Employee Turnover Prediction – Salifort Motors

Classifying who is at risk of leaving Salifort Motors and why

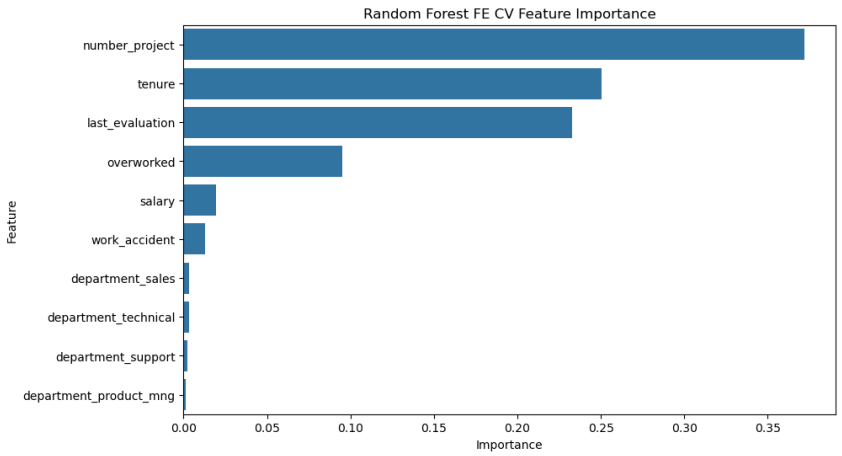
Project Overview

Salifort Motors wants to reduce costly employee turnover by identifying which employees are at a higher risk of leaving and why. Using historical HR data and guided by EDA, several Models were evaluated such as a Logistic Regression, Decision Tree, Random Forest and XGBoost both with and without feature engineering. The final solution is a feature-engineered Random Forest Classifier that reliably predicts who is likely to leave while remaining transparent and defensible for HR, Legal and Senior Leadership.

Key Insights

- Employee turnover is **not random** as it is strongly associated with workload, tenure, performance and project load.
- Employees with **many projects, sustained long hours (overworked), strong performance reviews** and **3-5 years of tenure** show the highest risk of leaving, especially when there is advancement or recognition lags.
- **Low misclassification of Leavers matters**, as the Tree-Based Models significantly reduce False Negatives compared with Logistic Regression, meaning that fewer high-risk employees are missed.
- **Feature engineering improved reliability** by removing the potential leakage and focusing on interpretable and business aligned signals (e.g. the `overworked` flag).
- Department and Salary level play a **secondary role** relative to workload and progression, thereby suggesting targeted retention interventions rather than broad pay or org-wide changes.
- The **Champion Model** achieved a 97% Test Accuracy with strong Recall and F1-Score for at-risk employees, providing a robust and trustworthy basis for targeted retention actions.

Details



The final champion Model is a feature-engineered Random Forest Classifier that was trained on cleaned, de-duplicated HR records with derived variables such as an `overworked` status. After rigorous Cross-Validation and comparison against Decision Tree, XGBoost and baseline Logistic Regression Models, it delivered the best balance of Accuracy and Recall for Leavers and generalisation on the hold-out test set. Its feature importance plot is intuitive and easily explained to HR, Legal and Leadership stakeholders, therefore making it suitable for operational use.

Next Steps

The next steps are to deploy the feature-engineered Random Forest Model in a secure retention dashboard, where HR and Managers can view risk tiers with clear explanations and use them to trigger supportive actions such as recognition, workload review or career conversations. Salifort should also run a focused pilot in selected departments as to monitor the impacts on employee turnover and manager/employee feedback and refine the thresholds and playbooks before scaling organisation-wide. Alongside this, the company should formalise clear governance usage policy such as access control, performance and bias monitoring and periodic retraining. All while gradually enhancing the Model with additional, privacy-safe data (e.g. internal mobility, training and engagement trends) to strengthen predictive power and alignment with broader people strategies.