INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
Vol 15, Issue 5, (2025) E-ISSN: 2222-6990

Machine Learning Solutions for Talent Loss: Predictive Insights and Strategic Countermeasures

Hanan Aldowah[1], Sohail Ahmed Shahani[2], Amira Saif[3], Ahmed Alsswey[4]
[1]School of Management- Universiti Sains Malaysia (USM), Penang, Malaysia, [2]Deparetement of Chemical Engineering, Universiti Teknologi PETRONAS (UTP), Perak, Malaysia, [3]Centre for instructional Technology and Multimedia- Universiti Sains Malaysia (USM), Penang, Malaysia, [4]Department of Multimedia Technology, Faculty of Architecture and Design, Al-Zaytoonah University of Jordan, Amman, Jordan

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v15-i5/25064 DOI:10.6007/IJARBSS/v15-i5/25064
Published Date: 25 May 2025

Abstract
In the context of global economic globalization, intense market competition across various industries emphasizes the crucial role of talent as core enterprise resources and competitive factors. This paper discusses the widespread issue of talent loss, which incurs significant operational challenges, including increased employment costs, diminished service quality, and reduced customer satisfaction. Utilizing a human resources dataset from the Kaggle platform, we employ a random forest algorithm to develop a talent loss prediction model. This model is designed to forecast potential resignations, allowing enterprises to implement timely retention strategies. Our analysis identifies key factors contributing to talent turnover and assesses the effectiveness of various countermeasures. The proposed strategies include integrating a talent-focused approach into corporate strategies, enhancing performance management systems, creating competitive compensation frameworks, developing internal training programs, and fostering a people-oriented culture. By predicting and understanding the dynamics of talent loss, this model provides a valuable tool for enterprise managers to mitigate its impacts effectively. The findings offer theoretical insights into preventing talent loss and reducing its adverse effects on company performance.
Keywords: Talent Loss, Countermeasures, Machine Learning

Introduction
In the era of economic globalization, the intense market competition has elevated the importance of talent as a core resource and a key competitive factor for enterprises. As career choices diversify and working methods become more specialized, the demand for talent has grown more complex, compelling industries to integrate talent strategies into their broader

development plans (Luna-Arocas & Danvila-del-Valle, 2022). Despite these efforts, talent loss remains a pervasive challenge globally, leading to increased employment costs, diminished after-sales service, and reduced customer satisfaction, which critically impact business operations. Particularly in high-tech industries, the competition for capital, emerging technologies, and skilled labor is fierce. Here, effective talent management is paramount as human resources, economic capital, and technological information form the three pillars of operational success. In such a dynamic and competitive business landscape, human resource management becomes a pivotal element for sustained enterprise success (Rustiawan et al., 2023). The ongoing global competition in economic development, technological advancement, and production is fundamentally a battle for acquiring and retaining talent.

This study focuses on the acute problem of talent loss which affects production and enterprise development significantly. Talent mobility, when managed well, optimizes workforce structure and boosts corporate vitality. However, when talent mobility primarily results in talent outflow, it signifies a loss, particularly of key personnel crucial for driving the enterprise forward (Vejsiu, 2019). This form of talent loss can be particularly detrimental when it involves highly specialized employees who cover essential professional areas. Talent loss not only hampers operational efficiency and productivity but also undermines team stability and cohesion, leading to increased costs in training and recruitment, and potentially damaging the company’s brand image and competitive position (Akinyomi, 2016; Shaikh et al., 2020). Furthermore, it sends a negative signal to the market, reflecting poorly on internal management and employee welfare, thereby affecting the company's market standing (Jing, 2022). Thus, understanding the dynamics of talent loss, its causes, and effects, and developing effective strategies to mitigate these issues is crucial.

Through this research, we aim to deepen understanding of why high-performing talents voluntarily leave organizations, how management can effectively prevent this loss, and the broader impacts of talent loss on organizational operations and performance. This paper utilizes a dataset from the Kaggle platform to apply a random forest algorithm in building a talent loss prediction model, exploring main causes and proposing countermeasures to prevent talent loss. Ultimately, this research seeks to provide actionable insights for enterprise managers to formulate effective talent retention strategies and ensure the long-term sustainability of their organizations. By anticipating potential resignations and understanding the underlying factors, the study will offer strategic interventions to mitigate these occurrences and enhance corporate retention capabilities.

The motivation for this study stems from the pressing need to support enterprises in proactively addressing talent loss before it disrupts operations and performance. As organizations increasingly rely on data-driven decision-making, the use of machine learning offers new opportunities to gain predictive insights into workforce dynamics. This research contributes to the field by integrating a predictive modeling approach, specifically using a random forest algorithm, with strategic human resource management practices. By combining predictive analytics with actionable countermeasures, the study provides a practical tool for anticipating employee resignations and delivers a comprehensive framework of organizational responses to reduce talent attrition. These contributions are particularly valuable for enterprises seeking to enhance talent retention strategies through evidence-based and technologically informed solutions.

Background of Study

The unstoppable trends of economic globalization have resulted in the dynamic flow of both capital and human resources, presenting both opportunities and challenges. This global exchange has offered major employers a broader range of talent choices, yet they face the dilemma of balancing talent shortages against surpluses of human capital. As the competition intensifies, the acquisition and retention of skilled talent have become paramount for securing a competitive edge and ensuring long-term organizational success. According to Singh (2021), effective talent management is now a critical source of competitive advantage and adds significant value to organizations, especially in the context of the current economic slowdown. Despite this high demand for skilled workers, many enterprises, particularly in traditional manufacturing, still prioritize production and sales metrics over strategic talent management. These outdated practices place undue performance pressure on employees and undermine the critical role of talent in organizational development, as noted by Resende & Coelho (2017). Research by Husain (2015) indicates that job satisfaction, trust relationships, and work stress are strongly correlated with employees' intentions to leave, with work pressure often being the most significant contributor. Moreover, the globalization of the economy and knowledge has raised talent's expectations of potential employers, increasing the standards by which they select their workplace. Samašonok (2024) points out that failures to meet job expectations, lack of motivation, and insufficient professional skills are primary personal factors influencing employees' decisions to change jobs. Many companies focus on material rewards but neglect the psychological fulfillment and sense of accomplishment that talents seek at different stages of their careers, failing to address key factors necessary for employee growth and development. Furthermore, the absence of long-term strategic planning for talent and organizational development means that potential high-value employees cannot see a clear direction for growth within their current companies, leading them to explore other opportunities. As the saying goes, "Good birds choose trees to roost, and good days choose masters to work for"; this wisdom underscores the necessity for companies to develop and refine their talent management systems to thrive in the global talent war of the 21st century.

Theoretical Background and Related Studies

Insights into the Macroscopic Phenomenon of Brain Drain

The phenomenon of talent loss within corporations has been extensively studied over recent decades, primarily within the framework of organizational behavior. Scholars have developed theories surrounding the perception of loss intention, incorporating elements such as job satisfaction and organizational commitment into a structured process framework that articulates how talent departs organizations.

James March (1958) approached the problem by linking labor market dynamics with individual employee behaviors, suggesting that the decision to leave an organization often hinges on two primary factors: satisfaction with the current company and the lure of new opportunities elsewhere. Building on this, Steers (1981) expanded the factors influencing talent loss to encompass organizational attributes, individual characteristics, and the broader labor market. He identified job satisfaction, organizational commitment, and work involvement as pivotal in determining the trajectory from the intention to actual talent loss. Steers noted that talent loss could either occur directly from intention or be mediated by the presence of alternative job opportunities.

Further contributing to the discourse, Price (1977) developed an employee turnover model that posits job satisfaction and career opportunities as key intermediaries, influenced by factors like work level, integration, and communication within the company. He demonstrated that job satisfaction negatively correlates with turnover rates—higher satisfaction typically leads to lower turnover and vice versa. However, he also observed that decreased employee satisfaction does not invariably lead to turnover if no alternative job opportunities exist.

Mobley (1977) emphasized the sequential nature of talent loss, where dissatisfaction leads employees to seek and secure alternatives before making a final decision to leave. However, his focus on the step-by-step process of departure was critiqued for not considering the broader, more holistic aspects of employee turnover, which could sometimes include impulsive decisions to leave without a backup plan.

Collectively, these scholars provide a multi-dimensional view of talent loss, highlighting the complex interplay of personal satisfaction, organizational dynamics, and external job market conditions that influence an employee’s decision to stay or leave.

Quantitative Analysis and Research on Brain Drain

The phenomenon of brain drain has been quantitatively analyzed through various models and theories to understand the underlying mechanisms and impacts of talent loss within organizations. Kuck (1978) introduced the Cook curve to illustrate the trajectory of employee creativity, suggesting that creativity peaks within two and a half years of employment before it begins to decline. This model advocates for job rotation and environmental changes to mitigate burnout and reduce turnover. Vascellaro (2007) developed a comprehensive model considering macro and micro elements of talent loss, categorizing variables into environmental, structural, and individual factors. This framework encompasses social opportunities, family responsibilities, work content engagement, training opportunities, emotional experiences, job complexity, autonomy, fairness in workload distribution, compensation, work pressure, and promotion opportunities. Despite its thoroughness, the model's complexity has limited its practical application due to the challenging nature of managing multiple variables simultaneously. Further, Porter et al. (2016) investigated the influence of internal and external networks on employee turnover. Their findings indicate that robust internal networks diminish the likelihood of voluntary turnover, whereas strong external networks can increase the propensity for employees to leave, offering insights into how organizations can better manage internal relations to retain talent. Mihajlov & Mihajlov (2020) proposed a fluctuation model that differentiates between voluntary and involuntary turnover, providing a framework to manage and reduce the costs associated with employee departures more effectively. Additionally, Silpa et al. (2023) developed the Enriched Employee Retention Analysis System (EERAS), combining feature selection models and machine learning techniques to identify and act on the critical factors influencing employee turnover. Chin (2017) utilized quantitative methods such as Pearson's r correlation coefficients and multiple regression analysis to explore the relationships between career development, supervisory relationships, and turnover intentions, establishing a significant negative correlation. This suggests that enhanced career development opportunities and effective supervision are pivotal in reducing the likelihood of employee turnover.

These diverse analytical approaches offer valuable insights into the factors driving brain drain, underscoring the need for targeted strategies to foster employee retention and mitigate the adverse effects of talent loss on organizational health.

Assessing the Economic and Operational Impact of Talent Turnover in Enterprises

The effects of talent turnover on enterprises have been a significant focus of scholarly research, particularly concerning its economic impact and influence on overall workplace dynamics. Scholars like Tapola (2016) have articulated that while employee turnover incurs costs and can degrade service quality within contact centers, it can also yield positive outcomes. In industries where high turnover rates are common, organizations must strive to strike a balance between minimizing turnover's detrimental effects and leveraging its potential benefits. This balanced approach can help companies harness the positive dynamics of personnel changes. Further exploring the repercussions of losing key personnel, Taye & Getnet (2020) suggest that such departures can stifle organizational innovation and disrupt the consistent delivery of services to primary users. Delays in service provision can result, diminishing work efficiency and quality of service, thereby escalating resource wastage as new employees acclimate. Such scenarios also risk eroding public confidence in the organization's operations. Liao Bing et al. (2020) emphasize the oft-overlooked costs associated with employee turnover. They point out that expenses related to hiring and training new staff, coupled with initial inefficiencies and the loss of technological expertise, significantly impact corporate finances. Recognizing these costs underscores the necessity for strategic management of talent turnover to mitigate its negative effects while enhancing organizational resilience and operational efficiency.

Exploring the Root Causes of Brain Drain in Organizations

Extensive research has examined the factors contributing to brain drain, identifying key influences such as workplace conditions, leadership styles, compensation structures, and organizational climate. Saberi et al. (2023) found a significant relationship between job stressors and turnover intention, particularly in high-pressure industries. Similarly, Ahmad & Shahbaz (2017) identified inadequate compensation, lack of recognition, job dissatisfaction, and low motivation as primary drivers of employee departure. Leadership and work environment also play a crucial role. Tangkudung (2015) concluded that poor leadership styles and unfavorable work environments significantly increase employees' intentions to leave. Husain (2015) further emphasized that turnover is closely linked to job satisfaction, trust, job security, and organizational commitment, though person-organization fit showed no strong correlation. Gialuisi & Coetzer (2013) found that unresolved workplace conflicts, limited career advancement, and misaligned job roles contribute to both voluntary and involuntary turnover. Structural factors also influence employee retention. Telly (1969) proposed the inequality theory, where perceived unfair treatment in supervision, workload distribution, and workplace conditions increases turnover rates. Oktavio & Kaihatu (2020) linked high turnover in the hospitality sector to employees' low commitment, driven by uncertain career prospects. Xiao-ming (2008) highlighted unfair wages, weak social security systems, and ineffective management mechanisms as key contributors to talent loss. Social dynamics within the workplace also impact turnover. Feeley et al. (2008) developed a social network-based model, suggesting that employees with strong workplace relationships exhibit lower turnover intentions. Similarly, KHAN et al. (2021) emphasized the impact of workplace harassment, which often goes unreported but significantly influences retention. Several theoretical models provide frameworks for understanding turnover. The turnover model by March & Simon (1958) remains a foundational approach, distinguishing between turnover intention and actual departure. Price (2011) developed the Price-Mueller model, identifying environmental, individual, structural, and mediating variables as determinants of employee turnover. Price (2017) further analyzed communication behaviors in organizations, showing that disengagement begins months before an employee formally resigns. Recent studies have also explored predictive models. Haldorai (2019) applied the Pull-Push-Mooring framework to examine hotel employees' turnover intentions, finding workload to be the strongest predictor. Chaudhary (2022) analyzed the banking industry in Nepal, highlighting leadership awareness, workplace environment, and employee facilities as critical determinants of turnover rates. Jiao Yang & Bai Xinwen (2017) emphasized the role of internal management practices and corporate structure in shaping turnover patterns, noting that different types of enterprises experience varied levels of employee attrition. Overall, these studies illustrate the complex, multifactorial nature of brain drain, reinforcing the need for organizations to adopt comprehensive strategies that address leadership, compensation, workplace climate, and career development to improve employee retention.

Analysis of Countermeasures Taken by Companies Against Brain Drain

Organizations have implemented various strategies to reduce employee turnover and enhance retention. Elsafty & Sayed (2023) found that improving perceived organizational support and work-life balance helps lower stress and turnover rates in small businesses. Syed & Wang (2018) emphasized the role of both financial and non-financial incentives in improving retention, while Saidu (2018) recommended a combination of compensation benefits such as allowances, pensions, and honoraria to enhance employee commitment. Leadership and organizational culture also play a crucial role. Bonsu (2020) suggested that fostering an employee-centered culture with trust, autonomy, and strong workplace relationships enhances creativity and innovation. Hauer et al. (2020) highlighted the importance of transformational leadership, emotional intelligence, and effective communication in motivating employees and reducing turnover. Widayati & Fiorincia (2021) stressed that reducing workload pressure and job insecurity through structured work processes can improve retention. A positive work environment is another key factor. Gunaprasida & Wibowo (2019) proposed that supportive colleagues and supervisors create a comfortable work atmosphere, reducing work-family conflicts, particularly for female employees. Additionally, Mutua (2017) recommended integrating innovation and stability into retention strategies, while Alsayyed & Braiki (2015) emphasized that effectively managing turnover is key to productivity and profitability.

Overall, these studies emphasize that a combination of strategic leadership, fair compensation, supportive workplace culture, and structured job roles is essential in reducing employee turnover and mitigating brain drain.

Classification and Prediction of Brain Drain Using Machine Learning

With the increasing availability of data and advancements in algorithms, machine learning has become a crucial tool for predictive analytics in human resource management. Its automation, speed, and high accuracy enable enterprises to anticipate employee turnover and implement retention strategies. Researchers worldwide have applied machine learning techniques to classify and predict talent loss, offering new insights for workforce management. Nagadevara et al. (2008) used CART to examine factors such as absenteeism, tenure, and demographics in predicting turnover. Saradhi et al. (2011) applied multiple models, including Naïve Bayes, support vector machines (SVM), logistic regression, decision trees, and random forests, recommending SVM as the most effective. Ajit (2016) improved turnover prediction using the XGBoost model, addressing challenges in handling noisy data. Similarly, Sisodia et al. (2017) evaluated five machine learning models, including KNN and Naïve Bayes, to optimize turnover prediction. More advanced approaches have emerged in recent years. Ozmen et al. (2022) introduced a convolutional neural network (CNN)-based hybrid model, ECDT-GRID, to enhance accuracy in predicting employee turnover in retail. Musanga et al. (2022) used feature selection techniques alongside random forest, logistic regression, and gradient boosting to refine predictions. Qadir et al. (2021) applied a deep learning method using B-LSTM, while Jain et al. (2020) proposed an achievement-based employee importance model (AEIM) for classification and retention strategies. Further, Al Abid et al. (2024) integrated a multi-criteria decision-making (MCDM) approach with random forests for classification. Chang (2009) introduced a subset selection method for long-term industry-wide turnover predictions. Chinese scholars, such as Li Jiahao et al. (2021), combined decision trees, random forests, and Adaboost in a Stacking-based LRA model, proving its robustness in turnover prediction.

These studies demonstrate the effectiveness of machine learning in identifying turnover patterns and improving predictive accuracy, offering enterprises valuable tools for proactive talent management.

Theoretical Foundations of Classification Models

Decision Tree

A decision tree is a widely used machine learning model that structures data hierarchically to facilitate classification and regression tasks. It consists of nodes representing decisions, branches indicating decision paths, and leaf nodes denoting outcomes. The algorithm recursively splits data based on selected attributes, ensuring a clear and interpretable decision-making process (Fürnkranz, 2020). The structure of a decision tree is illustrated in Figure 1.

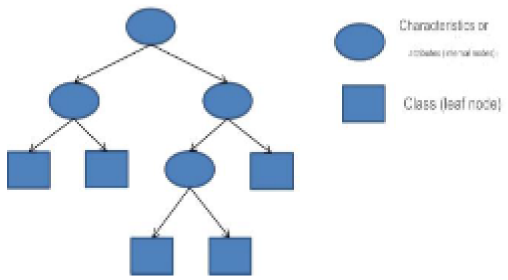

Figure 1. Decision tree

The process begins by assuming all samples belong to a single category. If they do not, the algorithm identifies the most significant feature and recursively partitions the data until reaching a final classification. Decision trees are intuitive and effective for handling both categorical and numerical data, making them suitable for probability analysis, risk assessment, and feasibility studies. However, they are prone to overfitting, sensitive to outliers, and may struggle with high-dimensional data (Song & Lu, 2015). Ensemble techniques such as Random Forest address these limitations.

Random Forest

Principles of the Random Forest Algorithm

The Random Forest algorithm, introduced by Breiman (2001), is an ensemble learning method designed to enhance the accuracy and robustness of decision trees. It constructs multiple decision trees using different training subsets obtained via bootstrapping and aggregates their predictions. This approach reduces overfitting and increases classification accuracy. The general structure of a Random Forest is shown in Figure 2.

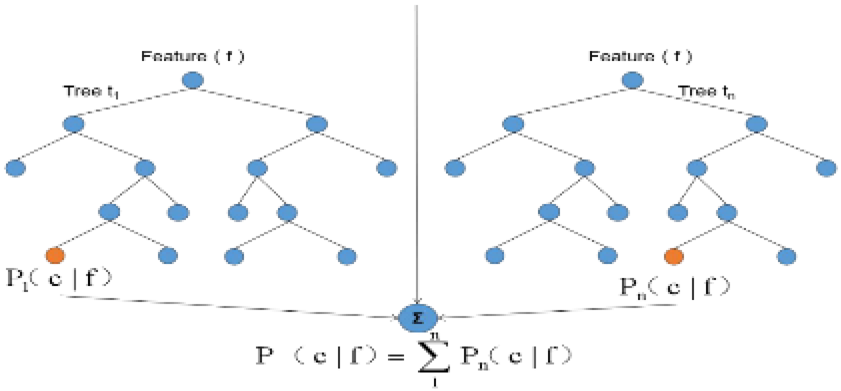

Figure 2. Random forest algorithm

Process of the Random Forest Algorithm

A decision tree in a **Random Forest** is a binary tree that recursively splits data from the **root node** downward. The root node contains the original dataset, which is progressively divided into **left and right nodes** based on the **Gini criterion**, ensuring that each split increases node purity. This process continues until a stopping condition is met. Figure 2.3 shows the tree structure and node splitting process.

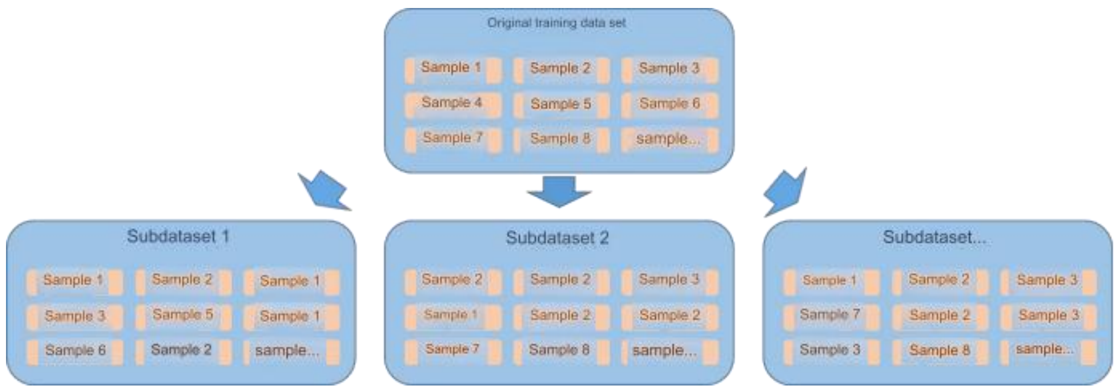

Figure 3. Random forest algorithm sampling process

To improve model robustness, **bootstrap sampling (self-help resampling)** is used to create **m new training subsets**, each containing duplicate samples from the original dataset. These subsets are used to construct **m classification trees**, while **out-of-bag (OOB) data** samples not included in a subset are used for validation. Figure 4 shows the sub-data construction and OOB sample process.

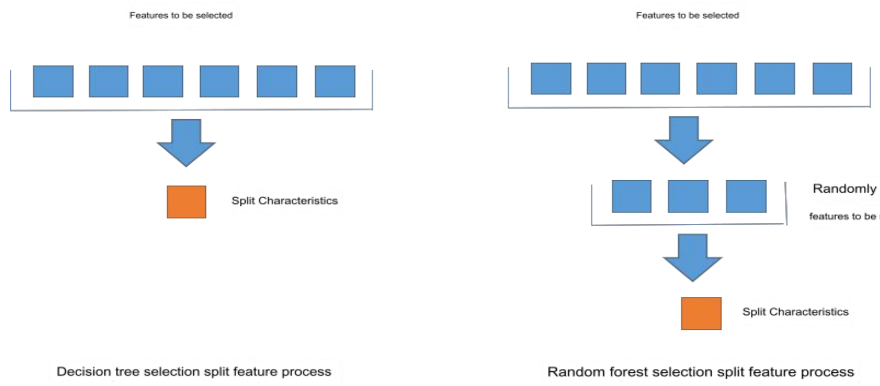

Figure 4. Random forest algorithm feature selection process

At each node, **k feature variables** are randomly selected, and the best classification feature is chosen to split the data. The classification threshold is determined through evaluation across candidate features. In the decision tree's feature selection process, **blue squares represent candidate features, while yellow squares indicate the selected split feature**. Figure 5 above shows an illustration of the feature selection process in a Random Forest model. By aggregating multiple trees, **Random Forest improves classification accuracy, reduces overfitting, and enhances model stability**, making it a widely used technique in predictive analytics.

Model Evaluation Method for Random Forest Algorithm

The performance of the **Random Forest prediction model** is evaluated using indicators such as **F-value, ROC curve, and AUC**, typically derived from the **confusion matrix**. The matrix variables include:

- **TP (True Positive):** Correctly predicted resignations.
- **FP (False Positive):** Incorrectly predicted resignations (still in service).
- **FN (False Negative):** Incorrectly predicted in-service employees (who actually resigned).
- **TN (True Negative):** Correctly predicted in-service employees.

These metrics help assess model accuracy and effectiveness in predicting employee turnover.

Table 1
Confusion Matrix

The true situation	forecast result	
	Resign	On the job
Resign	TP	NP
On the job	FP	TN

According to the confusion matrix, the precision rate P and recall rate R can be defined:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

Recall and precision are **trade-off measures**—as one increases, the other tends to decrease. The **balance point (BEP)** occurs when **recall = precision**, typically measured by **F1-score**:

$$F = \frac{2 \times P \times R}{P + R} = \frac{2 \times TP}{\text{Total number of samples} + TP - TN}$$

The **ROC curve** is similar to the **P-R curve** and visualizes model performance by plotting **True Positive Rate (TPR)** against **False Positive Rate (FPR)**:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{TN + FP}$$

Figure 5 shows the **ROC curve**, where the **diagonal line** represents a random model, and **(0,1)** corresponds to an ideal model.

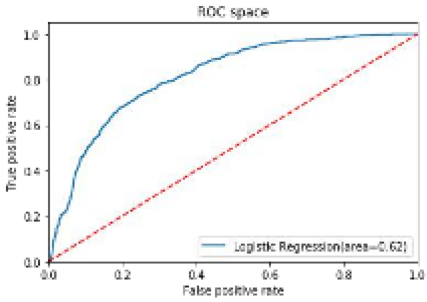

Figure 5. ROC curve diagram

Methods

This research adopts a Pragmatism/Critical Realism paradigm to understand the complex issue of talent attrition using practical and actionable insights derived from both qualitative and quantitative methods. Pragmatism is chosen because it emphasizes the practical application of research findings and the use of multiple methods to address research questions. This paradigm supports the integration of qualitative and quantitative methods, aligning well with the study's aim to provide comprehensive insights into talent attrition. This also study employs a Mixed Methods Research approach to explore the problem of talent attrition. This approach is justified by the need to leverage the strengths of both qualitative and quantitative data, offering a more complete understanding of the research problem. Specifically, the study utilizes the following methodologies:

Document Analysis: Reviewing existing literature and data to understand the current state of research on talent attrition and to identify relevant variables and hypotheses. Data Analysis: Using Python to analyze a dataset on talent attrition and to construct decision tree and random forest models. These models help identify key factors influencing talent attrition and predict future trends.

Qualitative Methods: Supplementing quantitative findings with qualitative insights to provide a more nuanced understanding of the reasons behind talent attrition.

This mixed methods approach allows for the development and testing of predictive models (decision tree and random forest) using Python, informed by document analysis and data analysis. These models help identify critical factors influencing talent attrition. Combining quantitative data (for prediction and pattern identification) with qualitative insights (for a deeper understanding of underlying reasons) ensures a comprehensive analysis.

Using a **Pragmatism/Critical Realism paradigm** and **mixed methods research**, this study provides **practical insights** into talent attrition. The blend of qualitative and quantitative approaches enables a comprehensive analysis for effective recommendations.

Data Collection and Procedure

The Random Forest algorithm was chosen for the prediction model due to its ability to analyze large datasets reliably and handle complex patterns efficiently. The dataset used in this study is sourced from Human Resources Analytics on the Kaggle platform, comprising 14,999 employee records containing essential workforce attributes. The dataset primarily reflects talent data from state-owned manufacturing enterprises, making it a suitable benchmark for studying talent loss in this sector. To streamline the analysis, the original "sales" variable has been replaced with "job" for clarity. Table below presents a sample of five relevant data entries used in the study.

Out [1]:

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	job	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low

Figure **Error! No text of specified style in document.**6. Data Display

Data preprocessing steps: After obtaining the data, we first need to clean and process the data to check whether there are missing values in the data and remove records with missing values to avoid interference with the results of subsequent data analysis and research. Through the code df.apply (lambda x: sum(x.isnull()),axis = 0) for cleaning and checking, there are no missing values in this data source, and the next step of analysis can be carried out. Secondly, we can check the data type from the data list. We can see that each record includes 10 variables. The data type and attribute type of each variable are shown in the table, where the attribute type includes numeric type and text type.

```
In  [3]: df.dtypes

Out[3]: satisfaction_level       float64
        last_evaluation          float64
        number_project             int64
        average_montly_hours       int64
        time_spend_company         int64
        Work_accident              int64
        left                       int64
        promotion_last_5years      int64
        job                       object
        salary                    object
        dtype: object
```

Figure **Error! No text of specified style in document.**7. Type of data

Finally, from the talent data attribute table, we can see that the content of the two columns of position and salary level is text type rather than data type, so we need to convert the text type content into numeric type, digitize and normalize the text type content, so as to facilitate the statistical analysis of the relationship between data columns later. Replace "low, medium, high" in the salary column with "1, 2, 3" respectively, and replace "sales, accounting, hr, technical, support, management, IT, product_mng, marketing, RandD" in the job column with "0, 1, 2, 3, 4, 5, 6, 7, 8, 9" respectively.

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	job	salary
0	0.38	0.53	2	157	3	0	1	0	[illegible]	1
1	0.80	0.86	5	262	6	0	1	0	[illegible]	2
2	0.11	0.88	7	272	4	0	1	0	[illegible]	2
3	0.72	0.87	5	223	5	0	1	0	[illegible]	1
4	0.37	0.52	2	159	3	0	1	0	[illegible]	1
...	...	...	...	...	...	...	...	...	...	...
14994	0.40	0.57	2	151	3	0	1	0	4	1
14995	0.37	0.48	2	160	3	0	1	0	4	1
14996	0.37	0.53	2	143	3	0	1	0	4	1
14997	0.11	0.96	6	280	4	0	1	0	4	1
14998	0.37	0.52	2	158	3	0	1	0	4	1

14999 rows × 10 columns

Figure 8. Encoded data list

Statistical Analysis

Descriptive Statistical Analysis

Human Resources Analytics collects data and information on a large number of talents in state-owned enterprises, including the department they work in, the projects they participate in, the length of their work, their salary, performance, and work-related injuries. These factors are of great value in analyzing the factors that cause talent loss. Descriptive statistical analysis can intuitively explore the trend of data distribution and find out whether there are extreme outliers in the data. Therefore, a descriptive statistical analysis is performed on the data set, and the results are shown in Table 2.

Table 2
Descriptive statistics Table

	count	mean	std	min	50%	max
satisfaction_level	14999	0.61	0.25	0.09	0.64	1
last_evaluation	14999	0.72	0.17	0.36	0.72	1
number_project	14999	3.8	1.23	2	4	7
average_montly_hours	14999	201.05	49.94	96	200	310
time_spend_company	14999	3.5	1.46	2	3	10
Work_accident	14999	0.14	0.35	0	0	1
left	14999	0.24	0.43	0	0	1
promotion_last_5years	14999	0.02	0.14	0	0	1
job	14999	3.34	2.82	0	3	9
salary	14999	1.59	0.64	1	2	3

From the descriptive statistics table, we can see that the average turnover rate of talents is 24%, the average satisfaction of talents is 61%, the performance evaluation of excellent talents is 72%, the average number of projects each talent participates in is 4, and the average working hours per month for each talent is about 201 hours. From the analysis of the table, we can see that the mean-average value and std-standard deviation value of each attribute are not much different, so there is no extreme outlier, and the data can be further modeled.

Inferential Statistical Analysis

Following **descriptive statistical analysis**, **correlation analysis** is conducted to examine relationships between key variables using a **correlation matrix**. Figure 9 highlights significant correlations:

- **Positive correlations:** Number of projects with both **average monthly working hours** and **performance evaluation**.

- **Negative correlations:** Resignation with **job satisfaction** and **work accidents**.

The analysis focuses on variables strongly linked to **talent loss**, ensuring meaningful insights for further investigation.

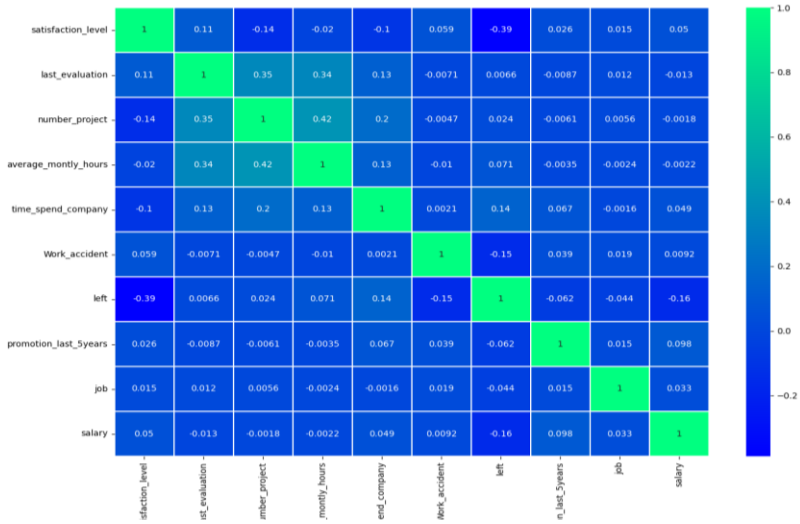

Figure 9. Correlation Matrix

Model Building and Evaluation

The **Random Forest algorithm** is a powerful classification model that effectively handles complex data, processing both **continuous and discrete attributes**. As an ensemble classifier, it outperforms **single decision tree models** by applying **random sampling** in both data rows and variable columns, generating multiple decision trees and aggregating their outputs for more reliable predictions. Additionally, it evaluates **feature importance**, identifying key factors influencing classification results (Scornet et al., 2014). This study applies **Random Forest** with **feature correlation analysis** to predict **corporate talent loss**, enabling companies to implement proactive retention strategies. The model identifies **employees at risk of leaving**, helping decision-makers **minimize talent attrition**. Introduced by Breiman (2001), the **Random Forest algorithm** is based on the **bagging technique**, offering advantages such as **minimal parameter tuning, strong generalization for high-dimensional data, and resistance to overfitting** (Gao, 2019). It efficiently ranks **feature importance**, handles both continuous and discrete datasets, and does not require **data normalization** before processing. Furthermore, **bootstrap resampling** enhances model robustness, particularly for **small datasets**. Unlike traditional **train-validation splits**, which reduce sample size and introduce bias, bootstrap resampling **preserves sample integrity**, improving prediction accuracy while allowing validation without sacrificing training data. By leveraging these advantages, the **Random Forest model** provides a **reliable framework for predicting talent loss**, offering actionable insights to optimize workforce retention and strategic planning.

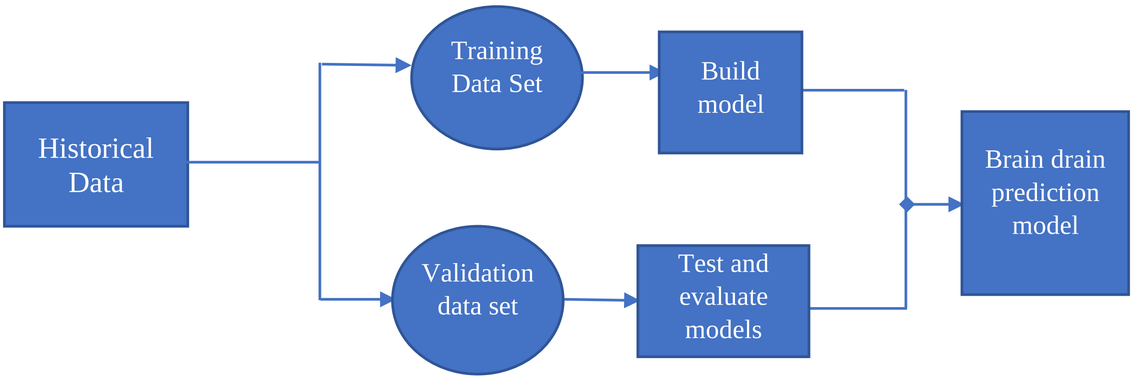

Figure 10. Brain drain prediction feature model

First, monitor and mine human resource data, analyze the correlation and regularity within the data, and establish a characteristic model for predicting talent turnover. The specific process is shown in Figure 10.

Dataset Division and Model Parameter Configuration

The software environment for the model building experiment in this paper is the python interactive experimental environment jupyter notebook. The data is randomly allocated into training sets and validation sets in proportion. The training set is used to adjust and fix the parameters in the model to achieve the best effect. The validation set is used to evaluate the fitting and convergence effect of the model. In this experiment, the parameter test_size=0.15 is defined.

Table **Error! No text of specified style in document.**

Distribution of samples in training set and test set

Sample type	Number of samples
Training set	12749
test set	2250

The training set of samples is mainly to initially grasp the basic prediction of the model for the phenomenon of population loss in the research object of this paper, and finally the conclusions of the validation set to verify the results of the training set, so as to confirm the final conclusions reached. Configure the random forest algorithm parameters:RandomForestClassifier(n_estimators=1000,max_depth=None,min_samples_split=10,oob_score=False,n_jobs=1, class_weigh t= "balanced"). The random forest algorithm is used for model training and construction based on the training dataset.

Model Evaluation

The **Random Forest talent loss prediction model** is evaluated and optimized using the **validation set**, requiring multiple **training iterations and parameter tuning** to achieve optimal performance. The **classification_report function** is used to display key metrics, including classification labels and support values. To compare performance, **decision tree results** are also included. **Table 3.3 and Table 3.4** show that **Random Forest outperforms the decision tree model** across all indicators.

Table **Error! No text of specified style in document.**
Random forest algorithm model

	precision	recall	f1-score	support
0	0.99	1.00	0.99	1720
1	0.99	0.96	0.97	530

Table 5
Decision tree algorithm model

	precision	recall	f1-score	support
0	0.98	0.99	0.98	1720
1	0.97	0.92	0.94	530

To further assess accuracy, **confusion matrices** visualize prediction results (**Figures 11 and 12**). The **Random Forest model correctly identified 509 out of 530 resignations**, achieving a **96% recall rate**, compared to **92%** (487/530) **for the decision tree model**.

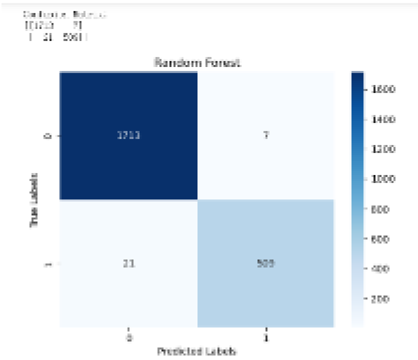

Figure **Error! No text of specified style in document.**. Random forest model confusion matrix

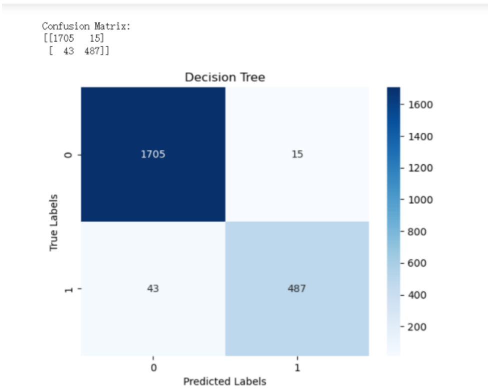

Figure **Error! No text of specified style in document.**. Decision tree model confusion matrix
The **ROC curve** (Receiver Operating Characteristic) evaluates **binary classifiers**, illustrating sensitivity across different thresholds. A well-performing model's curve deviates further from the baseline. Figure 13 shows that the **Random Forest model outperforms the Decision Tree model** in prediction accuracy.

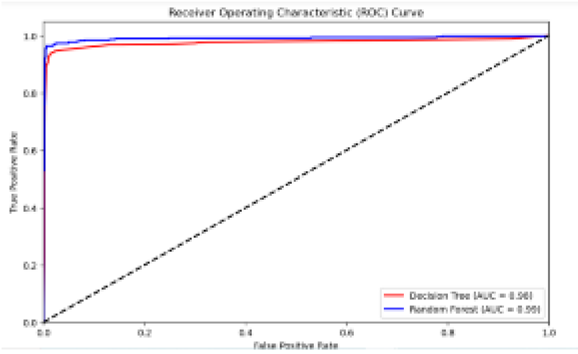

Figure 1**Error! No text of specified style in document.**. ROC curve

Importance Ranking of Attribute Features in Classification

The optimized random forest algorithm model is used to estimate the importance ranking of attribute features in classification, and the ranking result is shown in Figure 14. From the feature importance ranking results of the random forest algorithm model, it can be seen that the three feature items of employee satisfaction, company length, and number of projects are ranked in the top three. When using the feature analysis model, the results of the feature importance ranking can be referred to, which helps us understand the model and can also be used as a basis for analyzing the causes of talent loss.

Feature Importance Ranking:

	Feature	Importance
0	satisfaction_level	0.277448
4	time_spend_company	0.236403
2	number_project	0.161565
3	average_montly_hours	0.148803
1	last_evaluation	0.129320
7	salary	0.012724
5	Work_accident	0.011710
17	job_technical	0.004045
15	job_sales	0.003394
16	job_support	0.002652
6	promotion_last_5years	0.002201
11	job_hr	0.001826
9	job_RandD	0.001641
8	job_IT	0.001508
10	job_accounting	0.001437
12	job_management	0.001388
13	job_marketing	0.001108
14	job_product_mng	0.000836

Figure 14. Use random forest algorithm model to rank feature importance

Result and Discussion

Imbalance in Job Differences Leads to Talent Loss

Talent position and resignation data were extracted and visualized to analyze the relationship between job roles and turnover rates. **Figure 15** shows that **sales, technical, and support departments** have the highest turnover rates, while **management has the lowest**. Liu & Zhang (2019) note that sales personnel, though crucial to enterprise profits, are often undervalued due to their large numbers and lower knowledge levels, leading to frequent turnover.

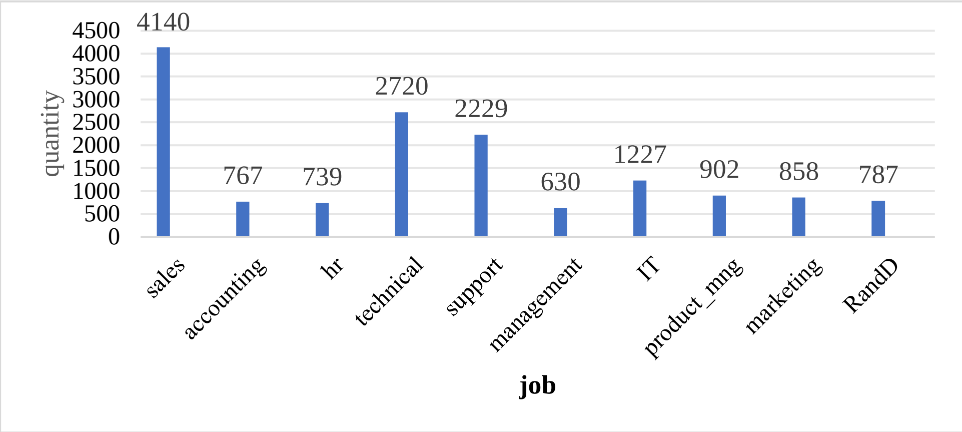

Figure 15. Distribution chart of number of people in each position

Figure 16 below shows that **sales staff make up 27.6%** of employees, with a **24.49% turnover rate**. Trojanowsk (2012) emphasizes that losing key sales staff disrupts customer relationships, as replacements struggle to adapt quickly, leading to **customer loss and weakened competitiveness** (Mbonwa, 2016). Similarly, **technical staff comprise 18.13%** of the workforce, with a **25.63% turnover rate**. Potgieter & Pretorius (2009) highlight that technical employees possess **critical expertise**, and frequent turnover can severely impact a company's technological progress. To mitigate this, organizations should implement **non-compete agreements** and structured talent retention strategies. **Service personnel account for 14.86%**, with a **24.90% turnover rate** (Zhong et al., 2022). Despite its impact on **service quality and customer retention**, many managers dismiss high service staff turnover as routine. While some turnover fosters innovation, excessive attrition signals **HR inefficiencies**, requiring **immediate intervention** to maintain workforce stability.

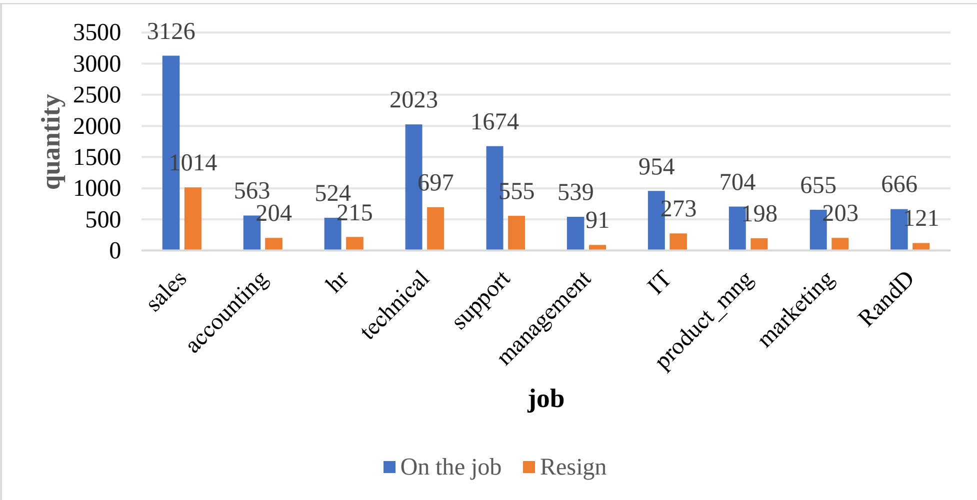

Figure 16. Distribution of Employee Turnover Across Positions

Career Bottlenecks for Employees with 3–5 Years of Experience

Figure 17 illustrates that employee with 3–5 years of experience face the highest turnover rates—10.57% (3 years), 5.93% (4 years), and 5.55% (5 years). Notably, over 50% of employees with 4–5 years of tenure resign, signaling a career development bottleneck. Early resignations (within six months) often stem from misaligned expectations regarding workplace culture, job roles, and working conditions. Employees leaving within 6–12 months typically cite issues with direct leadership, emphasizing the need for stronger team management and conflict resolution. For 3–5-year employees, stagnation in salary growth, promotions, and skill development drives resignations. To retain this critical talent, companies must address career progression challenges and implement targeted leadership and HR strategies to reduce turnover. Hammerberg (2002) found that replacing an experienced employee costs 1.5 times their salary, highlighting the financial impact of turnover. To minimize losses, leaders and HR managers must address career growth barriers while considering industry-specific factors to implement targeted retention strategies.

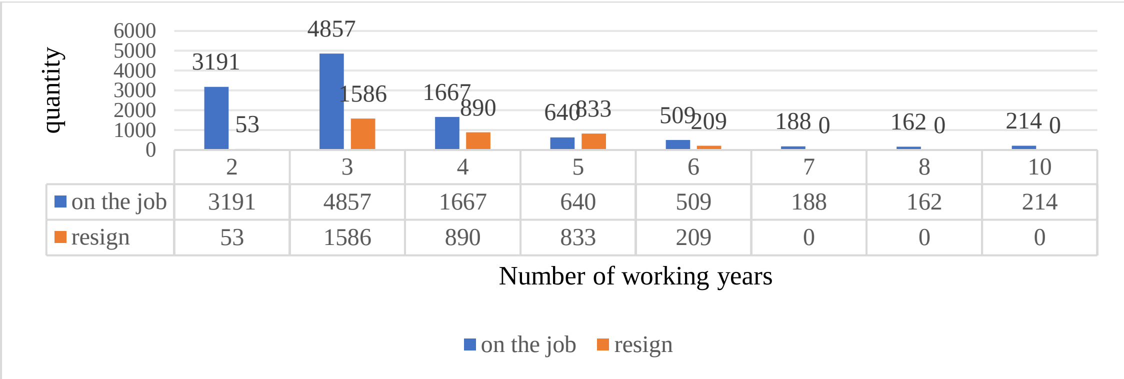

Figure 17. Resignation and Employment Distribution Across Age Groups

Imbalance between Salary Levels and Employee Input Leads to Talent Loss

Figure 18 shows that high-salary employees have a low resignation rate (0.55%), while low and medium-salary employees leave at 14.48% and 8.78%, respectively. When salaries fail to meet basic financial needs, other retention factors become ineffective (Arianti & Triyanto, 2020). A static pay structure fails to adapt to market demands. As industry needs grow, salaries must align with employee contributions to ensure fairness and job satisfaction (Zhou, 2022). Addressing pay imbalances is crucial for improving retention.

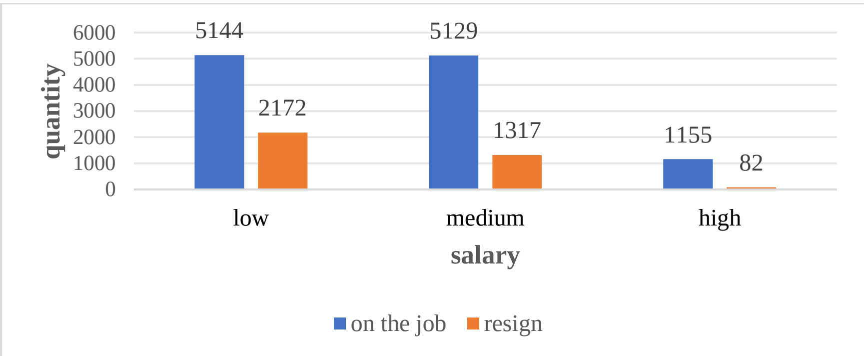

Figure 18: Resignation Distribution Across Different Salary Levels

Low Job Satisfaction Leads to Talent Loss

The Mobley talent loss model identifies job satisfaction as a key factor influencing employee turnover (Mobley, 1977). Employees with low job satisfaction tend to resign, regardless of external opportunities, impacting work quality. Kernel density estimation (KDE) is used to analyze the distribution of employee satisfaction levels, as shown in Figure 19. The probability density function estimates satisfaction levels for both retained and resigned employees, calculated using the following equation:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

Where h is a preset positive number, usually called window width or smoothing function, and Kh is a kernel function, which generally meets the following conditions:

$$K(-u) = K(u)$$

$$\int_{-\infty}^{\infty} K(u)du = 1$$

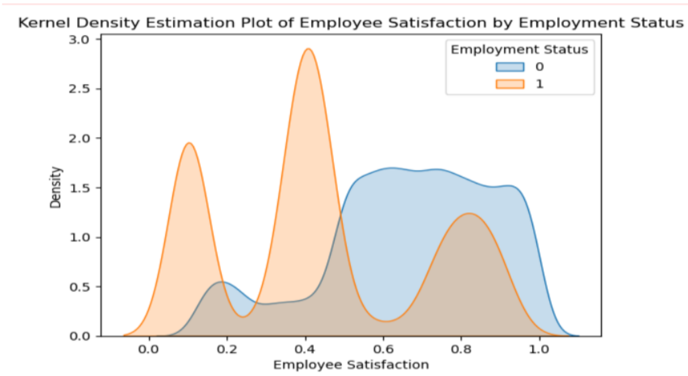

Figure 19. Talent satisfaction probability density distribution chart

Figure 19 illustrates that in-service employees have higher satisfaction levels (0.5–0.9), while 70% of resigned employees show dissatisfaction (0–0.5). Surprisingly, 20% of highly satisfied employees also resigned, indicating additional influencing factors beyond dissatisfaction. Understanding these patterns helps HR and management develop targeted retention strategies, improving job satisfaction and reducing turnover. To further explore talent loss, we conducted a regression analysis on talent satisfaction, performance evaluation, and resignation status, using the Implot regression chart to reveal deeper insights into the internal relationships of these factors. These findings provide valuable references for management, emphasizing the need for proactive measures to mitigate talent loss effectively.

Figure 20 categorizes resigned employees into three clusters. The first cluster consists of employees with low satisfaction (below 0.2) but high evaluations (above 0.75), indicating they work hard but feel overworked. The second cluster includes those with moderate dissatisfaction (0.35-0.45) and low evaluations (below 0.58), suggesting poor performance and dissatisfaction with company recognition. The third cluster represents employees with high satisfaction (0.7-1.0) and strong evaluations (above 0.8), showing that even high-performing, content employees may resign. These findings highlight the complexity of talent loss, necessitating further surveys to identify underlying causes and improve retention strategies.

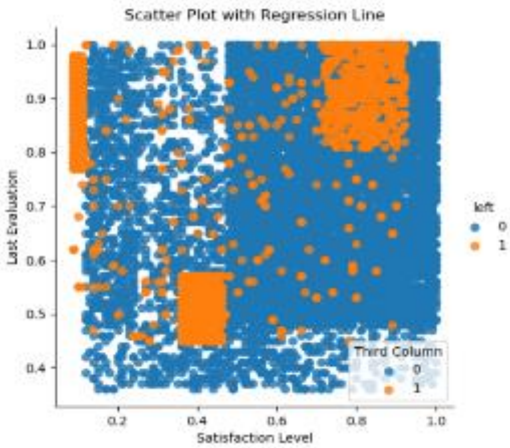

Figure 20: Talent Satisfaction vs. Performance Evaluation and Resignation Trend

Irrational Performance Evaluation Leads to Talent Loss

Talent performance evaluation and talent resignation data are extracted from the dataset and plotted as probability density distribution diagrams, as shown in Figure 21. The distribution reveals a bimodal pattern, indicating that departing employees are either low or high performers, with few in between. This suggests that employees with low evaluations are more likely to leave, while those with high evaluations may feel undervalued or overworked. Research by (Nabi et al., 2022) confirms that increased stress levels negatively impact employee performance. Employees with performance scores between 0.6 and 0.8 show greater stability within the company.

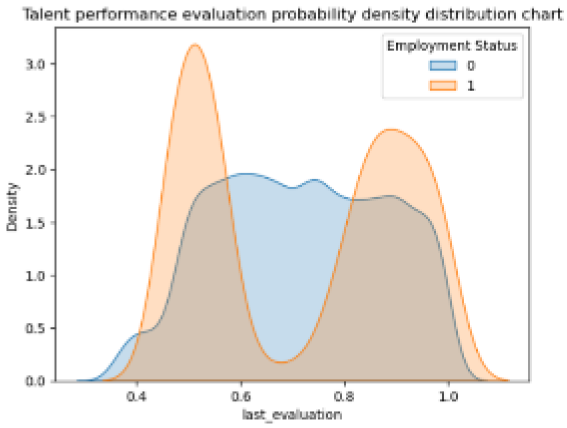

Figure 21. Probability Density Distribution of Talent Performance Evaluation

Recommendations and Conclusions

Establishing a Differentiated Talent Retention Strategy

Tailoring Strategies for Different Positions

Talent retention depends on effective job matching. Rather than selecting the best candidates for every role, companies should focus on hiring the most suitable ones based on skills and experience. This approach optimizes resource allocation, reduces costs, and enhances sustainable development. Job-specific training, career path planning, and customized incentives aligned with departmental needs are essential for retaining employees. For example, R&D teams should have compensation linked to project contributions to stimulate innovation. A well-defined promotion system based on performance rather than tenure enhances motivation, ensuring talent retention and attracting new employees.

Addressing Retention at Different Career Stages
Retention strategies should vary based on an employee's tenure. For new hires, clear communication about company culture prevents early dissatisfaction. Employees with 1–2 years of experience require leadership training and stronger managerial engagement to foster morale. Mid-career employees (3–5 years) often face career stagnation, requiring salary adjustments and career growth opportunities. Long-term employees (5+ years) should be given leadership roles or innovative tasks to maintain engagement. Strategic and humane retention policies ensure job satisfaction and loyalty.

Building a Competitive Salary System
Compensation directly influences employee retention. Companies should implement a dynamic performance evaluation system that aligns pay with contributions. A well-structured salary framework—combining base pay, position-based allowances, and performance-based rewards—ensures fair compensation. Additional incentives such as bonuses, stock options, and recognition programs enhance employee satisfaction. By integrating financial and non-financial rewards, companies can mitigate poaching risks and boost retention.

Establishing an Internal Talent Development System
Beyond salaries and job roles, employees value professional growth. A structured training system enhances skills, fosters career development, and boosts job satisfaction. Companies should define career progression pathways, provide job-specific training, and create opportunities for skill enhancement. Simplifying training procedures, ensuring relevance to corporate goals, and implementing post-training assessments maximize the impact of professional development programs.

Creating a People-Oriented Corporate Culture
Corporate culture significantly affects talent retention. Employees seek not only financial stability but also purpose, belonging, and professional growth. Leaders must integrate corporate values into everyday work, reinforcing a shared mission. Encouraging employee participation in discussions and initiatives strengthens engagement and loyalty. A well-defined behavioral framework fosters a positive work environment, enhancing long-term retention.

Establishing Effective Communication Channels
Efficient communication is critical for operational success. Many companies struggle with departmental silos, leading to inefficiencies and misalignment. Transparent, multi-directional communication—beyond hierarchical structures—ensures strategic clarity, swift issue resolution, and better team collaboration. Encouraging open dialogue between employees and management fosters a culture of trust and responsiveness.

In conclusion, talent loss is a critical challenge for business sustainability. This study utilizes a Kaggle-based dataset and applies a random forest algorithm to predict resignations and analyze influencing factors. Key drivers of turnover include job satisfaction, tenure, and workload. Addressing talent loss requires a multifaceted strategy: differentiated retention plans, competitive salaries, structured training, a strong corporate culture, and effective communication. Findings indicate that personalized retention approaches based on tenure and job function can significantly improve employee engagement. A fair and competitive compensation system, combined with professional development opportunities, enhances satisfaction and minimizes turnover. While this study provides a predictive framework for talent loss, its applicability across industries requires further research. Future studies should explore deeper industry-specific insights into salary structures, corporate culture, and individual employee motivations.

References

Ahmad, Z., & Shahbaz, M. (2017). Use of Cybernetic Approach to Minimize Employee Turnover in Turbulent Environment of Library & Information Science Profession. International journal of information and management sciences, 1, 49-58.

Al Abid, F. B., Bakri, A. B., Alam, M. G., Uddin, J., & Chowdhury, S. J. (2024). Simplified Novel Approach for Accurate Employee Churn Categorization using MCDM, De-Pareto Principle Approach, and Machine Learning. Baghdad Science Journal.

Alsayyed, B., & Braiki, F. A. (2015). Employee turnover, causes, the relationship between turnover and productivity and recommendations to reduce it.

Chang, H. (2009). Employee turnover: a novel prediction solution with effective feature selection.

Chaudhary, M. K. (2022). Rationale of Employee Turnover: An Analysis of Banking Sector in Nepal. International Research Journal of MMC.

Chin, C. (2017). The Influence of Job Satisfaction on Employee Turnover Intention in the Manufacturing Industry of Malaysia.

Elsafty, A., & Sayed, Y. R. (2023). Factors Affecting Employee Turnover in a Small Business in Egypt. International Journal of Social Science Studies.

Feeley, T. H., Hwang, J. M., & Barnett, G. A. (2008). Predicting Employee Turnover from Friendship Networks. Journal of Applied Communication Research, 36, 56 - 73.

Gialuisi, O., & Coetzer, A. (2013). An exploratory investigation into voluntary employee turnover and retention in small businesses. Small Enterprise Research, 20, 55 - 68.

Gunaprasida, N., & Wibowo, A. (2019). The effect of work-family conflict and flexible work arrangement on turnover intention: do female and male employees differ? Jurnal Siasat Bisnis.

Haldorai, K., Kim, W.G., Pillai, S. G., Park, T., & Balasubramanian, K. (2019). Factors affecting hotel employees' attrition and turnover: Application of pull-push-mooring framework. International Journal of Hospitality Management.

Hauer, G., Quan, T. A., & Liang, Y. (2021). Leadership as an influencing factor in employee retention - a case study analysis in East Asian multinational corporations in the digital age. Revista Română de Informatică și Automatică.

Husain, M. S. (2015). Causes of Employee Turnover Intention : A Study on Banking Industry of Pakistan.

Jain, N., Tomar, A., & Jana, P. K. (2020). A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. Journal of Intelligent Information Systems, 56, 279 - 302.

KHAN, R.I., Khan, M.N., & KHAN, I. (2021). The Impact of Workplace Bullying and Harassment on Employee Turnover in Banking Industry of Pakistan. Journal of Business & Tourism.

March, J.G., & Simon, H.A. (1958). Organizations. Wiley.

Mihajlov, S. R., & Mihajlov, N.S. (2020). An overview of employee turnover models.

Mobley, W. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 62, 237-240.

Musanga, V., Tarambiwa, E., & Zvarevashe, K. (2022). A Supervised Machine Learning Model to Optimize Human Resources Analytics for Employee Churn Prediction. 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT), 1-6.

Nagadevara, V., Srinivasan, V., & Valk, R. (2008). Establishing a link between employee turnover and withdrawal behaviours: application of data mining techniques.

Oktavio, A., & Kaihatu, T.S. (2020). Implication of Future Time Perspective and Hope on the Continuance Commitment of Hotel Employees. Jurnal Economia, 16, 161-172.

Porter, C.M., Woo, S.E., & Campion, M.A. (2016). Internal and External Networking Differentially Predict Turnover Through Job Embeddedness and Job Offers. Personnel Psychology, 69, 635-672.

Price, J.L. (2001). Reflections on the determinants of voluntary turnover. International Journal of Manpower, 22, 600-624.

Qadir, M., Noreen, I., & Shah, A.A. (2021). Bi-LSTM Deep Learning Approach for Employee Churn Prediction.

Saberi, H.R., Khoshakhlagh, A.H., Laal, F., Mirzahosseininejad, M., Hannani, M., & Bamel, U. (2023). Investigating the Reasons for Turnover Intention of Workers in the Spinning and Weaving Industries in Iran. Corporate Reputation Review, 26, 320-331.

Saidu, N. (2018). Factors Influencing Employee's Turnover and Their Effects in the Malaysian Private Sector.

Samašonok, K. (2024). Employee turnover: causes and retention strategies. Entrepreneurship and Sustainability Issues.

Saradhi, V.V., & Palshikar, G.K. (2011). Employee churn prediction. Expert Syst. Appl., 38, 1999-2006.

Shaikh, M., Shaikh, S.H., Benghal, G., Shaikh, H., & Shah, N.J. (2020). Impact of Turnover on Organizational Efficiency: A Case Study of Dawlance Company.

Silpa, N., Maheswara Rao, V.R., Subbarao, M.V., Kurada, R.R., Reddy, S.S., & Uppalapati, P.J. (2023). An Enriched Employee Retention Analysis System with a Combination Strategy of Feature Selection and Machine Learning Techniques. 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), 142-149.

Sisodia, D.S., Vishwakarma, S., & Pujahari, A. (2017). Evaluation of machine learning models for employee churn prediction. 2017 International Conference on Inventive Computing and Informatics (ICICI), 1016-1020.

Syed, J., & Wang, J. (2018). Shanghai Silk Pharma Ltd: Employee Turnover. Asian Journal of Management Cases, 15, S55 - S67.

Tapola, M. (2016). Effects of employee turnover on service quality and customer satisfaction in contact centers.

Taye, D., & Getnet, B. (2020). The Impact of Employee Turnover on Organizational Performance: A Case Study of Mada Walabu University, Bale Robe, Ethiopia. American Journal of Pure and Applied Biosciences.

Telly, C.S. (1969). Inequity and Its Relationship To Turnover Among Hourly Workers in the Major Production Shops of the Boeing Company.

Vascellaro, J. E. (2007). Social networking goes professional: doctors, salesmen, executives turn to new sites to consult, commiserate with peers; weeding out impostors. Wall street journal.