

Employee Attrition Analysis

IDSC Project Presentation

By:

S Harshini (CS23B1050)

R K Larika (CS23B1028)

Shwetha Ganesh (CS23B1028)



Introduction / Problem Statement

Analyzing factors influencing employee attrition in the organization

Understanding key drivers of employee turnover

Goal: Develop predictive model to identify attrition risk factors

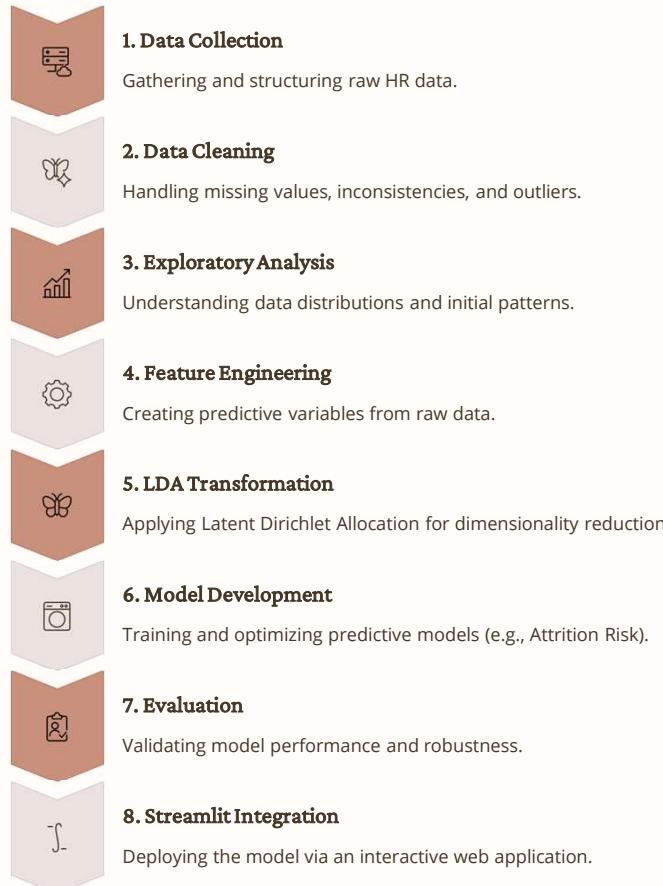
→ Helps organizations improve
retention and reduce
recruitment costs

→ Supports data-driven HR
decision-making

→ Focuses on understanding
patterns across demographics,
job roles, and satisfaction levels

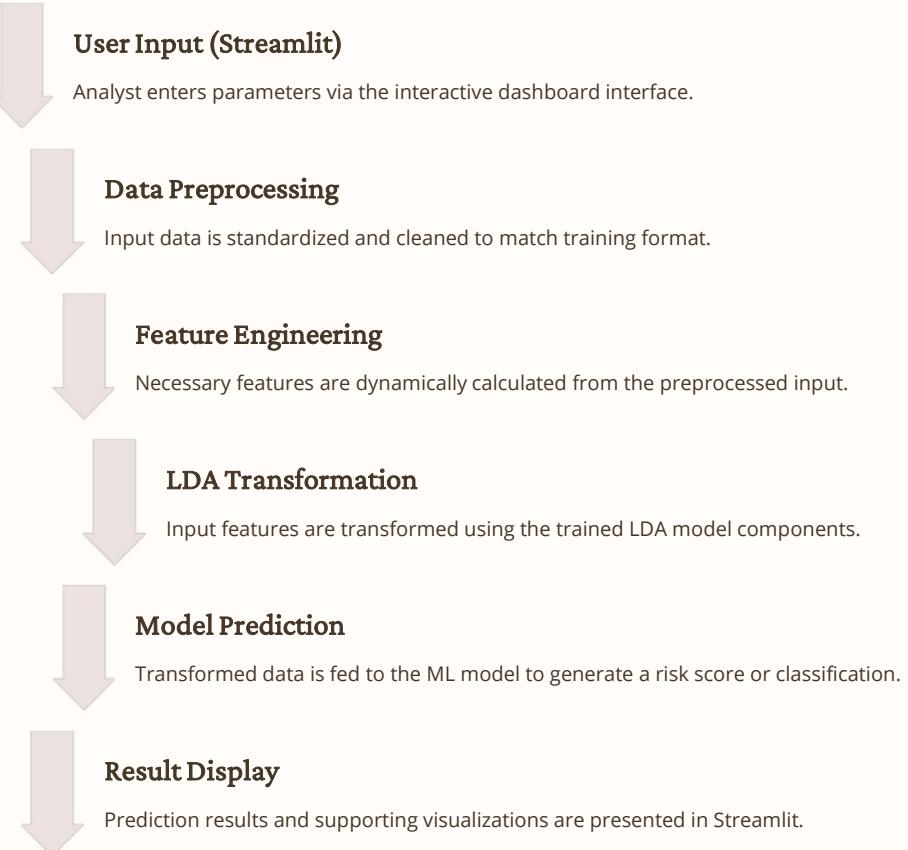
Project Pipeline Overview: From Raw Data to Insight

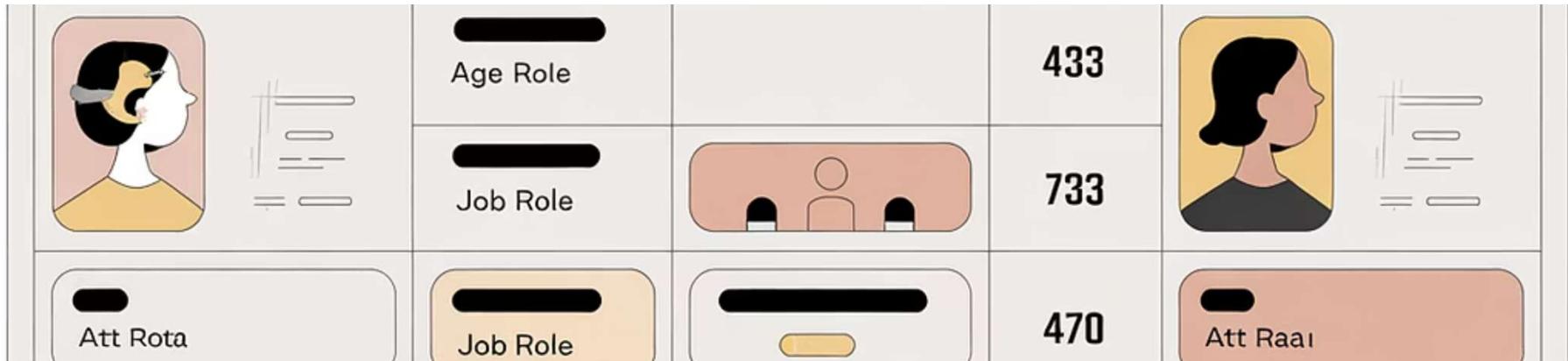
This pipeline outlines the sequential steps for developing and deploying the HR Analytics machine learning solution.



User Interaction Pipeline: Streamlit to Prediction

This details the real-time workflow for an HR analyst using the deployed Streamlit application.





Data Collection

- Dataset with employee information and attrition status
- 1470+ employee records with 35 features
- Mix of numerical and categorical variables
- Data sourced from IBM HR Analytics Employee Attrition & Performance dataset from kaggle

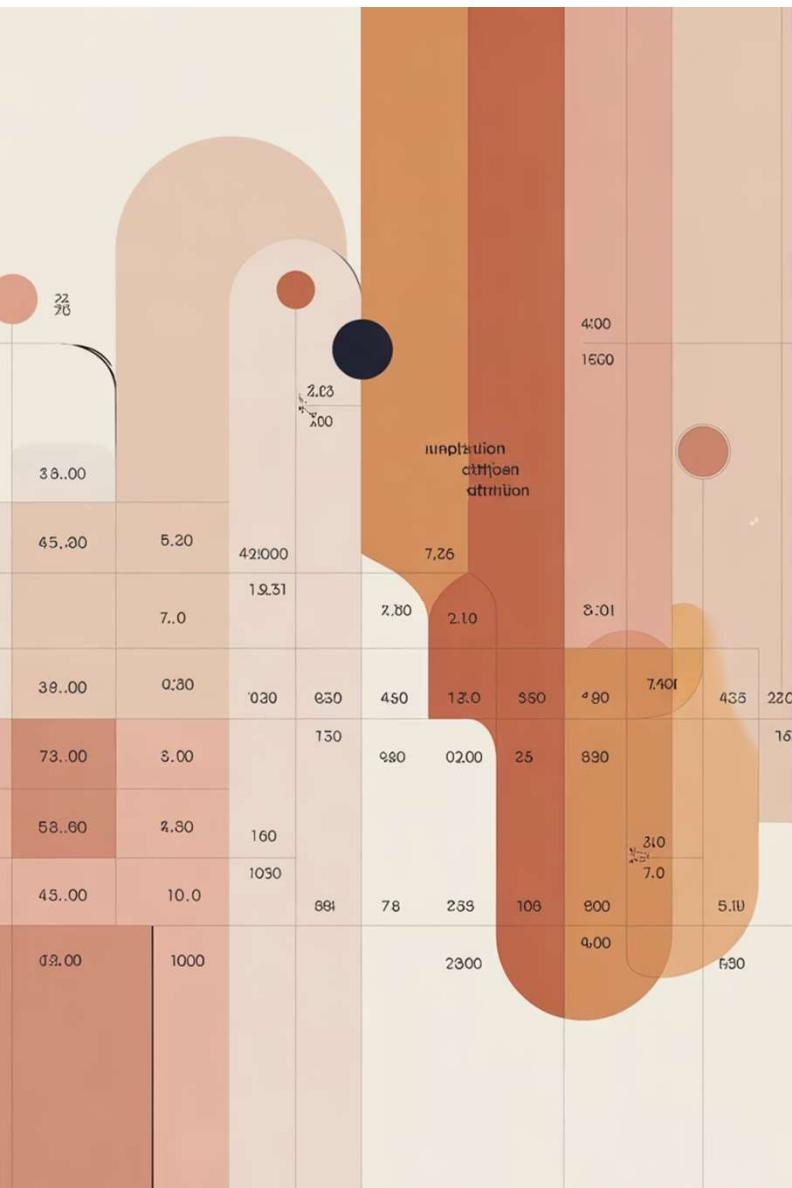
Data Cleaning and Preprocessing

- Checked for missing values (None found)
- Handled constant columns and duplicates
('EmployeeCount', 'Over18', 'StandardHours')
- Treated outliers using IQR method
- Standardized numeric features
- Encoded categorial variables
- Standardization for Numeric – Z scale
- Binning / Discretization
- Feature Engineering

Data Exploration

- Conducted univariate analysis of all features
- Analyzed feature distributions and patterns
- Identified key demographic patterns (Visualizations in Cells 21-22: histograms and categorical plots)





Data Analysis

Bivariate Analysis

Performed bivariate analysis with attrition

Correlation analysis between numeric features

Statistical Testing

Hypothesis testing (t-test, z-test)

(Key visualizations: Correlation heatmap in Cell 27, pairplot in Cell 28)

Data Modeling

Applied Linear Discriminant Analysis (LDA)

Feature engineering and transformation

Trained multiple classification models:

Logistic Regression

Random Forest

SVM

KNN



Model Evaluation and Deployment

Evaluation

- Compared performance of all models
- Metrics: Accuracy, Precision, Recall, F1-Score
- Identified best performing model

Deployment (Streamlit)

- Deployed the trained model using a Streamlit web app for interactive prediction
- Takes top 15 features from Ida as input from the user
- Goes through the same pipeline that we used for the test data.
- Gives the prediction

Strategic Roadmap: Future Work and Expansion



Real-Time Prediction

Implement continuous, event-driven attrition prediction for immediate insights.

Broader Metric Inclusion

Integrate additional HR metrics like performance reviews and engagement scores.

Recommendation Systems

Add a recommendation system to suggest retention for at-risk employees

Interactive HR Dashboard

Create a comprehensive, dynamic dashboard for deeper operational insights.

Focusing on **real-time capabilities** and **functional specialization** to maximize HR impact.

Thank You

Any Questions?

