***PROJECT REPORT ON EMPLOYEE WORK HOUR ABSENTEEISM***

This project has been built upon an enormous HR data set that focuses on employee absence. It contains a staggering 8335 rows and 13 columns of data.

The data set contains employee numbers and names, gender, city, job title, department, store location, business unit, division, age, length of service, and the number of hours absent.

**AIM:**

To create ML models to predict the ‘No. of hours absent’ with minimum MSE and RMSE and maximum R-Square score-

1. Multiple Linear Regressor (MLR)
2. Random Forest Regressor (RFR)
3. MLR with PCA
4. RFR with PCA

and an inference on the models that has been prepared.

**INFERENCES:**

**Multiple Linear Regressor (MLR)**

MSE-675.8427527850749

RMSE-25.996975839221665

R2-SCORE-0.702225519180218

**Random Forest Regressor (RFR)**

MSE- 657.8174158683769

RMSE- 25.647951494580944

R2-SCORE- 0.7101674336564072

**Multiple Linear Regressor (MLR) with PCA**

MSE-673.0690071102016

RMSE-25.94357352236198

R2-SCORE- 0.7034476239891515

**Random Forest Regressor (RFR) with PCA**

MSE-666.8069756283787

RMSE-25.822605903130277

R2-SCORE-0.7062066580480251

Comparing the **Multiple Linear Regressor (MLR)** model results with and without PCA:

*Without PCA:*

MSE-675.8427527850749

RMSE-25.996975839221665

R2-SCORE-0.702225519180218

*With PCA:*

MSE-673.0690071102016

RMSE-25.94357352236198

R2-SCORE- 0.7034476239891515

Both models, with and without PCA, seem to perform fairly similarly.

The model with PCA has slightly better metrics with a slightly lower MSE and higher R2 score, indicating better predictive performance and a better fit to the data.

However, the difference in performance between the two approaches is marginal.

The model with PCA slightly outperformed the model without PCA based on the provided metrics.

Comparing the **Random Forest Regression (RFR)** model results with and without PCA:

*Without PCA:*

MSE: 657.8174158683769

RMSE: 25.647951494580944

R-squared (R2): 0.7101674336564072

*With PCA:*

MSE: 666.8069756283787

RMSE: 25.822605903130277

R-squared (R2): 0.7062066580480251

Both models, with and without PCA, seem to perform fairly similarly.

The model without PCA has slightly better metrics with a slightly lower MSE and higher R2 score, indicating better predictive performance and a better fit to the data.

However, the difference in performance between the two approaches is marginal.

The model without PCA slightly outperformed the model with PCA based on the provided metrics.