Employee Turnover Predictor

Empowering Salifort Motors HR

Bangkit Adhi Wiguna
Research and Data Analytics Officer
Center for Digital Society UGM

Google Advanced Data Analytics Capstone Project

PACE Stage #1:
Plan

Project Background

- **Growing Concern**: Salifort Motors has observed a **noticeable increase in employee turnover** over the past few years. This isn't just a number, it represents a significant drain on resources and organizational knowledge.
- **Impact on Business**: High attrition rates lead to:
  - **Increased Recruitment Costs**: Time, effort, and money spent on advertising, interviewing, and onboarding new staff.
  - **Productivity Losses**: Gaps in teams, learning curves for new hires, and disruption to ongoing projects.
  - **Loss of Institutional Knowledge**: Experienced employees take valuable insights and expertise with them.
  - **Morale Impact**: Turnover can negatively affect the morale of remaining employees.
- **Traditional Methods Limitations**: Relying solely on exit interviews or reactive measures is often too late to retain valuable employees. We needed a more proactive approach.
- **Strategic Imperative**: To maintain a competitive edge and stable workforce, Salifort Motors needs a **data-driven strategy to understand and mitigate turnover**.

Data & Methodology: From Raw Data to Actionable Insights

Phase 1: Data Collection & Understanding

- **Objective**: Comprehensive data assembly for analysis.
- **Total Records**: Analyzed **11,991 unique employee records**.
- **Sources**: Integrated HR Information Systems (HRIS), performance reviews, compensation, and project management tools.
- **Key Data Points**: Employee satisfaction, last evaluation, # projects, avg. monthly hours, years at company, work accident, promotion (last 5 years), department, salary, and left (target variable: 1 for turnover, 0 for retained).
- **EDA & Preprocessing**: Assessed data quality, identified patterns (e.g., low salary & high projects as drivers), and prepared data for modeling.

Data & Methodology:
From Raw Data to Actionable Insights

Phase 2: Model Development & Evaluation

- **Objective**: Build and validate a robust predictive model.
- **Algorithms**: Explored Logistic Regression, Random Forest, and **XGBoost** (selected for superior Recall).
- **Training**: Historical data split into training and unseen test sets.
- **Evaluation (Recall-First)**:
    - **Recall (Primary Focus)**: Maximize detection of actual departures (minimize false negatives) for early intervention.
    - **Accuracy, Precision, F1-Score**: Used for comprehensive performance assessment.
- **Optimization**: Fine-tuned the chosen model for peak performance and interpretability.

PACE Stage #2:
Analyze

Salifort Motors faces an

Overall Turnover Rate:

16,6%

Total Attrition: **1,991** employees left out of 11,991 records.

This figure represents the baseline challenge that the predictive model aims to mitigate.

The Main Driver

Salary is the single most differentiating factor for turnover.

Low Salary:

- **20.5% Turnover Rate**
- **Observation:** This group represents the highest risk for attrition.

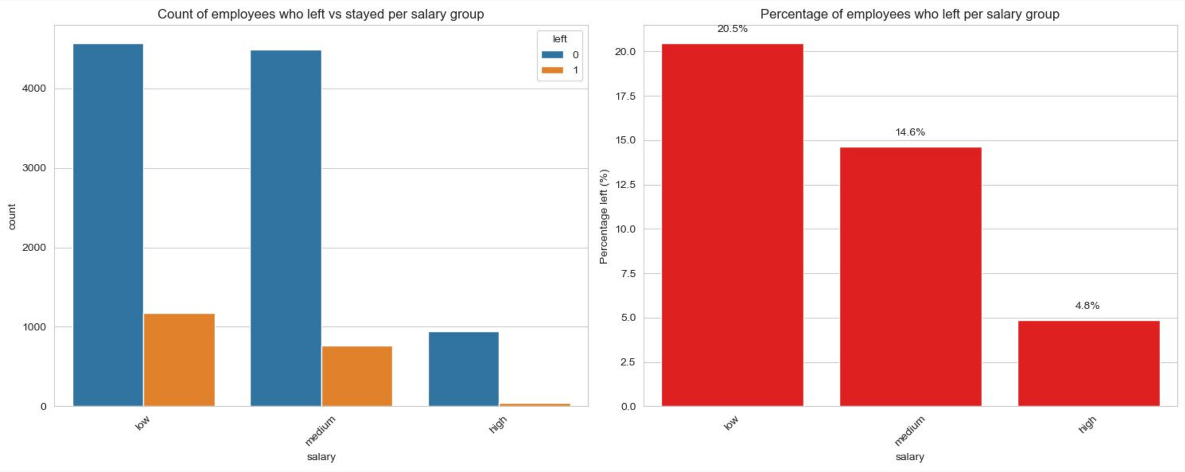
Medium Salary:

- **14.6% Turnover Rate**
- **Observation:** Shows a moderate risk of turnover.

High Salary:

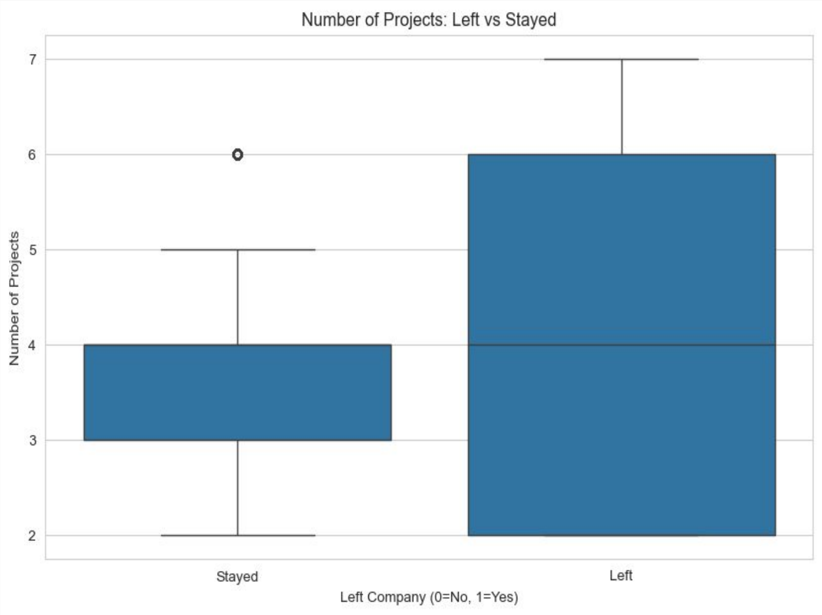
- **4.8% Turnover Rate**
- **Observation:** This group demonstrates the lowest risk for attrition.

Turnover Rate by Salary Level



Employees with a **Low Salary** leave at more than **4x the rate** of those with a High Salary. Compensation directly impacts retention.

Projects Number:
The Workload Extremes Paradox



Under-Utilized (High Risk): Employees assigned only **2 projects** exhibited an extremely high turnover rate of **54.2%**.

- Employees with very little work become bored, feel stagnant, or are poorly utilized, leading them to seek new employment.

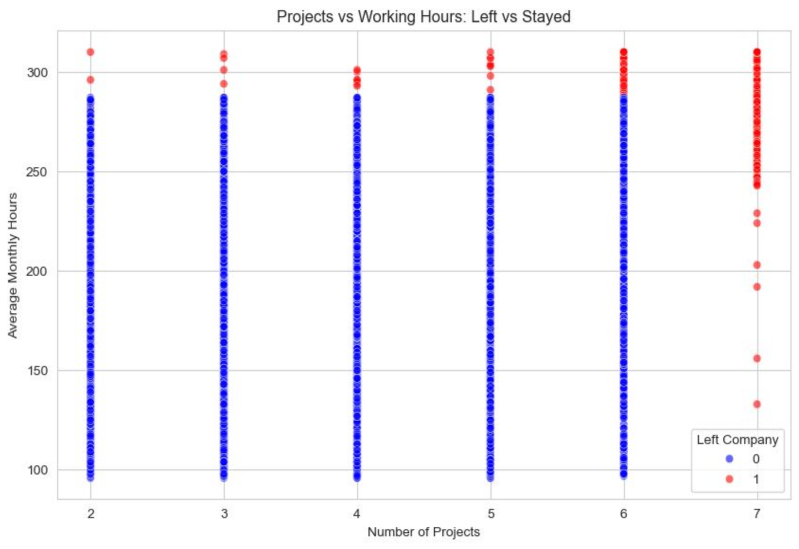
Optimal Zone (Low Risk): The lowest turnover rate was found among employees handling **3 projects 1.1%** turnover.

- This indicates the 'sweet spot' for balanced workload and engagement.

Burnout Zone (High Risk): Employees with **6 projects** had a very high turnover rate of **44.9%**, and those with **7 projects** saw **100.0%** turnover.

- Severe over-utilization, indicative of burnout and an impossible work-life balance, directly drives immediate attrition.

Workload & Hours: The High Cost of Extremes



Leavers Work More: Employees who **Left** worked a consistently higher number of hours, averaging **208 hours/month**.

Stayers Work Less: Employees who **Stayed** worked less, averaging **199 hours/month**.

Dual Peak Phenomenon: The distribution for employees who **Left** often shows **two distinct peaks**:

1. One peak around **$130-160$ hours/month** (The under-utilized/low-satisfaction group).
2. A massive peak around **$240-310$ hours/month** (The burnout/over-utilized group).

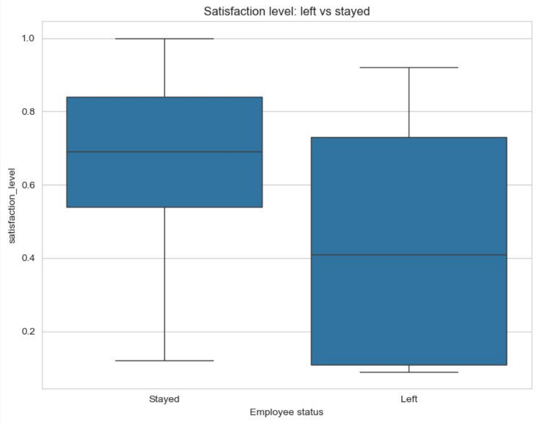
The data confirms that **excessive working hours are a major contributor to turnover**, alongside the risk of *under-utilization*.

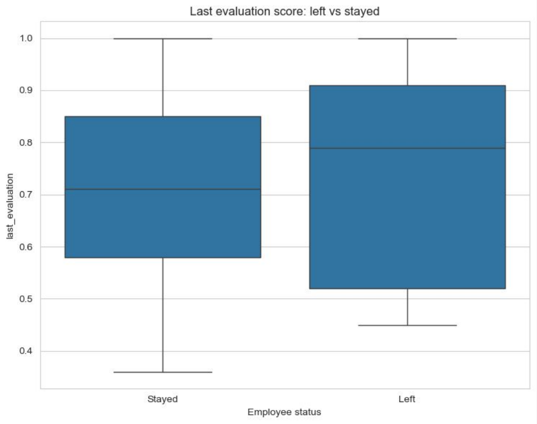
Performance vs. Satisfaction

Lowest Satisfaction, Highest Risk: Employees with **very low satisfaction (< 0.2)** are the highest risk group for turnover, regardless of their performance.

The High-Performer Trap: High Evaluation (> 0.8) + Low-Medium Satisfaction (< 0.5) = Critical Risk Group. These are valuable, productive employees likely seeking better opportunities due to unhappiness.

Under-Performer Risk: Low Evaluation (< 0.6) + Low Satisfaction (< 0.5) = Also a high-risk group. Unhappy and underperforming employees are prone to leave or be managed out.



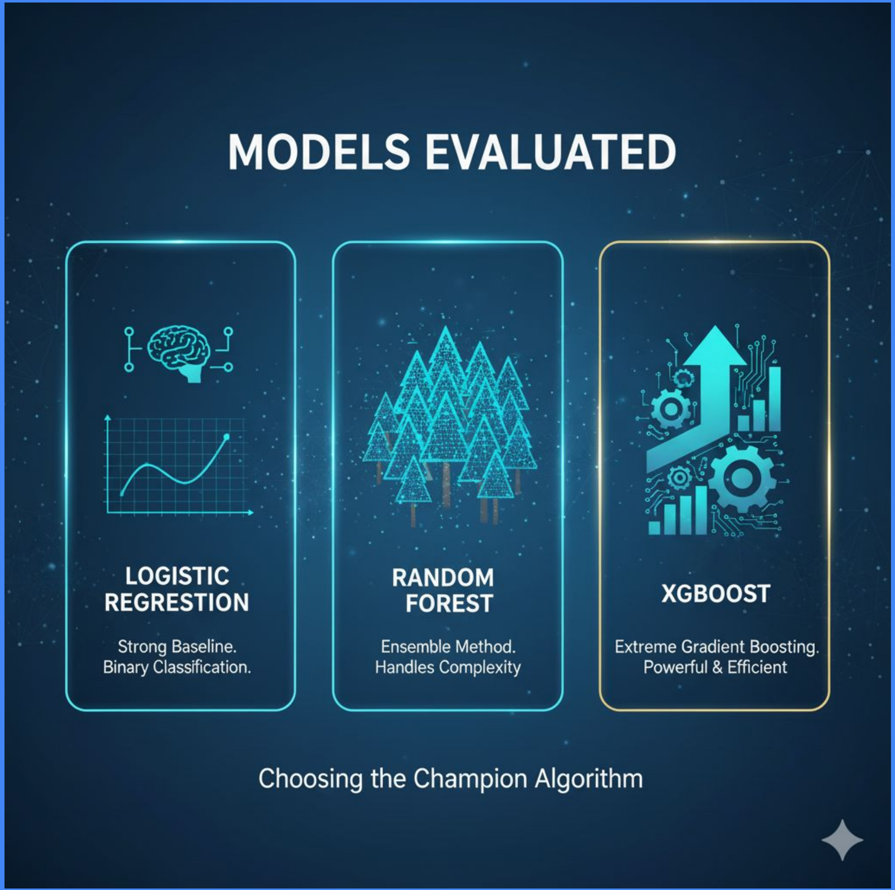


PACE Stage #3:
Construct

Model Training and Evaluation

Models Evaluated: We systematically assessed three robust machine learning algorithms:

- **Logistic Regression**: A strong baseline for binary classification.
- **Random Forest**: An ensemble method known for its accuracy and ability to handle complex relationships.
- **XGBoost (Extreme Gradient Boosting)**: Another highly efficient and powerful gradient boosting framework.



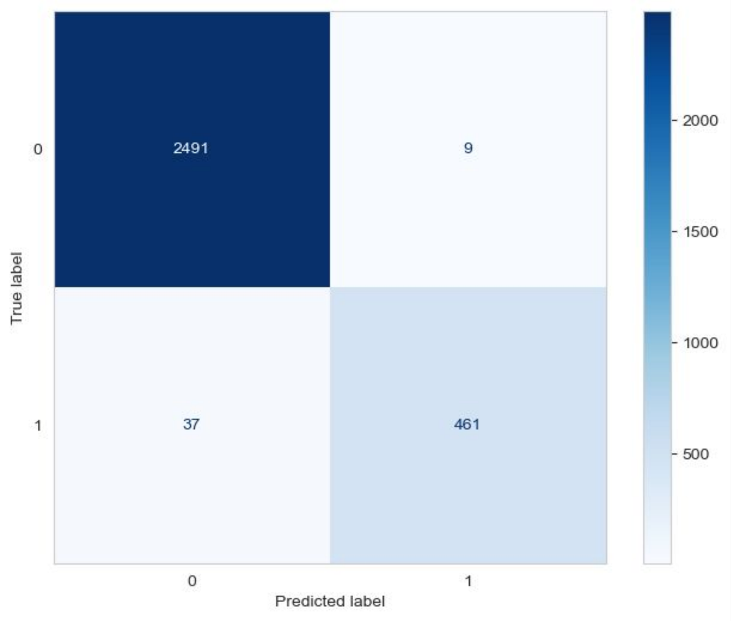
The Champion Model: XGBoost

After extensive training and optimization, the XGBoost model demonstrated exceptional performance on a completely **unseen test dataset** (approximately 3,000 employee records).

Key Performance Metrics: XGBoost on Unseen Data

- **Accuracy: 0.985**
- **Recall: 0.926**
- **Precision: 0.981**
- **F1-score: 0.952**
- **False Negatives**: Only **37** misclassified actual departures out of nearly 3,000 test cases.

XGBoost Confusion Matrix



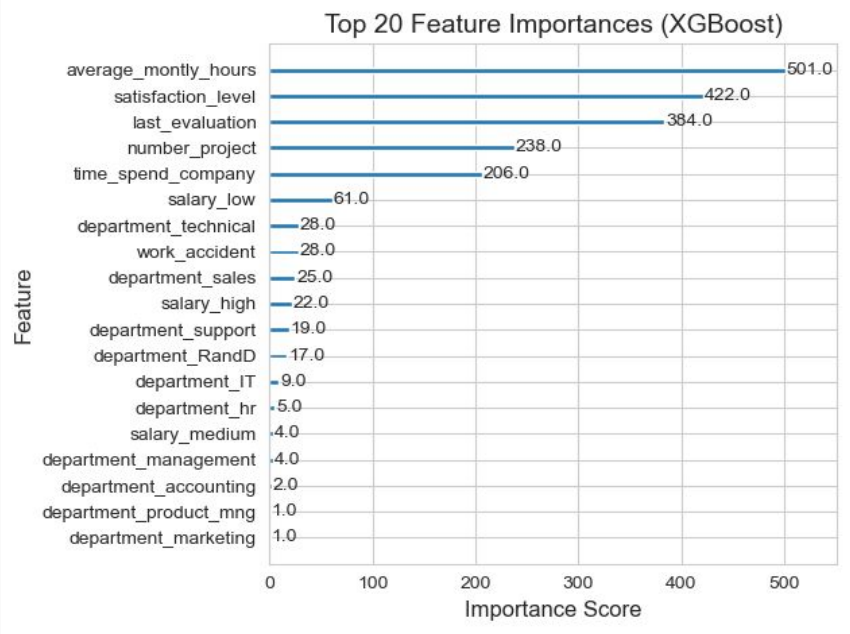
The full dataset of 11,991 employee records was split into an **80% training set (~9,593 records)** to train the model, and a **20% testing set (~2,398 records)** to objectively evaluate its performance on unseen data.

	Precision	Recall	F1-score	Accuracy
Random Forest	0.987715	0.913589	0.949172	0.983765
XGBoost	0.980791	0.920958	0.949912	0.983876

5 Fold Cross-validation results

PACE Stage #4:
Execute

Understanding Model Drivers: Feature Importance



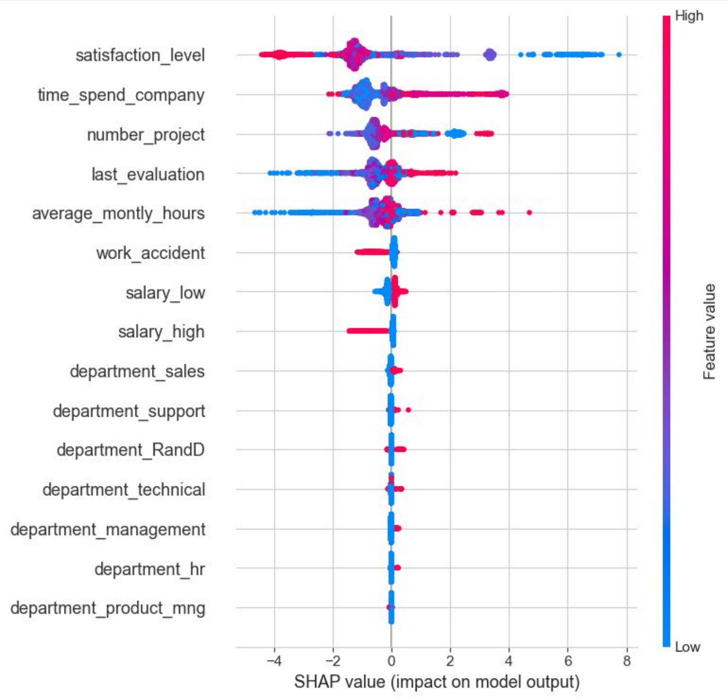
Workload & Engagement: average_montly_hours, satisfaction_level, last_evaluation, number_project, time_spend_company are the **top 5 most critical features**. This validates our earlier EDA insights about burnout and engagement.

Compensation: salary_low is a significant positive indicator of turnover, highlighting its critical role. salary_high is less important, suggesting a high salary acts more as a protective factor.

Departmental Influence: Department features (e.g., department_technical, department_sales) show relatively **lower importance scores**. This reinforces our finding that the turnover issue is systemic, not department-specific.

Actionable Insight: Focusing retention efforts on these top-ranked features will yield the most significant impact, as they are where the model finds the strongest predictive power.

Deeper Dive: SHAP Values for Individual Impact



Satisfaction Level: **Low satisfaction (blue)** strongly pushes employees **towards leaving** (positive SHAP values).

Time Spend Company: Both **very low and very high time spent** at the company push towards leaving, confirming the "U-shape" for tenure.

Number of Projects: Both **very low and very high project counts** push towards leaving.

Average Monthly Hours: **High hours (red)** significantly push towards **leaving**.

Salary: **Low salary (red)** strongly pushes towards **leaving**. **High salary (blue)** pushes towards **staying**.

Actionable Insight: SHAP values provide a granular view, allowing HR to understand *why* a specific employee is flagged. For example, an employee might be at risk due to high hours *and* low satisfaction.

Business Insights & Strategic Recommendations

Actionable HR Strategies Driven by Model Insights

- **Proactive Workload Management:**
  - Monitor number_project to avoid **under-utilization (2 projects)** and **burnout (6+ projects)**.
  - Flag employees consistently exceeding **200+ monthly hours**.
- **Boost Engagement & Satisfaction:**
  - Implement regular satisfaction_level check-ins, especially for scores **below 0.5**.
  - Address high-performers (last_evaluation $\ge$ 0.8) with low satisfaction to prevent "flight risk."
- **Targeted Compensation Review:**
  - Prioritize compensation adjustments for employees in the **salary_low** category, particularly when combined with other risk factors.
- **Leverage Predictive Alerts:**
  - Utilize the XGBoost model to generate an **"at-risk" employee dashboard**.
  - Enable **personalized HR interventions** (e.g., career counseling, workload rebalancing, recognition).
- **Overall Goal:** Shift from reactive hiring to proactive, data-informed employee retention, fostering a more stable and productive Salifort Motors.

Thank you!

Contact me:

Bangkit Adhi Wiguna
Research & Data Analytics Officer
Center for Digital Society
FISIPOL UGM

bangkitwiguna984@gmail.com