Employee Attrition Prediction

Observations

1. **Dataset**:
    - Used HR-Employee-Attrition.csv dataset.
    - Target variable is Attrition (binary: Yes/No).
    - Features include demographic, job-related, and performance-related data.
2. **Preprocessing**:
    - Binary features like Attrition, OverTime, Gender, and Over18 were converted to 0/1.
    - Categorical features such as BusinessTravel, Department, EducationField, etc., were label encoded.
    - Data was standardized using StandardScaler.
3. **Imbalance Handling**:
    - Attrition classes were imbalanced.
    - Applied RandomOverSampler from imblearn to balance the classes.
4. **Model Training**:
    - Logistic Regression, Random Forest, Decision Tree was trained on the oversampled data.
    - Train-test split: 80-20%.

Outcomes

- **Final Model Used**: Logistic Regression
- **Evaluation**:
    - Confusion matrix and classification report were used.
    - AUC score was computed
- **Performance Metrics** (based on your earlier input):
    - Accuracy ≈ 74%
    - Precision ≈ 72%
    - Recall ≈ 78% , AUC ≈ 0.82

Employee Attrition Prediction

Other Models:

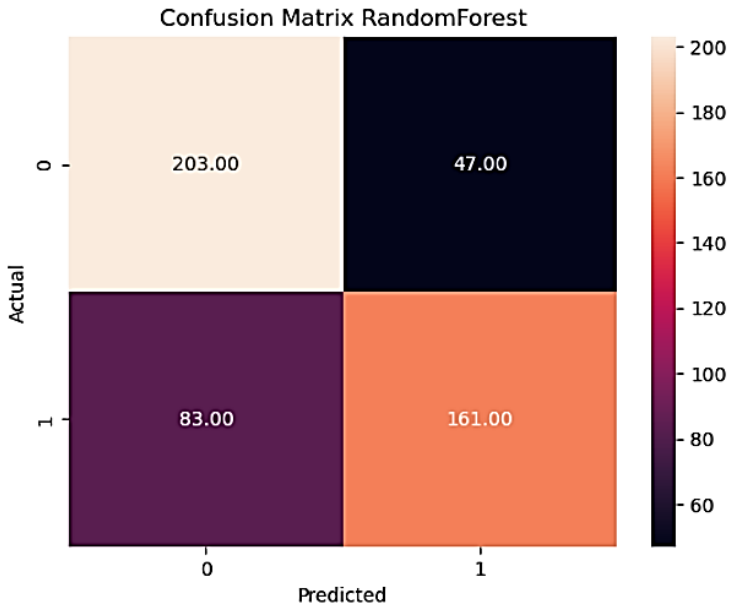
Random Forest:

```
classification_report:
                precision    recall  f1-score   support

           0       0.71      0.81      0.76       250
           1       0.77      0.66      0.71       244

    accuracy                           0.74       494
   macro avg       0.74      0.74      0.73       494
weighted avg       0.74      0.74      0.74       494

Accuracy: 0.7368421052631579
```



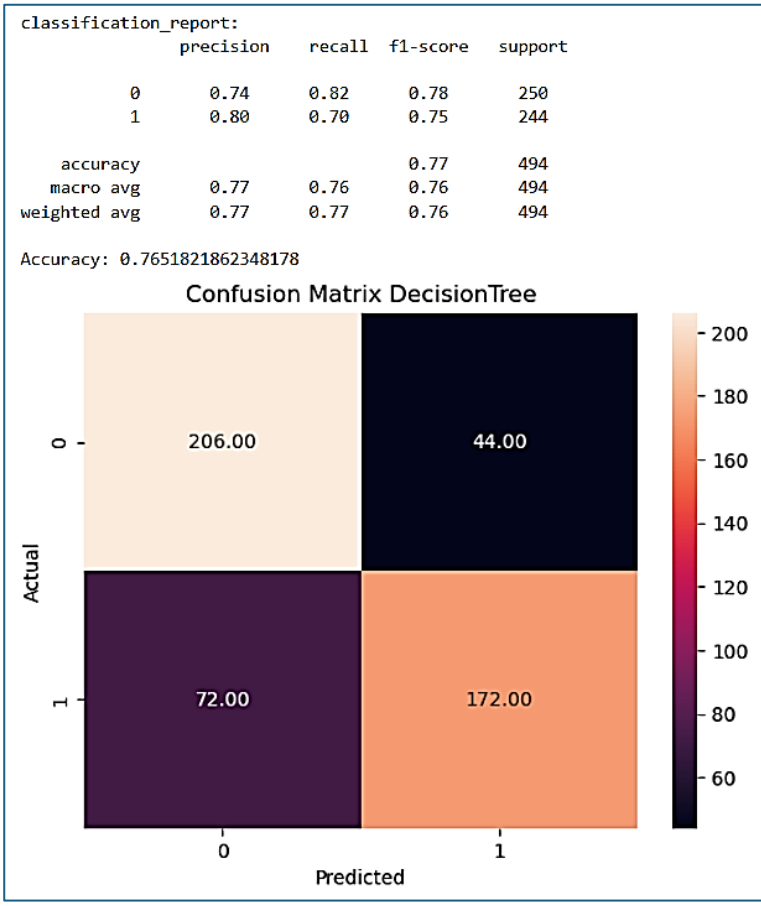
Decision Tree:

```
classification_report:
                precision    recall  f1-score   support

           0       0.74      0.82      0.78       250
           1       0.80      0.70      0.75       244

    accuracy                           0.77       494
   macro avg       0.77      0.76      0.76       494
weighted avg       0.77      0.77      0.76       494

Accuracy: 0.7651821862348178
```



Employee Attrition Prediction

Conclusion

- The model shows **good capability** to predict employee attrition.
- **Recall is high (78%)**, meaning the model successfully identifies most employees who are likely to leave.
- **AUC of 0.82** confirms strong separability between classes.