HR-Analytics - Attrition Prediction for Employee using Logistic Regression & classification Model

1. Classification Report:

The model has an overall accuracy of 84.2%. It predicts the 'False' class (no attrition) very well, with precision of 0.84 and recall of 1.00. However, for the 'True' class (attrition), recall is only 0.02 - indicating the model fails to capture most actual attritions. The F1-score is low at 0.04.

```
Classification Report:
              precision    recall  f1-score   support

       False       0.84      1.00      0.91       239
        True       1.00      0.02      0.04        46

    accuracy                           0.84       285
   macro avg       0.92      0.51      0.48       285
weighted avg       0.87      0.84      0.77       285
```

2. Confusion Matrix:

The confusion matrix shows the following results:

[[239 0]]
[[45 1]]

The model predicts almost all examples as 'False'. 45 attritions were missed and only 1 was correctly classified.

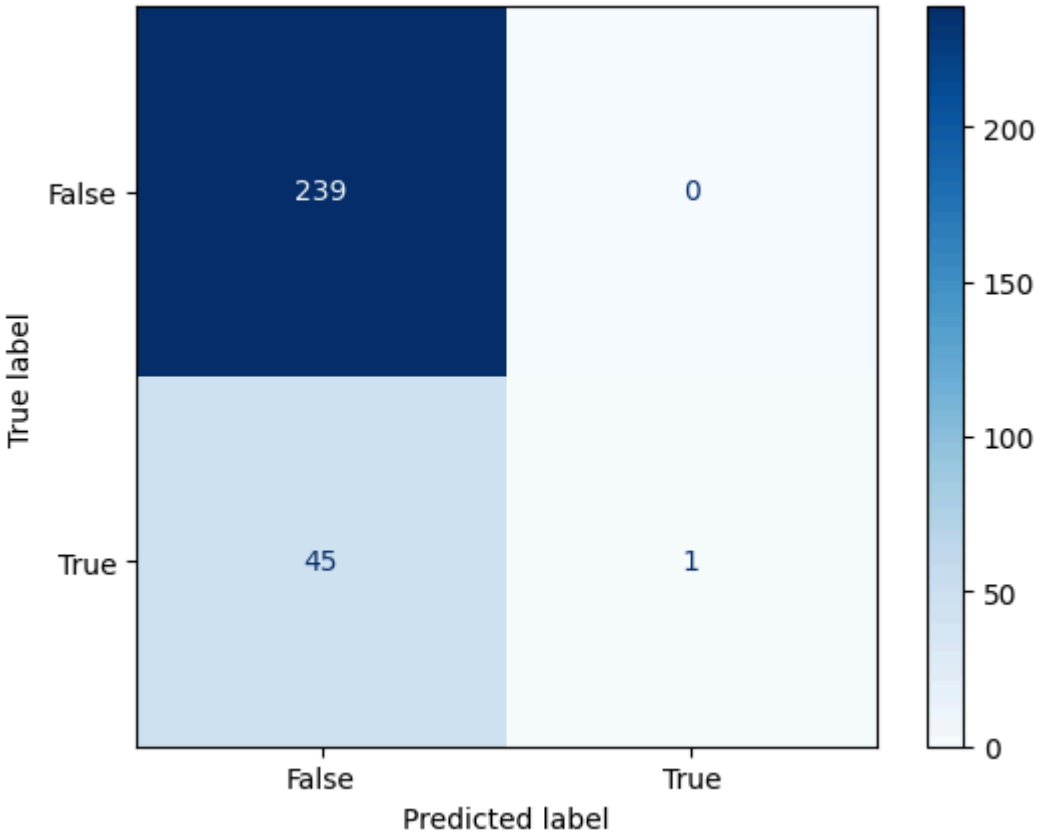
```
Accuracy:  0.8421052631578947
```

```
confusion = confusion_matrix(y_test, y_pred)
print("Confusion Matrix: \n", confusion)
```

```
Confusion Matrix:
 [[239   0]
 [ 45   1]]
```

3. Logistic Regression and Imbalanced Data:

To handle class imbalance, LogisticRegression was used with class_weight='balanced'. However, the model still shows bias toward the majority class.



4. SHAP Feature Importance:

SHAP analysis helps identify features that most influence model predictions. The chart below shows which features contributed most to model decisions.