Salifort Motors

Capstone Project Submission

Executive Summary & Project Proposal

Prepared by: Gerald Appau-Bonsu

Course: Google Advanced Data Analytics

Date: August 2025

Executive Summary (Updated): Predicting Employee Turnover at Salifort Motors

Objective

Use employee survey data to predict the likelihood of turnover and surface actionable drivers to improve retention and reduce replacement costs.

Data Overview

After de-duplication, the analytical dataset contains **11,991 employees**. Turnover is the minority class: **1,991 employees (16.6%)** left and **10,000 (83.4%)** stayed.

Approach

Built a baseline Logistic Regression for interpretability and an XGBoost classifier for performance. Both models used a preprocessing pipeline (scaling numeric features and one-hot encoding categorical features). Evaluation focused on precision, recall, F1, and ROC AUC on a stratified 80/20 train-test split.

Model Performance (Test Set)

Metric	Logistic Regression	XGBoost
Accuracy	0.787	0.975
Precision (leavers)	0.429	0.914
Recall (leavers)	0.844	0.937
F1 (leavers)	0.569	0.926
ROC AUC	0.849	0.983

Interpretation

• **Logistic Regression** achieves high recall but low precision, making it a good early-warning baseline that tends to over-flag risk. • **XGBoost** delivers a step-change improvement: high precision and high recall with an excellent ROC AUC, providing reliable, deployable predictions for targeted retention.

Key Drivers & Patterns Observed

• **Workload & intensity:** Higher average monthly hours and a greater number of projects are positively associated with turnover (burnout risk).
• **Tenure:** Increased time at the company is correlated with leaving, suggesting stagnation and limited advancement for some cohorts.
• **Compensation:** Turnover rates differ sharply by salary band (low ≈ 29.7%, medium ≈ 20.4%, high ≈ 6.6%), indicating pay competitiveness is a strong lever.

Recommendations

1) **Deploy XGBoost** as the production model for risk scoring; maintain Logistic Regression for executive explainability. 2) **Targeted retention** for flagged employees: rebalance workload, review compensation, and strengthen progression/mentoring—especially for mid-tenure cohorts. 3) **Threshold tuning** with HR to balance false positives vs. missed leavers based on intervention cost. 4) **Monitoring & fairness**: track performance by department/salary band; retrain quarterly with new data.

Conclusion

The updated modeling shows XGBoost provides both accuracy and actionable reliability for identifying at-risk employees, enabling Salifort Motors to focus interventions where they will have the greatest impact on retention and cost savings.

Project Proposal: Employee Turnover Prediction & Retention Insights

Organization: Salifort Motors
Analyst: [Your Name]

Milestones are aligned with the PACE framework:

1. Plan: Define scope & success criteria.
2. Analyze: Data preparation & exploration.
3. Construct: Model development and evaluation.
4. Execute: Deliver executive summary & recommendations.