PREDICTING EMPLOYEE TURNOVER

Erin Keough

22 October 2025

Prepared for MFS Investment Management interview for the Data Science Lead Analyst (People Analytics) position. Example project is intended for interview pusposes only.

CONTENTS

INTRODUCTION

At MFS Investment Management, understanding the drivers of employee retention and attrition is critical to sustaining long-term success. By analyzing workforce trends and uncovering data-backed insights, we can identify opportunities to strengthen engagement, improve outcomes, and maintain our competitive edge in the industry.

PROBLEM DEFINITION

Business Challenge	High employee turnover increases costs and reduces organizational performance
Analytical Goal	Predict which employees are likely to leave the organization
Approach	Use HR, demographic, and satisfaction data from the IBM HR dataset to train and compare ML models
Key Questions	What factors most influence attrition? Which model best predicts turnover risk?
Success Metric	Model performance evaluated by accuracy, recall, F1 score, and ROC AUC, prioritizing recall for early detection of at-risk employees

ASSUMPTIONS

Data Assumptions

- Representative sample of employees
- Sampling bias and completeness
- Temporal stability and stationarity

Modeling Assumptions

- Class balance
- Feature Independence
- Independent and Identically Distributed (IID) samples

Business Assumptions

- Actionable predictions
- Causality direction
- Ethical and privacy considerations

MODELING STRATEGY

Model Comparison:

- XGBoost, Random Forest, Gradient Boosted Tree, Logistic Regression, and Support Vector Machine

Approach:

- Train a classification model to predict attrition risk
- **Input features**: Tenure, age, job role, satisfaction scores, overtime status, etc.
- **Target**: Attrition (Yes/No)

Validation process:

- Train/test split, hyperparameter tuning, cross-validation
- Metrics: Accuracy, Precision, Recall, F1 score, ROC-AUC

Strategy Overview:

1. Load & examine IBM HR dataset
2. Exploratory Data Analysis (EDA)
3. Preprocessing & feature engineering
4. Model training & evaluation
5. Feature importance for explainability
6. Metrics comparison & model selection
7. Distill findings into Tableau dashboard

FINDINGS

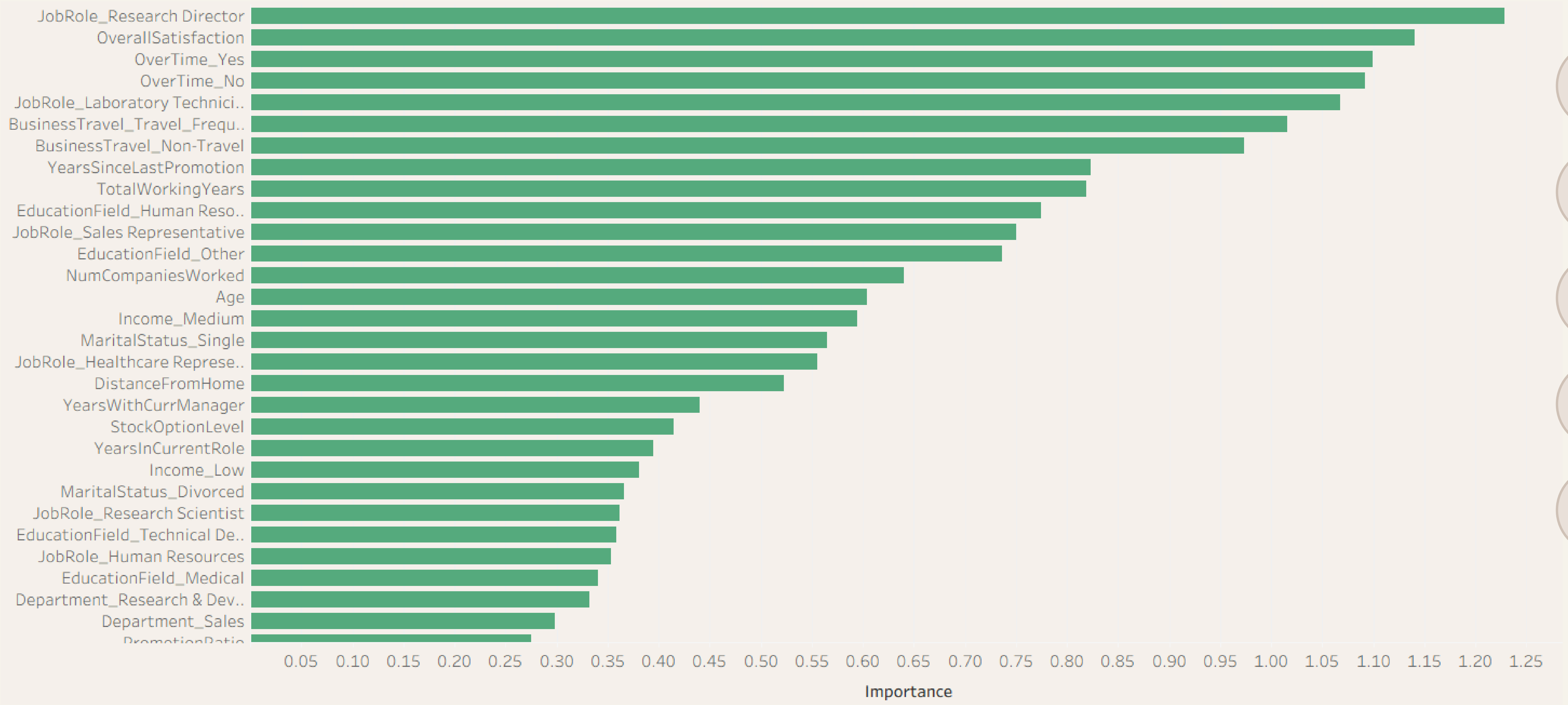
Model performance:

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Logistic Regression	0.871	0.680	0.362	0.472	0.819	0.589
Support Vector Machine	0.871	0.800	0.255	0.387	0.831	0.566
Random Forest	0.847	0.583	0.149	0.237	0.801	0.468
Gradient Boosting	0.837	0.481	0.277	0.351	0.763	0.459
XGBoost	0.857	0.609	0.298	0.400	0.793	0.489

Logistic Regression selected for recall, F1 score, and overall interpretability

FINDINGS

Feature Importance:



1. Job Role
2. Overall Satisfaction
3. Overtime
4. Business Travel
5. Years Since Last Promotion

BUSINESS IMPLICATIONS

- Logistic regression model is most effective at identifying employees who are at risk of leaving
  - **Recall:** Missing an employee who is likely to leave (a false negative) is costlier than incorrectly flagging someone who stays (a false positive)
  - **F1 Score:** Balance between correctly identifying attrition cases & minimizing false alarms
- Across all models, top features contributing to attrition prediction include job role, overtime, overall satisfaction, business travel, and years since last promotion
  - These features can be used to help HR decision-makers catch as many at-risk employees as possible

RECOMMENDATIONS

Recommendations

- Operationalize Logistic Regression model & implement quarterly attrition monitoring
- Build targeted retention programs for employees flagged as “high risk”
- Enhance overall employee satisfaction:
  - Prioritize career development opportunities to reduce attrition
  - Job satisfaction pulse checks with manager

Deployment Strategy

- **Pipeline**: Ingest data from Workday → preprocessing/ETL pipeline → prediction service
- **Integration**: Tableau dashboards showing real-time attrition risk
- **Monitoring**: Track model drift, fairness metrics, and refresh model quarterly
- **Ethics**: Ensure compliance with labor laws, stress fairness & transparency, and avoid bias in promotion/hiring decisions

RESOURCES

1. IBM Dataset - Kaggle:
    - https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset
2. GitHub Repository: IBM dataset, Jupyter notebook, modeling artifacts, Tableau dashboard, presentation:
    - https://github.com/erinkeough/MFS-Interview-2025
3. Public Tableau Dashboard Link:
    - https://public.tableau.com/views/MFS-EmployeeTurnover-Dashboard/MainPage?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

THANK YOU

Erin Keough

erinkeough@ymail.com