IBM HR Analytics – Survival Modeling for Attrition

Student: Quach Lan Huong - ID: 1137131 • Group 8

Summary

We model employee attrition over time using Cox Proportional Hazards and Random Survival Forest, with an XGBoost + SHAP companion model for interpretability. The final deliverable is a 3/6/12 month risk table that prioritizes proactive HR actions (check-ins, workload adjustments, rotation/promotion paths, targeted compensation).

Repository

https://github.com/Jennyyzu227/Final-assignment-HR-attrition-survival

Run Instructions

1. Install: pip install -r requirements.txt
2. Place dataset at data/hr_data.csv; place probabilities at data/employee_risk_table_3_6_12m.(csv|xlsx) if provided.
3. Export risk table: python src/make_risk_table.py --probs data/employee_risk_table_3_6_12m.csv --out results/employee_risk_table_3_6_12m.xlsx
4. Open the XLSX and sort by 6 month probability to prioritize top 10–20% employees for intervention.

Notes

- Risk-tier thresholds (6 month): High ≥ 0.30; Medium 0.10–0.30; Low < 0.10. Edit CLI flags to adjust.
- Dataset is synthetic/demo; retrain and recalibrate thresholds for production use.