Machine Learning Approach for Employee Performance Prediction

Project Guide Summary

Milestone 1: Project Initialization and Planning Phase

Activity 1: Define Problem Statement

A production manager in a garments factory needs to predict employee productivity based on various operational and HR metrics such as department, team size, SMV, idle time, and overtime. Accurate prediction will enable better resource planning, productivity improvement, and targeted support to low-performing teams.

Document: [Click here](#)

Activity 2: Project Proposal (Proposed Solution)

The project aims to develop a machine learning model to predict actual employee productivity. The dataset is sourced from Kaggle and includes 14 relevant features. The proposed solution uses data preprocessing, feature engineering, and model comparison to build the best predictor. The model is deployed using Gradio for ease of use.

Document: [Click here](#)

Activity 3: Initial Project Planning

Project planning involved understanding the dataset, setting model objectives (predicting actual productivity), selecting relevant tools like Python, Pandas, and XGBoost, and organizing the workflow across multiple deliverables such as preprocessing, modeling, tuning, and deployment.

Document: [Click here](#)

Milestone 2: Data Collection and Preprocessing Phase

Activity 1: Raw Data Source and Collection

Dataset: Garment Workers Productivity Dataset from Kaggle
It includes fields like department, team number, targeted productivity, SMV, idle time, etc.

Documents: [Click here](#)

Activity 2: Data Quality and Cleaning

Steps included checking for missing values, dropping irrelevant features, handling categorical data (quarter, department, day), and encoding them properly.

Documents: [Click here](#)

Activity 3: Preprocessing and Transformation

Converted date to month, encoded categorical values, and split data into training and testing sets using train_test_split. Handled missing values using SimpleImputer.

Document: [Click here](#)

Milestone 3: Model Development Phase

Activity 1: Feature Selection

All 14 features were used, but importance was analysed using correlation heatmaps and model-based selection.

Document: [Click here](#)

Activity 2: Model Selection

Three models were tested:

- Linear Regression
- Random Forest Regressor
- XGBoost Regressor

XGBoost gave the best results.

Document: [Click here](#)

Activity 3: Model Evaluation

Used MAE, MSE, R^2 score, and visualized Actual vs Predicted scatter plots. F1-score was calculated by converting regression output into classification (above/below target).

Document: [Click here](#)

Milestone 4: Model Optimization and Tuning

Activity 1: Hyperparameter Tuning

Hyperparameters of Random Forest and XGBoost were tuned using grid search and manual testing.

Activity 2: Metrics Comparison

Model	R²	F1-Score
Linear Regression	0.72	0.74
Random Forest	0.79	0.81
XGBoost	**0.83**	**0.86**

Activity 3: Final Model Justification

XGBoost was selected for its better generalization, performance, and training time.

Document: [Click here](#)

Milestone 5: Project Files Submission

All project assets are uploaded to GitHub:

- Notebook (Employee_Performance_Prediction.`ipynb`)
- Saved model (`gwp.pkl`)
- Output screenshots
- 10 PDF documentation files
- Demo video

GitHub Link: https://github.com/Tanmayy-k/Employee-Performance-ML-Prediction.git

Milestone 6: Project Demonstration

A 1-minute screen recording has been prepared showing:

- Colab code execution
- Final model results
- Gradio web UI in action
- Prediction output and suggestions

Demo Video Link: https://drive.google.com/file/d/1V7wVZK4b-kQ1d3HX8K32z5QIcaBaR0kf/view?usp=sharing