Employee Performance Predictor

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: A company seeks to predict employee performance based on various factors such as experience, education, training hours, and department. The challenge lies in accurately identifying high-performing individuals to guide promotions and training investments.

Problem Statement Report: Click Here

Activity 2: Project Proposal (Proposed Solution)

Develop a machine learning model that predicts employee performance categories (e.g., low, medium, high) using structured HR data. This will enhance decision-making in talent management and improve organizational efficiency.

Project Proposal Report: Click Here

Activity 3: Initial Project Planning

- Define project scope and objectives.
- Identify stakeholders (HR team, data scientists, management).
- Allocate resources and set timelines.
- Understand the dataset and plan preprocessing and modeling workflows.

Project Planning Report: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant workers productivity

data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

Source employee performance data from Kaggle or internal HR databases. Ensure data includes relevant features like department, education, training hours, and performance scores.

Data Collection Report: Click Here

Activity 2: Data Quality Report

The dataset undergoes rigorous quality checks to identify and resolve missing or inconsistent entries. Techniques such as imputation and validation against expected distributions are applied to maintain data integrity. Ethical handling of employee data is prioritized to ensure responsible usage.

Data Quality Report: Click Here

Activity 3: Data Exploration and Preprocessing

Data exploration involves analyzing the employee dataset to uncover patterns, distributions, and outliers. Preprocessing includes:

- Handling missing values
- Descriptive Analysis
- Encoding categorical variables (e.g., department, day)
- Correlation Analysis

Data Exploration and Preprocessing Report: Click Here

Milestone 3: Model Development Phase

This phase focuses on building and validating predictive models to classify employee performance levels.

Activity 1: Feature Selection Report

TThe Feature Selection Report outlines the rationale behind choosing specific features such as SMV, department, incentive, and number of workers. These features are evaluated for relevance and impact on predictive accuracy, ensuring the model captures key indicators of employee performance.

Feature Selection Report: Click Here

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing algorithms like Random Forest, Decision Tree, Linear Regression, and XGBoost. Each model is assessed for its ability to handle categorical and numerical data, interpretability, and predictive performance. The selection aligns with the project's goal of accurate and actionable performance prediction.

Model Selection Report: Click Here

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

Initial model training involves applying selected algorithms to the processed dataset. The Model Validation and Evaluation Report assesses performance using metrics such as MAE, MSE and R2 score. This ensures the model's reliability in predicting employee performance and supports informed HR decision-making.

Model Development Report : Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Documentation

The XGBBoost model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the XGBBoost model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification outlines the rationale for choosing XGBoost Regressor. Its ability to capture nonlinear relationships, robustness to multicollinearity, and superior performance across multiple regression metrics make it the optimal choice for predicting employee performance scores.

Model Optimization and Tuning Phase Report: Click Here

Milestone 5: Project Files Submission and Documentation

For the documentation, Kindly refer to the link. Click Here

Milestone 6: Project Demonstration

individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.