

Employee Turnover Prediction – Salifort Motors

Project Overview

The HR department at Salifort Motors is looking to improve employee satisfaction levels within the company. They have collected data from employees and are seeking data-driven insights to understand what factors might lead an employee to leave the company. The business question is essentially: what is likely to make an employee leave the company?

This project was completed as part of the **Capstone Project in the Google Advanced Data Analytics Certificate on Coursera**. It applies skills in data wrangling, exploratory analysis, and machine learning.

The goal of this project is to analyze the collected HR data and build a model that predicts whether an employee will leave the company. By identifying employees who are likely to quit, we can pinpoint contributing factors and take proactive measures to increase employee retention. This is beneficial to the company considering the significant time and cost involved in finding, hiring, and training new employees.

Dataset

This project uses an employee HR analytics dataset (file: HR_capstone_dataset.csv), originally from a Kaggle HR analytics challenge. The dataset contains **14,999** employee records with **10** features related to each employee's workplace experience.

Key variables include each employee's self-reported satisfaction level (0 to 1), their last performance evaluation score (0 to 1), the number of projects they worked on, their average monthly hours, tenure at the company (in years), whether they had a work accident, whether they had a promotion in the last 5 years, their department, and their salary level (categorized as low, medium, high).

The target column left indicates whether the employee left the company (1 means the employee left, 0 means they stayed).

Before analysis, the data was checked for quality issues. It was relatively clean with no missing values, so minimal data cleaning was required. Basic preprocessing included ensuring correct data types and addressing a few outliers (for example, extremely high tenure values were investigated to ensure they were reasonable).

Analysis and Modeling Approach

To tackle the problem, a structured data analysis approach was followed:

1.Exploratory Data Analysis (EDA): We conducted EDA to understand the data distribution and relationships between features and employee attrition (left). This involved visualizing features like satisfaction level, monthly hours, and evaluating how they differ between employees who left vs. those who stayed. We identified

trends such as lower average satisfaction among those who left, and examined department-wise and tenure-wise attrition patterns.

2.Feature Engineering: Based on initial insights, we created or transformed features to improve modeling. For instance, we binned certain continuous variables (like tenure) to see if that improved predictive power and removed any outliers that could unduly influence the model (e.g., very unusual combinations of low satisfaction but extremely high hours).

3.Model Selection: We built and evaluated several machine learning classification models to predict employee turnover:

- **Logistic Regression** – as a baseline linear model.
- **Decision Tree** – to capture nonlinear relationships and feature interactions.
- **Random Forest** – an ensemble of decision trees to improve generalization.

4.Hyperparameter Tuning: For the tree-based models (Decision Tree, Random Forest), we performed hyperparameter tuning using cross-validation to find the best model settings (such as tree depth, number of estimators, learning rate for XGBoost, etc.). Grid search and cross-validated scoring (using metrics like AUC and F1-score) were used to avoid overfitting while maximizing predictive performance.

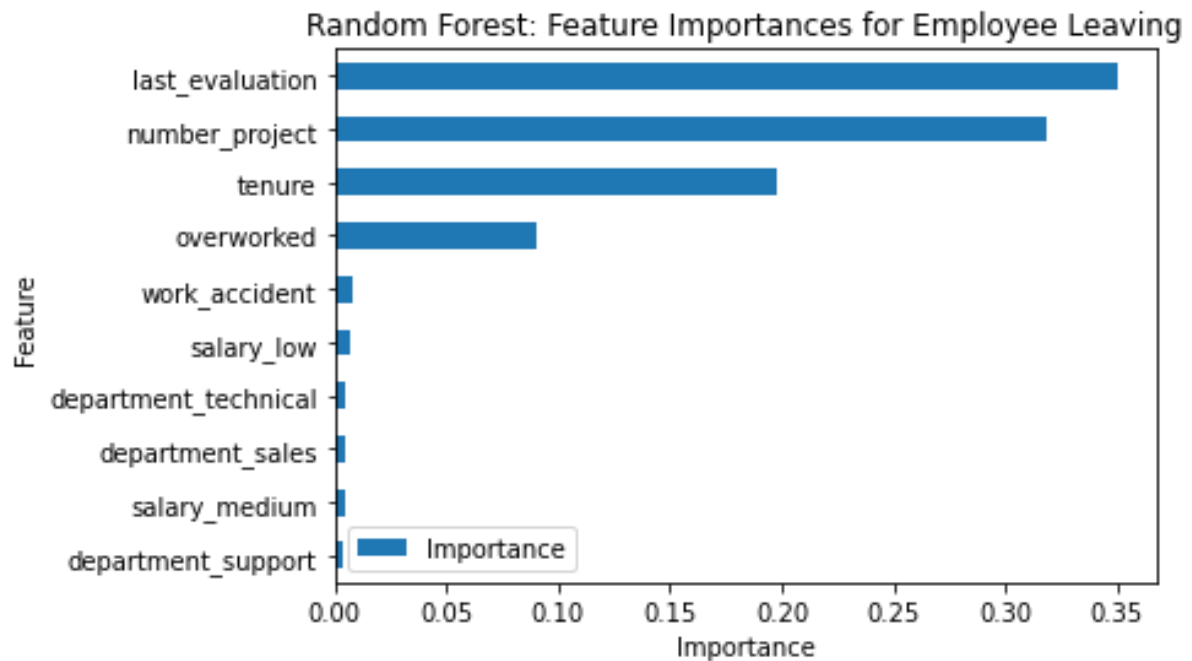
5.Model Evaluation: We evaluated each model on a held-out test set using **accuracy, precision, recall, and F1-score**. This allowed us to compare how well the models identify employees who leave versus those who stay. We paid special attention to **recall** for the “leaving” class (to ensure we catch as many at-risk employees as possible) and balanced performance overall.

6.Model Selection: Finally, we selected the best-performing model based on the evaluation metrics and complexity. We also interpreted the model to understand which features were most influential in predicting turnover.

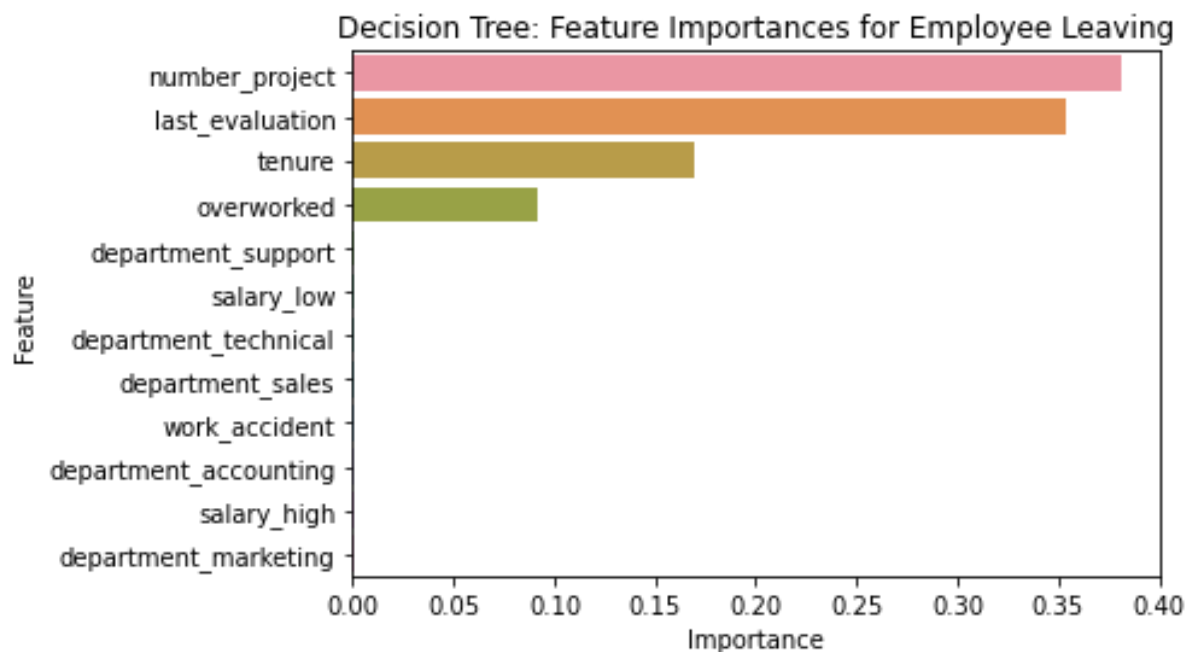
Results and Key Findings

•**Model Performance:** Among the models tested, the Random Forest classifier performed the best in predicting employee turnover. After tuning, it achieved a test **accuracy** of around **98%**, with a **precision** and **recall** in the **92–94%** range for identifying employees who leave. This outperformed the **logistic regression** (which had ~**83% accuracy**) and the single decision tree model. The XGBoost model had performance comparable to the Random Forest, also achieving very high accuracy, but with a slightly longer training time; hence, the Random Forest was chosen for its balance of accuracy and simplicity. These results indicate that the model can very reliably distinguish employees who are likely to leave from those who will stay, given the features in the dataset.

- **Important Factors Influencing Turnover:** The machine learning models and EDA together highlighted several key factors:



In the random forest model above, `'last_evaluation'`, `'tenure'`, `'number_project'`, `'overworked'`, `'salary_low'`, and `'work_accident'` have the highest importance. These variables are most helpful in predicting the outcome variable, `'left'`.



Barplot above shows the most relevant variables: `'last_evaluation'`, `'number_project'`, `'tenure'` and `'overworked'`.

Recommendations

Based on these findings, a number of actions are recommended for Salifort Motors' HR and management:

1.Improve Job Satisfaction: Conduct regular employee satisfaction surveys and act on feedback. Focus on improving workplace factors (like management practices, team culture, work-life balance) that drive satisfaction. Employees who report low satisfaction should be engaged by managers to address their concerns before they decide to leave.

2.Manage Workload and Overtime: Ensure that no employees are consistently overworked. Consider hiring additional staff or redistributing tasks in teams where many employees have extremely high average monthly hours. Encouraging employees to take time off and promoting a healthy work-life balance can reduce burnout.

3.Recognize and Reward High Performers: Since even high-performing employees were leaving when overburdened or if they felt unrewarded, establish stronger recognition programs. This could include bonuses, salary raises, or promotions for those with strong performance evaluations, to make them feel valued and motivated to stay.

4.Competitive Compensation: Review the salary structure, especially for roles or departments with high turnover. If low-salary employees are leaving at higher rates, consider adjustments to ensure compensation is competitive with the market and commensurate with the employees' contributions. This will help retain talent that might otherwise leave for better-paying opportunities.

5.Career Development Opportunities: Provide clear career paths and professional development. Employees who see opportunities for growth and have a chance for promotion are more likely to stay. Mentorship programs, training, and internal recruitment for higher positions can alleviate feelings of stagnation.

6.Department-Focused Interventions: Investigate the root causes of high turnover in departments like Sales. This might involve speaking with employees and managers in those departments to identify specific issues (e.g., sales targets pressure, commission structure, workplace environment). Tailored interventions can then be implemented, such as extra support or resources for high-stress departments.

Conclusion

In this capstone project, we successfully built a predictive model for employee turnover and identified key factors contributing to attrition at Salifort Motors. The analysis provides valuable insights: low satisfaction, excessive overtime, lack of advancement, and low pay are all linked to higher employee turnover. By addressing these areas with targeted HR strategies and policies, Salifort Motors can improve its employee retention. Ultimately, retaining talent will save costs on rehiring and training and contribute to a more experienced and stable workforce, giving the company a competitive edge. The project demonstrates how data analytics and machine learning can inform decision-making in HR, turning raw employee data into actionable business insights.