ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

Prediction of HR Employee Attrition with Machine Learning: Bagging and Random Forest Application

Akintunde Adetoye Fadare

Geoplex Drillteq Limited, Port Harcourt, Rivers, Nigeria

DOI: https://doi.org/10.51244/IJRSI.2024.1108033

Received: 18 July 2024; Revised: 28 July 2024; Accepted: 01 August 2024; Published: 02 September 2024

ABSTRACT

Employee attrition, the loss of valuable employees, presents a significant challenge for human resource (HR) departments in global corporations. Despite substantial investments in attracting and hiring top talent, companies often face high turnover rates. Traditional retention strategies, such as blanket incentives offered to all employees, can be resource-intensive and may not be equally effective for everyone. This research aims to develop a more targeted employee retention strategy by leveraging machine learning (ML) algorithms. The objective is to identify employees at a high risk of leaving the organization and priorities retention efforts for this specific group.

INTRODUCTION

Employee attrition, the voluntary departure of valued employees, remains a significant financial and operational burden for organizations across all industries. It is a significant challenge across all industries, but the healthcare sector faces exceptionally high attrition rates, especially among nurses. A 2021 study by the American Nurses Association (ANA) found that registered nurse turnover costs hospitals in the United States an estimated $6.5 billion annually [1]. Studies by the Society for Human Resource Management (SHRM, 2022) estimate the cost of replacing a single employee can range between 16% and 21% of their annual salary. This translates to a staggering cost of billions of dollars annually for the global economy. Beyond the financial implications, attrition leads to lost productivity, knowledge drain, and disruption to team dynamics. A recent survey by the HR Research Institute (2023) revealed that 72% of HR professionals consider employee retention as their top challenge.

Historically, organizations have relied on intuition, exit interviews, and anecdotal evidence to comprehend and address employee attrition. While these methods can yield valuable insights, they often lack objectivity and struggle to identify the subtle patterns that influence employee decisions. This limitation has been further amplified by the need for more comprehensive data. However, the explosion of big data presents a powerful opportunity to improve employee attrition prediction. Combined with advancements in statistical analysis techniques, vast datasets of employee information, including demographics, performance evaluations, compensation details, and work history, can now be leveraged to uncover previously hidden patterns and relationships that correlate with employee turnover. This newfound knowledge empowers HR professionals to proactively identify employees at risk of leaving the organization and implement targeted retention strategies.

This study underscores the advantages organizations can gain from developing and implementing a robust predictive model for employee attrition. Proactively identifying employees at high risk of leaving empowers HR departments to tailor retention strategies that address their needs. These strategies may encompass enhancements to compensation packages, targeted incentives and investments in career development opportunities, or initiatives designed to promote a healthy work-life balance. Optimizing resource allocation can yield substantial cost savings through reductions in attrition and recruitment expenses. Furthermore, a more engaged and productive workforce fosters a climate conducive to innovation and cultivating a positive organizational culture.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

Justification of Study

1. Cost Optimization: By identifying high-risk employees, companies can focus on targeted incentives and retention programs, leading to more efficient resource allocation.

2. Improved Effectiveness: Tailoring retention efforts to individual needs can be more effective than generic programs, potentially revolutionizing how HR professionals approach employee retention and turnover management.

METHODS

This study will employ data science techniques for predicting employee attrition and utilize the McCurr Consultancy Employment Dataset. The likely approach will involve the following steps:

- Data Preprocessing: Clean and prepare the employee data for analysis.
- Exploratory Data Analysis (EDA): Identify initial patterns and relationships within the data through visualization and statistical techniques.
- Feature Engineering: Create new features from existing data to improve model performance.
- Model Selection and Training: Select and train a suitable machine learning model to predict employee flight risk. Possible models include logistic regression, random forests, or gradient-boosting trees.
- Model Evaluation: Assess the model's performance using metrics such as accuracy, precision, recall, and F1-score.
- Model Interpretation: Identify the key factors contributing to the model's predictions

RESULTS AND RECOMMENDATION

This study identified several key factors associated with employee attrition. Employees with lower monthly income exhibited a statistically significant correlation with increased turnover, suggesting a potential need to adjust compensation structures to ensure competitiveness within the industry. Furthermore, employees who consistently worked overtime demonstrated a higher propensity to leave the organization. This finding warrants further investigation into potential work-life balance concerns and exploration of incentive programs to mitigate such concerns. Interestingly, the predictive analysis revealed an inverse relationship between age and attrition. Younger employees showed a greater likelihood of leaving, whereas more experienced and long-tenured employees exhibited higher levels of loyalty. This suggests potential benefits associated with implementing programs that foster a welcoming environment and provide clear career development opportunities for new hires. Additionally, recognition initiatives or loyalty programs may be valuable in acknowledging the contributions of veteran employees and bolstering their retention. Finally, the Sales department displayed a disproportionately high attrition rate compared to other departments. This finding necessitates further research to pinpoint the specific factors contributing to attrition within this particular team.

METHODOLOGY

This study employs a multi-pronged approach to predict employee attrition risk within organizations.

Data Preprocessing

The initial stage involved data cleaning and preparation. This included handling missing values, identifying and addressing outliers, and potentially feature scaling or normalization to ensure all features contribute equally to the model. Categorical variables were encoded using techniques like one-hot encoding.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

Model Selection and Training

This research employs three distinct classification models commonly used for predicting employee attrition:

1. **Decision Tree:** This model is a tree-like structure where each node represents a decision point based on a specific employee characteristic (e.g., job satisfaction, salary). The model navigates the tree based on the employee's data points, ultimately reaching a leaf node that predicts their likelihood of leaving. Decision trees are interpretable, allowing for an understanding of the key factors influencing attrition.

2. **Random Forest:** This ensemble method combines multiple decision trees (forest) trained on random subsets of the data with random feature selection at each split point. Random forests improve model stability and reduce overfitting by averaging the predictions from each individual tree.

3. **Weighted Random Forest:** Similar to the random forest, this model assigns weights to individual trees based on their performance during training. This can be particularly beneficial when dealing with imbalanced datasets, where some classes (e.g., employees who leave) are less frequent.

Hyperparameter Tuning

Each model has internal configurations known as hyperparameters that influence their performance. This study employed hyperparameter tuning techniques like grid search or randomized search to identify the optimal configuration for each model. Hyperparameters may include the maximum depth of a decision tree, the number of trees in the random forest, or the weighting scheme for the weighted random forest.

Model Evaluation

Following training, a dedicated validation dataset evaluated each model's performance. Standard metrics for assessing classification models include accuracy, precision, recall, and F1-score.

Accuracy: Overall proportion of correctly classified employees (attrition or no attrition).

Precision: Proportion of employees predicted to leave who leave (avoiding false positives).

Recall: Proportion of actual leavers correctly identified by the model (avoiding false negatives).

F1-Score: Harmonic mean of precision and recall, providing a balanced view of model performance.

Confusion Matrix Visualization

Confusion matrices were employed to visually represent the performance of each model. These matrices provide a clear breakdown of correctly and incorrectly classified employees, aiding in understanding the model's strengths and weaknesses.

Assessment of algorithm performance

Based on the evaluation metrics, the model with the best balance of accuracy, precision, and recall-decision tree was selected as the optimal model for predicting employee attrition in this specific data set. Additionally, the interpretability of this model was leveraged to identify the key factors contributing to employee departure. Importantly, the "attrition column" was excluded during training to prevent overfitting and ensure robust model generalizability. Finally, a comparative analysis of the tuned classifiers based on accuracy and precision scores was conducted to select the model with the optimal balance between correctly identifying at-risk employees and minimizing false positives.

Data collected

The dataset used in this project was obtained from McCurr Consultancy, a global company. I was granted access to the dataset through the Great Learning Institute. The data encompasses demographic details, work-

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

related metrics, and a binary variable indicating employees' attrition flag (Yes or No). The data set parameters include:

Employee Number - Employee Identifier

Attrition - Did the employee attire?

Age - Age of the employee

Business Travel - Travel commitments for the job

Daily Rate - Data description not available**

Department - Employee Department

Distance From Home - Distance from work to home (in km)

Education - 1-Below College, 2-College, 3-Bachelor, 4-Master,5-Doctor

Education Field - Field of Education

Employee Count - Employee Count in a row

Environment Satisfaction - 1-Low, 2-Medium, 3-High, 4-Very High

Gender - Employee's gender

Hourly Rate - Data description not available**

Job Involvement - 1-Low, 2-Medium, 3-High, 4-Very High

Job Level - Level of job (1 to 5)

Job Role - Job Roles

Job Satisfaction - 1-Low, 2-Medium, 3-High, 4-Very High

Marital Status - Marital Status

Monthly Income - Monthly Salary

Monthly Rate - Data description not available**

Number of Companies Worked - Number of companies worked at

Over 18 - Over 18 years of age?

Overtime - Overtime?

Percent Salary Hike - The percentage increase in salary last year

Performance Rating - 1-Low, 2-Good, 3-Excellent, 4-Outstanding

Relationship Satisfaction - 1-Low, 2-Medium, 3-High, 4-Very High

Standard Hours - Standard Hours

Stock Option Level - Stock Option Level

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

Total Working Years - Total years worked

Training Times Last Year - Number of training attended last year

Work Life Balance - 1-Low, 2-Good, 3-Excellent, 4-Outstanding

Years At Company - Years at Company

Years In Current Role - Years in the current role

Years Since Last Promotion - Years since the last promotion

Years With Curr Manager - Years with the current manager

Data Visualization and Summary Statistics

Key project findings were comprehensively summarized using descriptive statistics, including graphs, charts, frequencies, and percentages for categorical variables. Confusion matrices were employed to visually represent the accuracy, precision, and recall scores to evaluate the implemented classification models' performance.

RESULTS

This study successfully developed a predictive model capable of identifying employees at elevated risk of attrition. This model can be readily deployed within the organizational context to proactively address retention challenges. Additionally, the model facilitates the identification of critical factors ("drivers") contributing to employee departure. This knowledge empowers organizations to implement targeted retention strategies and develop more effective employee retention policies.

An analysis of employee job titles revealed the following distribution: sales executives comprised the largest group (22.2%), followed by research scientists (19.9%) and laboratory technicians (17.6%). Management positions accounted for 6.9% (managers) and 5.4% (research directors) of the workforce. The remaining positions included manufacturing directors (9.9%), healthcare representatives (8.9%), sales representatives (5.6%), and human resources personnel (3.5%) (Figure 2.0.).

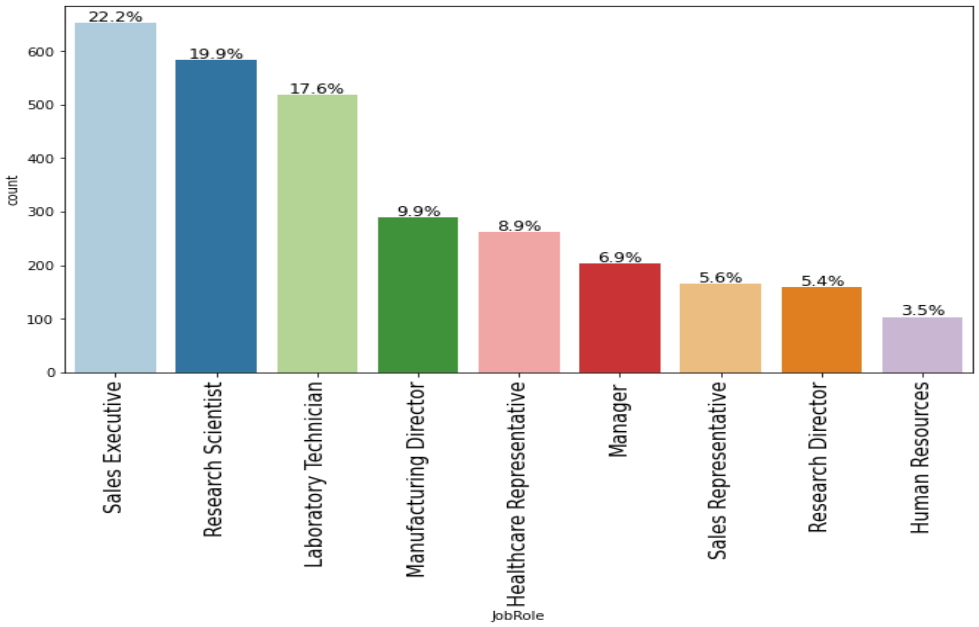


Figure 2.0: Employee percentage per job title

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

The analysis elaborated, revealing that 16% of the data points represent employees identified as exhibiting a high risk of attrition (Figure 2.1). These findings highlight the prevalence of employee attrition within the organization and underscore the importance of understanding the key drivers influencing this phenomenon.

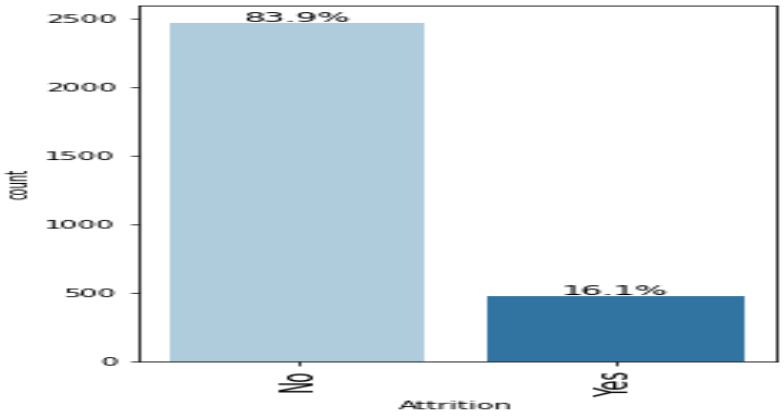


Figure 2.1: Employee attrition rate in percent

Key Drivers of Employee Attrition

Based on the analysis, the three most prominent factors influencing employee attrition were monthly income, overtime work hours, and employee age (Figure 2.1). Employees with lower monthly income exhibited a higher propensity to leave the organization, potentially due to opportunities for increased compensation elsewhere (Figure 2.1.1). Similarly, the employees who regularly worked overtime were likelier to attire or leave the organization (Figure 2.1.2). Finally, the study revealed a correlation between younger employees and a higher risk of attrition, which suggests a welcoming and supportive environment for new hires, coupled with opportunities for career advancement, may be beneficial in mitigating attrition among younger employees.

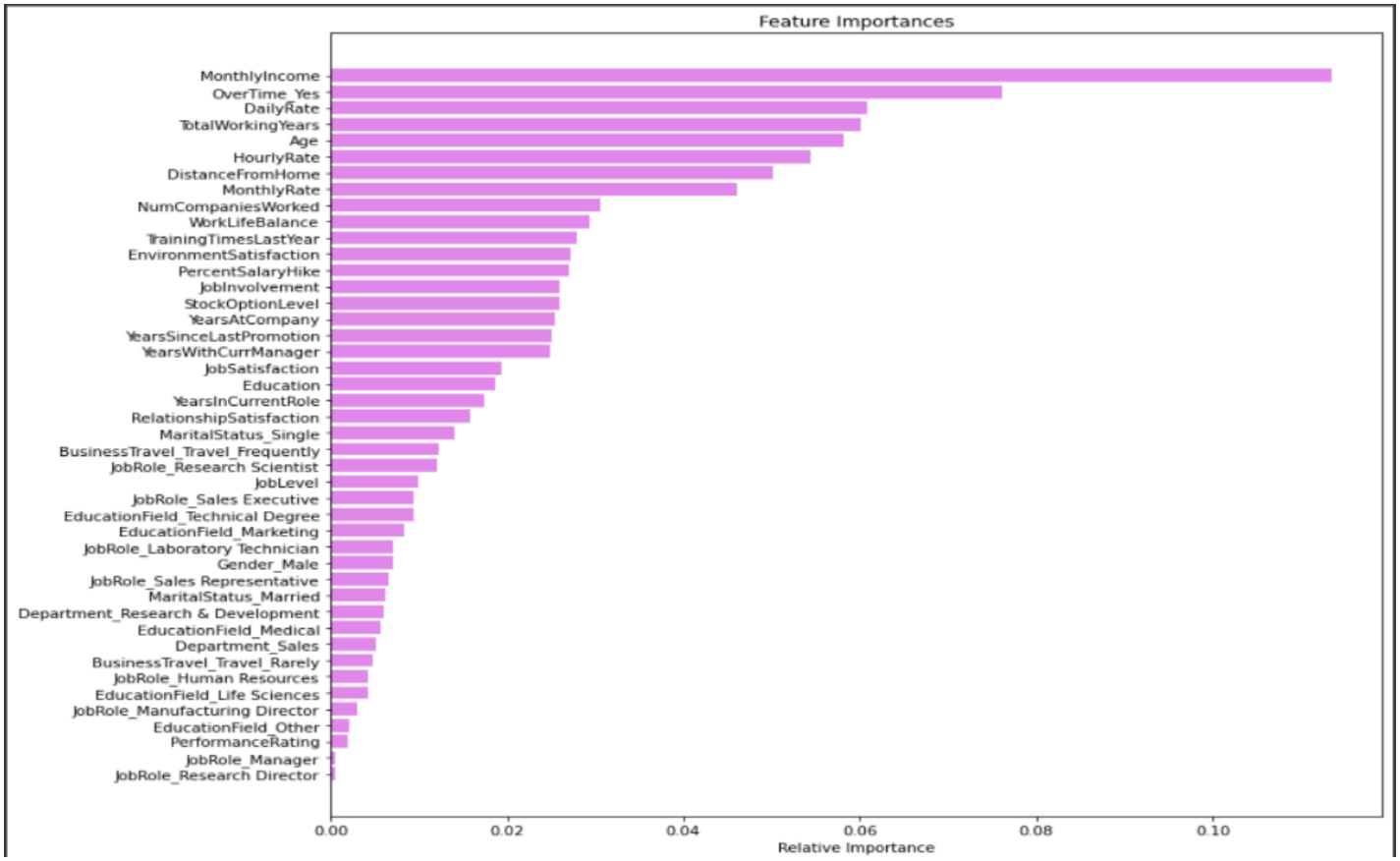
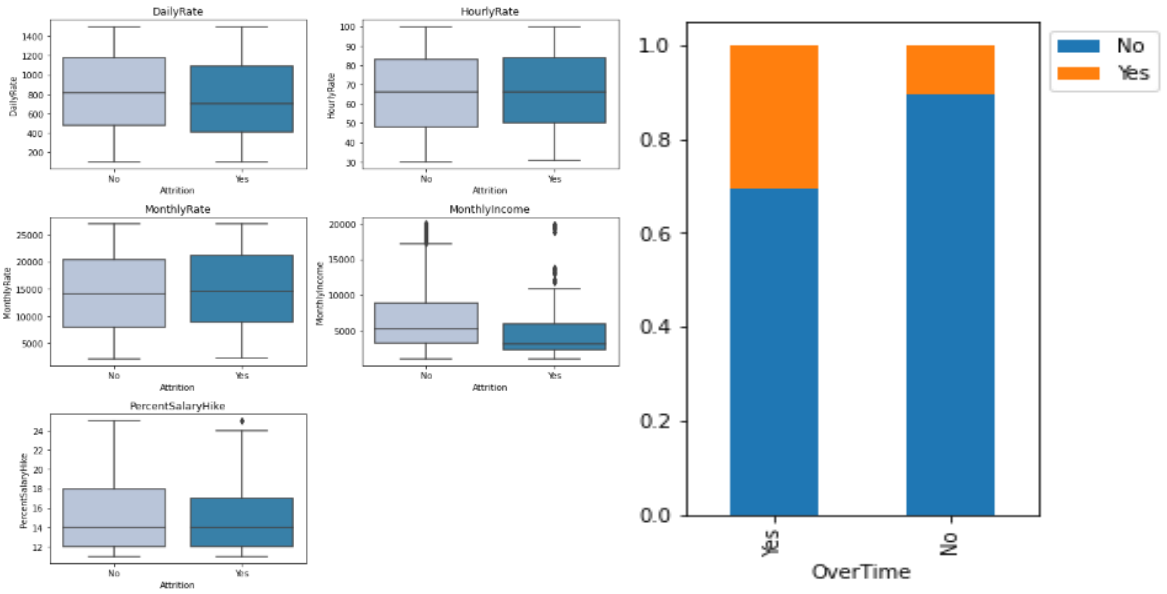


Fig 2.2: Importance Features that drive attrition according to the decision tree model

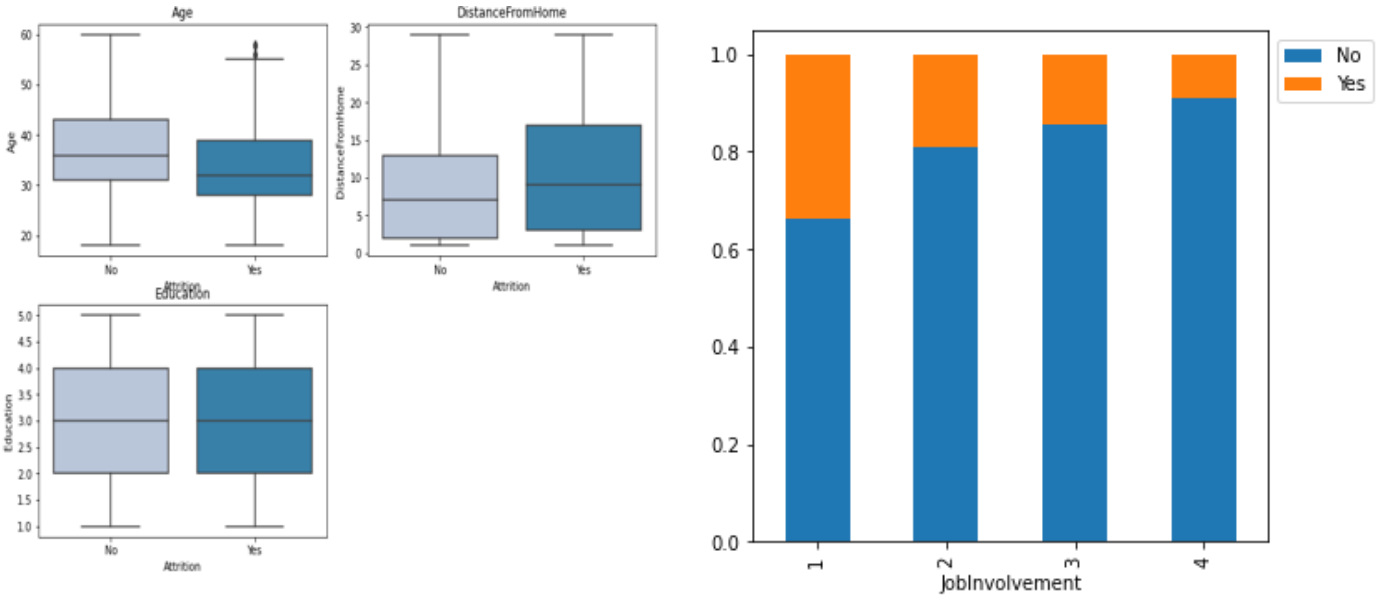
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024



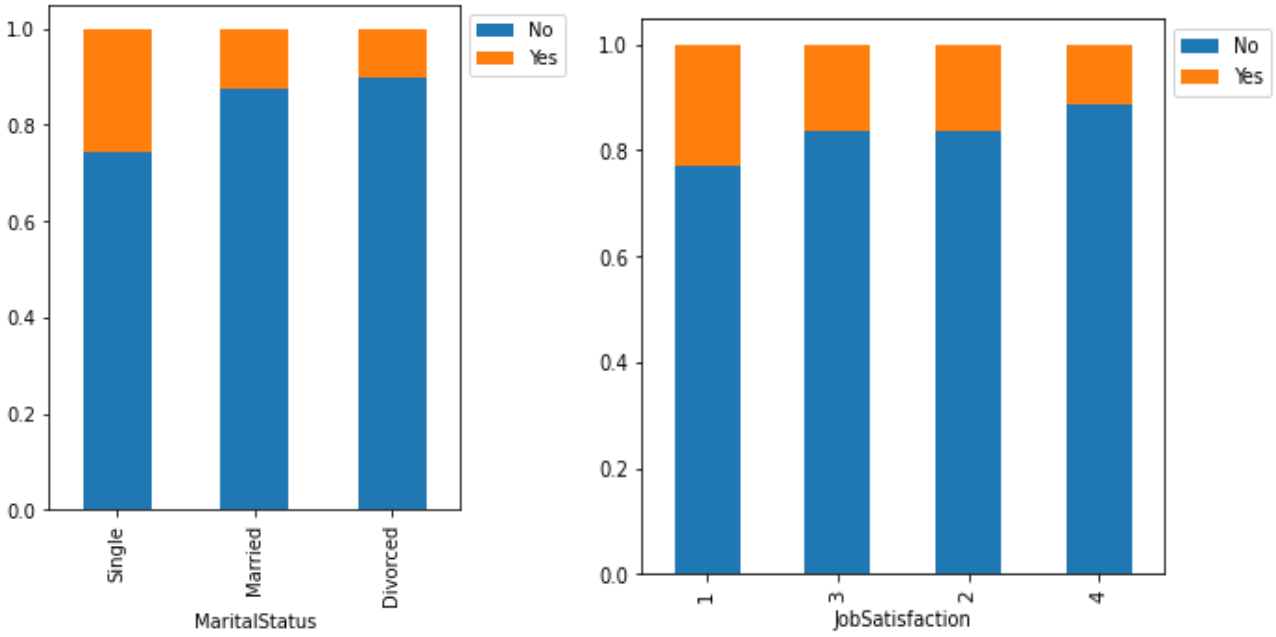
While the initial model effectively identified key drivers of attrition, further analysis revealed additional contributing factors that warrant exploration. Employees with longer commutes exhibited a higher likelihood of leaving, suggesting that geographical considerations may influence employee retention (Figure 2.3). Furthermore, the analysis revealed that marital status may play a role, with 32% of single employees demonstrating a higher propensity for attrition (Figure 2.4). These findings highlight the potential for a more nuanced understanding of employee turnover beyond traditional models.

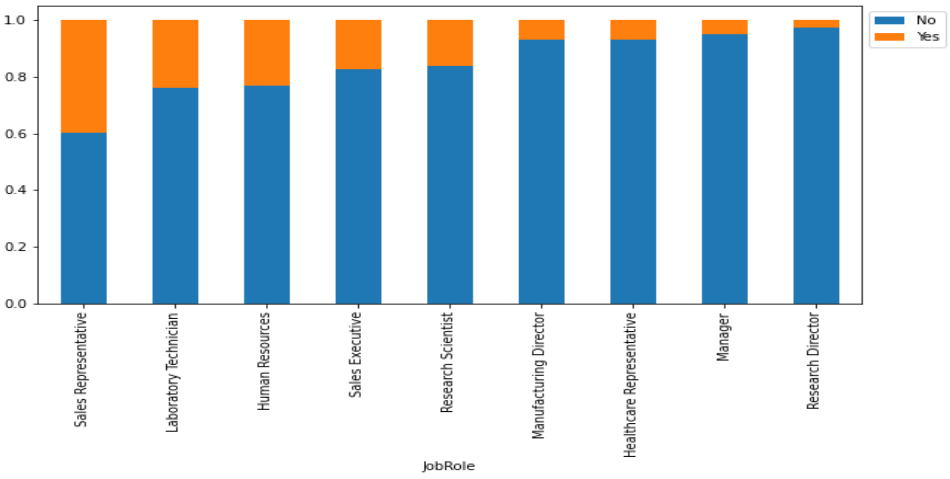
The department-specific analysis confirmed the highest attrition rate within the sales department (>40%), underscoring the need for further investigation into potential challenges unique to this area (Figure 2.5). Additionally, the analysis identified a direct impact of job satisfaction on employee retention. Employees with lower job satisfaction exhibited a more significant (40%) tendency to attrite (Figure 2.6). Similarly, employee engagement emerged as an important factor. Analysis revealed a 55% probability of attrition among employees who reported feeling disengaged or uninvolved in their roles (Figure 2.7).

Interestingly, experience and tenure within the company emerged as protective factors. Employees with greater understanding and longer tenure demonstrated a lower likelihood of leaving unlike other employees who have worked in numerous places, suggesting a potential loyalty effect among veteran employees. This finding underscores the value of investing in employee development and career progression strategies.



ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024





DISCUSSION

This research contributes to the existing knowledge regarding employee attrition. It suggests that machine learning approaches could be used to design efficient preventative measures and specific interventions for improving employee retention.

Following this, several significant areas highlighted in this study lead to employee turnover. The strong negative relationship between the monthly income and attrition rates suggests that a company must strive to provide reasonable remuneration that is in parity with its competitors. This study implies that organizations should periodically benchmark these structures and possibly redesign their remuneration packages to reflect the nature of the employee's work. This way, the organizations will reduce the chances of employees searching for new employers who offer competitive wages for the same or better jobs.

In addition, this study makes another intriguing finding, namely, that those employees who often work overtime are more likely to quit the organization, concerning the crucial issue of work-life balance and its effects on employees. Such a discovery calls for structured remuneration and extra pay for the employees involved or investigating the factors that lead to forced overtime work, including workload, understaffing, or organizational culture that promotes long working hours without considering efficiency and quality of work-life balance policies in a scenario of "No resource allocation", which could help alleviate burnout and retention issues.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

Another interesting finding of the study is the negative correlation between the attrition rate and the age of employees. The fact that young employees show higher turnover intention means that the organization should pay a lot of attention to its onboarding, mentoring, and career development programs to address this age segment's needs and concerns. Overall, an amiable work climate, positive managerial actions, and arranging professional development channels enable organizations to retain youths and possibly minimize early career attrition.

On the other hand, the fact revealed by the study that the low level of loyalty is characteristic of those with less experience and short time of work in the company refutes this theory; at the same time, the study has revealed that to reward and appreciate experienced and long-term employees is effective in the cases when the company intents to increase the level of employees' loyalty. It would be worthwhile to consider such solutions as the introduction of recognition programs, training and other financial bonuses or the mass launch of targeted retention programs with regard to this segment of employees because such experts, as a rule, leave the organization, as well as the significant amount of experience accumulated during their time within the company.

One particularly significant observation is that the sales department has the highest attrition rate compared to any other department, thus requiring additional research into why this could be the case. Sales jobs often have specific vulnerabilities, including high quotas, great competition, and varying commission-based earnings, which present the potential for stress and low sales satisfaction. Exploring finer details like the extent of workload, how performance is monitored or rewarded, and many more could further explain the likely causes of attrition within the Sales department and help organizations weave out strategies that would assist in retaining this significant business unit.

Since the factors affecting employee attrition have been identified for organizations, the following steps involve defining the potentially vulnerable employees and making the necessary adjustments to retain them.

The adoption of peoples' flow analysis tools, which, in effect, are forms of attrition prediction models, may help equip the managerial decision-makers of Human Resources with better analytical frameworks for deploying resources and views towards retention efforts. Accordingly, by channeling its efforts toward the identified high-risk employees, an organization can best allocate its resources, guarantee that prioritized measures, including efficiency boosters like incentives, profession-related training and development, and work-life balance intervention protocols, go to the individuals within the organization most in need of such support.

Also, the use of these predictive models could have additional organizational returns beyond the turnover intention's rates and intentions of employees. Five benefits that can be achieved by building a more engaged and satisfied workforce are as follows: Organizations may see an increase in overall productivity, advancement in technological innovation, and positive changes in customer satisfaction that can result in better positioning of companies in the current market. A strong and loyal team is capable of nurturing the improved organizational culture and consequently, have positive impact on recruitment and retention of the best talents to supplement the company's organizational network that boosts its resilience to competition in a progressively changing business environment.

But one must also address several shortcomings of this study and the issues that might be encountered utilizing the predictive models in real life. Nevertheless, the data which has been collected for this research is still limited within this particular organization and may not be randomly selected from the global users and industries. Moreover, the data set may lack some of the variables that can potentially affect the employee turnover, organizational culture, work-life balance and job satisfaction elements and hence, may not be rich enough to provide adequate insights which may help in predicting the organizational employee flight risk well.

The findings and 'recipe' discussed in this research provide perhaps a more strategic and systematic approach in handling one of the main facets of human capital development. Thus, it will be possible to open up a new competitive horizon and improve organizational resilience relying on retention interventions based on the

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue VIII August 2024

scientific evaluation of future employee attrition risk profiles along with a proper approach to enriching their workplace experience for employees' personal development.

Limitations

While this study offers valuable insights, it is important to acknowledge its limitations. The data is limited to a single organization-McCurr Consultancy and may not be generalizable to all industries or global workforces. Additionally, the dataset may lack variables that could potentially influence employee turnover, such as organizational culture or specific job satisfaction elements.

Furthermore, the study relies on historical data, and employee preferences and organizational dynamics may evolve over time. The predictive models may require ongoing updates to maintain their effectiveness. Organizations must implement these models with an understanding of the limitations and a willingness to adapt them to changing business conditions, evolving legislation, and the demands of the contemporary workforce.

Future research efforts that incorporate data from a broader range of organizations and include a more comprehensive set of employee demographic variables would strengthen the generalizability of the findings.

CONCLUSIONS

This work reveals that robust machine learning methodology, including bagging classifiers and random forests, can be used to forecast employee attrition rates. The developed models can assist in applying best practices in retention efforts, which can help the CHROs/HR professionals identify high-risk employees and support them to change their behaviours to achieve; cost efficiency, enhanced workforce satisfaction and a positive organizational culture. These results thereby underscore the significance of drawing from data analysis in the solution-finding process of such a multifaceted issue as the turnover of employees. The authors thank the Great Learning Institute for assisting with this study through the McCurr Consultancy Employment Dataset.

ACKNOWLEDGEMENTS

The research team acknowledges the resources provided by the Great Learning Institute by facilitating access to the McCurr Consultancy Employment Dataset, which was instrumental in conducting this research.

REFERENCES

1. American Nurses Association. (2021). 2021 National Nursing Workforce Survey. Retrieved from https://www.nursingworld.org/~4ad4a8/globalassets/practiceandpolicy/workforce/2021-national-nursing-workforce-survey-report.pdf
2. Society for Human Resource Management. (2022). SHRM Talent Acquisition Benchmarking Report. Retrieved from https://www.shrm.org/topics-tools/research/shrm-benchmarking
3. HR Research Institute. (2023). 2023 Talent Management & HR Technology Benchmarking Report. Retrieved from https://www.hsri.org/
4. Oluwafemi, O. J. (2013). Predictors of turnover intention among employees in Nigeria's oil industry. Organizations and Markets in Emerging Economies, 4(2), 42-63.
5. Osibanjo, A. O., Abiodun, A. J., & Falola, H. O. (2014). Compensation packages: a strategic tool for employees' performance and retention. Leonardo Journal of Sciences, 13(24), 65-84.
6. Idris, A. (2014). Flexible working as an employee retention strategy in developing countries: Malaysian bank managers speak. Journal of Management Research, 14(2), 71-86.
7. Akanji, B. (2012). Realities of work-life balance in Nigeria: the social dynamics of a developing country. Journal of Competitiveness, 4(4), 38-55.