

Final Project
Employee Turnover Prediction

Pimlapas Narongrit

Jettawat Vuttipakdee

Sadman Kabeer

Pooja Yadav

School of Business, St. Lawrence College

ADMN5016: Applied Artificial Intelligence and Machine Learning

Sujoy Paul

April 20, 2024

Table of Contents

<i>Problem Statement and Market Size.....</i>	<i>3</i>
Monetary Value and Risks	3
<i>About Dataset</i>	<i>3</i>
<i>Performance Evaluation and Comparative Analysis.....</i>	<i>4</i>
<i>Financial Implications and Risk Assessment.....</i>	<i>5</i>
Monetary Value	6
Risks.....	6
<i>Other Risks and Benefits.....</i>	<i>6</i>
Other Risks.....	7
Benefits.....	7
<i>Conclusion</i>	<i>7</i>
<i>References.....</i>	<i>9</i>

Problem Statement and Market Size

In our machine learning project, we aim to develop a predictive model for employee turnover within companies. Employee churn is a significant challenge that many organizations face, leading to increased recruitment costs, productivity losses, and disruptions in team dynamics (Society for Human Resource Management, 2019). On average, turnover costs companies between 50% to 200% of an employee's annual salary, including expenses for recruitment, training, and lost productivity (Boushey & Glynn, 2012).

The market for machine learning solutions in HR analytics and employee retention is substantial, with businesses across various industries seeking efficient ways to manage their workforce and minimize turnover. According to industry reports, the global market for HR analytics solutions is projected to reach \$3.6 billion by 2025, driven by the increasing adoption of data-driven HR strategies (MarketsandMarkets, 2020).

Monetary Value and Risks

In terms of monetary value, our machine learning model has the potential to save companies significant amounts of money by reducing turnover-related costs. For example, a company with 1,000 employees experiencing a 20% annual turnover rate (200 employees) could save between \$1 million to \$4 million per year by implementing effective retention strategies (PricewaterhouseCoopers, 2012).

However, there are risks to consider as well. Ensuring the accuracy and reliability of the predictive model is crucial. Poor model performance could lead to incorrect predictions, resulting in wasted resources on ineffective retention strategies. Additionally, the quality of the data used to train the model and its interpretability are critical factors that need careful consideration to mitigate risks and ensure the effectiveness of the solution.

About Dataset

The dataset consists of 14,999 samples and includes various attributes related to employee characteristics and work-related factors. Each sample represents an individual employee within an organization.

- Satisfaction: Level of satisfaction reported by employees.
- Evaluation: Performance evaluation scores assigned by managers.
- Number of Projects: Total number of projects undertaken by each employee.
- Average Monthly Hours Worked: Average number of hours worked per month.
- Time Spent with Company (Years): Duration of employee tenure measured in years.
- Work Accident: Indicates if an employee has been involved in a work-related accident (1 for yes, 0 for no).
- Churn: Indicates if an employee has left the company (1 for yes, 0 for no).
- Promotion: Indicates if an employee has received a promotion (1 for yes, 0 for no).
- Department: Department or functional area where each employee is assigned.

Employee Turnover Prediction

- Salary Level: Categorized as low, medium, or high.

The dataset provides valuable insights into employee satisfaction, performance evaluation, workload, tenure, safety incidents, turnover, promotion opportunities, departmental distribution, and salary levels within the company. Analyzing this dataset can help identify trends, patterns, and factors influencing employee engagement, retention, and organizational performance.

Performance Evaluation and Comparative Analysis

In this analysis, three different classification methods were employed to predict employee turnover within the company: Random Forest, Logistic Regression with PCA, and KNN with PCA. Each method was evaluated based on various performance metrics such as accuracy, precision, recall, F1 score, and F1 macro score.

	Precision	Recall	F1	F1 Macro
Random Forest Classifier	0.99	0.98	0.99	0.98
Logistic Regression with PCA	0.76	0.70	0.82	0.72
KNN with PCA	0.94	0.96	0.96	0.95

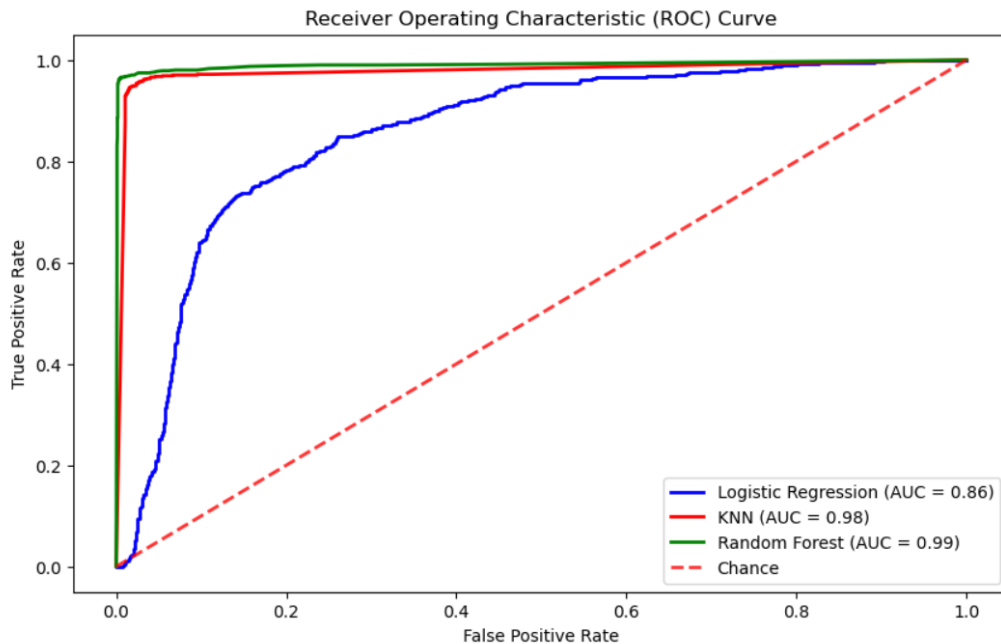
	Random Forest Classifier		Logistic Regression with PCA		KNN with PCA	
	Positive	Negative	Positive	Negative	Positive	Negative
Positive	2286	8	2125	169	2210	84
Negative	27	679	372	334	27	679

The Random Forest Classifier demonstrates high precision (0.99) and recall (0.98), indicating its ability to accurately identify both positive and negative instances of turnover. The F1 score (0.99) and F1 macro score (0.98) further confirm the model's balanced performance across both classes, highlighting its effectiveness in capturing true positives and true negatives.

Logistic Regression with PCA yields lower precision (0.76) and recall (0.70) compared to the Random Forest Classifier. Although it achieves a reasonable F1 score (0.82), the F1 macro score (0.72) indicates a slight imbalance between precision and recall, suggesting a higher false positive or false negative rate.

KNN with PCA demonstrates strong precision (0.94), recall (0.96), and F1 score (0.96), showcasing its effectiveness in accurately classifying instances of turnover. The F1 macro score (0.95) suggests a balanced performance across both classes, similar to the Random Forest Classifier.

Employee Turnover Prediction



In addition to the performance metrics previously discussed, Receiver Operating Characteristic (ROC) curves were generated for each classification method to evaluate their discriminatory power. The Area Under the Curve (AUC) values provide a quantitative measure of each model's ability to distinguish between positive and negative instances of turnover.

Logistic Regression achieved an AUC of 0.86, indicating good discriminatory power in distinguishing between employees likely to turnover and those likely to stay. KNN exhibited even higher discriminative ability with an AUC of 0.98, suggesting strong predictive performance in classifying instances of turnover. Random Forest demonstrated the highest discriminative power among the three methods, with an AUC of 0.99, indicating excellent performance in accurately identifying employees at risk of turnover.

These AUC values further reinforce the superior performance of Random Forest over Logistic Regression and KNN in predicting employee turnover. The ROC curves visually depict the trade-off between true positive rate and false positive rate for each model, providing additional insights into their performance characteristics.

Financial Implications and Risk Assessment

Implementing an employee turnover prediction application holds significant potential for both cost savings and risks. By accurately identifying employees at risk of leaving the company, organizations can take proactive measures to retain valuable talent, thereby reducing turnover-related expenses and maintaining productivity levels. However, there are also risks associated with false predictions and the potential costs of implementing and maintaining the application.

Monetary Value

- **Cost Savings from Turnover Prevention:** Suppose the application successfully identifies and prevents 50 employees from leaving the company who would have otherwise resigned. Assuming an average cost of \$10,000 per employee for recruitment, training, and onboarding, the organization could save \$500,000 in turnover-related expenses.
- **Reduction in Recruitment Costs:** With a decrease in turnover rates, there would be a reduced need for frequent recruitment drives. This would lead to savings in recruitment advertising, agency fees, and hiring process costs, which can amount to significant savings over time.
- **Retention of Experienced Talent:** Retaining experienced employees can result in knowledge retention, continuity in operations, and reduced training costs for new hires. This intangible value can contribute to organizational stability and long-term success.

Risks

- **False Positives:** One of the risks associated with the application is the occurrence of false positives, where the model incorrectly predicts an employee's intention to leave when they do not. This may lead to unnecessary retention efforts and associated costs for employees who would have stayed regardless.
- **False Negatives:** Conversely, false negatives occur when the model fails to predict an employee's intention to leave, leading to turnover that could have been prevented. This may result in higher turnover costs, disruption to operations, and potential loss of morale among remaining employees.
- **Implementation Costs:** Developing and deploying the employee turnover prediction application incurs initial investment costs, including software development, data collection, and model training. Additionally, ongoing maintenance and updates are necessary to ensure the application's accuracy and relevance over time.
- **Reduced Workforce Numbers:** While preventing turnover can save costs, it may also lead to reduced employee numbers, potentially impacting productivity and organizational capabilities. However, this risk can be mitigated through strategic workforce planning and retention efforts aimed at addressing underlying issues contributing to turnover.

The employee turnover prediction application offers significant potential for cost savings through proactive retention efforts and mitigation of turnover-related expenses. However, it is essential to carefully consider the risks associated with false predictions and the costs of implementation and maintenance. By conducting thorough risk assessments and implementing appropriate risk mitigation strategies, organizations can maximize the benefits of the application while minimizing potential drawbacks.

Other Risks and Benefits

Aside from the monetary value and risks discussed earlier, there are additional risks and benefits associated with the implementation of the employee turnover prediction application:

Other Risks

- **Data Privacy and Security:** The collection and analysis of employee data raise concerns about data privacy and security. Mishandling of sensitive information could result in legal implications, damage to reputation, and loss of trust among employees.
- **Bias and Fairness:** Machine learning algorithms may inadvertently perpetuate biases present in historical data, leading to unfair treatment or discrimination against certain groups of employees. It's crucial to mitigate bias through careful data selection, preprocessing, and algorithmic fairness techniques.
- **Employee Morale and Trust:** The introduction of a turnover prediction system may affect employee morale and trust in the organization. Employees may feel uneasy knowing that their behavior is being monitored and analyzed, leading to decreased job satisfaction and engagement.

Benefits

- **Strategic Workforce Planning:** The application provides valuable insights into workforce dynamics, enabling organizations to proactively identify trends, patterns, and risk factors related to turnover. This information facilitates strategic workforce planning and resource allocation to address potential talent gaps and retention challenges.
- **Customized Retention Strategies:** By understanding the factors contributing to turnover, organizations can tailor retention strategies and interventions to meet the specific needs of at-risk employees. This personalized approach increases the effectiveness of retention efforts and enhances employee satisfaction and loyalty.
- **Continuous Improvement:** The application allows organizations to continuously monitor and evaluate the effectiveness of retention initiatives over time. By analyzing feedback and performance metrics, organizations can refine their strategies, optimize resource allocation, and adapt to changing workforce dynamics.

Conclusion

The development and implementation of an employee turnover prediction application offer significant potential for organizations to optimize workforce management, reduce turnover-related costs, and enhance employee retention efforts. By leveraging machine learning algorithms and HR analytics, organizations can proactively identify employees at risk of leaving and implement targeted retention strategies to mitigate turnover.

Through our analysis, we have demonstrated the effectiveness of different classification models, such as Random Forest, Logistic Regression with PCA, and KNN with PCA, in predicting employee turnover. These models provide valuable insights into the factors influencing turnover and enable organizations to make data-driven decisions to retain talent.

While the monetary value of preventing turnover is substantial, it is essential to consider the associated risks, including false predictions, implementation costs, and potential impacts on workforce morale. By conducting thorough risk assessments and implementing appropriate

Employee Turnover Prediction

mitigation strategies, organizations can maximize the benefits of the application while minimizing potential drawbacks.

Overall, the employee turnover prediction application represents a strategic investment in human capital management, enabling organizations to foster a more engaged, productive, and resilient workforce. By harnessing the power of data and predictive analytics, organizations can proactively address turnover challenges and build a culture of retention and organizational success.

References

- Boushey, H., & Glynn, S. J. (2012). *There Are Significant Business Costs to Replacing Employees*. Center for American Progress. Retrieved from <https://www.americanprogress.org/issues/economy/reports/2012/11/16/44464/there-are-significant-business-costs-to-replacing-employees/>
- MarketsandMarkets. (2020). *HR Analytics Market by Component (Solutions and Services), Application (Workforce Management, Recruitment, and Employee Development), Organization Size, Deployment Type, Vertical (BFSI, Manufacturing), and Region - Global Forecast to 2025*. Retrieved from <https://www.marketsandmarkets.com/Market-Reports/hr-analytics-market-124068045.html>
- PricewaterhouseCoopers. (2012). *Retaining talent: A guide to analyzing and managing employee turnover*. Retrieved from <https://www.pwc.com/m1/en/services/consulting/documents/retaining-talent-a-guide-to-analyzing-and-managing-employee-turnover.pdf>
- Society for Human Resource Management. (2019). *Employee Turnover: The Real Cost of Losing an Employee*. Retrieved from <https://www.shrm.org/hr-today/news/hr-magazine/summer2019/pages/employee-turnover-the-real-cost-of-losing-an-employee.aspx>