Employee Performance Prediction using Machine Learning

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Problem Statement:
An organization faces challenges in predicting employee performance accurately due to reliance on manual evaluations, leading to inefficiencies and potential bias in HR decisions. The project aims to build a machine learning model to predict employee productivity based on historical performance and work-related data.

Activity 2: Project Proposal (Proposed Solution)

The proposed project, **"Employee Performance Prediction using ML,"** aims to leverage machine learning techniques to predict employee productivity. Using a dataset containing features such as department, targeted productivity, overtime, idle time, and incentives, the project seeks to create a predictive model to aid HR teams in improving evaluation efficiency, workforce planning, and overall productivity management.

Activity 3: Initial Project Planning

Initial Project Planning involves setting objectives, defining scope, identifying stakeholders, and outlining a workflow for data processing, model training, and deployment. This ensures proper resource allocation and a systematic approach to achieving accurate performance predictions.

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves gathering employee productivity data (e.g., from Kaggle HR datasets). It includes data cleaning, encoding categorical variables, handling missing values, and preparing the data for machine learning model development.

- **Activity 1: Data Collection Plan & Raw Data Sources:**
  Dataset sourced from Kaggle HR analytics repository with features like quarter, department, overtime, idle time, etc.

- **Activity 2: Data Quality Report:**
  Ensuring clean, validated data by addressing missing values, standardizing formats, and checking for outliers.
- **Activity 3: Data Exploration & Preprocessing:**
  EDA performed to identify trends, correlations, and outliers, followed by encoding and scaling for modeling.

Milestone 3: Model Development Phase

This phase involves building machine learning models for performance prediction using selected features. Models include **Linear Regression, Decision Tree, Random Forest, and XGBoost**, with validation through appropriate metrics like R^2, RMSE, and MAE.

- **Activity 1: Feature Selection Report:**
  Includes selected features like quarter, department, targeted productivity, overtime, incentive, idle time, etc.
- **Activity 2: Model Selection Report:**
  Compares multiple models based on accuracy and efficiency, finalizing **XGBoost** as the best-performing model.
- **Activity 3: Model Training and Validation:**
  Splitting the dataset (80:20), training models, and validating results to ensure reliability and accuracy.

Milestone 4: Model Optimization and Tuning Phase

Hyperparameter tuning was performed on **XGBoost**, improving accuracy from baseline metrics.

- Documented hyperparameter changes (learning rate, n_estimators, max_depth).
- Compared performance metrics before and after optimization.
- Justified XGBoost as the final model due to its superior predictive power.

Milestone 5: Project Files Submission and Documentation

All project files (datasets, code, model files, documentation) were uploaded to GitHub.

Milestone 6: Project Demonstration

The final phase includes a recorded video demonstration explaining the dataset, preprocessing, model training, evaluation metrics, web deployment using Flask, and prediction examples.
