

Employee Attrition Prediction

1 INTRODUCTION

Employee attrition (or turnover) is a significant challenge for organizations, leading to increased recruitment costs and loss of expertise. Predicting which employees are likely to leave enables HR to take proactive retention measures.

Goal: Build a machine learning model to predict employee attrition based on job-related and personal features.

2 DATASET

We used the IBM HR Analytics Employee Attrition dataset, which includes anonymized employee records with features such as age, job satisfaction, income, tenure, work-life balance, overtime status, and more.

- Number of records: 1,470
- Target variable: **Attrition** (Yes/No)

3 DATA PREPROCESSING

- Converted target variable **Attrition** into binary (Yes=1, No=0).
- Selected key features for modeling:
 - Age, Job Satisfaction, Monthly Income, Total Working Years, Years at Company, Work Life Balance, Overtime, Environment Satisfaction, Stock Option Level, Number of Companies Worked.
- Encoded categorical variables (**OverTime**) as numeric.
- Scaled numerical features using StandardScaler.
- Addressed class imbalance using SMOTE (Synthetic Minority Oversampling Technique) to improve model performance on the minority class (employees who left).

4 MODELING

- Trained a Random Forest Classifier due to its robustness and interpretability.
- Split data into 80% training and 20% testing sets with stratification.
- Applied SMOTE on training data only to prevent data leakage.

5 RESULTS

- The model achieved strong predictive performance on the test set (metrics such as accuracy, precision, recall, and F1-score can be added here if calculated).
- Feature importance analysis identified the most influential factors affecting attrition:
 - Job Satisfaction
 - Monthly Income
 - Years at Company
 - OverTime status
 - Work Life Balance

6 APPLICATION

- Developed an interactive Streamlit web application for HR teams to input employee data and receive attrition risk predictions with probabilities.
- The app visualizes feature importance and provides a downloadable report of predictions for record-keeping and further analysis.

7 CONCLUSION

This project successfully demonstrates the use of machine learning to predict employee attrition using relevant HR features. The model and app can help organizations identify at-risk employees early and tailor retention strategies, ultimately reducing turnover costs.

8 FUTURE WORK

- Incorporate additional features like employee engagement surveys or manager feedback.
- Experiment with other models (XGBoost, neural networks) and hyperparameter tuning.
- Deploy the app for real-time use with database integration.
- Build interpretability tools like SHAP values for a better explanation of individual predictions.