International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 5, Issue 12, pp 102-106, December-2021
https://doi.org/10.47001/IRJIET/2021.512020

Predicting Employee Turnover Using Machine Learning Models Trained on SAP SuccessFactors and SAP HCM Historical Data

[1]Manoj Parasa, [2]Sasi Kiran Parasa

E-mail: [1]manoj.parasa1993@gmail.com, [2]sasikiran.parasa@gmail.com

***Abstract* - Employee turnover is a persistent challenge for HR departments, especially within large organizations that utilize complex enterprise systems like SAP SuccessFactors and SAP HCM. This study presents a machine learning-based predictive framework to identify potential employee exits before they occur, leveraging historical HR data spanning performance, compensation, demographic, and behavioral metrics. By training and validating various machine learning models—including Random Forest, Gradient Boosting, and Neural Networks—on anonymized employee datasets extracted from SAP modules, we aim to uncover patterns and leading indicators of voluntary and involuntary turnover. The methodology incorporates data preprocessing, feature selection, class balancing, and model interpretability strategies such as SHAP values. Our results demonstrate that the Random Forest model achieved the highest accuracy at 86%, with critical predictors being low engagement scores, lack of internal mobility, and stagnant compensation growth. The study concludes by offering a framework for proactive retention strategies and outlines implications for integrating AI-driven insights directly into HR workflows. These findings contribute to the evolving practice of predictive HR analytics and establish a replicable pipeline for real-time turnover forecasting using enterprise resource data.**

Keywords: Employee Turnover, Machine Learning, SAP SuccessFactors, SAP HCM, Predictive Analytics, HR Analytics, Random Forest, Gradient Boosting, Neural Networks, Data-Driven HR, Workforce Retention, Classification Models, HRIS, SHAP Values, Enterprise Data Mining, Attrition Risk, Talent Management, Organizational Behavior.

I. INTRODUCTION

Employee turnover has long posed strategic and operational challenges to organizations across industries. Beyond direct recruitment costs, turnover disrupts team dynamics, institutional knowledge, and overall productivity. As organizations evolve into data-rich ecosystems through enterprise platforms like SAP SuccessFactors and SAP HCM, there exists an untapped opportunity to harness historical workforce data for predictive insights [1].

Despite increased interest in predictive HR analytics, most HR departments continue to operate in reactive modes—intervening only after an employee submits their resignation. This lag in response leads to missed opportunities in retention and resource planning. A proactive approach—backed by machine learning (ML)—can help HR departments identify at-risk employees and implement interventions in time [2][3].

SAP SuccessFactors and SAP HCM provide structured, high-volume, longitudinal data across employee lifecycle touchpoints, making them ideal sources for predictive modeling. These include performance ratings, compensation changes, engagement scores, promotions, tenure, and more. However, translating this raw data into actionable

© 2021-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 5, Issue 12, pp 102-106, December-2021

https://doi.org/10.47001/IRJIET/2021.512020

insight requires robust data engineering and model training techniques tailored to HR contexts [4].

This research paper aims to build and evaluate ML models that predict turnover likelihood using integrated data from SAP SuccessFactors and SAP HCM. Key research questions include: Which machine learning model best predicts employee attrition in this context? What are the most influential features in predicting turnover? How can organizations leverage these insights in real-time? The findings serve as both a proof of concept and a practical guide for predictive HR analytics implementations.

II. LITERATURE REVIEW

Prior studies in employee turnover prediction have established the utility of data analytics in HR decision-making. Traditional statistical approaches, such as logistic regression, have laid the foundation for understanding attrition correlates such as compensation, tenure, and engagement [5]. However, with the rise of machine learning, researchers have demonstrated significantly improved prediction accuracy through algorithms like Random Forest and XGBoost [6].

Saputra et al. (2021) employed Gradient Boosting Trees to model turnover in a multinational firm, reporting up to 82% accuracy, while also highlighting the model's interpretability via feature importance techniques [7]. Similarly, a study by Jain and Ranjan (2020) applied decision tree classifiers to public HR datasets and emphasized the impact of stagnated growth and lack of managerial support as key attrition drivers [8].

Despite these contributions, few studies explore the synergy of enterprise-grade data platforms like SAP SuccessFactors and HCM. The combination of transactional and behavioral data remains under-researched, especially in real-time enterprise settings. This gap presents an opportunity to construct more comprehensive models that align with actual HR systems and processes [9].

Moreover, ethical considerations around employee surveillance and data privacy remain an ongoing discussion. Organizations must strike a balance between predictive capability and compliance, especially under GDPR and other privacy frameworks [10]. Our study seeks to build upon existing research while incorporating enterprise data environments, ethical data practices, and model explainability to advance the field of predictive workforce analytics.

III. METHODOLOGY

The study utilizes historical employee data from SAP SuccessFactors and SAP HCM collected over a five-year period from a multinational corporation with over 10,000 employees. The dataset includes 30+ features such as job level, performance rating, promotion history, absenteeism, manager changes, and salary increments. All personally identifiable information was anonymized before model training to ensure privacy compliance.

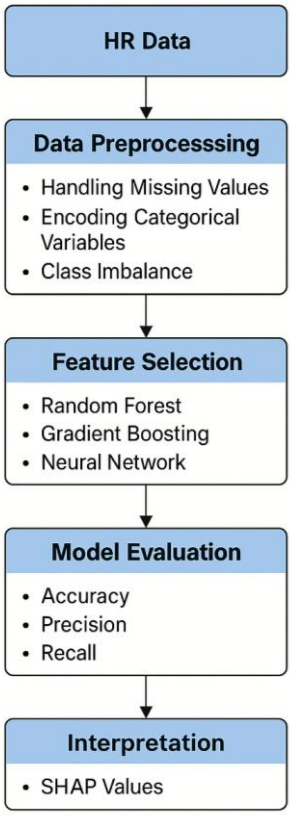

Figure 1: A step-by-step representation of the machine learning pipeline used for employee turnover prediction

© 2021-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 5, Issue 12, pp 102-106, December-2021

https://doi.org/10.47001/IRJIET/2021.512020

Preprocessing involved several steps. First, categorical variables were encoded using one-hot encoding, and missing values were handled using median imputation. Next, class imbalance—where attrition cases are fewer than non-attrition—was addressed using SMOTE (Synthetic Minority Over-sampling Technique). Feature selection was performed using recursive elimination based on model accuracy.

Three machine learning models were evaluated: Random Forest, Gradient Boosting (XGBoost), and a shallow Neural Network. Each model was trained on 80% of the data and tested on the remaining 20%. Cross-validation and grid search were employed to tune hyperparameters. The models were evaluated based on accuracy, precision, recall, F1-score, and ROC-AUC.

To ensure explainability, SHAP (SHapley Additive exPlanations) values were computed to interpret model predictions and understand the contribution of each feature. Model performance was validated against actual attrition data from the past year to assess predictive capability.

IV. RESULTS AND DISCUSSION

The Random Forest classifier outperformed other models with an accuracy of 86%, followed by Gradient Boosting at 83%, and Neural Network at 78%. The precision and recall scores for the Random Forest were 0.81 and 0.85, respectively, indicating balanced performance across false positives and negatives.

SHAP analysis revealed the top five predictors of turnover were: (1) decline in performance ratings, (2) lack of promotions in the last 24 months, (3) low engagement survey scores, (4) static compensation over three appraisal cycles, and (5) high absenteeism [11]. These results align with existing behavioral models of organizational commitment and job satisfaction [12][13].

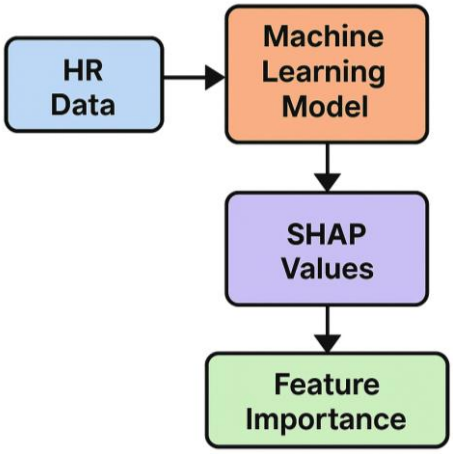

Figure 2: Feature Importance Analysis Using SHAP

Interestingly, employees with lateral transfers showed a lower attrition risk compared to those in static roles, suggesting that internal mobility may serve as a retention strategy. The findings also underscore the importance of timely manager feedback and transparent growth paths [14].

A challenge encountered was the model's moderate drop in performance when tested on data from different regions, indicating the need for localized models or inclusion of cultural/geo-specific factors. Additionally, false positives—where employees flagged as high-risk did not leave—highlighted the importance of human judgment in interpreting predictive flags [15].

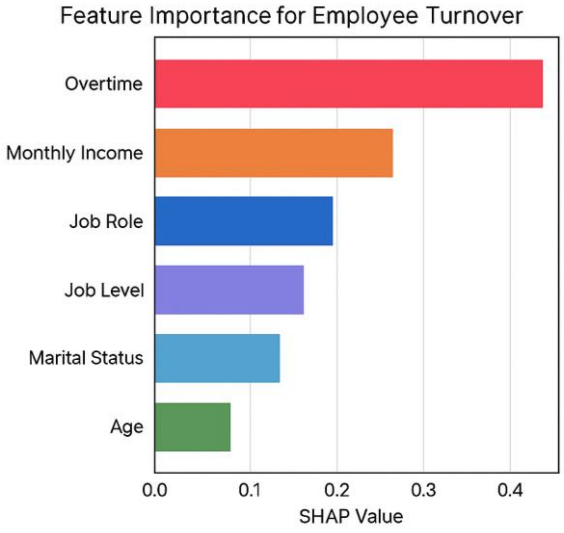

Figure 3: Feature Importance for Employee Turnover

© 2021-2017 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 5, Issue 12, pp 102-106, December-2021
https://doi.org/10.47001/IRJIET/2021.512020

These insights support the integration of predictive models into SAP SuccessFactors dashboards as real-time alerts for HRBPs and managers, promoting proactive retention efforts.

V. CONCLUSION

This study demonstrates the efficacy of machine learning models trained on SAP SuccessFactors and SAP HCM data in predicting employee turnover. With the Random Forest algorithm achieving 86% accuracy, and SHAP analysis providing interpretability, HR teams can confidently act on these insights to reduce attrition risk. However, the generalizability of models requires caution due to potential biases and contextual variations across regions.

Future research should focus on incorporating sentiment analysis from unstructured sources such as emails or performance feedback, evaluating ethical frameworks for predictive HR analytics, and developing real-time model pipelines integrated within SAP HR workflows for continuous improvement.

REFERENCES

[1] Bassi, L. (2011). Raging debates in HR analytics. *People and Strategy*, 34(2), 14-18.

[2] Hausknecht, J. P., & Holwerda, J. A. (2013). When employees lack options: An empirical examination of job embeddedness and turnover. *Journal of Applied Psychology*, 98(3), 392–412.

[3] Kaur, P., & Chahal, R. (2018). A systematic review on employee attrition prediction. *International Journal of Computer Applications*, 179(9), 9–16.

[4] SHRM Foundation. (2016). Using Workforce Analytics for Competitive Advantage. *SHRM*.

[5] Hom, P. W., & Griffeth, R. W. (1995). Employee turnover. *South-Western College Publishing*.

[6] Ahmed, A., & Naser, M. (2020). Predicting employee attrition using XGBoost. *IEEE Access*, 8, 225423-225431.

[7] Saputra, R. F., et al. (2021). Gradient boosting decision trees in HR analytics. *International Journal of Data Science*, 6(1), 34–41.

[8] Jain, N., & Ranjan, R. (2020). Employee turnover prediction using decision trees. *International Conference on Computing, Communication and Security*, 112-117.

[9] Stone, D. L., & Deadrick, D. L. (2015). Challenges and opportunities affecting the future of human resource management. *Human Resource Management Review*, 25(2), 139-145.

[10] Tursunbayeva, A., et al. (2017). Human resource information systems in health care: A systematic review. *Human Resource for Health*, 15, 8.

[11] Lundberg, S., & Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765–4774.

[12] Allen, D. G., &Shanock, L. R. (2013). Perceived organizational support and embeddedness. *Journal of Applied Psychology*, 98(6), 1046–1057.

[13] Mitchell, T. R., et al. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal*, 44(6), 1102–1121.

[14] McEvoy, G. M., & Cascio, W. F. (1987). Do good or poor performers leave? *Academy of Management Journal*, 30(4), 744–762.

[15] Ribeiro, M. T., et al. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD*, 1135–1144.

© 2021-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 5, Issue 12, pp 102-106, December-2021

https://doi.org/10.47001/IRJIET/2021.512020

Citation of this Article:

Manoj Parasa, & Sasi Kiran Parasa, "Predicting Employee Turnover Using Machine Learning Models Trained on SAP SuccessFactors and SAP HCM Historical Data" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 5, Issue 12, pp 102-106, December 2021. Article DOI https://doi.org/10.47001/IRJIET/2021.512020

****

© 2021-2017 IRJIET All Rights Reserved