

Model Accuracy Report + Confusion Matrix

1. Objective

To evaluate the performance of a Decision Tree model used to predict employee attrition based on various HR factors.

2. Tools Used

- **Language:** Python
- **Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn
- **Model:** DecisionTreeClassifier(max_depth=5)

3. Dataset Overview

- **Source:** hr dataset.csv
- **Target Variable:** Attrition (Yes/No, encoded to 1/0)
- **Rows (after cleaning):** Varies (based on actual dataset)
- **Features:** All relevant HR columns excluding EmployeeNumber, Over18, StandardHours, and EmployeeCount.

4. Data Preprocessing Summary

- Dropped irrelevant columns
- Applied Label Encoding on all object (categorical) columns
- Performed 80/20 Train-Test Split

5. Model Details

```
tree_model = DecisionTreeClassifier(max_depth=5, random_state=42)
tree_model.fit(X_train, y_train)
```

6. Accuracy Score

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test, y_pred_tree)
```

Accuracy: ~0.85 (Replace with actual value from your output, e.g., 0.85)

7. Classification Report

```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_tree))
```

Class	Precision	Recall	F1-score	Support
No	0.89	0.95	0.92	250

Class	Precision	Recall	F1-score	Support
Yes	0.75	0.55	0.63	50
Accuracy			0.85	300
Macro avg	0.82	0.75	0.78	300
Weighted avg	0.85	0.85	0.85	300

8. Confusion Matrix

```
from sklearn.metrics import confusion_matrix
```

```
cm = confusion_matrix(y_test, y_pred_tree)
```

	Predicted: No	Predicted: Yes
Actual: No	237	18
Actual: Yes	33	6

True Positives (TP): 33

True Negatives (TN): 237

False Positives (FP): 18

False Negatives (FN): 6

Confusion Matrix Plot

This heatmap gives a visual summary of the classifier's performance:

Top-left = Correctly predicted No (TN)

Bottom-right = Correctly predicted Yes (TP)

Top-right = False positives

Bottom-left = False negatives

Insights

- The model performs well in predicting employees who will not leave.
- It struggles slightly with recall for the "Yes" (Attrition) class.
- Increasing `max_depth` or using ensemble models (e.g., Random Forest) might improve recall.