Employee Turnover Prediction in Chinese Private Manufacturing: An Integrated Approach

Huijun Hao and Wei Chen[*]

School of Information Engineering, Wenzhou Business College
Wenzhou, Zhejiang, PR China
Correspondence author: Wei Chen
Email: 392116074@qq.com; cnchw@qq.com

Abstract. In recent years, employee turnover has become a significant concern for private manufacturing enterprises in China. It has led to an increase in human resource costs, such as those associated with recruitment, training, and new employee on-boarding. Additionally, the disruption of work processes and the obstruction of knowledge transfer among teams due to high turnover have negatively impacted operational continuity. Unfortunately, there is a lack of in-depth research specifically addressing the employee turnover issue within this industry. Drawing on 2,516 personnel records—1,566 former employees and 950 still on staff—from a Wenzhou manufacturer, we ran correlation checks, machine learning models, weight ranking, and survival-time analysis. The Random Forest classifier reached 95.6 % accuracy, pinpointing work environment satisfaction as the strongest driver of turnover. This framework offers managers a clear path to lift retention and support steady growth.

Keywords: Employee Turnover; Machine Learning; Survival Analysis

1   Introduction

Private manufacturing enterprises are vital to China's economy, making significant contributions to employment and GDP [1]. However, they now face intense market competition, rising labor costs, and the imperative for technological innovation. In this context, employee turnover has emerged as a major challenge [2].

Cech et al. noted in their study on human resource dilemmas in China's manufacturing industry that reducing employee turnover has long been a persistent issue [3]. To remain competitive, these companies must attract and retain talented employees. However, due to factors such as insufficient pay, poor working conditions and limited career development paths, these factors lead to high employee turnover [4]. In particular, the loss of technical personnel has an especially serious impact: it reduces production efficiency, weakens practical experience, and disrupts daily operations, having a serious impact on the operation and development of enterprises [5].

It is most important to know the ways in which workers leave their jobs, for it is bound up with many parts of the work of managing those within the company, the growth of the group, and the increase of its fruit [6]. First, the company must gather

and keep the best workers, so that it may grow without fail and stand firm in its doing [7]. Second, the knowledge, skill, and wisdom of the workers are the strength by which the company is made strong. When workers leave, it strikes at the speed of their labor and the power to make new things[8]. Also, when many workers leave, the cost of finding new ones, teaching them, and bringing them into the work grows. This brings harm to the company's wealth [9].

Traditional methods of exploring the reasons for employee turnover cannot fully gain insight into the complex causes behind this phenomenon. However, the development of information technology, especially the application of machine learning, has opened up a new approach to studying employee turnover [10]. With the help of machine learning algorithms, we can build a model of employee turnover and predict the future trend, so as to provide enterprises with more targeted employee retention methods[11].

Previous studies have studied the turnover of employees and found that salary satisfaction, job satisfaction and the business model of the enterprise are important factors affecting the turnover of employees [12]. However, there are few related studies on private manufacturing enterprises in China. In particular, there are still relatively few studies using machine learning techniques to study the turnover situation in this industry, which is a key gap in current research.

This study aims to bridge this gap by using machine learning models to predict employee turnover in Chinese private manufacturing enterprises. Specifically, we attempt to determine the key factors leading to turnover and evaluate the performance of using integrated machine learning algorithms in predicting turnover.

2   Methods

2.1   Datasets

We collected data from Company A, a manufacturing company in Wenzhou, one of the birthplace of private enterprises in China. The data includes employees' basic information, job characteristics, work environment evaluation, and human resource policies. Employees' basic information covers personal background information such as age, gender, education, household registration, and marital status. Job characteristics include specific information such as employment location, days of employment, employment location, job grade, person-job fit, salary expectations, and opportunities for promotion . Work environment evaluation includes data on job autonomy, job stress, leadership, team atmosphere, and fairness of performance evaluation. Data on human resource policies are used to analyze factors such as employee incentives, salary, and job requirements.

2.2   Data Mining and Analytics Methods

We conducted data processing on various aspects such as employees' basic information, job characteristics, job environment evaluation, and human resources policies within the collected companies. Utilizing the pandas library in Python, we computed

the Pearson's correlation coefficient matrix and leveraged the seaborn library to generate a visually appealing heatmap illustrating the correlations among these factors. This heatmap uses color variations and block sizes to visualize the strength of the relationships between different variables. Thus, it provides a clear and insightful presentation of the data, facilitating our initial analysis of important associations. Specifically, this step aids in identifying key factors that significantly impact employee turnover, thereby establishing a solid foundation for subsequent in-depth analysis.

We performed essential data preparation steps such as missing value handling, standardization, and partitioning the dataset into training and testing sets. Implementation-wise, we opted for the Random Forest algorithm as our preferred modeling technique and utilized the scikit-learn library in Python for model training. To gauge the performance of our model, we employed techniques like k-fold cross-validation to evaluate accuracy, precision, and recall metrics. Through these efforts, we successfully developed a well-validated Random Forest model, which yields actionable data-driven insights to inform evidence-based employee retention strategies.

To translate the model into practical guidance, we extracted feature importance weights via the model's feature importance attribute. By interpreting these weights, we identified key determinants of employee turnover—specifically, salary and benefits packages, job satisfaction (operationalized as feedback scores on work conditions), and career development opportunities. This granular analysis provides a robust empirical basis for designing targeted retention interventions tailored to the manufacturing context.

To explore temporal dynamics of turnover, we compiled employee tenure data, including hire dates and departure timelines (with a focus on production and technical roles). Using the lifelines library in Python, we generated Kaplan-Meier curves to visually depict how turnover rates evolve over time—for instance, highlighting elevated attrition within the first 6 months of employment, a critical window for retention efforts. Complementing the descriptive Kaplan-Meier curves, we employed the Cox proportional hazards model to quantify associations between employee attributes (e.g., tenure, technical certification status) and turnover hazard ratios. These analyses collectively revealed nuanced patterns in turnover behavior, offering empirical support for refining retention strategies to address stage-specific and role-specific risks.

3   Results

3.1   Correlation Study of Employee Turnover Variables

To better identify factors driving employee turnover in manufacturing, we used correlation analysis to explore links between key variables and turnover risk. Results, visualized as a correlation heatmap in Figure 1, lay groundwork for deeper analysis. Through this analysis, we can identify significant factors that influence employee turnover and provide reliable insights for model construction and prediction.
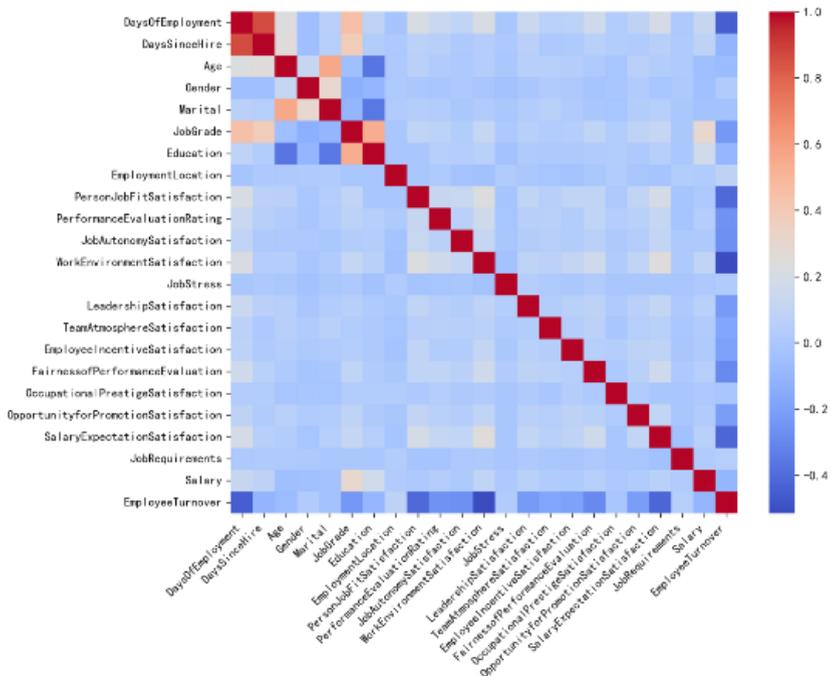
Figure 1. Correlation analysis of risk factors for employee turnover in Chinese private manufacturing enterprises

3.2 Model Construction for Employee Turnover Prediction

During the model selection stage, we evaluated the suitability of various machine learning algorithms, including Random Forest (RF), Artificial Neural Network (ANN), Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), etc., and selected the algorithm most suitable for employee turnover prediction. In Figure 2, we compared the accuracy rates of different models, and the results clearly demonstrate that random forest outperformed other models with a remarkable accuracy rate of 0.956. As shown in Figure 3, commonly used evaluation metrics such as accuracy, precision, recall, F1 score, and ROC curve can be used to measure the performance of the model.

After building the machine learning model (Random Forest) for employee turnover, it is important to use model interpretation to understand the model's predictive ability and decision-making process. After performing a meticulous feature weight analysis on our predictive model, we have managed to determine the relative significance of various features in projecting employee turnover (as shown in figure 4). Among these features, work environment satisfaction emerged as the most critical, with a weight value of 0.152138. This highlights the importance of employee satisfaction with their work environment in predicting the risk of turnover.
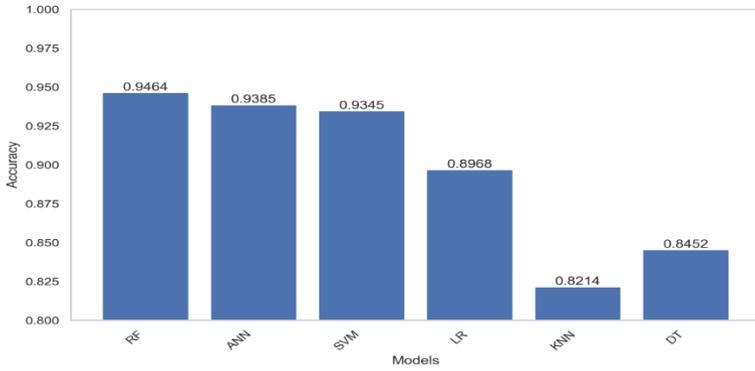
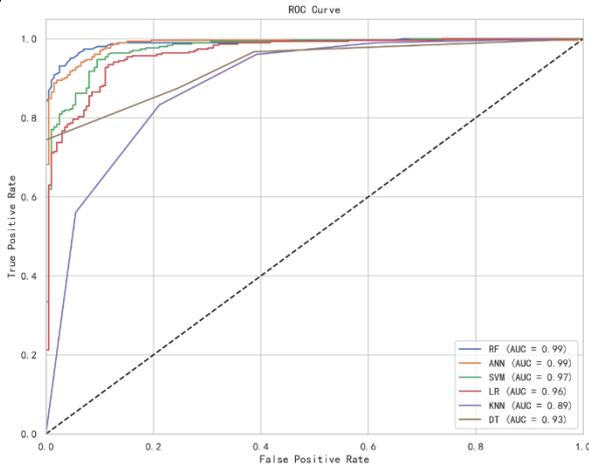Figure 2. Comparison of accuracy for machine learning models.

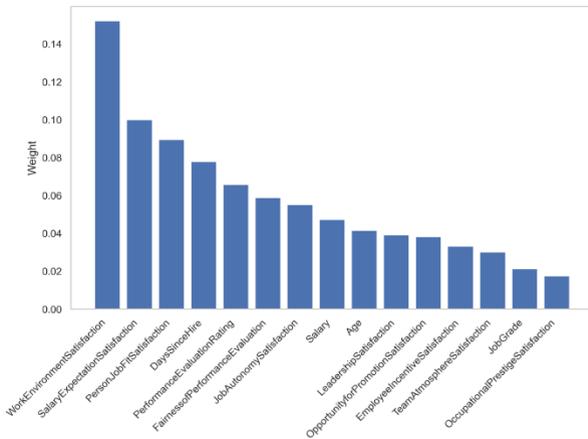Figure 3. Comparison of ROC curves for machine learning models.

Figure 4. Analysis of the weights of factors influencing employee turnover.

3.3   Survival Analysis of Employee Turnover Risk

We performed an survival analysis to examine employee turnover and compare it across two risk groups (Figure 5). Initially, we applied the Cox proportional hazards model to our dataset. In this model, the duration variable represented the number of days since an employee's employment, while the event of Employee Turnover indicated whether an employee had departed from the company. Then, based on different risk levels, we divided employees into high-risk groups and low-risk groups. By comparing the survival curves of the high-risk group and the low-risk group, we observed significant differences in turnover rates. This implies that the risk level of employees has a pronounced impact on the occurrence of employee turnover. By considering the survival analysis results of different risk groups comprehensively, we can develop targeted retention strategies for the organization, with a focus on addressing potential issues within the high-risk group, in order to reduce employee turnover and enhance organizational stability and performance.

Further analysis and examination of the various covariates in survival analysis has yielded significant and golden insights into the multitude of factors that directly influence employee turnover within an organization (Figure 6). Among these influential factors, the shining star is job grade as it emerged with great prominence, demonstrating that lower job grade hold a direct correlation with heightened turnover risk. Delving deeper, the exploration uncovered additional noteworthy variables that impact turnover likelihood, namely satisfaction with salary expectations, work environment, and person-job fit. Encouragingly, it was discovered that higher levels of contentment in these areas are consistently associated with reduced turnover rates, suggesting that employees who feel adequately compensated and inherently fulfilled within their workplace are less likely to entertain thoughts of seeking greener pastures elsewhere. The study shows that fairness in assessments of production quota fulfillment, satisfaction with frontline supervisors' management of production teams, work autonomy (e.g., flexibility in scheduling assembly line tasks), and team atmosphere are key factors influencing turnover risk. These findings suggest that for enterprises retaining talent, prioritizing fair production assessments, cultivating capable frontline leadership, enhancing task autonomy, and fostering positive team dynamics are critical focuses. Notably, the analysis also considered other variables—work location, gender, marital status—though their influence on turnover was marginal.
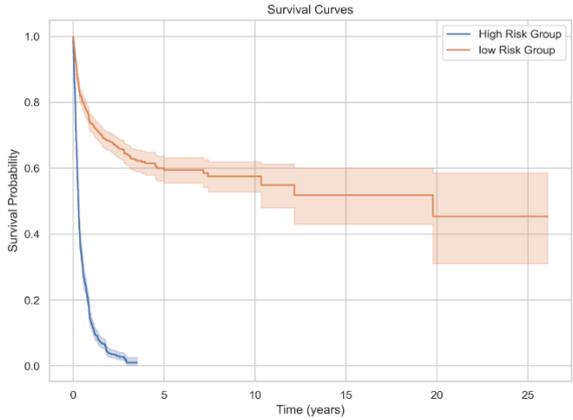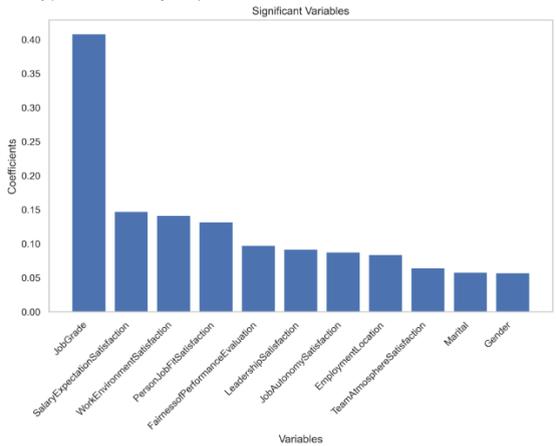
Figure 5. Employee turnover risk survival curve

Figure 6. Covariate analysis of survival analysis

4   Discussion

In this study, we performed a comprehensive analysis of employee turnover using correlation analysis, machine learning, weight analysis and survival analysis. Firstly, correlation analysis helps to identify the relationship between each factor and turnover, which lays the foundation for subsequent analysis. Second, a machine learning model (random forest model) uses the dataset to predict employee turnover. In addition, the weight analysis allows us to specify the importance of each factor in predicting turnover. Finally, survival analysis deeply explores the time point of employee turnover and provides valuable information for employee retention patterns. Utilizing survival analysis, organizations can formulate strategies to enhance retention rates at critical junctures during an employee's tenure. These methods complement each other, providing diverse perspectives for a comprehensive understanding of employee turn-

over. By integrating these methods, we analyzed the turnover situation from multiple angles, deepening our understanding of the underlying driving factors and providing strong support for formulating targeted retention strategies.

Our research results indicate that factors such as job level, salary satisfaction, work environment satisfaction, and job-person fit are of great significance in predicting turnover. These findings are consistent with previous studies. For instance, Raza also regarded job level as a critical factor and pointed out that senior positions usually offer more promotion opportunities, which aligns with our observation that a clear promotion path can reduce the turnover risk. However, the uneven distribution of promotion opportunities may lead to dissatisfaction. To solve this problem, organizations should establish a transparent and fair promotion process based on performance and talent, implement a structured career development framework, and provide regular feedback to employees regarding their progress and future potential within the company to enhance their perception of fairness and future prospects.

5   References

1. Yang YX, Yu HM, Su X & Wang R (2023) Exploring the role of green finance and natural resource policies in carbon emission efficiency of China's manufacturing industry in the context of post-COVID-19 period. Resour Policy 86.
2. Li ZT & Fox E (2023) Prediction and optimization of employee turnover intentions in enterprises based on unbalanced data. Plos One 18.
3. Cech M, Yao W, Samolejová A, Li J & Wicher P (2016) Human Resource Management in Chinese manufacturing companies. Perspectives in Science 7: 6-9.
4. Ren ZY & Hussain RY (2022) A mediated-moderated model for green human resource management: An employee perspective. Front Env Sci-Switz 10.
5. Ozmen EP & Ozcan T (2022) A novel deep learning model based on convolutional neural networks for employee churn prediction. J Forecasting 41: 539-550.
6. Chowdhury S, Joel-Edgar S, Dey PK, Bhattacharya S & Kharlamov A (2023) Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. Int J Hum Resour Man 34: 2732-2764.
7. Alsheref FK, Fattoh IE & Ead WM (2022) Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms. Comput Intel Neurosc 2022.
8. Mitchell TR, Holtom BC, & Lee TW (2004). How to keep you Best Employees: Developing an effective retention policy. In Academy of Management Executives,15(4), 96–109.
9. Alnehabi M & Al-Mekhlafi AA (2023) The Association between Corporate Social Responsibility, Employee Performance, and Turnover Intention Moderated by Organizational Identification and Commitment. Sustainability-Basel 15.
10. Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, & Rigol-Sanchez JP(2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS journal of photogrammetry and remote sensing, 67, 93–104.
11. Awad, M, & Khanna, R (2015). Machine learning in action: examples. In efficient learning machines, 209-240.
12. Akasheh MA, Malik EF, Hujran O & Zaki N (2024) A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review. Expert Syst Appl 238.