Employee Productivity Analysis

The HR team at SkillPro seeks to enhance workforce productivity by identifying and addressing key factors influencing employee performance. By predicting employee productivity scores using a comprehensive dataset, they aim to design targeted interventions such as tailored training programs and strategic role assignments.

Objective

Predict employee productivity scores based on multiple factors such as experience, training hours, work-life balance, team dynamics, and work environment.

Tasks

1. Data Preparation:

- **Exploratory Data Analysis (EDA):**
  - Perform descriptive analytics and visualizations to uncover data patterns, trends, and anomalies.
  - Identify and visualize potential outliers for key features.
- **Preprocessing:**
  - Handle missing values, if any, using appropriate imputation techniques.
  - Normalize or standardize continuous features like experience and training hours.
  - Encode categorical variables such as team assignment or job role using one-hot encoding or label encoding.
  - Address class imbalance in categorical data if applicable (e.g., using SMOTE).

2. Feature Engineering:

- Create interaction terms (e.g., experience × training hours) to capture nonlinear relationships.
- Generate new features such as average productivity within the team or department.
- Evaluate feature importance to identify the most influential predictors.

3. Model Building:

- Split the dataset into training (70%), validation (15%), and testing (15%) sets.
- Train a Multilayer Perceptron (MLP) regression model with:
  - An optimized architecture using hyperparameter tuning for the number of layers, neurons, activation functions, and dropout rates.
  - Early stopping to prevent overfitting.

4. Model Evaluation:

- Evaluate the model using the following metrics:
    - Mean Squared Error (MSE).
    - Mean Absolute Error (MAE).
- Compare the MLP model with simpler baseline models (e.g., linear regression, decision trees).

5. Advanced Insights and Recommendations:

- Assess how factors like work-life balance or team dynamics influence productivity.
- Use SHAP (SHapley Additive exPlanations) to interpret feature importance and model predictions.
- Identify employees or groups at risk of low productivity and recommend interventions:
    - Additional training for specific skill gaps.
    - Changes in work environment or team structure.

Data Description

The dataset `employee_productivity.csv` ([DataSet Link](#)) includes the following columns:

- **Experience (years):** Years of work experience.
- **Training Hours:** Number of training hours attended annually.
- **Work-Life Balance (1-5):** A score indicating the balance between work and personal life.
- **Team Dynamics Score (1-10):** A score reflecting collaboration within the team.
- **Environment Score (1-10):** A score reflecting the work environment quality.
- **Role:** Employee's job role (categorical).
- **Team ID:** Identifier for the team to which the employee belongs.
- **Productivity Score:** The target variable, indicating employee productivity (numerical).

Practicalities

1. Visualizations:

- Correlation heatmaps to identify relationships between numerical features.
- Boxplots and histograms to detect and analyze outliers.
- Scatterplots for key relationships, such as experience vs. productivity.

2. Additional Considerations:

- Discuss the limitations of the model (e.g., generalizability or data availability).
- Recommend further data collection or studies to refine the analysis.
- Provide an actionable report summarizing findings and next steps for improving productivity.