

Employee Attrition Prediction

CCDS-223 (Data Mining)

November 08, 2025

Group2 :

Abdulelah Abdullah Nasser 2340466

Abdullah Atiah Alzahrani 2340088

Mohannad Abdullah Alamri 2340047

Tariq Zaid Alqhtani 2340458

Instructor: Abdullah Alghoson

Distribution of Duties:

Abdullah Alzahrani: Defined the business problem, described the project background, and explained the importance and relevance of the issue.

Mohannad Alamri: Formulated the project objectives and developed testable hypotheses based on the identified business problem.

Tariq Alqhtani: Designed and structured the main research and analytical questions that the project aims to answer.

Abdulelah Nasser: Identified and summarized the dataset, including its source and main features, and coordinated the final submission.

Problem Statement:

Employee turnover is a major challenge facing organizations worldwide, including in Saudi Arabia. High employee turnover leads to increased recruitment costs, decreased productivity, and a negative impact on overall company morale. By using data mining and predictive analytics techniques, organizations can identify the key factors driving employee departures and the reasons behind them, enabling them to take proactive steps to retain employees. This project aims to analyze employee data to predict the likelihood of employee turnover using machine learning models such as decision trees and K-means analysis. The project aligns with the goals of Saudi Vision 2030, which aims to improve company efficiency and sustainability, enhance employee satisfaction, and increase productivity.

Objectives

The primary objectives of this project are as follows:

- Identify key factors that influence employee attrition.
- Build a predictive model to classify employees as 'likely to stay' or 'likely to leave'.
- Perform Exploratory Data Analysis (EDA) to uncover trends and correlations.
- Compare different machine learning models (Decision Tree, Logistic Regression) to select the best one.
- Provide actionable insights for HR departments to improve retention strategies.

Hypotheses

- **H1:** Job satisfaction has a negative correlation with employee attrition.
- **H2:** Employees with lower monthly income are more likely to leave.
- **H3:** Longer distance from home increases the likelihood of attrition.
- **H4:** Employees who frequently work overtime have a higher chance of leaving.
- **H5:** Lack of promotions over long periods increases attrition probability.

Project Questions

- What are the main factors influencing employee attrition?
- Can we accurately predict which employees are likely to leave the organization?
- How do salary, work-life balance, and satisfaction affect attrition rates?
- What data-driven HR policies could reduce employee turnover?

Data Source

We selected the IBM HR Analytics Employee Attrition Dataset from Kaggle because it provides comprehensive HR data suitable for data-mining analysis with both numerical and categorical features, ideal for classification and clustering tasks.

- **Demographic Information:** Age, Gender, MaritalStatus, EducationField.

Work-related Features: Department, JobRole, OverTime, YearsAtCompany, TotalWorkingYears.

- **Performance & Satisfaction:** JobSatisfaction, WorkLifeBalance, EnvironmentSatisfaction, PerformanceRating

- **Financial Metrics:**

MonthlyIncome, NumCompaniesWorked, YearsSinceLastPromotion.

- **Target Variable:** Attrition (Yes/No).

In the next phase (Milestone 2), we will perform data cleaning and keep 20 relevant variables that best represent employee behavior and attrition factors.

The Dataset Link: <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>