Minor Project (21CSA697A)

Proposal (Individual Mode)

Title: Employee Attrition Prediction

Student Name: Amruta Kumbar

Roll No.: AA.SC.P2MCA24074031

Address of the Company: Not Applicable

Name, Designation and Communication details of the Guide: NA

Abstract

This project builds an explainable employee attrition prediction system using advanced classification algorithms and ensemble learning techniques. The system aims to identify employees at risk of leaving the organization by analyzing HR data and deriving actionable insights. Logistic Regression, Random Forest, XGBoost, and Gradient Boosting models are employed to improve predictive accuracy. The model integrates SHAP/LIME analysis to explain feature importance, providing HR managers with interpretable insights. A Streamlit-based dashboard is developed for interactive visualization and decision-making support.

Assumptions/Declarations:

 - The dataset used (IBM HR Analytics Dataset) is assumed to be clean and representative of HR attrition patterns.
- All model evaluations will be based on historical data.
- The project will focus on technical model development and explainability, not live HR integration.
- Declaration: This is original work developed solely for academic purposes under institutional guidance.

Main Objective/Deliverable:

> To develop an explainable machine learning model that predicts employee attrition and provides interpretable insights into key factors influencing employee turnover.
>
> Deliverables:
> - Predictive model trained on IBM HR Analytics dataset
> - Explainability analysis using SHAP/LIME
> - Interactive HR dashboard using Streamlit or Flask
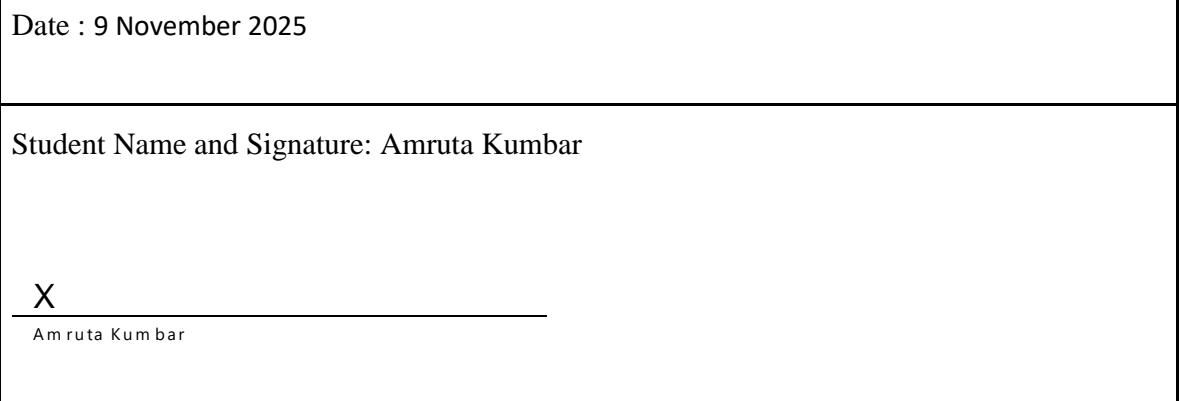
Timeline and Milestones:

	Milestones	Timeline
1.	Literature Review and Data Collection:	Week 1–2
2.	Data Preprocessing and Feature Engineering:	Week 3–4
3.	Model Development (Logistic Regression Random Forest, XGBoost):	Week 5–6
4.	Model Evaluation and Explainability Analysis (SHAP/LIME):	Week 7
5.	Dashboard Development using Streamlit:	Week 8
6.	Final Report and Presentation:	Week 9–10

Tools to be used.

Software/Hardware Tools	Specifications
• Python: • Scikit-learn, XGBoost, LightGBM: • Pandas, NumPy, Matplotlib, Seaborn: • SHAP / LIME: • Streamlit / Flask: • Jupyter Notebook:	➤ Programming Language ➤ Machine Learning Libraries ➤ Data Analysis and Visualization ➤ Explainability Tools ➤ Interactive Dashboard Deployment ➤ Development Environment

Learning involved:

Topic	Description
- Data Preprocessing - Machine Learning Algorithms - Model Evaluation - Explainable AI - Dashboard Development	- Handling missing data, encoding categorical variables, and feature scaling. - Logistic Regression, Decision Tree, Random Forest, Gradient Boosting. - Accuracy, Precision, Recall, ROC-AUC metrics. - Applying SHAP and LIME to interpret model outputs. - Deploying interactive visualization using Streamlit or Flask.

Date : 9 November 2025

Student Name and Signature: Amruta Kumbar

X _____

Amruta Kumbar