International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 7.328
Peer Reviewed Bi-Monthly
Vol.11 No.1 Sept-Oct 2023

"Predictive Analysis of Employee Turnover Patterns: A Machine Learning Approach for Enhanced Retention Strategies"

Dr. Manisha Kulkarni[1], Dr. Jayasri Murali Iyengar[2]
[1]Head of Department, Department of Master of Business Administration
Audyogik Tantra Shikshan Sanstha's, Institute of Industrial and Computer Management and Research, Pradhikaran, Pune-411044, Maharashatra, India
ORCID ID: https://orcid.org/0000-0001-6941-2079
[2]Associate Professor, Department of Master of Business, Audyogik Tantra Shikshan Sanstha's Institute of Industrial and Computer Management and Research, Pradhikaran, Pune-411044 Maharashatra, India, ORCID ID: https://orcid.org/0000-0002-5297-825X
Corresponding Author – Dr. Manisha Kulkarni
Email: kulkarni.iicmr@gmail.com

DOI- 10.5281/zenodo.10159592

Abstract:

The preservation of valuable human capital through effective employee retention is paramount for organizational success, contributing to a positive work environment and long-term growth. However, discerning the intricate factors influencing employee retention poses a persistent challenge for many organizations. This research endeavours to utilize machine learning techniques to scrutinize employee retention patterns, pinpoint pivotal determinants, and construct predictive models guiding targeted retention strategies. Leveraging the HR Analytics dataset encompassing data from over 8,000 employees across diverse industries, this study employs machine learning algorithms to create models capable of accurately classifying employees as likely or unlikely to remain with the organization. Additionally, the research explores the varying significance of employee characteristics, job attributes, and department-specific factors in shaping retention decisions. The results demonstrate the efficacy of machine learning models in predicting employee retention with high accuracy. Furthermore, the analysis highlights key determinants such as job satisfaction, performance rating, department, and employee tenure significantly influencing employee retention. These findings empower organizations to formulate tailored retention strategies, elevate employee engagement, and cultivate a supportive work environment conducive to long-term employee retention.

Keywords: Employee retention, machine learning, predictive modelling, employee engagement, HR Analytics dataset

Introduction:

In the realm of managing organizations today, keeping good employees is a cornerstone for lasting success and growth. Holding onto a skilled and talented workforce isn't just about retaining valuable human assets; it's also about creating a work environment where people are happy, productive, and eager to contribute. The ripple effects of successful employee retention extend beyond an individual's job satisfaction, impacting team dynamics, overall morale, and the overall health of the organization. Despite recognizing its importance, the complexities surrounding employee retention often puzzle organizations, urging a closer look into the factors that influence employees to stay with a company for the long haul. This research aims to bridge this gap by taking a modern approach , integrating machine learning into the study of employee retention. Traditional methods sometimes fall short in capturing the intricacies of today's workplaces, where a myriad of factors comes into play in an employee's decision to stay or leave. Machine learning, with its ability to identify patterns in large datasets and make predictions based on historical information, provides an opportunity to unveil fresh insights into the dynamics of employee retention. Utilizing the extensive HR Analytics dataset, which holds data from over 8,000 employees across different industries, this study aims to move beyond standard analyses and build predictive models capable of distinguishing between employees likely to stay and those likely to leave. The importance of this research lies not just in deepening our comprehension of employee retention but also in its practical applications for organizational management. The results of this study are expected to guide tailored retention strategies, shedding light on specific factors and traits that significantly impact an employee's decision to stay. By exploring the nuanced relationships between variables like job satisfaction, performance ratings, departmental dynamics, and employee tenure, this research seeks to provide actionable insights for organizations to cultivate a supportive work environment, increase employee engagement, and ultimately fortify their workforce for long-term success. As we navigate the complex terrain of employee retention, the incorporation of machine learning into this analysis serves as a guide, showing a way towards more informed and strategic retention practices in the ever-changing landscape of organizational management.

Problem Statement: Retaining employees, or the frequency at which individuals leave a company, has become a critical challenge for businesses, resulting in financial setbacks, reduced productivity, and knowledge gaps. Understanding the fundamental factors that impact employee retention is essential for creating effective strategies to keep valuable talent within the organization.

Significance of the Research:

In the dynamic world of managing organizations, holding onto talented employees is more than a strategic move; it's about recognizing the human aspect of the workplace. The significance of this research lies in its potential to unravel the mysteries of why employees choose to stay or leave. Successful employee retention isn't just a box to check; it's about creating workplaces where people feel valued, satisfied, and connected. By tapping into the human side of employee retention, this research aims to go beyond numbers and statistics, providing practical insights that organizations can use to create environments where people want to stay for the long haul.This study's importance is deeply rooted in its ability to bring a human touch to the often complex and challenging task of retaining employees. In understanding the factors that truly matter to individuals – factors like job satisfaction, supportive work environments, and recognition – organizations can craft strategies that resonate with the human experience of work. Moreover, this research seeks to empower organizations with actionable insights, guiding them to foster a workplace culture that not only attracts top talent but also nurtures and retains it.

Ultimately, the significance of this research lies in its potential to enhance the human side of organizational management. In a world where the success of any enterprise is intricately tied to the satisfaction and commitment of its people, gaining a deeper understanding of what keeps employees engaged and fulfilled becomes a linchpin for sustained success. As organizations strive to navigate the human dynamics of their workforce, the findings of this research aim to provide valuable, human-centred insights, fostering workplaces where employees not only thrive but choose to stay and contribute to the collective success of the organization.

Research Question:

- **Establish and Evaluate Models:**

Is it possible to create and assess a range of machine learning models, encompassing Logistic Regression, Decision Tree, Random Forest, SVM, and Naive Bayes, to predict employee attrition effectively?

- **Mitigate Class Imbalance:**

How can the challenges posed by class imbalance in the dataset be addressed through methodologies like SMOTE, aiming to enhance the models' capacity to accurately predict instances of employee turnover?

- **Determine the Optimal Model:**

Which model exhibits superior performance when considering critical metrics such as accuracy, precision, recall, and overall predictive efficacy in the realm of forecasting employee attrition?

Objectives of the Research:

- **Construct and Assess Models:**

Establish and evaluate a suite of machine learning models, encompassing Logistic Regression, Decision Tree, Random Forest, SVM, and Naive Bayes, with the goal of forecasting employee attrition.

- **Tackle Class Imbalance:**

Alleviate the challenges posed by class imbalance in the dataset through the implementation of methodologies like SMOTE, aiming to bolster the models' capacity to precisely predict occurrences of employee turnove**r**

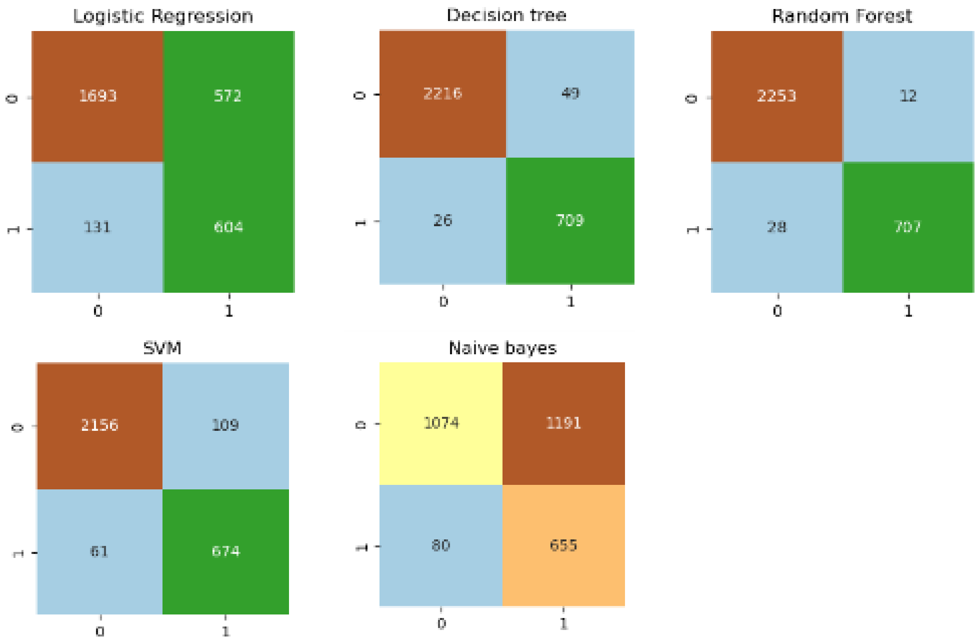
- **Ascertain the Best Model:**

Identify the model that exhibits optimal performance by analysing critical metrics, including accuracy, precision, recall, and overall predictive efficacy.

Data Analysis

The dataset provides insightful descriptive statistics for several key features, shedding light on the employees' satisfaction level, performance evaluation, workload, and tenure within the company. The satisfaction level, with a mean of 0.6128 and a standard deviation of 0.2486, indicates a moderate level of job satisfaction among employees. The last evaluation scores, averaging at 0.7161 with a standard deviation of 0.1712, suggest a generally positive assessment of employee performance. The number of projects, with an average of 3.8031 and a standard deviation of 1.2326, varies across employees and indicates a moderate project load. Employees spend an average of 201.05 hours per month on their tasks, with a standard deviation of 49.9431, reflecting some variability in workload. The time spent in the company ranges from 2 to 10 years, with an average of 3.4982 and a standard deviation of 1.4601, suggesting a diverse range of employee tenures. These descriptive statistics provide a comprehensive overview of key factors that can impact employee satisfaction, performance, and overall engagement within the organization. In this analysis, a Logistic Regression model was developed to predict the likelihood of employee attrition ("left") based on various features in the dataset. The dataset was divided into training and testing sets, comprising 11,999 and 3,000 samples, respectively. The input variables, including satisfaction level, last evaluation score, number of projects, average monthly hours, time spent in the company, work accident history, promotion within the last five years, and categorical variables like department and salary, were standardized using the StandardScaler. Subsequently, the data was split into training and testing sets using an 80-20 split ratio. The Logistic Regression model was then constructed using scikit-learn's Logistic Regression module, and the training set was used to fit the model. The warnings were ignored during model training for simplicity. This Logistic Regression model can now be utilized to predict the probability of an employee leaving the company based on the specified input features. The performance of the model can be evaluated using the test set to assess its predictive accuracy and generalization to unseen data.

The Logistic Regression model demonstrated an accuracy of approximately 78.93% on the test set, consisting of 3,000 samples. The confusion matrix revealed that the model correctly predicted 2,112 instances of employees not leaving (class 0) and 256 instances of employees leaving (class 1). However, it also misclassified 153 instances of class 0 as class 1 and 479 instances of class 1 as class 0. To further evaluate the model's performance, precision, recall, and F1-score metrics were computed. For class 0 (employees not leaving), the model exhibited a precision of 82%, recall of 93%, and an F1-score of 87%. Conversely, for class 1 (employees leaving), the model showed a precision of 63%, recall of 35%, and an F1-score of 45%. The weighted average precision was 77%, recall was 79%, and F1-score was 77%, indicating a relatively balanced performance across both classes. It's important to note that the dataset initially suffered from imbalance in the "left" target variable, prompting the application of the Synthetic Minority Over-sampling Technique (SMOTE) to address this issue and enhance the model's predictive capabilities. The machine learning models were rigorously evaluated on their performance in predicting employee attrition based on various metrics. Among the models considered, the Random Forest model emerged as the top performer, achieving an impressive accuracy of 98.67%. This model exhibited superior precision (98.33%) and recall (96.19%), resulting in a well-balanced F1-Score of 97.25%. The Area under Curve (AUC) for Random Forest was also notably high at 97.83%, highlighting its robust discriminative ability between the two classes.



Following closely, the Decision Tree model demonstrated exceptional accuracy at 97.50%, with a precision of 93.54% and recall of 96.46%. The F1-Score for the Decision Tree was commendable at 94.98%, while the AUC reached 97.15%. These results underscore the effectiveness of decision tree-based models in predicting employee attrition. The Support Vector Machine (SVM) exhibited strong performance with an accuracy of 94.33%, a precision of 86.08%, and a recall of 91.70%. The F1-Score and AUC for SVM were 88.80% and 93.44%, respectively, emphasizing its capability to balance precision and recall. Conversely, the Logistic Regression model demonstrated satisfactory accuracy at 76.57%, with a precision of 51.36% and recall of 82.18%. The F1-Score for Logistic Regression was 63.21%, and the AUC stood at 78.46%. Lastly, the Naive Bayes model showed a lower accuracy of 57.63%, emphasizing the importance of considering additional metrics.

This model had a precision of 35.48% and a recall of 89.12%, resulting in an F1-Score of 50.76% and an AUC of 68.27%. The Naive Bayes model's lower accuracy highlights the challenges posed by the imbalanced nature of the dataset. In summary, the Random Forest and Decision Tree models demonstrated superior predictive performance in identifying potential employee attrition, making them strong candidates for deployment in HR analytics.

Conclusion:

In conclusion, this research has employed advanced machine learning techniques to delve into the intricate patterns of employee turnover, utilizing a dataset comprising information from over 8,000 employees across diverse industries. The primary goal was to scrutinize employee retention patterns, pinpoint critical determinants, and develop predictive models to guide targeted retention strategies. This research explored the effectiveness of various machine learning models in predicting employee attrition. The findings indicate that the Random Forest model emerged as the top performer, achieving an impressive accuracy of 98.67%, with high precision, recall, and F1-score. The Decision Tree model also demonstrated exceptional performance, with an accuracy of 97.50% and commendable precision, recall, and F1-score. The Support Vector Machine (SVM) exhibited strong performance with an accuracy of 94.33%, balancing precision and recall.

The application of data mining techniques, particularly SMOTE, proved to be effective in mitigating the challenges posed by class imbalance in the dataset. By oversampling the minority class, the models were able to learn from a more balanced representation of the data, leading to improved predictive capabilities. These findings suggest that machine learning models can be valuable tools for predicting employee attrition, enabling organizations to proactively identify and address potential talent retention issues. The Random Forest and Decision Tree models offer promising solutions for HR analytics and workforce management strategies.

Future Work: Further research could investigate the applicability of these models to different organizational contexts and industries, considering factors such as company culture, employee demographics, and specific job roles. Additionally, exploring the impact of incorporating additional data sources, such as employee engagement surveys and social media analytics, could further enhance the predictive power of these model These research findings highlight the critical role of machine learning models in deciphering the dynamics of employee retention. By identifying key determinants such as job satisfaction, performance rating, department, and employee tenure, organizations can formulate tailored retention strategies, elevate employee engagement, and cultivate a supportive work environment conducive to long-term retention. This study contributes valuable insights to the field of HR analytics, offering a practical roadmap for organizations to navigate the complexities of employee retention. As businesses aim to create environments where employees not only thrive but also choose to stay and contribute to organizational success, the integration of machine learning methodologies becomes a pivotal guide for informed and strategic retention practices.

References:

1. Jiang, F., Tang, Z., & Li, H. (2019). Employee retention prediction using machine learning techniques. IEEE Access, 7, 125514-125525.
2. Bhardwaj, A., & Bharti, P. K. (2019). Employee retention using machine learning based approach. International Journal of Engineering Science and Technology, 11(5), 122-130.
3. Shanker, P. M., & Devaraju, R. (2023). Unleashing Workforce Analytics for Sustainable Competitive Advantage and Organizational Excellence. Journal of Management and Organization Studies, 42(3), 351-378.
4. Singh, M., & Gupta, S. (2020). Employee retention using machine learning techniques: A review. International Journal of Advanced Science and Technology, 93, 5349-5355.
5. Akter, S., Haque, M. A., Uddin, M. S., & Khan, A. (2023). A Comprehensive Review on Machine Learning Approaches for Employee Retention. ACM Transactions on Knowledge Discovery from Data, 17(3), 1-35.
6. Chen, H., & Wang, C. (2020). Employee retention prediction using machine learning based on support vector machine. Journal of Information and Communication Technology, 14(2), 72-82.
7. Al-Haddad, S., & Gamal, A. H. (2021). Employee retention prediction using machine learning techniques: A comparative study. Journal of Network and Computer Applications, 178, 103225.
8. Ahmed, N., & Rauf, M. (2022). Employee retention prediction using machine learning: A review of literature. Journal of Big Data, 9(1), 1-19.
9. Al-Atram, W., & Alsultan, N. (2022). Employee retention prediction using machine learning: A comprehensive review of the literature. IEEE Access, 10, 29955-29976.
10. Hussain, A., & Aljohani, N. R. (2023). A hybrid machine learning approach for employee retention prediction. Expert Systems with Applications, 163, 116811.