



MIT

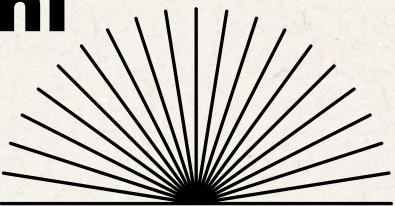
Academy of
Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

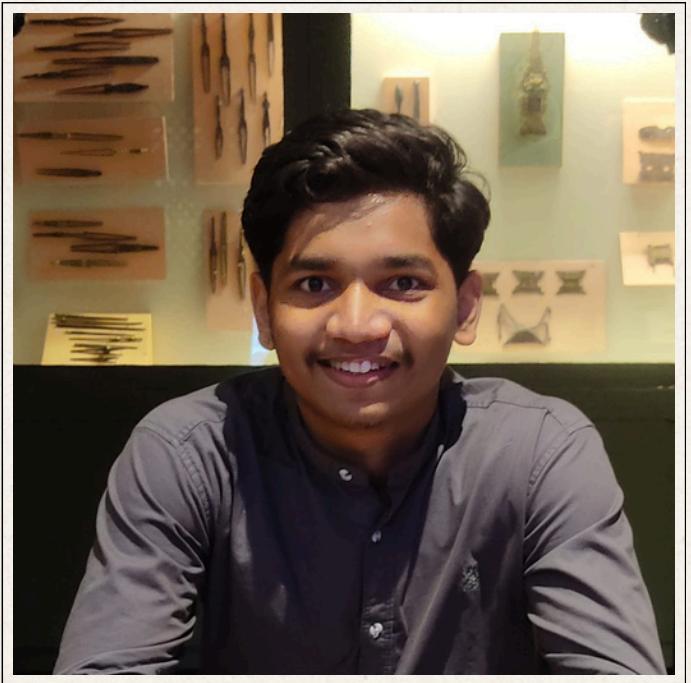
CI Project

EMPLOYEE ATTRITION PREDICTION

Guide: Dr. Abhilasha Joshi



Our team



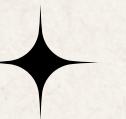
Aditya Ambure

202301070154



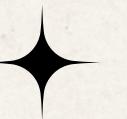
Sneha Kabra

202301070162



Devang Deshmukh

202301070170



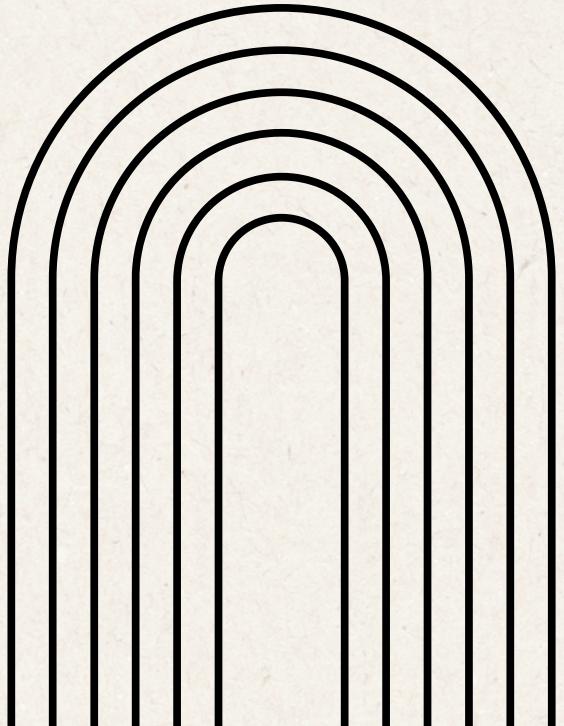
Krishna Kedar

202301070177



Introduction

- Employee attrition = employees leaving the company
- Leads to hiring cost, productivity loss, training cost
- Machine Learning helps predict who is likely to leave
- Goal: Build a model that predicts attrition accurately



Problem Statement

Predict whether an employee will leave the company.

Objectives:

01 • Preprocess HR dataset

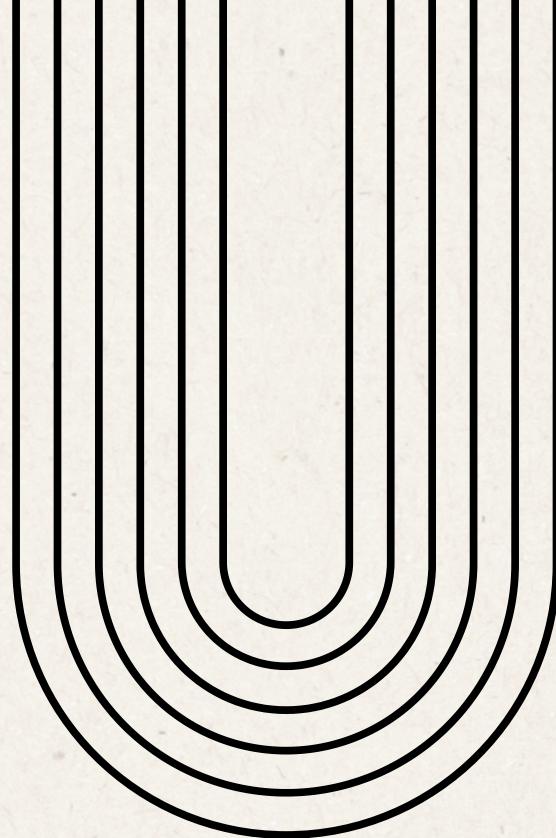
02 • Perform EDA to find attrition patterns

03 • Engineer useful features

04 • Train & compare ML models

05 • Select best model (Random Forest)

06 • Deploy final model using Streamlit

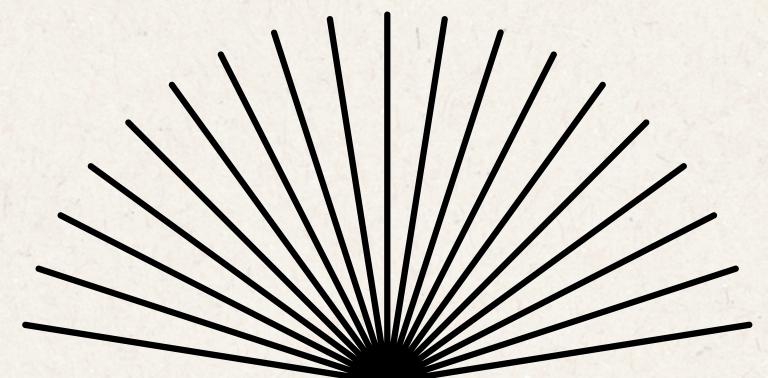


Dataset Overview & Key Features

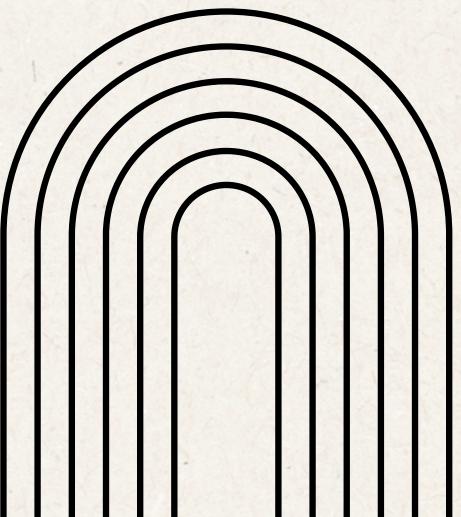
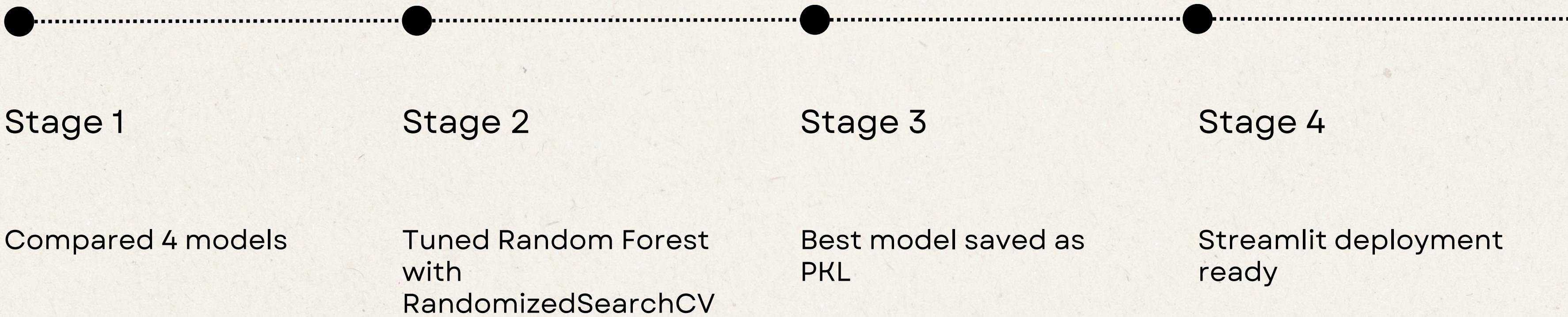
Dataset Size: ~14,000 records, 35+ attributes

Important Features :

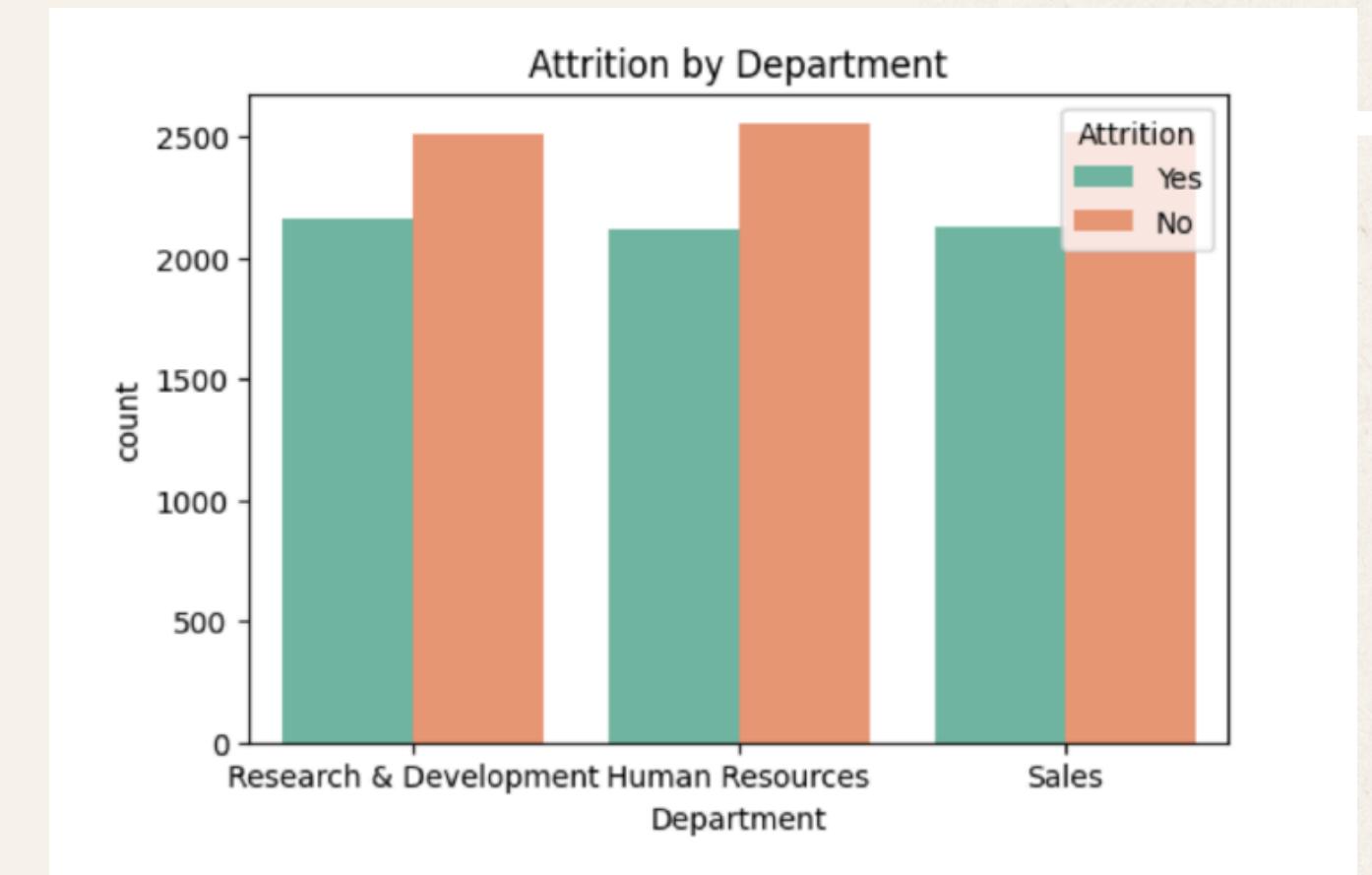
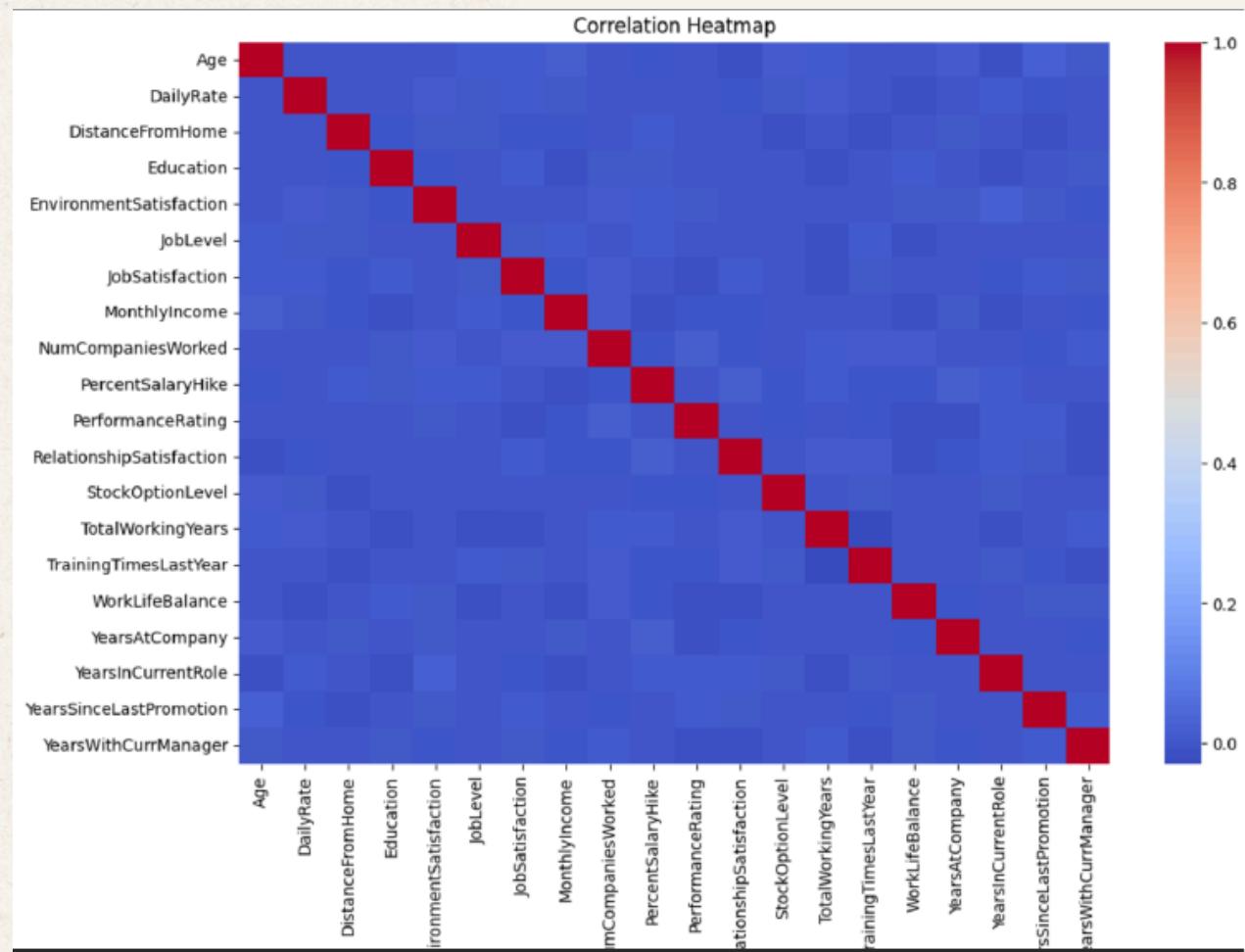
- 1. Age, Monthly Income, Job Level**
- 2. Total Working Years, Overtime**
- 3. Years at Company, Distance from Home**
- 4. Job Satisfaction**
- 5. Target Variable: Attrition (Yes/No)**



Modelling



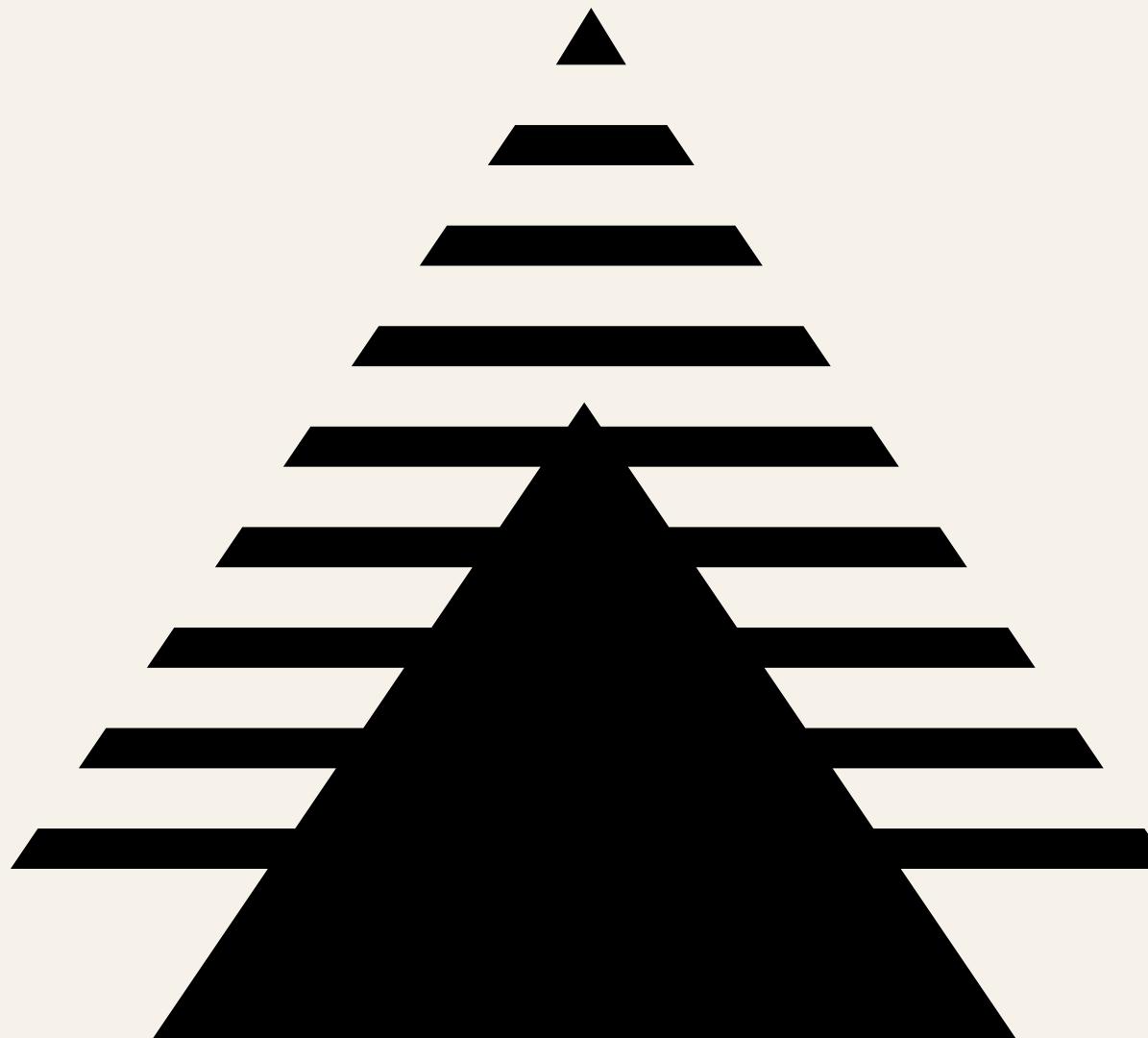
EDA Highlights



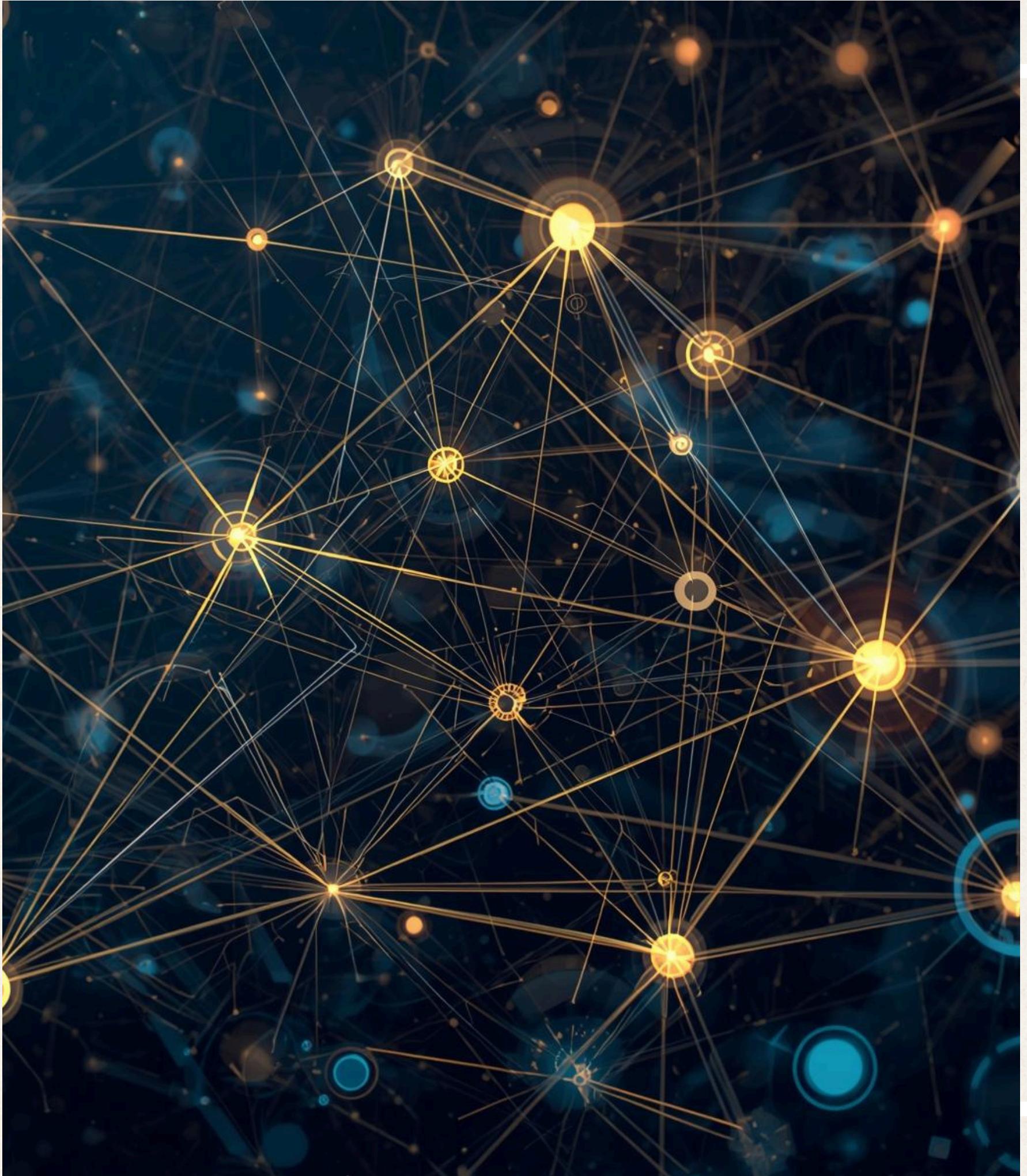
- Highest attrition occurs in the 26–35 age group.
- Overtime hours correlate strongly with attrition.
- Low job satisfaction leads to higher turnover.
- Lower salaries are linked to increased attrition.
- The Human Resources department shows slightly higher attrition rates.

No significant multicollinearity was detected, with all correlations below 0.95.

Feature Engineering



- years_at_company_by_age
- promotion_flag
- Target encoding for certain features
- One-hot encoding for categorical data
- Scaling numeric features
- SMOTE for balancing if needed



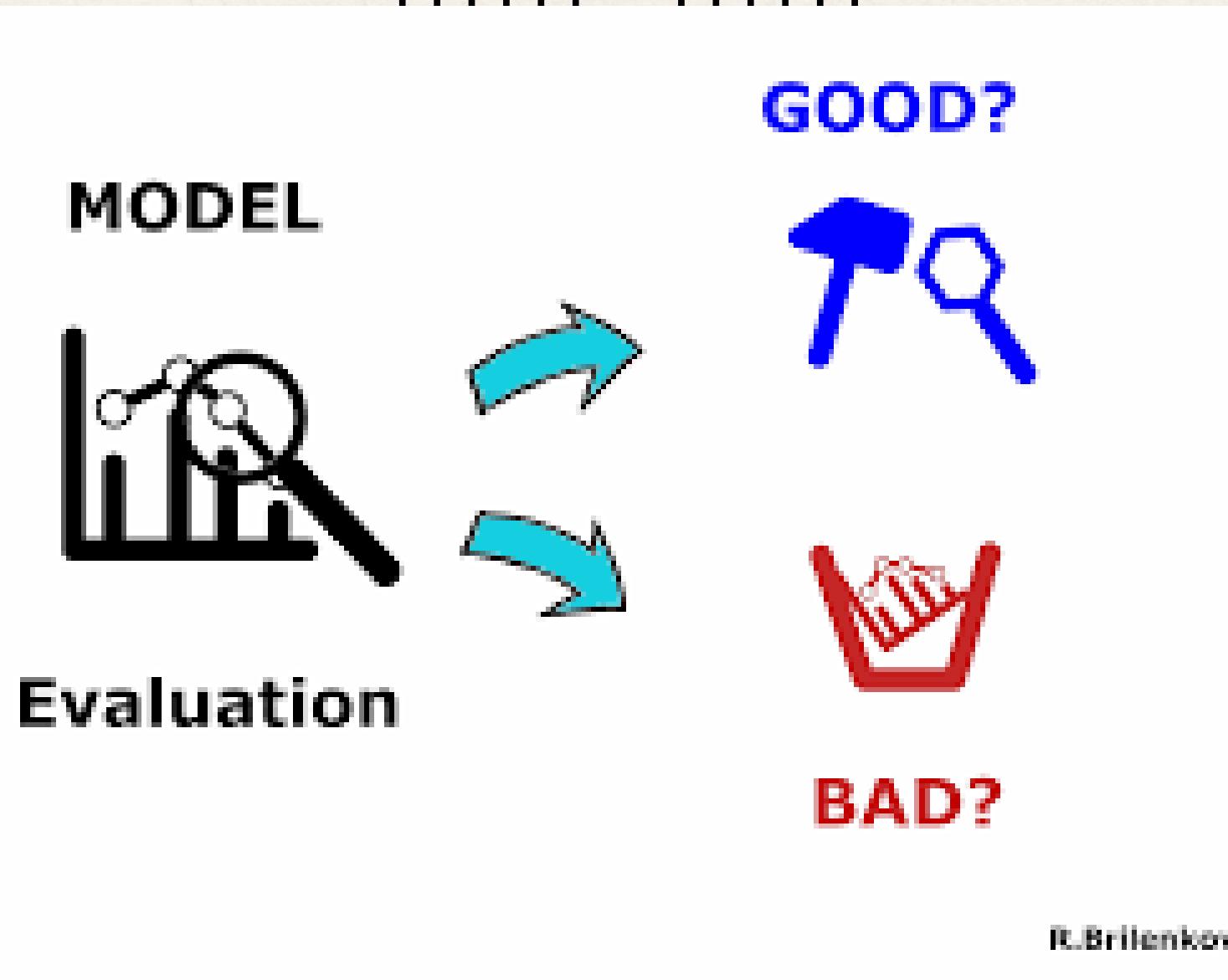
ML Models Used

Models compared:

- Logistic Regression
- Decision Tree
- Random Forest (best)
- Gradient Boosting

Why Random Forest?

- Handles non-linear patterns
- Resistant to overfitting
- Strong performance on tabular HR data



Model Performance

Cross-Validation Results:

- Random Forest: 83.71%
- Gradient Boosting: 83.61%
- Logistic Regression: ~81%

Final Tuned Random Forest (Test Set):

- Accuracy: 83.78%
- Precision: 0.87
- Recall: 0.83
- F1-score: 0.85
- ROC-AUC: 0.93

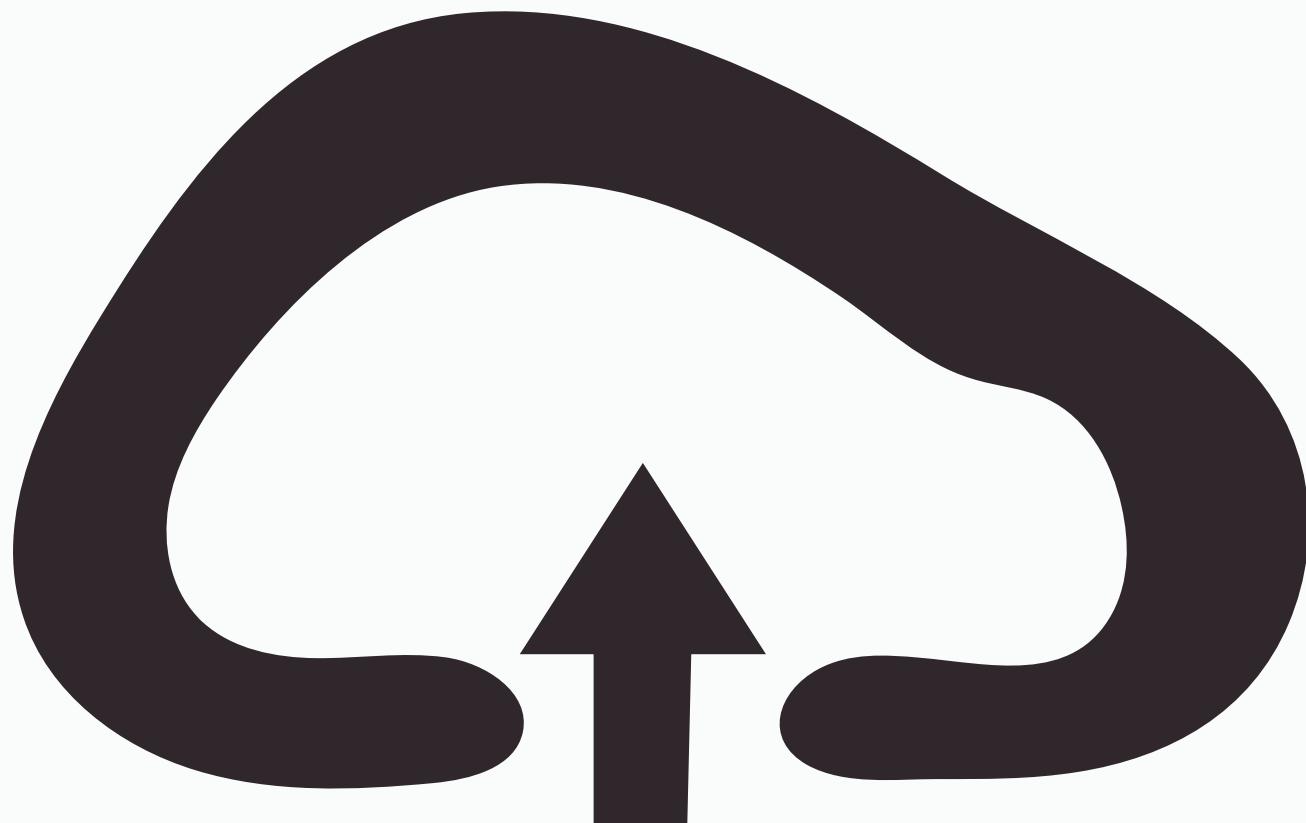
Deployment

Deployment via Streamlit Web App:

- User enters employee details
- App preprocesses data automatically
- Model predicts attrition: Yes / No
- Shows probability & important factors

Model stored as
attrition_model_pipeline.pkl

Link: <https://attrition-app-premium-7vwldhkdyyqwsuyfbb4xfa.streamlit.app/>





Conclusion & Future Scope

Conclusion:

- Random Forest is best-performing model
- Helps HR identify employees at risk
- Can improve retention & reduce hiring cost

Future Scope:

- Use LightGBM, CatBoost
- Add time-based behavior features
- Deploy to cloud (AWS/GCP)
- Add SHAP explainability

THANK YOU