

IBM HR ANALYTICS: ATTRITION & PERFORMANCE

AN END-TO-END DATA SCIENCE PROJECT





OBJECTIVE

- Analyze and predict employee attrition at IBM to assist HR in retaining top talent.



DATASET DESCRIPTION

- Source: IBM HR Analytics Employee Attrition
- Records: 1470 employees
- Features: Demographics, Job-related, Performance
- Target: Attrition (Yes/No)

EXPLORATORY DATA ANALYSIS (EDA)

- Younger employees showed higher attrition
- Sales and HR departments had higher attrition
- Singles were more likely to leave
- Low job and environment satisfaction correlated with attrition



VISUALIZATIONS

- KDE Plot: Age vs Attrition
- Bar Plot: Gender-wise Attrition Rate
- Department-wise and JobSatisfaction plots



DATA PREPROCESSING

- Label Encoding & OneHotEncoding
- Feature scaling using StandardScaler
- No missing values



MODEL BUILDING

- Train-Test Split: 70/30
- Algorithms: Logistic Regression, Random Forest, Decision Tree, KNN
- Best model: Random Forest

BEST MODEL: RANDOM FOREST CLASSIFIER

Metric	Class 0 (No Attrition)	Class 1 (Attrition)
Precision	0.88	0.80
Recall	1.00	0.10
F1-score	0.93	0.18
Support	255	39

FEATURE IMPORTANCE

- OverTime
- MonthlyIncome
- TotalWorkingYears
- Age
- JobLevel

CONCLUSION

- Overtime & Low income are major attrition indicators
- Younger & less experienced employees are more at risk
- Predictive modeling aids proactive HR planning



PROJECT STRUCTURE

- README.md
- HR_Attrition_EDA.ipynb
- Attrition_Prediction_Model.ipynb
- data/HR-Employee-Attrition.csv
- images/feature_importance.jpg

FUTURE WORK

- Add external surveys & performance history
- Explore deep learning/stacked models
- Create Streamlit dashboard for HR

THANKYOU !

VISMAYA VT