SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 4799-4808
Open Access. Online & Print

DOI: 10.5281/zenodo.12426525

ANALYSIS OF TALENT ACQUISITION AND TALENT RETENTION IN HOSPITALITY INDUSTRIES IN TAMIL NADU USING MACHINE LEARNING ALGORITHM (RANDOM FOREST)

Guru P[1], Bhuvaneswari J[2*], Amirtha A[3], Swathi S[4], Sindhu P[5], Sathiya N R[6]

[1,2,3,4,5]*Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India*

Received: **15/11/2025**
Accepted: **13/03/2026**

Corresponding Author: Bhuvaneswari J
(bhuvaneswari@pmu.edu)

ABSTRACT

The research study analyzes talent acquisition and retention within the hospitality sector in Tamil Nadu through a machine learning model with a special emphasis to the random forest algorithm. Hospitality business is labor intensive and the success of an organization majorly is determined by the supply and their retention of skilled employees. Nevertheless, employee turnover, skills shortage, and rising competition have been troubling the industry, which makes it important to employ data-driven human resource management approaches. The study is quantitative in nature and involves primary numerical data (collected using structured questionnaires) of employees and HR professionals of hospitality establishment in the state of Tamil Nadu. The most important variables examined are the effectiveness of recruitment, the compensation and benefits, the training and development, the career growth opportunities, the work-life balance, job satisfaction, and organizational support. These are variables that are analyzed to determine the impact on employee retention. The application of the machine learning algorithm of the Random Forest is used to classify employees, regarding their probability of being retained, and the most important predictors of retention results. To ascertain the reliability and robustness of the model, standard measures of accuracy, precision, recall, F1-score, and ROC_AUC are used to assess the model. The Random Forest model can therefore offer high predictive accuracy and feature importance analysis as it handles complex and multi-dimensional HR data. It is hoped that the research will prove the relevance of machine learning in improving human resource analytics as applied to the hospitality industry. The results should provide useful, evidence-based information to the management and human resource practitioners of the hotels to enhance the talent acquisition process, retention policies, employee turnover, and quality of services provided. In the end, this study will make its contribution to the increased use of machine learning in human resource management and will aid with sustainable workforce development of the hospitality industry in Tamil Nadu.

KEYWORDS*:* *Talent Acquisition, Talent Retention, Hospitality Industry, Four- and Five-Star Hotels, Tamil Nadu.*

1. INTRODUCTION

Tourism, business travel, medical tourism, and rich cultural heritage are some of the key drivers that have made the hospitality business an important part of the economy of Tamil Nadu. Being a service-based industry most of its success is pegged on the quality, dedication, and consistency of the human resource [1]. Hotels always aim at giving out the best guest experiences and such goal can only be fulfilled with skilled, motivated and service oriented employees. Nevertheless, there are endemic problems in capturing and retaining such talent within the industry because of the high turnover rates, working long hours, competition in the labor market, and changing employee-related expectations. Talent recruitment and talent retention has thus turned out to be a strategic consideration to hospitality organizations. Good practices within recruitment methods make sure that the right technical abilities, attitude, and service orientation are employed, whereas good retention methods keep the companies steady in terms of workforce, minimize the recruiting expenses and keep the service standards. Compensation and benefits, training and development, career growth opportunities, work life balance, job satisfaction and organizational support are some of the factors, which have a significant impact on whether employees will join and stay in an organization or not [2]. It is also necessary to understand the relationship between these factors to come up with effective human resource policies. Conventional HR methods are usually based on descriptive analysis and managerial hunch that might not adequately reflect the multifaceted interaction between several workforce variables. As data analytics evolved, machine learning technologies can be used to provide powerful tools to study large numerical data and expose hidden patterns that impact employee behavior [3]. The Random Forest algorithm is one of the most popular methods, and it is especially effective to work with multidimensional data, has high predictive accuracy, and can determine the most significant variables by measuring feature importance.

It is on this basis that the current study will examine the practices of talent acquisition and talent retention in the hospitality sector of Tamil Nadu based on the machine learning algorithm of the Random Forest. Combining HR analytics and machine learning, the study aims at forecasting the retention outcomes and determining the significant drivers, which can impact the workforce stability [4]. It is believed that the findings will offer factual information to the hotel management and the HR professionals to improve recruitment strategies, reinforce retention policies and overall organizational performance within the hospitality industry in Tamil Nadu.

2. RELATED WORK

The literature highlights that the competitiveness in the Indian hospitality business is based on effective recruitment and retention recruitment. Research shows that machine learning could be used to determine the performance of retention policies, including competitive wages, career development, rewards, safe work conditions, flexible work hours, and ongoing training. These strategies are affected by demographic factors and proper retention strategies lower the turnover intention of employees. Employee satisfaction and retention are known to be the result of compensation, appreciation, safety at workplace, professional growth and development, work-life balance and healthy relationships at work.

Kaja Mytheen et al., in (2025) carried research among Malaysian corporate bodies that have adopted AI in HRM, the authors have studied the impact of AI-enhanced HR functions on the data-driven decision making (DDM). The research had a good response rate of 83.5 per cent of 376 participants and used partial least squares Structural Equation Modeling (PLS-SEM) to identify the direct and indirect relationship between workforce planning, learning and development, employee engagement and retention, performance management, and talent acquisition. The results confirm that AI-based HR activities play a critical role in enhancing information-driven decision-making, which implies the importance of advanced technologies in streamlining HR operations and organizational performance. Li et al., (2024) conducted a review to exploratory research and descriptive study that explores the present situation in the use of AI and machine learning in talent acquisition. It mentions the importance of ML algorithms in enhancing the recruitment procedures, analyzes their advantages, and addresses the related issues and shortcomings. The results show that even though there is an increasing use of technology in talent acquisition, there are a number of questions and obstacles that do not contribute to the widespread application. The paper ends with real-life recommendations and suggestions that stakeholders should adopt to optimize the use of AI and ML in the recruitment practices.

Sinha, B.C et al., (2025) examined how the digitalisation of the recruitment process, attractive pay, training, and employee engagement programs of five-star hotels in Kerala affect the satisfaction of employees with the organisation through a sample size of 290 respondents. It was found that digital recruiting services and high wages are the most successful means of accessing talent, and they have high mean scores (4.4 and 4.3). Training programs were observed to be very effective in increasing employee engagement, motivation and job satisfaction. The paper also places importance on such practices as innovative and region-specific HR practices and states that to enhance the levels of talent acquisition and retention in the hospitality industry advanced HR technologies and cross-regional comparisons are necessary. Afna, A.S., (2025) shown that the hotel sector uses a wide range of retention initiatives, such as high and competitive wages and compensation, career development, recognition, safe working conditions, flexible working schedules, job security, and life-long learning. These strategies show a demographic difference in their influence on the employees and there is a small negative correlation between the intention of employees to quit and the retention practices adopted. Compensation, appreciation, workplace safety, and professional development are all very strong factors that affect job satisfaction. The research also emphasizes on the significance of welfare programs in work-life balance and building professional relationships amongst hospitality industry employees to enhance employee retention. Ramasamy, K (2025) analyzes the current recruitment and retention procedures in Pradeep Stainless India Private Limited in order to determine the problems and recommend viable recommendations to achieve long-term employee satisfaction. The research conducted an analysis of aspects affecting recruitment channels, employer branding, compensation, career development, and workplace culture using primary data on surveys and interviews with HR professionals, managers and employees. The results reveals that good leadership, good working environment, and remuneration are some of the critical motivators of retention. The paper also emphasizes how the engagement of the employees, their skill development, and open communication have contributed to the loyalty of the workforce and presents real-world ideas of how better talent management and organizational performance can be achieved by means of these factors.

Hospitality business in Tamil Nadu is very competitive in terms of service-oriented business with the quality of the guest service being directly related to the competency and stability of the working force. Based on the literature review above, it can be noted that the gap in the research is as follows: High Employee Turnover: Tamil Nadu Hospitality establishments have been experiencing high employee turnover that interferes with the service quality, and this raises the training and recruitment expenses.

- Lack of Ability to acquire Competent Talent: Hotels find it hard to attract talented individuals with the necessary technical abilities, service orientation and cultural assimilation.
- Weak Assessment of HR Practices: The current talent acquisition and retention strategies are evaluated through conventional approaches that cannot reflect the intricate nature of the workforce.
- Absence of Data-Driven Decision Making: Advanced machine learning and analytical tools are underexploited to learn and forecast the results of employee retention.
- Lack of Empirical ML-Based Study in Hospitality HR: The use of numerical analysis algorithms such as Random Forest to perform talent management in relation to the numerical HR data has not been studied extensively in the hospitality industry of Tamil Nadu.

3. PROPOSED WORK

The proposed work is aimed at examining talent acquisition practice, and talent retention practice in the hospitality industry in Tamil Nadu through a machine learning perspective. Considering the human-oriented environment of the hospitality industry and a high turnover rate, the research will be designed to take a data-based approach to facilitate strategic human resource decisions [9]. Primary numerical data will be obtained by the use of structured questionnaires to the employees and HR professionals, including such factors as the effectiveness of recruitment, compensation and benefits, training and development, career growth opportunities, work life balance, job satisfaction, and organizational support.

Random Forest machine learning algorithm will be used in analyzing the data collected and predicting the results of employee retention. The reason behind selecting this algorithm is that it can work with multi-dimensional numerical variables, as well as model non-linear relationships and determine the most influential variables on retention by analyzing feature importance. Standard measures like accuracy, precision, recall, F1-score and ROC-AUC will be used to test model performance [10].

The proposed work will provide important factors that lead to employee retention and also evaluate the effectiveness of current talent acquisition practices in the hospitality sector. The study results are projected to offer practical and evidence-based solutions to the hotel management and HR professionals to minimize the levels of attrition, increase workforce stability, and to improve the quality of services [11]. Finally, the study will aim at showing the feasibility of machine learning methods in the human resource management of the hospitality industry in Tamil Nadu. The flow of proposed work is shown below in figure 1.

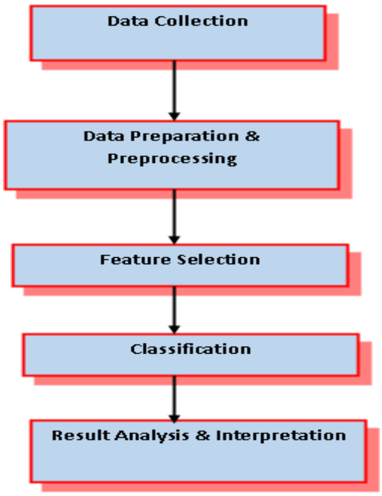

Figure 1: Work flow of proposed work

3.1 Data Collection

In order to collect data on this study, the structured questionnaire will be used in order to obtain numerical and perceptual data on the topic of talent acquisition and talent retention practices in hospitality organizations within Tamil Nadu. The survey shall be conducted personally to employees and HR managers of hotels, resorts and restaurants by visiting and where a necessity arises, online form shall be used to administer the survey. The questionnaire will be subdivided into various sections that will address the demographics of employees, talent acquisition approaches, talent retention approaches, reasons why organisation retains talent, employee engagement, training and development, job performance, and the issues that HR managers have when acquiring and retaining talents. The answers to the majority of questions will be evaluated on a five-point Likert scale (Strongly Agree to Strongly Disagree), which would make it possible to transform the perceptions into numerical data that can be analyzed by machines. Also, ranking-type questions represent managerial problems of talent management. They will be seeking the prior consent of the hospitality organizations and will assure that respondents will be confidentially informed to participate forthright. The responses obtained will be summarized into a dataset, coded into numbers and ready to undergo preprocessing and analysis with the help of the Random Forest algorithm to determine the critical retention drivers among the employees.

3.2 Data Preparation and Preprocessing

A systematic data preprocessing process is accepted to make sure that the survey information obtained among hospitality employees can be processed in machine learning using the Random Forest algorithm [12]. Because the questionnaire is composed of demographic variables, Likert-scale questions, and questions with a ranking, the raw answers have to be converted to structured numerical characteristics prior to the model implementation.

i. Numerical Coding of Responses

All Likert responses (Sections II- VII) are translated into numerical numbers to measure perceptions of employees:

Table 1: Sample coding for Responses

Response	Code
Strongly Agree (SA)	5
Agree (A)	4
Neutral (N)	3
Disagree (DA)	2
Strongly Disagree (SD)	1

Label encoding is utilized to transform categorical demographic variables, i.e. gender, education, job position, income level, and residential area, with each category corresponding to the specific integer. Table 1 above shows the sample coding for responses collected through survey.
Questions in Section IX and Section VI are ranked in numerical form and are directly used.

ii. Handling Missing Values

In order to sustain the integrity of the datasets, the missing values are handled in the following way:
In case the number of missing responses in each record is below 10, then they are filled with the mean of that variable, which can be computed by making use of the following formula:

$$x_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{kj} \quad (1)$$

Where x_{ij} represents the missing value in column j and n represents the number of valid responses.

The records having over 10% missing responses are dropped to eliminate bias.

iii. Feature Construction & Dimensionality Reduction

Section-wise composite scores are also calculated as an average of responses in each section as opposed to examining all individual questions as individual inputs. In a section consisting of m questions.

$$S = \frac{1}{m}\sum_{i=1}^{m} q_i \quad (2)$$

with q_i denoting the coded response to each question.

This brings out the following features.

- Acquisition_Score
- Retention_Strategy_Score
- Retention_Factor_Score
- Engagement_Score
- Training_Score
- Performance_Score

This procedure minimizes noise as well as increases the interpretability of models.

iv. Target Variable Encoding

The target variable is the Employee Retention Rate which is coded as Low, Moderate and High (0,1 and 2).

v. Data Normalization

In order to normalize feature ranges, Min-max normalization is used:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

where X is the initial value and X ' is the normalized value of between 0 and 1.

vi. Outlier Detection

The Z-score method is used to find the outliers:

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

Where μ is the mean and σ standard deviation. $|z| > 3$ are considered outliers and removed.

This preprocessing approach will guarantee that qualitative survey data is automatically converted into quantitative and noise-cancellable and machine-learning-ecological starts, enhancing the predictive power and reliability of the Random Forest model in assessing talent acquisition and retention in the Tamil Nadu hospitality sector [13].

3.3 Feature Selection

The Chi-Square (χ^2) test of independence is used in this research as a filter-based feature selection method before the application of the Random Forest classifier. Because the data is the result of a structured questionnaire of Likert-scale answers and categorical demographic data, and the desired outcome (Employee Retention Rate) is categorical in nature (Low, Moderate, High), the (χ^2) test is suited well to assess the statistical significance of each of these input features and the desired outcome (retention).

The main goal of applying the (χ^2) method is to retain only the variables that are significantly connected to the employee retention which in turn will reduce the dimensions, irrelevant attributes and then enhance the classification model efficiency and accuracy.

The Chi-Square test is used to understand whether two categorical variables have a significant relationship or not by comparing the observed frequencies and the projected frequencies with the assumption that they are independent.

The hypotheses are the following:

H_0: The level of feature f and retention of employees Y are independent.

H_1: Y is a dependent feature and employee retention f.

A feature that is important is one in which the null hypothesis is rejected.

The χ^2 statistic is computed as

$$\chi^2 = \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (5)$$

Where:

- O_{ij} = Observed frequency of cell i, j.
- E_{ij} = Expected frequency in cell i, j <|human|>E ij = Expected frequency in cell i,j

The expected frequency is computed as:

$$E_{ij} = \frac{Row\ Total_i\ X\ Column\ Total_j}{Grand\ Total} \quad (6)$$

A contingency table is created concerning the Employee Retention Rate categories and each questionnaire feature (e.g., Acquisition Score, Retention Strategy Score, Engagement Score, demographic variables) against each other. The χ^2 measure is calculated on each feature at a time.

A p-value is then obtained from the χ^2 distribution table:

- If $p < 0.05$, the feature has a statistically significant association with retention and is selected.
- If $p \geq 0.05$, the feature is considered less relevant and removed from the model input.

The χ^2 distribution table then yields a p-value:

- In case p is below 0.05, the feature is chosen as it is statistically significantly associated with retention.
- When p is greater than or equal to 0.05, the feature is said to be less relevant and it is excluded as an input to the model.

Table 2: Feature Selection using p-value

S.No	Feature/ Variable	χ^2 Value	p-Value	Decision
1	Acquisition_Score	42.58	0.0001	Selected
2	Retention_Strategy_Score	55.73	0	Selected
3	Retention_Factor_Score	61.29	0	Selected
4	Engagement_Score	38.14	0.0003	Selected
5	Training_Score	33.87	0.0012	Selected
6	Training_Score	29.45	0.0038	Selected
7	Working Experience	18.62	0.0170	Selected
8	Monthly Income	16.25	0.0224	Selected
9	Job Position	14.73	0.0416	Selected
10	Education Level	9.12	0.1675	Rejected
11	Gender	3.84	0.4280	Rejected
12	Marital Status	4.29	0.3682	Rejected
13	Residential Area	6.15	0.1884	Rejected

Based on the table 2 above Features like Retention Factors, Retention Strategies, Acquisition Practices, Engagement, and Training have very high χ 2 but a very small p-values, which means that it is associated with employee retention. The demographic variables such as Gender, Marital status, and Education exhibit weak correlation, and are dropped off the model. The inputs to the Random Forest classifier are only passed as statistically significant features.

Chi-Square analysis showed that Retention Factor Score ($\chi 2 = 61.29$, $p = 0.001$), Retention Strategy Score ($\chi 2 = 55.73$, $p = 0.001$), and Acquisition Score (2 = 42.58, $p = 0.001$) are the most important predictors of employee retention. The gender and marital status were considered statistically insignificant and were not included in the analysis.

3.4 Classification

The algorithm used in this research is the Random Forest algorithm as a supervised classification method to forecast the rate of employee retention (Low, Moderate, High) using statistically sampled features in the survey data. Random Forest is an ensemble learning algorithm, which builds several decision trees and uses the results of the decision trees to create a more precise and strong classification. The figure 2 below shows the flow chart for the classification process.

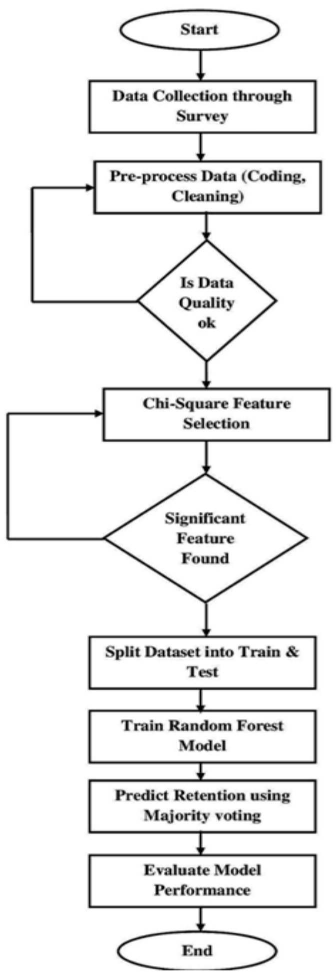

Figure 2: Flow chart of Classification using RF

i. Input to the model

Following the preprocessing and Chi-Square feature selections, the resulting dataset comprises significant predictors, which include Acquisition_Score, Retention_Strategy_Score, Retention_Factor_Score, Engagement _Score, Training_Score, Performance_Score and Selected demographic variables (e.g., experience, income, job position)

The target variable is:

$$Y \in \{0,1,2\} \quad (7)$$

In this 0 = Low Retention, 1 = Moderate Retention, 2 = High Retention.

ii. Bootstrap Sampling

Random Forest takes bootstrap samples (subsets) of the training set of size N. A decision tree grows on each of the subsets.

iii. Random Feature Selection

On every node split in a tree, rather than using all features, m (by random) features out of the total p features are picked.

$$m < p \quad (8)$$

This randomness enhances the correlation between trees as well as generalization.

iv. Node Splitting using Gini Impurity

Every decision tree is broken down into nodes according to Gini Impurity that quantifies the ability of a feature to divide the classes.

The Gini impurity at a node is:

$$Gini = 1 - \sum_{k=1}^{K} p_k^2 \quad (9)$$

in which p_k is the fraction of sample in class k at that node and $K = 3$ in this work.

v. Tree Construction

Every tree is trained to grow to the fullest depth without pruning, which provides a variety of decision boundaries to be classified.

vi. Majority Voting for Final Classification

Suppose that there are T trees in the forest. A tree is a predictor of a class that is assigned to a particular employee record. Majority voting is the final determination of prediction:

$$Y = mode\{h_1(x), h_2(x), \ldots\ldots h_T(x)\} \quad (10)$$

$h_t(x)$ is the prediction of the t^{th} tree.

Algorithm 1: Classification using RF

Step 1: Feed the input data

- Gather responses of the survey and generate dataset D.
- Numerical code (1 5) Likert responses.
- Nominalize the demographic variables that are categorical.
- Treat cases where there are no values and eliminate invalid cases.
- Select Features (e.g. Chi-Square) → get final set X.

Step 2: Split the dataset

Divide the data into a training set D_{train} and testing set D_{test}.

$$D = D_{train} \cup D_{test} \quad (11)$$

Step 3: Initialize Random Forest

- Instantiate decision tree number T.
- Specify the random features of each node m, such that m<p (total features).

Step 4: Build the forest (Training)

For each tree $t = 1$ to T:

- Sample N records of D_{train} train with replacement to form a bootstrap sample D_t.
- Grow a decision tree using D_t

At each node

At each node

i. Choose randomly m features out of the p features.

ii. Split quality where this feature is selected is evaluated by Gini impurity.

$$Gini = 1 - \sum_{k=1}^{K} p_k^2 \quad (12)$$

iii. Choose the feature and divide the one that gives the greatest decrease in impurity.

iv. Split further until some terminating criterion is met (e.g. maximum depth, minimum samples per node).

v. Store the trained tree h_t

Step 5: Classification (Testing)

For each test instance $\in D_{test}$:

Get the widely spread predictions of all trees.

$$h_1(x), h_2(x), \ldots. h_T(x) \quad (13)$$

Assign the final predicted class using majority voting:

$$y' = mode\{h_t(x)\} \forall t = 1\ to\ T \quad (14)$$

4. Performance Analysis

The proposed system performance analysis is aimed to determine the effectiveness of the proposed system in predicting the employee retention level (Low, Moderate, High) based on the processed survey data of the employees of the hospitality industry in Tamil Nadu. Once the data has been preprocessed and the Chi-Square selection of features is made, the narrowed dataset is divided into training and testing subsets (usually, 80:20). RF model is trained with the help of the training data and tested with the help of the unseen test data to estimate the predictive capacity.

4.1 Quality Parameters

In order to determine the performance of the model, conventional measures of classification are employed:

i. Accuracy

Accuracy is a measure of the general validity of the model to predict retention classes.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

ii. Precision

Precision represents the number of accuracy cases of the predicted retention cases.

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

iii. Recall

Recall is the capacity of the model to distinguish the correct cases of retention.

$$Recall = \frac{TP}{TP+FN} \quad (17)$$

iv. F1-Score

F1-Score is a harmonic mean between Precision and Recall and it is particularly effective when the number of classes is distorted.

$$F1 = 2X \frac{Precision\ X\ Recall}{Precision+Recall} \quad (18)$$

v. Confusion Matrix

The confusion matrix is generated to illustrate the correct and incorrect predictions that are made in the three retention classes (Low, Moderate, High). It assists in the appreciation of misclassification patterns.

vi. ROC Curve and AUC

The ROC curve is a measure of the trade-off between True Positive Rate and False Positive Rate with multi-class classification. An increase in the AUC values means that the model has a better discriminative power.

4.2 RESULT ANALYSIS

i. Confusion Matrix of Proposed work

The three-class confusion matrix of retention prediction model in Figure 3 (Low, Moderate, High) has shown that the model has performed well with an overall accuracy of 91.6% on 500 survey responses. The model was able to identify 137 low-retention, 178 moderate-retention, and 143 high-retention employees which represented 458 correct predictions out of 500. There are low misclassifications and they are evenly distributed amongst classes meaning that the model is well balanced and does not attribute any particular retention category. The findings indicate that the model is effective especially in the prediction of the moderate retention category and highly reliable in identifying low and high retention worker, which proves the strength of the proposed method of classifications.

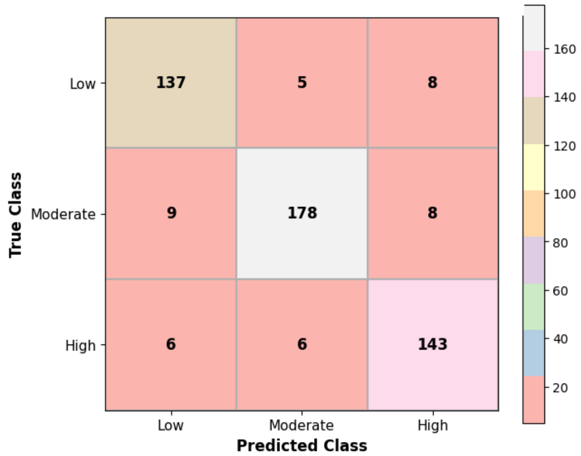

Figure 3: Confusion Matrix

ii. Overall Performance

In this chapter, a comparative table 3 presents the results of using the proposed Random Forest model on the data related to employee retention in comparison to some of the existing classification techniques (common ranges are reported in HR analytics and survey-based ML research).

Table 3: Overall performance Measure

S.No	Method	Accuracy	Precision	Recall	F1-Score
1	Logistic Regression	78	77	76	76
2	Decision Tree	83	82	81	81
3	K-NN	81	80	79	79
4	SVM	87	86	85	85
5	Random Forest (Proposed)	91.6	91.7	91.6	91.6

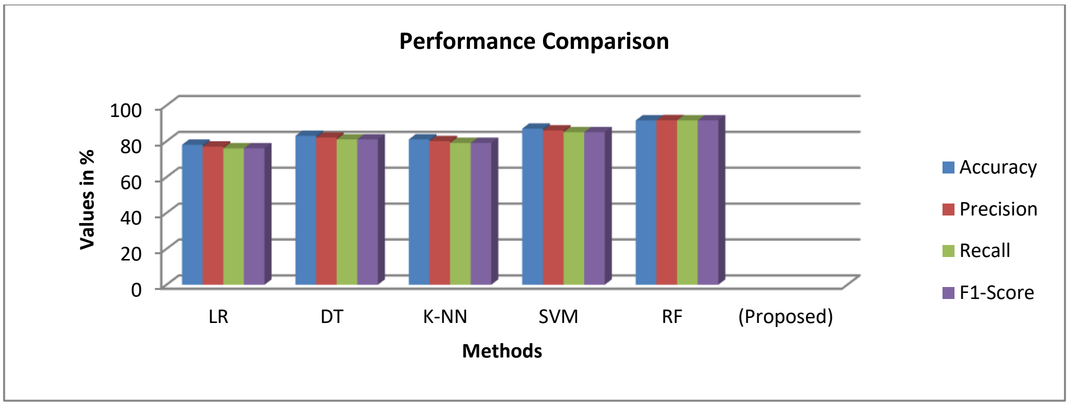

Figure 4: Overall comparison

Performance comparison graph in Figure 4 shows the comparison of five methods of classification, including Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbour (K-NN), Support Vector Machine (SVM), and the proposed method of classification, Random Forest (RF), in terms of Accuracy, Precision, Recall, and F1-Score. It is evident that the performance between LR and RF has improved gradually as indicated in the graph. The lowest scores are registered by Logistic Regression, then moderate improvements are registered by decision tree and K-NN. SVM shows superior predictive capacity among larger metric values, however, the suggested Random Forest model operates higher than any of the current approaches in all four measurement metrics. The relative high bars in Random Forest show that it is more efficient in classifying employee retention levels accurately and recalling it at equal measures, which proves its validity and dependability in conducting an analysis of survey-based hospitality HR data.

iii. Accuracy Analysis

In figure 5, the training and validation accuracy indicate a naturalistic learning behavior of the Random Forest model with different epochs. The training accuracy is a gradually increasing curve with slight fluctuations and this is a good sign that the model is slowly learning as it goes through the data.

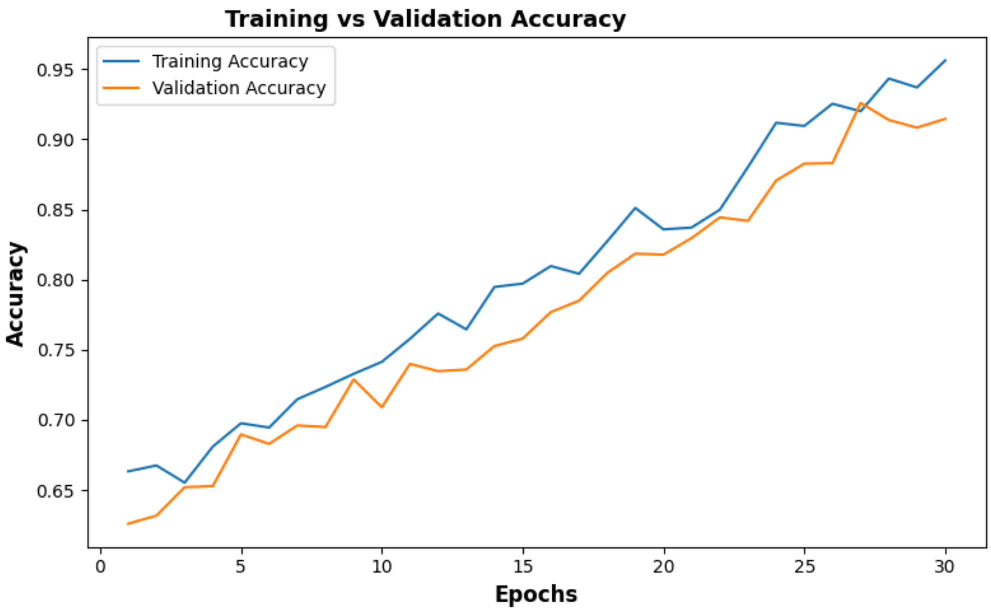

Figure 5: Accuracy analysis on training and validation

The validation accuracy also exhibits a similar tendency and is also near the training curve during the process, indicating that there is no serious overfitting to the process and the validation is quite satisfactory. These minor differences between the two curves are the expected behavior of normal learning which proves that the model is stable and even more effective in its predictive capabilities with unknown data.

iv. Loss Analysis

Figure 6 below illustrates that the training and validation loss curve is gradually reducing with increase in the epochs, a fact that indicates that the model is learning effectively on the data. The training loss decreases continuously with slight variations whereas the validation loss also reduces in a similar way and follows the training curve without being too far. Such similar performance indicates that this model is not overfitting and generalizing to untrended data. It is the progressive decline of the loss that will testify to the increased model performance and stability during the learning process.

Figure 6: Loss analysis on training vs Validation

5. CONCLUSION

This paper has provided a quantitative method to examine the 4- and 5-star hospitality organizations in Tamil Nadu to determine the talent acquisition and talent retention using a Random Forest classification model. The information derived in surveys through demographic, acquisition, retention, engagement, training and job performance were pre-processed and filtered using Chi-Square feature selection that was later classified. The model proposed had a high accuracy of 91.6, balanced accuracy with regards to recall and F1-score, and equal accuracy among Low, Moderate, and High retention classes, and thus, strong predictive ability. It was demonstrated by comparing results with traditional algorithms, including Logistic Regression, Decision Tree, KNN, and SVM that the Random Forest performed better than these algorithms in multi-class classification of survey-based data with fewer overfitting and more stability. The findings affirm that the offered practice is an effective way to determine critical variables that contribute to employee retention and give significant information to HR managers working in the hospitality industry to develop improved talent management practices.

REFERENCES

1. A. Singh, "Quality of work-life practices in the Indian Hospitality sector," in *Human Relations Management in Tourism*, 2024, pp. 208–224.
2. W. Chang, J. Busser, and A. Liu, "Authentic leadership and career satisfaction: the meditating role of thriving and conditional effect of psychological contract fulfillment," International Journal of Contemporary Hospitality Management, vol. 32, no. 6, pp. 2117–2136, Jun. 2020, doi: 10.1108/IJCHM-06-2019-0551.
3. M. N. Khuong, N. T. N. Khai, P. T. Huyen, N. V. H. Thuong, and N. T. M. Phuong, "Factors affecting employee performance through a mediation of job satisfaction. An empirical study of hospitality industry in Ho Chi Minh City, Vietnam," *HCMCOUJS - economics and business administration*, vol. 6, no. 1, Mar. 2020, doi: 10.46223/HCMCOUJS.econ.en.6.1.110.2016.
4. Kaja Mytheen, B.K., Jeyakumar, M., Ramasamy, K., Mani, G., Jayamurugan, P., & Ananthan, B. (2025). Machine learning based strategies for managing employee retention: determining factors in hospitality industry. *Indonesian Journal of Electrical Engineering and Computer Science*.
5. Li, W., Hussain, W.M., Jia, X., Na, M., & Alam, S.S. (2024). Analyzing the Impact on Talent Acquisition and Performance Management: HR and Data Analysis. J. Organ. End User Comput., 36, 1-30.
6. Sinha, B.C., Sharma, N., Sharma, P., Sharma, A., Sinha, H.S., Shukla, B., Sharma, U., Agrawal, B., & Kumari, A. (2025). An analysis of the talent acquisition process using artificial intelligence and machine learning tools among hiring professionals in India. Multidisciplinary Reviews.
7. Afna, A.S. (2025). Assessing the Impact of Talent Acquisition on Employee Retention in Star Hotels in Kerala. Journal of Information Systems Engineering and Management.
8. Ramasamy, K, (2025). Machine learning based strategies for managing employee retention: determining factors in hospitality industry. Indonesian Journal of Electrical Engineering and Computer Science. https://doi.org/10.11591/IJEECS.V38.I3.PP1652-1660.
9. J. He, A. M. Morrison, and H. Zhang, "Improving millennial employee well-being and task performance in the Hospitality industry: the interactive effects of HRM and responsible leadership," *Sustainability*, vol. 11, no. 16, p. 4410, Aug. 2019, doi: 10.3390/su11164410.
10. Geetha A et al., (2025). A study on talent acquisition and retention strategies in Pradeep Stainless India (P) Ltd., MEPZ, *International Journal of Research Publication and Reviews*, Vol (6), Issue (4), April (2025), Page – 14892-14906.
11. M. N. Khuong, N. T. N. Khai, P. T. Huyen, N. V. H. Thuong, and N. T. M. Phuong, "Factors affecting employee performance through a mediation of job satisfaction. An empirical study of hospitality industry in Ho Chi Minh City, Vietnam," *HCMCOUJS - economics and business administration*, vol. 6, no. 1, Mar. 2020, doi: 10.46223/HCMCOUJS.econ.en.6.1.110.2016.
12. E. Gamor, E. F. Amissah, A. Amissah, and E. Nartey, "Factors of work-family conflict in the hospitality industry in Ghana," *Journal of Human Resources in Hospitality and Tourism*, vol. 17, no. 4, pp. 482–501, 2018, doi: 10.1080/15332845.2017.1328263.
13. W. Chang, J. Busser, and A. Liu, "Authentic leadership and career satisfaction: the meditating role of thriving and conditional effect of psychological contract fulfillment," *International Journal of Contemporary Hospitality Management*, vol. 32, no. 6, pp. 2117–2136, Jun. 2020, doi: 10.1108/IJCHM-06-2019-0551.