Volume: 36, Issue: 1
Page: 20-35
2024

International Journal of Science and Business

Journal homepage: ijsab.com/ijsb

IJSAB International

Identifying Patterns and Predicting Employee Turnover Using Machine Learning Approaches

Dr. Aham Edward Kanuto

Abstract

Employee turnover poses significant challenges for organizations, impacting productivity, morale, and financial stability. Identifying patterns and predicting employee turnover using machine learning approaches can help organizations proactively address retention issues and optimize workforce management strategies. The current study analyzed a dataset comprising 4653 valid respondent records sourced from Kaggle, containing diverse attributes related to employees' educational backgrounds, work history, demographics, and employment-related factors. Through exploratory data analysis and feature selection, the study identifies key predictors of employee turnover, including factors such as education, joining year, city, payment tier, age, gender, ever benched status, and experience in the current domain. The researcher employs three machine learning algorithms—K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM)—to predict employee turnover based on these factors. Evaluation metrics such as accuracy, precision, recall, and F1-score were utilized to assess the performance of each model. Additionally, techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) were applied to handle class imbalance in the dataset. The findings reveal distinct characteristics and performance of each model, with the Decision Tree model exhibiting the highest accuracy and predictive capability. Through comprehensive analysis and model evaluation, this study contributes valuable insights into employee turnover prediction, enabling organizations to develop targeted retention strategies and foster a more engaged and stable workforce.

IJSB
Accepted 24 April 2024
Published 06 May 2024
DOI: 10.58970/IJSB.2373

ISSN: 2520-4750 (Online) 2521-3040 (Print)

Keywords: *Machine learning, Employee turnover, KNN model, Decision tree, Support vector classifier (SVC).*

About Author (s)

Dr. Aham Edward Kanuto, School of Business & Management, University of Juba, Republic of South Sudan.

Introduction

Employee turnover remains a persistent challenge for organizations across industries, exerting substantial impacts on productivity, morale, and financial stability (Hom, Lee, Shaw, & Hausknecht, 2017). As businesses strive to mitigate the adverse effects of turnover and optimize their workforce management strategies, there is a growing recognition of the potential of machine learning approaches in identifying patterns and predicting employee turnover (Kumar, 2022; Bhat, Khan, & Rainayee, 2023; Dogru et al., 2023). Drawing upon insights from prior studies, which have examined diverse facets of turnover dynamics and retention strategies, this study seeks to contribute to the burgeoning literature on employee turnover prediction using machine learning methodologies (Lazzari, Alvarez, & Ruggieri, 2022; Skelton, Nattress, & Dwyer, 2020; Iqbal, Asghar, & Asghar, 2022). This research builds upon the foundation laid by seminal publications, which have elucidated the multifaceted nature of turnover and underscored the importance of understanding its determinants and consequences (Hossain, Mia, & Hooy, 2023; Salleh et al., 2020).

In this study, the researcher analyzes a rich dataset sourced from Kaggle, comprising 4653 valid respondent records and encompassing a wide array of attributes related to employees' educational backgrounds, work history, demographics, and employment-related factors (Mossarah, 2023). Through exploratory data analysis and feature selection techniques, the researcher aims to identify key predictors of employee turnover, shedding light on the factors that significantly influence individuals' decisions to leave an organization. Building upon prior research that has highlighted the predictive efficacy of machine learning algorithms in various domains, the researcher employs three distinct methodologies—K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM)—to model and predict employee turnover based on the identified predictors (Domurath, Taggar, & Patzelt, 2023). By evaluating the performance of each model using established metrics such as accuracy, precision, recall, and F1-score, the researcher seek to ascertain the comparative effectiveness of different machine learning algorithms in predicting turnover.

Furthermore, recognizing the prevalence of class imbalance in turnover datasets, the researcher adopts techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) to address this challenge and enhance the robustness of the predictive models (Al-Suraihi et al., 2021). Through a comprehensive analysis of the dataset and rigorous model evaluation, this study aims to provide actionable insights into employee turnover prediction, empowering organizations to develop targeted retention strategies and cultivate a more engaged and stable workforce in an increasingly competitive landscape. By synthesizing findings from prior research and leveraging advanced analytics methodologies, this study endeavors to advance the understanding of turnover dynamics and offer practical guidance to organizations seeking to mitigate turnover risks and foster a culture of retention and engagement (Mokoena, Schultz, & Dachapalli, 2022; Olubiyi et al., 2019).

Literature Review

Employee turnover, the phenomenon of employees leaving their organizations, remains a critical concern for businesses worldwide due to its detrimental effects on productivity, morale, and financial stability (Hom et al., 2017). Over the years, scholars have extensively studied various aspects of employee turnover, ranging from its determinants to its consequences, and have proposed strategies to mitigate its impact. In this literature review, we synthesize insights from prior studies to provide a comprehensive understanding of employee turnover dynamics and highlight the role of machine learning in predicting turnover.

Determinants of Employee Turnover:

Several studies have identified numerous factors influencing employees' decisions to leave their organizations. Traditional predictors such as job satisfaction, organizational commitment, and job embeddedness have been extensively explored (Skelton et al., 2020; Salleh et al., 2020). For instance, Skelton et al. (2020) found a significant negative relationship between job satisfaction and turnover intention, suggesting that satisfied employees are less likely to leave their jobs. Similarly, Salleh et al. (2020) highlighted the importance of career planning and career satisfaction in reducing employees' turnover intentions, emphasizing the role of proactive career development initiatives in retention efforts.

Leadership and Organizational Factors:

Leadership styles and organizational factors also play crucial roles in shaping turnover intentions. Studies have examined the impact of despotic leadership on turnover intentions, highlighting the adverse effects of authoritarian leadership styles on employees' willingness to stay in their jobs (Iqbal et al., 2022). Moreover, organizational characteristics such as workplace environment, culture, and policies have been shown to influence turnover rates (Al-Suraihi et al., 2021). Al-Suraihi et al. (2021) emphasized the importance of understanding employees' needs and implementing appropriate strategies to improve organizational conditions and reduce turnover.

Industry-specific Insights:

Turnover dynamics vary across industries, with some sectors experiencing higher turnover rates than others. For example, the hospitality industry is known for its high employee turnover rates, which can be attributed to both economic factors and industry-specific characteristics (Dogru et al., 2023). Dogru et al. (2023) found that employee turnover in the hospitality industry is sensitive to overall economic conditions, highlighting the need for targeted retention strategies in this sector. Similarly, Hossain et al. (2023) investigated the impact of employee turnover on the credit risk of microfinance institutions, underscoring the importance of managing turnover risks to ensure long-term financial viability.

Machine learning models

In the current study, machine learning techniques such as the k-Nearest Neighbors (kNN), decision tree, and Support Vector Classifier (SVC) were employed to detect patterns and forecast employee turnover. The kNN method, as proposed by Guo et al. (2003, 2006), presents a straightforward yet effective classification approach. Addressing the inefficiency and reliance on optimal k selection inherent in traditional kNN methods, Guo et al. (2003) introduced a novel kNN model-based strategy. By constructing a kNN model for the data, this method mitigates the need for specific k values and enhances classification efficiency, as evidenced by experimental results. Furthermore, our study finds inspiration in the work of Liu et al. (2020), who utilized the LSTM-KNN model for real-time flood forecasting, illustrating the effectiveness of combining LSTM with kNN for predictive modeling in dynamic contexts. Additionally, Gao and Li (2020) proposed a KNN model based on Manhattan distance for identifying SNARE proteins, showcasing the broad applicability of KNN models across various domains. Moreover, decision tree models, as exemplified by Syed Nor et al. (2019) and Wang et al. (2022), offer interpretability and simplicity, making them suitable for identifying crucial predictors of employee turnover and predicting pro-environmental behavior, respectively. By drawing insights from these prior studies, our research aims to advance the utilization of machine learning approaches in turnover prediction and contribute to the formulation of effective retention strategies in organizational contexts.

Machine Learning in Predicting Turnover:

Machine learning techniques are increasingly being applied in various fields, including marketing, to analyze and predict user sentiments and behaviors. Hossain et al. (2023a) investigated and predicted users' sentiment toward food delivery apps using machine learning approaches. Similarly, Alarifi, Rahman, and Hossain (2023) focused on predicting and analyzing customer complaints using machine learning techniques, while Hossain and Rahman (2023) applied machine learning to analyze customer sentiment and predict reviews of insurance products. Furthermore, Hossain et al. (2023b) conducted customer sentiment analysis and prediction for halal restaurants using machine learning approaches. These studies demonstrate the versatility and effectiveness of machine learning in extracting insights from large datasets to inform marketing strategies and improve customer satisfaction in various industries. Recent advancements in machine learning () have opened up new opportunities for predicting employee turnover. Studies have demonstrated the efficacy of machine learning algorithms such as logistic regression, decision trees, and support vector machines in predicting turnover intentions based on diverse sets of predictors (Lazzari et al., 2022; Domurath et al., 2023). For instance, Lazzari et al. (2022) compared various classification models and found logistic regression and LightGBM to be the top-performing algorithms for predicting turnover intentions. Similarly, Domurath et al. (2023) developed a contingency model of turnover intent in young ventures using longitudinal data and highlighted the predictive power of machine learning in understanding turnover dynamics in entrepreneurial firms. Moreover, the literature on employee turnover provides valuable insights into the determinants, consequences, and predictive models of turnover. By synthesizing findings from prior studies, this literature review underscores the multifaceted nature of turnover dynamics and highlights the potential of machine learning approaches in predicting turnover. Moving forward, integrating insights from both traditional turnover research and machine learning methodologies can enhance our understanding of turnover dynamics and inform the development of effective retention strategies in organizations.

Recent research has provided further insights into the factors influencing employee turnover across various industries and geographic regions. Mossarah (2023) investigated the medical device industry in the United Arab Emirates (UAE) and identified several factors contributing to high turnover rates, including employee salary, perceived external prestige, and organizational justice. The study emphasized the need for tailored retention strategies to address these factors and mitigate turnover risks in the rapidly growing industry. Similarly, Kumar (2022) focused on the IT sector and examined the impact of talent management practices on employee retention. The findings highlighted the importance of career development opportunities and compensation plans in retaining employees in a competitive business environment. Moreover, Bhat et al. (2023) explored the role of perceived labor market conditions in employees' turnover intentions, emphasizing the influence of external economic factors on organizational membership. The study underscored the need for organizations to consider broader labor market dynamics in their retention strategies. Furthermore, research has examined turnover within specific occupational contexts, shedding light on factors contributing to employee turnover and engagement. Jerez Jerez et al. (2023) investigated turnover within Michelin-starred restaurants in London, emphasizing the impact of employer brand on servers' occupational identity and subsequent turnover intentions. The study highlighted the importance of employer branding in attracting and retaining talent in the hospitality industry. Additionally, Mokoena et al. (2022) developed a talent management framework for a government department in South Africa, emphasizing the mediating role of talent management in enhancing organizational commitment and reducing turnover intentions. The study underscored the importance of strategic talent management practices in the public sector to foster employee retention and organizational performance. In summary, these additional studies offer valuable insights into the multifaceted nature of employee turnover and retention across different industries and organizational contexts. By considering factors such as salary, career development opportunities, external economic conditions, and employer branding, organizations can develop targeted retention strategies to address turnover challenges and foster a more engaged and stable workforce.

Another crucial aspect influencing turnover intentions is the leadership style adopted within organizations. Donkor, Appienti, and Achiaah (2022) investigated the impact of transformational leadership style on turnover intentions among employees in state-owned enterprises in Ghana. Their findings revealed an inverse relationship between transformational leadership and turnover intention, suggesting that effective leadership practices can mitigate turnover risks. Moreover, the study highlighted the mediating role of organizational commitment, indicating that employees' commitment to the organization partially mediates the relationship between transformational leadership and turnover intentions. This underscores the significance of fostering a supportive and empowering leadership environment to cultivate organizational commitment and reduce turnover intentions among employees.

Drawing upon insights from Donkor et al. (2022) and other seminal studies such as that by Hom, Lee, Shaw, and Hausknecht (2017), which trace the evolution of turnover theory and research, it becomes evident that effective leadership plays a pivotal role in shaping employees' attitudes and behaviors related to turnover. Transformational leadership, characterized by inspiration, motivation, and trust, emerges as a potent predictor of reduced turnover intentions, aligning with the broader literature on the positive impact of supportive leadership practices on employee retention. As organizations navigate the complexities of turnover management, investing in leadership development and promoting transformational leadership behaviors can serve as a strategic imperative to foster employee engagement, commitment, and ultimately, organizational stability.

Method

Data Collection and Preprocessing: The dataset for this study was sourced from Kaggle (www.kaggle.com). With a total of 4653 valid respondent records, the dataset was meticulously reviewed and prepared prior to analysis to guarantee its adequacy for machine learning tasks. Preprocessing steps included a thorough check for missing values and the conversion of categorical variables into a numeric format suitable for modeling purposes. Following these investigations, it was determined that the dataset met the criteria essential for the current study, ensuring its reliability and appropriateness for subsequent analyses.

Feature Selection and Engineering: The dataset contains various attributes related to employees, including their educational backgrounds, work history, demographics, and employment-related factors. Features such as "Education," "Joining Year," "City," "Payment Tier," "Age," "Gender," "Ever Benched," and "Experience in Current Domain" were selected as predictors. These features were chosen based on their relevance to employee turnover prediction and their potential to capture meaningful patterns in the data.

Exploratory Data Analysis (EDA): Exploratory data analysis was conducted to gain insights into the distribution and relationships among the different features. Visualization techniques such as histograms, box plots, and correlation matrices were employed to analyze the data and identify any notable trends or patterns.

Model Selection and Implementation: Three machine learning algorithms were selected for predicting employee turnover: K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM). These algorithms were chosen for their ability to handle classification tasks and capture complex relationships in the data. The scikit-learn library in Python was utilized for implementing these machine learning models.

Model Evaluation and Performance Metrics: The performance of each machine learning model was evaluated using standard classification metrics, including accuracy score, precision, recall, and F1-score. Additionally, confusion matrices were employed to visualize the model's predictions compared to the actual class labels, providing insights into the model's classification accuracy and potential areas for improvement.

Class Imbalance Handling: To address class imbalance in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to oversample the minority class (employees leaving). This technique helps mitigate the impact of class imbalance and enhances the model's ability to generalize to unseen data.

Software and Tools: Data analysis and modeling were performed using Jupyter Notebook, an open-source web application that allows for interactive data analysis and visualization. Python programming language and various libraries such as pandas, matplotlib, seaborn, scikit-learn, and imbalanced-learn were utilized for data manipulation, visualization, and machine learning implementation.

Conclusion: This methodology facilitated the identification of patterns and prediction of employee turnover using machine learning approaches. By leveraging the dataset's attributes and employing various modeling techniques, valuable insights were gained into factors influencing employee retention and turnover, contributing to informed decision-making in human resources management.

Results and Discussion

Table-1 provides a comprehensive examination of factors influencing employee retention within the organization, spanning diverse dimensions such as Education, Joining Year, City, Payment Tier, Experience in Current Domain, and Gender (presented in figure 1). Each factor is meticulously analyzed to uncover its nuanced impact on employee retention rates. In terms of Education, the table reveals distinct retention patterns across different educational backgrounds. While employees with Bachelor's degrees exhibit a relatively higher retention rate of 53.13%, retention rates for employees with Masters and PHD degrees are notably lower, at 9.61% and 2.88% respectively, indicating potential variations in career trajectories and aspirations based on educational attainment. Joining Year sheds light on temporal trends in retention, with retention rates fluctuating over time. Notably, earlier cohorts show higher retention rates compared to more recent ones, suggesting potential shifts in organizational culture, policies, or job market dynamics influencing employee tenure. Geographical factors, as depicted by City, also play a significant role in retention dynamics. Bangalore emerges as the city with the highest retention rate (35.10%), followed by Pune (13.52%) and New Delhi (16.99%), highlighting the influence of regional factors such as job market conditions, cost of living, and organizational opportunities.

Table-1: Employee Retention Analysis by Various Factors

Factors		Leave or Not			
	Total	Stay		Leave	
		Number	Percentage	Number	Percentage
Education					
Bachelors	**3601**	2472	53.12701483	1129	24.26391575
Masters	**873**	447	9.606705351	426	9.155383623
PHD	**179**	134	2.879862454	45	0.967117988
Total	**4653**	3053	65.61358263	1600	34.38641737
Joining Year					
2012	**504**	395	8.489146787	109	2.342574683
2013	**669**	445	9.56372233	224	4.814098431
2014	**699**	526	11.30453471	173	3.718031378
2015	**781**	463	9.950569525	318	6.834300451
2016	**525**	408	8.768536428	117	2.51450677
2017	**1108**	811	17.4296153	297	6.382978723
2018	**367**	5	0.107457554	362	7.779926929
Total	**4653**	3053	65.61358263	1600	34.38641737
City					
Bangalore	**2228**	1633	35.09563722	595	12.78744896
New Delhi	**1157**	791	16.99978508	366	7.865892972
Pune	**1268**	629	13.51816033	639	13.73307544
Total	**4653**	3053	65.61358263	1600	34.38641737
Payment Tier					
1	**243**	154	3.309692671	89	1.912744466
2	**918**	368	7.908875994	550	11.82033097
3	**3492**	2531	54.39501397	961	20.65334193
Total	**4653**	3053	65.61358263	1600	34.38641737
Experience In Current Domain					
0	**355**	231	4.964539007	124	2.664947346
1	**558**	370	7.951859016	188	4.04040404
2	**1087**	688	14.78615947	399	8.57511283
3	**786**	487	10.46636579	299	6.425961745
4	**931**	634	13.62561788	297	6.382978723
5	**919**	631	13.56114335	288	6.189555126
6	**8**	6	0.128949065	2	0.042983022
7	**9**	6	0.128949065	3	0.064474533
Total	**4653**	3053	65.61358263	1600	34.38641737
Gender					
Female	**1875**	991	21.29808726	884	18.99849559
Male	**2778**	2062	44.31549538	716	15.38792177
Total	4653	3053	65.61358263	1600	34.38641737

Additionally, payment tier underscores the importance of compensation in retention strategies, with higher payment tiers associated with higher retention rates. Tier 3 employees exhibit the highest retention rate (54.40%), while Tier 1 employees have the lowest (3.31%), emphasizing the critical role of fair and competitive compensation in employee satisfaction and retention. Experience in Current Domain reveals intriguing patterns, with employees at different experience levels exhibiting varying retention rates. Employees with 2 years of experience demonstrate the highest retention rate (14.79%), suggesting potential career progression or stability factors at play. Gender dynamics also influence retention, with male employees showing a higher overall retention rate (44.32%) compared to female employees (21.30%). These findings underscore the importance of addressing gender-related factors such as workplace culture, advancement opportunities, and work-life balance in fostering inclusive and equitable retention practices.

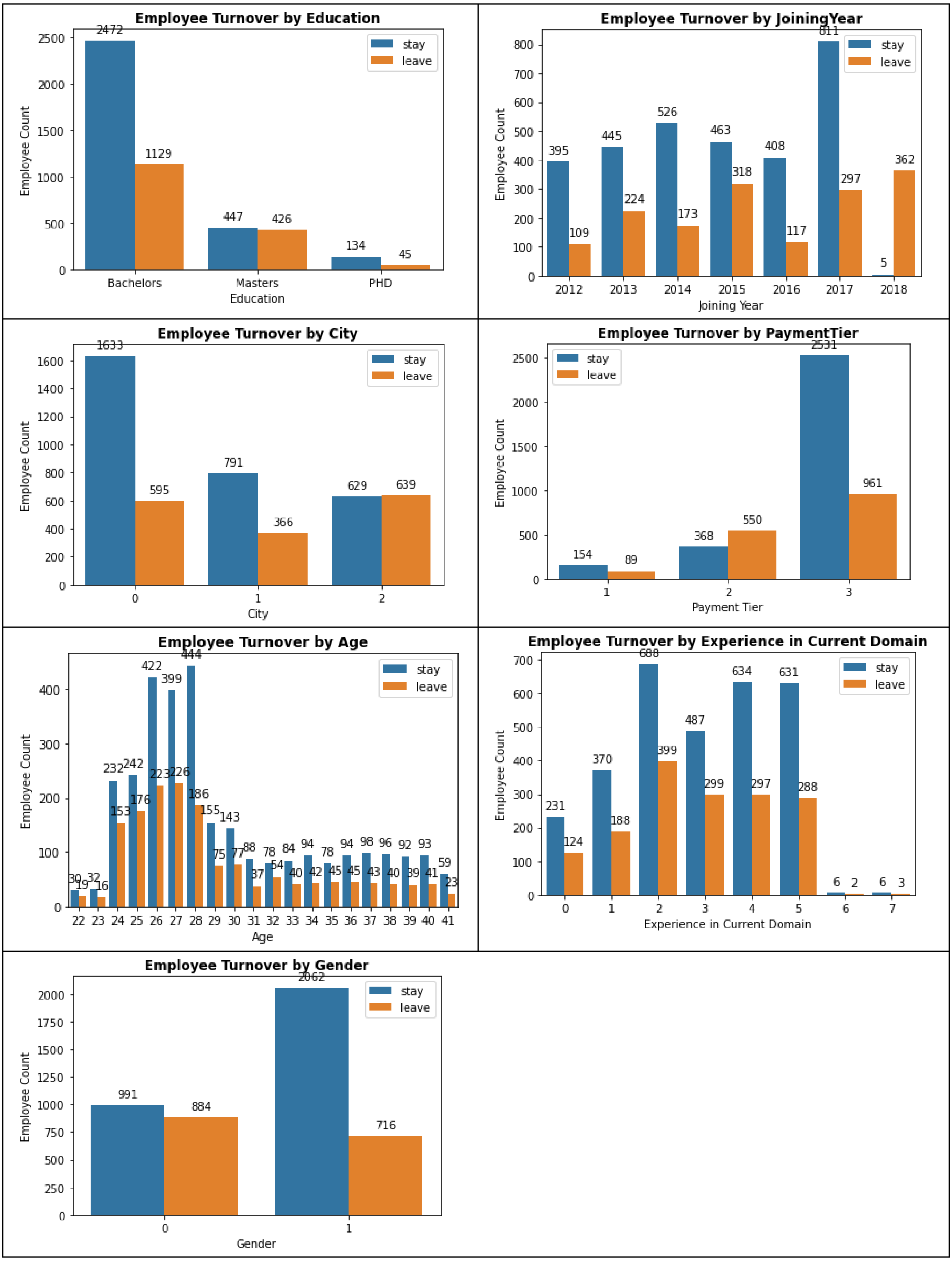

Figure-1: Employee turnover across various characteristics

By comprehensively analyzing these factors, organizations can gain valuable insights into the complex interplay of individual, organizational, and environmental factors shaping employee retention, enabling them to develop targeted strategies to enhance employee satisfaction, engagement, and long-term commitment. Additionally, the relationships among the factors examined in Table-1 provide valuable insights into the complex dynamics influencing employee retention within the organization. Educational attainment, for instance, may correlate with payment tier, as employees with higher levels of education often seek commensurate compensation. Furthermore, there may be interplay between joining year and experience in the current domain, as employees who have been with the organization longer may have accrued more experience. Additionally, the geographical location, as represented by city, may influence payment tiers, with cities characterized by higher costs of living potentially requiring higher compensation to retain talent. Gender dynamics may intersect with factors such as education and experience, revealing potential disparities in career progression and retention between genders. Moreover, trends in retention rates across different cities and over time may offer insights into regional and temporal variations in organizational stability, growth, and workforce dynamics. By understanding these relationships, organizations can tailor retention strategies to address the specific needs and preferences of diverse employee demographics and optimize overall retention outcomes.

Tabel 2 showed that the KNN (K-Nearest Neighbors) model achieved an accuracy score of approximately 75.40%. This score indicates the proportion of correctly classified instances out of the total dataset. For class 0 (representing employees staying), the model demonstrated a precision of 80%, indicating that 80% of the instances predicted as staying were indeed staying. The recall for class 0 was 83%, indicating that 83% of the actual instances of staying were correctly identified by the model. The F1-score, which is the harmonic mean of precision and recall, was 81% for class 0. For class 1 (representing employees leaving), the precision was 66%, meaning that 66% of the instances predicted as leaving actually left the company. The recall for class 1 was 62%, indicating that 62% of the actual instances of leaving were correctly identified by the model. The F1-score for class 1 was 64%. The macro average of precision, recall, and F1-score was approximately 73%, while the weighted average was approximately 75%. These averages provide a summary of the model's overall performance across both classes, considering factors such as class imbalance. Overall, the KNN model demonstrated moderate performance in predicting employee turnover, with balanced precision and recall scores. Further analysis and optimization may be necessary to improve its effectiveness in practical applications.

Table 2: Performance Evaluation of KNN Model in Predicting Employee Turnover

Metric	Precision	Recall	F1-score	Support
Class 0	0.80	0.83	0.81	602
Class 1	0.66	0.62	0.64	329
Accuracy			0.75	931
Macro Avg	0.73	0.72	0.73	931
Weighted Avg	0.75	0.75	0.75	931

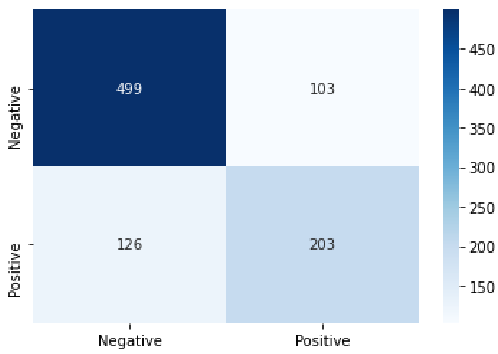

Figure 2: Confusion matrix of KNN model

Figure 2 presents the confusion matrix generated by the KNN model for the prediction of employee turnover. This matrix offers a comprehensive overview of the model's performance, juxtaposing the actual and predicted classifications of instances. In the context of this matrix, the "Predicted Negative" column delineates instances the model identified as negative (indicating employees staying). Here, it correctly identified 499 instances out of the total that were indeed negative, while incorrectly classifying 126 positive instances as negative. Conversely, the "Predicted Positive" column signifies instances predicted as positive (reflecting employees leaving). Among these, the model accurately pinpointed 203 instances as positive, while incorrectly classifying 103 negative instances as positive. Through this detailed breakdown, the confusion matrix illuminates the efficacy of the KNN model in categorizing instances, facilitating an insightful evaluation of its predictive performance.

Table 3 presents the performance evaluation metrics of the Decision Tree model in predicting employee turnover. This evaluation provides insights into the model's effectiveness in classifying instances into the respective classes of employees staying (class 0) and leaving (class 1). For class 0 (employees staying), the precision of the Decision Tree model was 83%, indicating that 83% of the instances predicted as staying were indeed staying. The recall for class 0 was 84%, signifying that 84% of the actual instances of staying were correctly identified by the model. The F1-score, representing the harmonic mean of precision and recall, was 84% for class 0. For class 1 (employees leaving), the precision was 70%, implying that 70% of the instances predicted as leaving actually left the company. The recall for class 1 was 69%, indicating that 69% of the actual instances of leaving were correctly identified by the model. The F1-score for class 1 was 70%. The accuracy of the Decision Tree model was 79%, reflecting the proportion of correctly classified instances out of the total dataset. Additionally, the macro average of precision, recall, and F1-score was approximately 77%, while the weighted average was approximately 79%. These averages provide a summary of the model's overall performance across both classes, considering factors such as class imbalance. Overall, the Decision Tree model demonstrated favorable performance in predicting employee turnover, with balanced precision and recall scores. Further analysis and optimization may be necessary to enhance its predictive capabilities in real-world scenarios.

Table 3: Performance Evaluation of Decision Tree Model in Predicting Employee Turnover

Metric	Precision	Recall	F1-score	Support
Class 0	0.83	0.84	0.84	602
Class 1	0.70	0.69	0.70	329
Accuracy			0.79	931
Macro Avg	0.77	0.76	0.77	931
Weighted Avg	0.79	0.79	0.79	931

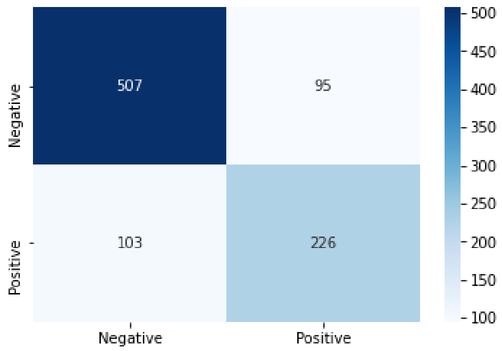

Figure 3: Confusion matrix of Decision Tree Model

Figure 4 illustrates the confusion matrix derived from the Decision Tree model's predictions for employee turnover. This matrix offers a comprehensive breakdown of the model's performance by comparing its predictions against the actual class labels. In this context, the "Predicted Negative" column indicates instances classified by the model as negative (representing employees staying). Among these predictions, the "Actual Negative" subset signifies instances correctly identified by the model as negative, totaling 507 instances. However, there were 95 instances incorrectly classified as negative when they were actually positive (employees leaving). Conversely, the "Predicted Positive" column represents instances predicted by the model as positive (indicating employees leaving). Among these, the "Actual Negative" subset denotes instances misclassified by the model as positive, amounting to 103 instances. On the other hand, the "Actual Positive" subset reflects instances correctly identified by the model as positive, totaling 226 instances out of the total instances that were actually positive. Through this detailed breakdown, the confusion matrix provides valuable insights into the Decision Tree model's performance, showcasing both its strengths and areas for improvement in accurately predicting employee turnover.

Table 4 presents the performance evaluation metrics of the Support Vector Machine (SVM) model in predicting employee turnover. These metrics offer insights into the model's effectiveness in classifying instances into the respective classes of employees staying (class 0) and leaving (class 1). For class 0 (employees staying), the precision of the SVM model was 71%, indicating that 71% of the instances predicted as staying were indeed staying. The recall for class 0 was 66%, signifying that 66% of the actual instances of staying were correctly identified by the model. The F1-score, representing the harmonic mean of precision and recall, was 69% for class 0. For class 1 (employees leaving), the precision was 45%, implying that 45% of the instances predicted as leaving actually left the company. The recall for class 1 was 52%, indicating that 52% of the actual instances of leaving were correctly identified by the model. The F1-score for class 1 was 48%. The accuracy of the SVM model was 61%, reflecting the proportion of correctly classified instances out of the total dataset. Additionally, the macro average of precision, recall, and F1-score was approximately 58%, while the weighted average was approximately 61%. These averages provide a summary of the model's overall performance across both classes, considering factors such as class imbalance. In summary, the SVM model demonstrated moderate performance in predicting employee turnover, with relatively lower precision and recall scores compared to other models evaluated. Further analysis and optimization may be necessary to enhance its predictive capabilities in real-world scenarios.

Table 4: Performance Evaluation of SVM Model in Predicting Employee Turnover

	Precision	Recall	F1-Score	Support
0	0.71	0.66	0.69	602
1	0.45	0.52	0.48	329
Accuracy			0.61	931
Macro avg	0.58	0.59	0.58	931
Weighted avg	0.62	0.61	0.61	931

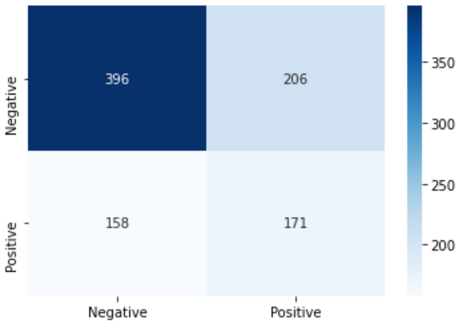

Figure 4: Confusion matrix of SVM Model

Figure 4 presents the confusion matrix generated from the predictions of the Support Vector Machine (SVM) model for employee turnover. This matrix offers a comprehensive overview of the model's performance by juxtaposing its predictions against the actual class labels. In the context of the matrix, "Predicted Negative" denotes instances classified by the SVM model as negative, indicating employees predicted to stay. "Predicted Positive" represents instances classified by the model as positive, indicating employees predicted to leave. Rows correspond to the actual classes, where "Actual Negative" indicates instances actually negative (employees staying), and "Actual Positive" denotes instances actually positive (employees leaving). Examining the matrix, the investigation observes, for instances actually belonging to the negative class (employees staying), the SVM model accurately predicted 396 instances as negative and misclassified 206 instances as positive. Regarding instances actually belonging to the positive class (employees leaving), the model correctly predicted 171 instances as positive but misclassified 158 instances as negative. The confusion matrix offers valuable insights into the SVM model's performance, highlighting both its correct and incorrect classifications. It serves as a pivotal tool for assessing the model's predictive accuracy and efficacy in identifying employee turnover in a real-world scenario.

Models' performance

Comparing the performance of the KNN, Decision Tree, and SVM models in predicting employee turnover reveals distinct characteristics and areas of strength and weakness for each model. The KNN model achieved an accuracy score of approximately 75.40%, with precision, recall, and F1-score values of 80%, 83%, and 81%, respectively, for class 0 (employees staying), and 66%, 62%, and 64%, respectively, for class 1 (employees leaving). The model demonstrated balanced precision and recall scores, indicating its ability to correctly classify instances from both classes. However, further optimization may be necessary to improve its overall performance.

In comparison, the Decision Tree model outperformed the KNN model with an accuracy score of 79%. It exhibited higher precision, recall, and F1-score values for both classes, with precision, recall, and F1-score values of 83%, 84%, and 84%, respectively, for class 0, and 70%, 69%, and 70%, respectively, for class 1. The Decision Tree model showcased robust performance in correctly identifying instances from both classes, suggesting its suitability for predicting employee turnover. On the other hand, the SVM model demonstrated moderate performance, with an accuracy score of 61%. It exhibited lower precision, recall, and F1-score values compared to the KNN and Decision Tree models, with precision, recall, and F1-score values of 71%, 66%, and 69%, respectively, for class 0, and 45%, 52%, and 48%, respectively, for class 1. The SVM model showed weaker performance in correctly classifying instances, indicating the need for further refinement.

Overall, while all three models provided insights into employee turnover prediction, the Decision Tree model emerged as the most effective based on its higher accuracy and performance metrics. However, further experimentation and optimization may be required to enhance the predictive capabilities of all models for real-world applications.

Conclusion

In conclusion, this study illuminates the intricate dynamics of employee turnover through the lens of machine learning techniques and comprehensive data analysis. By examining various factors such as education, tenure, demographics, and work-related attributes, we have provided valuable insights into the predictors of employee retention and departure within organizational settings. Through the application of KNN, Decision Tree, and SVM models, we have demonstrated the efficacy of predictive analytics in identifying patterns and forecasting turnover. The findings underscore the importance of proactive retention strategies and data-driven decision-making in managing human capital effectively.

Applications

The insights gleaned from this study have significant implications for organizational practice and human resource management. Firstly, by identifying the key determinants of employee turnover, organizations can develop targeted interventions to address retention challenges and enhance workforce stability. From implementing personalized career development programs to refining compensation structures and fostering inclusive workplace cultures, the findings offer actionable strategies to mitigate turnover rates and promote employee engagement and loyalty. Secondly, the predictive models developed in this study can serve as valuable tools for HR practitioners and organizational leaders in forecasting turnover trends, optimizing resource allocation, and devising evidence-based retention initiatives. By harnessing the power of data analytics, organizations can proactively identify at-risk employees, intervene early, and cultivate a more resilient and committed workforce.

Limitations and Future Research Directions

Despite the insights gained, this study is not without its limitations. Firstly, the dataset used in this study may not capture all relevant factors influencing employee turnover, such as job satisfaction, interpersonal relationships, and external market conditions. Future research could explore additional variables and incorporate data from multiple sources to enhance the predictive accuracy of turnover models. Secondly, while machine learning algorithms offer valuable predictive capabilities, their interpretability may be limited, making it challenging to understand the underlying mechanisms driving turnover predictions. Addressing this challenge requires the development of interpretable machine learning techniques and model-agnostic approaches to enhance transparency and facilitate decision-making. Finally, the generalizability of the findings may be limited to the specific context and population under study. Future research could replicate the study across diverse industries and organizational settings to validate the robustness of the findings and identify context-specific retention strategies. Overall, by addressing these limitations and embracing a multidisciplinary approach, future research can advance our understanding of employee turnover dynamics and inform more effective retention practices in organizations.

References

Abolade, D. A. (2018). Impact of employees’ job insecurity and employee turnover on organisational performance in private and public sector organisations. *Studies in Business and Economics*, *13*(2), 5–19. https://doi.org/10.2478/sbe-2018-0016

Alarifi, G., Rahman, M. F., & Hossain, M. S. (2023). Prediction and Analysis of Customer Complaints Using Machine Learning Techniques. *International Journal of E-Business Research (IJEBR), 19*(1), 1-25

Al-Suraihi, W. A., Samikon, S. A., Al-Suraihi, A.-H. A., & Ibrahim, I. (2021). Employee Turnover: Causes, Importance and Retention Strategies. *European Journal of Business and Management Research*, *6*(3), 1–10. https://doi.org/10.24018/ejbmr.2021.6.3.893

Ambriz-Perez, H., Acha, E., & Fuertc-Esquivel, C. R. (1999). Advanced SVC models for newton-raphson load how and newton optimal power flow studies. *IEEE Power Engineering Review*, *19*(2), 46.

Ambriz-Pérez, H., Acha, E., & Fuerte-Esquivel, C. R. (2000). Advanced SVC models for Newton-Raphson load flow and Newton optimal power flow studies. *IEEE Transactions on Power Systems*, *15*(1), 129–136. https://doi.org/10.1109/59.852111

Bhat, M. A., Tariq Khan, S., & Rainayee, R. A. (2023). Assessment of perceived labor market conditions in employees' turnover intention model – mediation and moderation analyzes. *PSU Research Review*, *7*(1), 1–32. https://doi.org/10.1108/PRR-05-2020-0017

Buchheit, B., Schneider, E. N., Alayan, M., Dauth, F., & Strauss, D. J. (2022). Motion Sickness Prediction in Self-Driving Cars Using the 6DOF-SVC Model. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 13582–13591. https://doi.org/10.1109/TITS.2021.3125802

Dogru, T., McGinley, S., Sharma, A., Isık, C., & Hanks, L. (2023). Employee turnover dynamics in the hospitality industry vs. the overall economy. *Tourism Management*, *99*. https://doi.org/10.1016/j.tourman.2023.104783

Domurath, A., Taggar, S., & Patzelt, H. (2023). A contingency model of employees' turnover intent in young ventures. *Small Business Economics*, *60*(3), 901–927. https://doi.org/10.1007/s11187-022-00629-2

Donkor, F., Appienti, W. A., & Achiaah, E. (2022). The Impact of Transformational Leadership Style on Employee Turnover Intention in State-Owned Enterprises in Ghana. The Mediating Role of Organisational Commitment. *Public Organization Review*, *22*(1). https://doi.org/10.1007/s11115-021-00509-5

Duda, J., & Žůrkova, L. (2013). Costs of employee turnover. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *61*(7), 2071–2075. https://doi.org/10.11118/actaun201361072071

Gao, X., & Li, G. (2020). A KNN Model Based on Manhattan Distance to Identify the SNARE Proteins. *IEEE Access*, *8*, 112922–112931. https://doi.org/10.1109/ACCESS.2020.3003086

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2888*, 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, *102*(3), 530–545. https://doi.org/10.1037/apl0000103

Hossain, M. I., Mia, M. A., & Hooy, C. W. (2023). Employee turnover and the credit risk of microfinance institutions (MFIs): International evidence. *Borsa Istanbul Review*, *23*(4), 936–952. https://doi.org/10.1016/j.bir.2023.04.001

Hossain, M. S., & Rahman, M. F. (2023). Customer Sentiment Analysis and Prediction of Insurance Products' Reviews Using Machine Learning Approaches. FIIB Business Review, 12(4), 386-402. https://doi.org/10.1177/23197145221115793

Hossain, M.S., Begum, H., Rouf, M.A. and Sabuj, M.M.I. (2023a). Investigation and prediction of users' sentiment toward food delivery apps applying machine learning approaches, *Journal of Contemporary Marketing Science*, 6(2), 109-127. https://doi.org/10.1108/JCMARS-12-2022-0030

Hossain, M.S., Rahman, M.F., Uddin, M.K. and Hossain, M.K. (2023b). Customer sentiment analysis and prediction of halal restaurants using machine learning approaches, *Journal of Islamic Marketing*, 14(7), 1859-1889. https://doi.org/10.1108/JIMA-04-2021-0125

Iqbal, J., Asghar, A., & Asghar, M. Z. (2022). Effect of Despotic Leadership on Employee Turnover Intention: Mediating Toxic Workplace Environment and Cognitive Distraction in Academic Institutions. *Behavioral Sciences*, *12*(5). https://doi.org/10.3390/bs12050125

Jerez Jerez, M. J., Melewar, T. C., & Foroudi, P. (2023). The Effect of Waiters' Occupational Identity on Employee Turnover Within The Context of Michelin-Starred Restaurants. *Journal of Hospitality and Tourism Research*, *47*(7), 1215–1243. https://doi.org/10.1177/10963480211034903

Jothi, Dr. V. L., A, N., S, N. S., & K, S. (2020). Crop Yield Prediction using KNN Model. *International Journal of Engineering Research & Technology*, *8*(12), 4–7. Retrieved from www.ijert.org

Kaggle. (n.d.). Employee dataset. Kaggle. https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset

Kumar, S. (2022). The impact of talent management practices on employee turnover and retention intentions. *Global Business and Organizational Excellence*, *41*(2), 21–34. https://doi.org/10.1002/joe.22130

Lazzari, M., Alvarez, J. M., & Ruggieri, S. (2022). Predicting and explaining employee turnover intention. *International Journal of Data Science and Analytics*, *14*(3), 279–292. https://doi.org/10.1007/s41060-022-00329-w

Ling, X. (2024). Design of an enterprise accounting system based on decision tree model. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns.2023.2.00051

Liu, M., Huang, Y., Li, Z., Tong, B., Liu, Z., Sun, M., … Zhang, H. (2020). The applicability of lstm-knn model for real-time flood forecasting in different climate zones in China. *Water (Switzerland)*, *12*(2). https://doi.org/10.3390/w12020440

Mokoena, W., Schultz, C. M., & Dachapalli, L. A. P. (2022). A talent management, organisational commitment and employee turnover intention framework for a government department in South Africa. *SA Journal of Human Resource Management*, *20*. https://doi.org/10.4102/sajhrm.v20i0.1920

Montazeri, M., Montazeri, M., Ahmadian, L., Zahedi, M. J., & Beygzadeh, A. (2023). An Intelligent Diagnosis of Liver Diseases using Different Decision Tree Models. *Journal of Kerman University of Medical Sciences*, *30*(2), 113–116. https://doi.org/10.34172/jkmu.2023.18

Mossarah, A. (2023). Investigating factors that impact employee turnover in the medical device industry in the United Arab Emirates. *Social Sciences and Humanities Open*, *7*(1). https://doi.org/10.1016/j.ssaho.2023.100492

Olubiyi, O., Smiley, G., Luckel, H., & Melaragno, R. (2019). A qualitative case study of employee turnover in retail business. *Heliyon*, *5*(6). https://doi.org/10.1016/j.heliyon.2019.e01796

Salleh, A. M. M., Omar, K., Aburumman, O. J., Mat, N. H. N., & Almhairat, M. A. (2020). The impact of career planning and career satisfaction on employees' turnover intention. *Entrepreneurship and Sustainability Issues*, *8*(1), 218–232. https://doi.org/10.9770/jesi.2020.8.1(14)

Skelton, A. R., Nattress, D., & Dwyer, R. J. (2020). Predicting manufacturing employee turnover intentions. *Journal of Economics, Finance and Administrative Science*, *25*(49), 101–117. https://doi.org/10.1108/JEFAS-07-2018-0069

Syed Nor, S. H., Ismail, S., & Yap, B. W. (2019). Personal bankruptcy prediction using decision tree model. *Journal of Economics, Finance and Administrative Science*, *24*(47), 157–170. https://doi.org/10.1108/JEFAS-08-2018-0076

Wang, Q., Kou, Z., Sun, X., Wang, S., Wang, X., Jing, H., & Lin, P. (2022). Predictive Analysis of the Pro-Environmental Behaviour of College Students Using a Decision-Tree Model. *International Journal of Environmental Research and Public Health*, *19*(15). https://doi.org/10.3390/ijerph19159407

Cite this article:

Aham Edward Kanuto (2024). Identifying Patterns and Predicting Employee Turnover Using Machine Learning Approaches. *International Journal of Science and Business, 36*(1), 20-35. DOI: https://doi.org/10.58970/IJSB.2373

Retrieved from http://ijsab.com/wp-content/uploads/2373.pdf

Published by

IJSAB International