with attrition, with younger employees and those with shorter tenures being more likely to leave. These findings align with existing research on factors affecting employee retention in healthcare and provide a solid foundation for developing targeted retention strategies.

Conclusion

AI and machine learning models offer a powerful tool for predicting employee attrition in healthcare settings. By leveraging historical data and key performance indicators, healthcare organizations can identify at-risk employees early and take proactive steps to reduce turnover. Models like Random Forests and Gradient Boosting Machines have proven to be effective in this domain, providing high accuracy and actionable insights. Future work could focus on integrating real-time data (such as employee satisfaction surveys and feedback) and exploring the use of natural language processing to analyze unstructured data, such as open-ended responses.

Ultimately, predictive models can empower healthcare organizations to foster a supportive, engaging work environment, improving employee retention and patient care outcomes.

References

- [1] Mahoney CB, Lea J, Schumann PL, Jillson IA. Turnover, burnout, and job satisfaction of certified registered nurse anesthetists in the United States: Role of job characteristics and personality. *AANA Journal*. 2020; 88: 39-48.
- [2] Sisodia DS, Vishwakarma S, Pujahari A. Evaluation of machine learning models for employee churn prediction. 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India. 2017; 1016-1020.
- [3] Albugami AS, Almazan JU, Cruz JP, Alquwez N, Alamri MS, et al. Factors affecting nurses' intention to leave their current jobs in Saudi Arabia. *International journal of health sciences*. 2020; 14: 33-40.
- [4] Phillips, Carswella DNP, APRN, AGPCNP-BC. Relationships between workload perception, burnout, and intent to leave among medical–surgical nurses. *International Journal of Evidence-Based Healthcare*. 2020; 18: 265-273.
- [5] Akhila Rao, Shailashri VT. Work-Life Balance of Women Medical Professionals in the Healthcare Sector-A Systematic Literature Review. *International Journal of Health Sciences and Pharmacy (IJHSP)*. 2021; 5: 54-79.
- [6] Mesfin D, Woldie M, Adamu A, Bekele F. Perceived organizational culture and its relationship with job satisfaction in primary hospitals of Jimma zone and Jimma town administration, correlational study. *BMC Health Services Research*. 2020; 20: 1-9.
- [7] Mohamed R, Abd-El-Aal N, Atalla A. Relationship between Career Planning and Career Advancement as Perceived by Professional nurses. *Int J Novel Res Healthc Nurs*. 2021; 8: 353-370.
- [8] BOOTS RC. Comparing The Prediction Power Of Synthetic Data With Practical Data On Employee Turnover (Doctoral dissertation, Tilburg University). Available at: <https://arno.uvt.nl/show.cgi?fid=1>.

- [9] Pratibha G, Hegde NP. HR Analytics: Early Prediction of Employee Attrition using KPCA and Adaptive K-means based Logistic Regression. *Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS)*, Chennai, India. 2022; 11-16.
- [10] Widyananda MA, Palupi I. Implementation of the Spiral Optimization Algorithm in the Support Vector Machine (SVM) Classification Method (Case Study: Diabetes Prediction). *International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Bali, Indonesia. 2021; 1-6.
- [11] Wu C. Real Time Attrition Prediction Mechanism Based on Deep Learning. *2nd International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, ON, Canada. 2022; 128-131.
- [12] Christian Hildebrand, Fotis Efthymiou, Francesc Busquet, William H Hampton, Donna L. Hoffman, et al. Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications. *Journal of Business Research*. 2020; 121: 364-374.
- [13] Rajendran S, Srinivas S, Pagel E. Mining voice of customers and employees in insurance companies from online reviews: a text analytics approach. *Benchmarking: An International Journal*. 2020; 30: 1-22.
- [14] Fallucchi F, Coladangelo M, Giuliano R, William De Luca E. Predicting Employee Attrition Using Machine Learning Techniques. *Computers*. 2020; 9: 86.
- [15] Jain, P.K., Jain, M. & Pamula, R. Explaining and predicting employees' attrition: a machine learning approach. *SN Appl. Sci.* 2, 757 (2020). <https://doi.org/10.1007/s42452-020-2519-4>
- [16] Yadav S, Jain A, Singh D. Early Prediction of Employee Attrition using Data Mining Techniques. *2018 IEEE 8th International Advance Computing Conference (IACC)*, Greater Noida, India. 2018; 349-354.
- [17] Bhartiya N, Jannu S, Shukla P, Chapaneri R. Employee Attrition Prediction Using Classification Models. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, Bombay, India. 2019; 1-6.
- [18] Alduayj SS, Rajpoot K. Predicting Employee Attrition using Machine Learning. *2018 International Conference on Innovations in Information Technology (IIT)*, Al Ain, United Arab Emirates. 2018; 93-98.