

Minor Project Proposal

(Individual mode)

Title: Employee Attrition Prediction

Student Name: Pawar Mayur Sharad

Roll No.: AA.SC.U3BCA2307009

Email Id: aa.sc.u3bca2307009@ahead.students.amrita.edu

Address of the Company: Not Applicable (Academic Individual Project)

Name, Designation and Communication details of the Guide: To be filled by the department

Abstract: This project focuses on predicting employee attrition — whether an employee is likely to leave the company — based on various HR-related factors such as age, job satisfaction, salary, years at company, and work-life balance. Using the IBM HR Analytics Employee Attrition Dataset, the project applies data preprocessing, feature encoding, and logistic regression techniques to identify the most influential factors contributing to attrition. The expected outcome is a trained machine learning model that can help HR departments make data-driven decisions to reduce employee turnover and improve retention strategies.

Assumptions / Declarations: - The dataset used is assumed to be accurate, complete, and publicly available for educational purposes. - It is assumed that the dataset reflects real-world employee behaviour trends.

Main Objective / Deliverable: To build a predictive machine learning model that determines the likelihood of employee attrition based on HR data. Deliverables include: - Cleaned and pre-processed dataset - Logistic Regression model for attrition prediction.

Timeline and Milestones:

Milestone	Timeline
Topic selection and proposal submission	Oct 17, 2025
Data collection and preprocessing	Nov 2025
Model building and training	Dec 2025
Final report writing	Jan 5, 2026
Final review, presentation, and viva	Jan–Mar 2026

Tools to be Used:

Category	Tools / Libraries
Programming Language	Python
Libraries	Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
IDE / Environment	Jupyter Notebook / Google Colab
Dataset IBM HR	Analytics Employee Attrition Dataset (Kaggle)

Learning Involved:

Topic	Description
Data Preprocessing	Handling missing values, encoding categorical data, feature selection
Machine Learning (Classification)	Applying Logistic Regression for binary prediction
Evaluation Metrics	Using accuracy, confusion matrix, precision, recall
Visualization	Understanding and plotting feature importance for HR insights
Project Documentation	Writing structured reports and presenting results effectively

Date: 17/10/2025

Student Name and Signature: Pawar Mayur Sharad