Model Development Phase Template

Date	10 July 2025
Team ID	NONE
Project Title	Employee Performance Prediction using Machine Learning

Model Selection Report

The Model Selection Report presents a comparative analysis of three regression models developed to predict employee productivity: Linear Regression, Random Forest, and XGBoost. The primary evaluation metric used was the R^2 (coefficient of determination) score, which measures how well the model explains the variance in the target variable (actual_productivity)

Model	**Description**	**Hyperparameters**	**Performance Metric** R^2
Linear Regression	served as the baseline model but underperformed ($R^2 \approx 0.168$) due to its assumption of linear relationships, which do not fully capture the complexity of workforce behavior.	- No specific hyperparameters (default settings) -	0.168168
XGBoost	showed improved performance ($R^2 \approx 0.3538$) by leveraging gradient boosting, but exhibited signs of overfitting and required more computational resources	Defualt(best model)	0.44671974
Random Forest	achieved the highest R^2 score of ~0.46, demonstrating strong generalization and robustness to non-linear patterns and feature interactions (e.g., over_time, idle_time, department).	Defualt	0.353859