

Chapter 4

HR Analytics and Employee Attrition Prediction Using Machine Learning

N. Krishnamoorthy

Vellore Institute of Technology, India

V. Vinoth Kumar

Vellore Institute of Technology, India

Chinchu Nair

*Dr. MGR Educational and Research
Institute, India*

A. Maheswari

 <http://orcid.org/0009-0002-2578-5265>

*Dr. MGR Educational and Research
Institute, India*

Sonali Mishra

Vellore Institute of Technology, India

Ayush Sinha

Vellore Institute of Technology, India

ABSTRACT

The high rate of staff turnover in knowledge-based organizations is a key challenge. Invaluable tacit information, which is frequently the source of a company's competitive advantage, is often taken with departing personnel. If a company wants to keep a better competitive edge over its competitors, it should prioritize lowering staff churn. By forecasting attrition based on demographic and job-related factors, this study finds employee traits that aid in predicting employee turnover in organizations. With an IBM dataset of 14,999 samples and 10 features, up-sampling techniques and various machine learning algorithms are employed to find the best predictive model. Data visualization and analysis reveal significant factors and correlations. Further, the authors have used models to predict and analyze employee attrition and turnover. Using classifiers like k-nearest neighbors (KNN), support vector machine (SVM), decision trees (DT), and random forest classifiers (RF), four main tests were run on the IBM dataset to predict employee attrition.

I. INTRODUCTION

The pace at which employees depart from an organization is referred to as employee attrition, also known as turnover. It can be caused by various factors such as job dissatisfaction, lack of advancement opportunities, poor leadership, or better job offers elsewhere. High attrition rates can be costly for businesses due to the expenses associated with recruiting and training new employees (Kumar, D et al., 2021). Employee attrition prediction involves using data analysis and predictive modeling techniques to estimate the likelihood of employees leaving an organization (Basheer, S at al., 2020). By analyzing various data sources, organizations can identify patterns and factors contributing to attrition and implement proactive strategies to retain valuable employees. This can lead to higher retention, employee satisfaction, and cost savings for the company (Maithili, K at al., 2018).

II. LITERATURE SURVEY

(Joseph, R et al., 2021) proposed a model taking into account a variety of factors and has a prediction accuracy of 86.0% (Random Forest Classifier) for attrition and mental health for a specific employee. (Rajeswari G. R et al., 2022) proposed four machine learning models by getting the best accuracy of 90% in logistic regression model. (Yedida, R et al., 2018) proposed a method in which the impact of voluntary attrition on organizations was discussed, along with the significance of foreseeing it with an accuracy of 94.32%. (Jain R., & Nayyar A, 2018) constructed a ML model utilizing the XGBoost strategy with the accuracy of 89%. On the IBM USA dataset, (Bhuva, K., & Srivastava, K, 2018) applied a variety of classification algorithms, including logistic regression model, DT, LDA, ridge classification, and RF, to make predictions about the departure of employees for a variety of reasons. The purpose of (Fallucchi, F et al., 2020) was to better anticipate employee attrition by identifying the contributing components and root reasons. With an accuracy of 82.5%, the Naive Bayes model was the highest accurate model. (George, S et al, 2022) compared the performance of four ML models, ETC, SVM, Logistic regression, and DTC, for predicting employee attrition When the dataset was multiplied by ten times, the SVM, LR, DTC, and ETC achieved accuracy scores of 88%, 74%, 84%, and 93%, respectively. (Mohbey, K. K, 2020, pp. 121-128) using the HR dataset proposed the effectiveness of several machine learning algorithms is assessed in this study. For this dataset, logistic regression worked well based on the accuracy measurement. Using models like SVM, Random forest, and decision tree, (Yadav, S., Jain, A., & Singh, D, 2018) provided a model that concluded that salary and additional financial considerations, such as promotions, are not among the primary causes of employee attrition. (Jain, P. K., Jain, M., & Pamula, R, 2020) and (Krishnamoorthy, N & Parameswari, V. L, 2018) proposed a method in which HR data has been subjected to predictive analysis approaches. Predictive models like SVM, DT, and RF were then applied to the cleaned data to determine best model for the dataset.

III. PROBLEM DESCRIPTION

Employee attrition prediction refers to the process of analyzing data related to employees in an organization to identify those who are likely to leave their jobs voluntarily or involuntarily. This problem is important for organizations to address, as high levels of employee turnover can lead to reduced

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/hr-analytics-and-employee-attrition-prediction-using-machine-learning/351259

Related Content

AI-Powered Teachers: The Role of Educators in a Data-Driven World

Pawan Kumar Goel (2026). *Transforming Education With Data Science in the AI Era* (pp. 191-210).

www.irma-international.org/chapter/ai-powered-teachers/389378

Vision Conflict Scale: Instrument Profile

J. Louis Spencer and W. David Winner (2021). *Handbook of Research on Advancements in Organizational Data Collection and Measurements: Strategies for Addressing Attitudes, Beliefs, and Behaviors* (pp. 191-209).

www.irma-international.org/chapter/vision-conflict-scale/285197

Boosting of Deep Convolutional Architectures for Arabic Handwriting Recognition

Mohamed Elleuch and Monji Kherallah (2019). *International Journal of Multimedia Data Engineering and Management* (pp. 26-45).

www.irma-international.org/article/boosting-of-deep-convolutional-architectures-for-arabic-handwriting-recognition/245262

An Automatic Video Reinforcing System for TV Programs using Semantic Metadata from Closed Captions

Yuanyuan Wang, Daisuke Kitayama, Yukiko Kawai, Kazutoshi Sumiya and Yoshiharu Ishikawa (2016). *International Journal of Multimedia Data Engineering and Management* (pp. 1-21).

www.irma-international.org/article/an-automatic-video-reinforcing-system-for-tv-programs-using-semantic-metadata-from-closed-captions/149229

The Role of Renewable Energy Consumption in Promoting Sustainability and Circular Economy: A Data-Driven Analysis

Lucio Laureti, Alberto Costantiello, Alessandro Massaro and Angelo Leogrande (2024). *Data-Driven Intelligent Business Sustainability* (pp. 360-386).

www.irma-international.org/chapter/the-role-of-renewable-energy-consumption-in-promoting-sustainability-and-circular-economy/334755