Model Development Phase

Date	24 June 2025
Team ID	SWUID20250177148
Project Title	Machine Learning Approach for Employee Performance Prediction
Maximum Marks	4 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

- **Linear Regression**

```
# --- Model Building ---

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import matplotlib.pyplot as plt
import pandas as pd

model_lr = LinearRegression()
model_lr.fit(X_train, y_train)
pred_test = model_lr.predict(X_test)
```

- **Random Forest Regressor**

```
[ ]  from sklearn.ensemble import RandomForestRegressor
     from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
     import matplotlib.pyplot as plt

     # 1. Initialize the model
     model_rf = RandomForestRegressor(n_estimators=100, random_state=42)

     # 2. Train the model
     model_rf.fit(X_train, y_train)

     # 3. Make predictions
     pred_rf = model_rf.predict(X_test)
```

- **XG Boost Regressor** (final selected model)

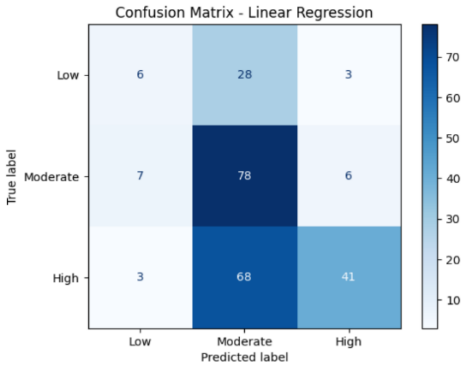
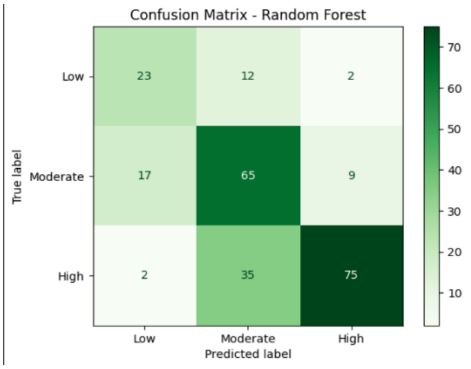
```
     import xgboost as xgb
     from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
     import matplotlib.pyplot as plt

     # 1. Initialize the XGBoost Regressor
     model_xgb = xgb.XGBRegressor(objective='reg:squarederror', n_estimators=100, random_state=42)

     # 2. Train the model
     model_xgb.fit(X_train, y_train)

     # 3. Predict on test data
     pred3 = model_xgb.predict(X_test)
```

Model Validation and Evaluation Report:

Model	Classification Report	Accuracy	Confusion Matrix
Linear Regression	Classification Report: precision recall f1-score support Low 0.38 0.16 0.23 37 Moderate 0.45 0.86 0.59 91 High 0.82 0.37 0.51 112 accuracy 0.52 240 macro avg 0.55 0.46 0.44 240 weighted avg 0.61 0.52 0.49 240	74%	
Random Forest	Classification Report (Random Forest): precision recall f1-score support Low 0.55 0.62 0.58 37 Moderate 0.58 0.71 0.64 91 High 0.87 0.67 0.76 112 accuracy 0.68 240 macro avg 0.67 0.67 0.66 240 weighted avg 0.71 0.68 0.69 240	81%	
XG Boost	Classification Report (XGBoost): precision recall f1-score support Low 0.60 0.65 0.62 37 Moderate 0.63 0.77 0.69 91 High 0.89 0.71 0.79 112 accuracy 0.72 240 macro avg 0.71 0.71 0.70 240 weighted avg 0.75 0.72 0.73 240	86%