Employee Turnover Analysis — Model Selection Report

Salifort Motors • HR Attrition Project

ISSUE / PROBLEM

- The company faces **17% employee turnover**, impacting productivity and retention costs.
- Need a reliable model to **identify high-risk employees early**.
- Target variable is **imbalanced** → "left = 1" is difficult to predict.

RESPONSE

Built and compared four models:

1. **Logistic Regression** (baseline)
2. **Decision Tree**
3. **Random Forest**
4. **XGBoost**

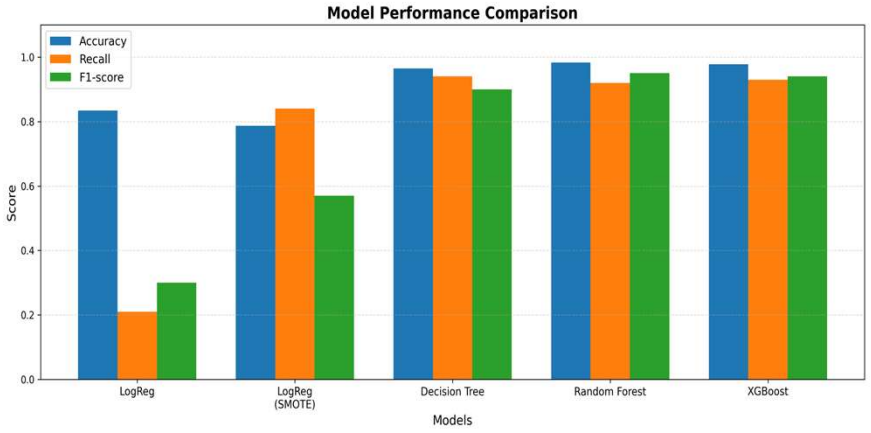
- Applied **SMOTE** to fix class imbalance.
- Standardized numeric features and evaluated models using: **Accuracy, Precision, Recall, and F1-Score**.

IMPACT

1. SMOTE improved minority-class recall from 21% → 84%.
2. Decision Tree reached 96.5% accuracy, F1=0.90.
3. Random Forest performed best: 98.3% accuracy, F1=0.95.
4. XGBoost is also strong with 97.8% accuracy, F1=0.94.

Key Findings on Drivers of Turnover

- Satisfaction level is the strongest predictor of turnover — employees with low satisfaction are significantly more likely to leave.
- Salary level, especially low salary, has a major impact on attrition and is consistently ranked high in feature importance.
- Overwork indicators such as high average monthly hours and heavy project load increase the likelihood of leaving due to burnout.
- Lack of advancement (no promotion in 5 years) and mid-level tenure without career growth contribute to higher turnover risk.
- Department has only a minor influence on predictions compared to satisfaction, salary, and workload-related factors.
- **Compare ML models and select the strongest predictor of turnover:**



Model	Precision	Recall	F1-Score	Accuracy	Notes
Random Forest	**0.98**	**0.92**	**0.95**	**0.983**	⭐ Best overall (high precision + recall)
XGBoost	0.94	0.93	0.94	0.978	Extremely strong, stable model
Decision Tree	0.86	0.94	0.90	0.965	Simple, interpretable but less robust
Logistic Regression (SMOTE)	0.43	0.84	0.57	0.787	Improved with SMOTE but still weak
Logistic Regression (Raw)	0.50	0.21	0.30	0.834	Performs poorly, especially recall

KEY INSIGHTS

- **Random Forest is the best-performing model**, offering the highest and most balanced precision, recall, and F1-score.
- SMOTE significantly improved all models, especially Logistic Regression.
- Important predictors across models: **satisfaction_level, salary_low, tenure, last_evaluation, average_monthly_hours**
- Tree-based models (Decision Tree, Random Forest, XGBoost) outperform linear models due to non-linear relationships in turnover.