EMPLOYEE PERFORMANCE PREDICTION :USING MACHINE LEARNING

PROGRESS REPORT

PRESENTED BY: B.GAYATHRI

COURSE: BUSINESS ANALYTICS | DATE;24/4/2025

WHY I CHOSE THIS PROBLEM

Employee performance is one of the most critical factors determining an organization's success. In today’s competitive business environment, organizations are constantly looking for ways to retain top talent, improve productivity, and make data-driven HR decisions. However, traditional performance evaluations are often

Subjective

Time-consuming

With my background in law and business, This problem combines HR strategy with advanced technologies like machine learning, offering a meaningful solution to improve workforce planning and talent management. It can also support equity and transparency in performance evaluations, which is a major challenge in modern workplaces.

c

Author Name	Year of Publication	Ti tle	Method ology	Accuracy	Research Gaps
Pablo Robles-Granda et al.	2023	Jointly Predicting Job Performance, Personality, Cognitive Ability, Affect, and Well-Being	Behavioraltracking, mobile and wearable data, psychological surveys	Not explicitly mentioned	Limited to U.S.-based knowledge workers; may not generalize globally
MD Rokibul Hasan et al.	2022	Employee Performance Prediction: An Integrated Approach of Business Analytics and Machine Learning	Business analytics, machine learning (feature engineering, model tuning)	High (exact % not stated)	Focuses on structured data; ignores unstructured data like feedback or commu ni cati on
Dr. Kishan Singh & Ayesha Khuteja	2021	Employee Performance Prediction Using Machine Learning Algorithms	CRISP-DM framework, Random Forest classification	87%	Limited algorithm diversity; does not explore ensemble or deep learning
–(Conference Paper)	2022	Application of Machine Learning in HRM: Employee Performance Prediction Model	ML models (Linear Regression, SVM, Decision Tree, Random Forest)	Random Forest performed best	Based on a single company’s data; lacks model interpretability and transparency

FORNT-END PLANS

Dashboard: Overview of employee performance predictions.

Employee Profile: Detailed view with performance trends.

Upload Feature: Upload employee data (CSV/XLSX).

Search & Filter: Find employees by role, department, or performance level.

Reports: Option to download individual performance reports (PDF).

jupyter Untitled17 Last Checkpoint: 52 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[5]:

```
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
import gradio as gr

# Load dataset
df = pd.read_csv("train_dataset.csv")

# Handle missing values
df.dropna(inplace=True)

# Separate features and target
X = df.drop("actual_productivity", axis=1)
y = df["actual_productivity"]

# Train/test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Model training
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Evaluation
mse = mean_squared_error(y_test, model.predict(X_test))

# Prediction function
def predict_productivity(*features):
    input_df = pd.DataFrame([features], columns=X.columns)
    prediction = model.predict(input_df)[0]
    return round(prediction, 4)

# Create Gradio inputs dynamically
inputs = []
for col in X.columns:
    if X[col].nunique() < 10:
```

Conclusion **- this project demonstrates the use of mashine learning to predict emplyee productivity based on historical data using a random forest regressor. This project highlights how companies can tranform raw employee data into actionable insights that improve productivity ,efficiency and decision making.**