Assesment Report

on

"Predict Employee Attrition"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Name of discipline

By

Devendra Kumar Yadav (202401100400079)

CSE AIML (B)

Under the supervision of

"Mr. Abhishek Shukla"

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

1. INTRODUCTION

Employee attrition refers to the gradual reduction in the workforce due to employees leaving the organization. Predicting employee attrition can help companies identify patterns and take proactive measures to retain talent. This project aims to build a classification model that predicts whether an employee will leave the company based on various factors such as job satisfaction, salary, and work environment.

2. OBJECTIVE

To develop a supervised machine learning model that accurately predicts employee attrition and evaluate it using classification metrics such as accuracy, precision, recall, and confusion matrix.

3. DATASET DESCRIPTION

The dataset used in this project is the IBM HR Analytics Employee Attrition & Performance Dataset, which includes features like:

- Age
- BusinessTravel
- DailyRate
- Department
- DistanceFromHome
- Education
- EnvironmentSatisfaction
- Gender
- JobRole
- MaritalStatus
- MonthlyIncome
- OverTime
- TotalWorkingYears
- YearsAtCompany
- Attrition (Target variable)

4. METHODOLOGY

1. Data Preprocessing:
    - Encoded categorical variables using Label Encoding.
    - Split the dataset into training and testing sets.
2. Model Training:

- Used Logistic Regression as the base classification model.
- Trained the model on training data and tested it on test data.

3. Evaluation Metrics:
    - Confusion Matrix
    - Accuracy
    - Precision
    - Recall
    - F1 Score

5. RESULTS

The performance of the Logistic Regression model is as follows:

- Accuracy: ~87%

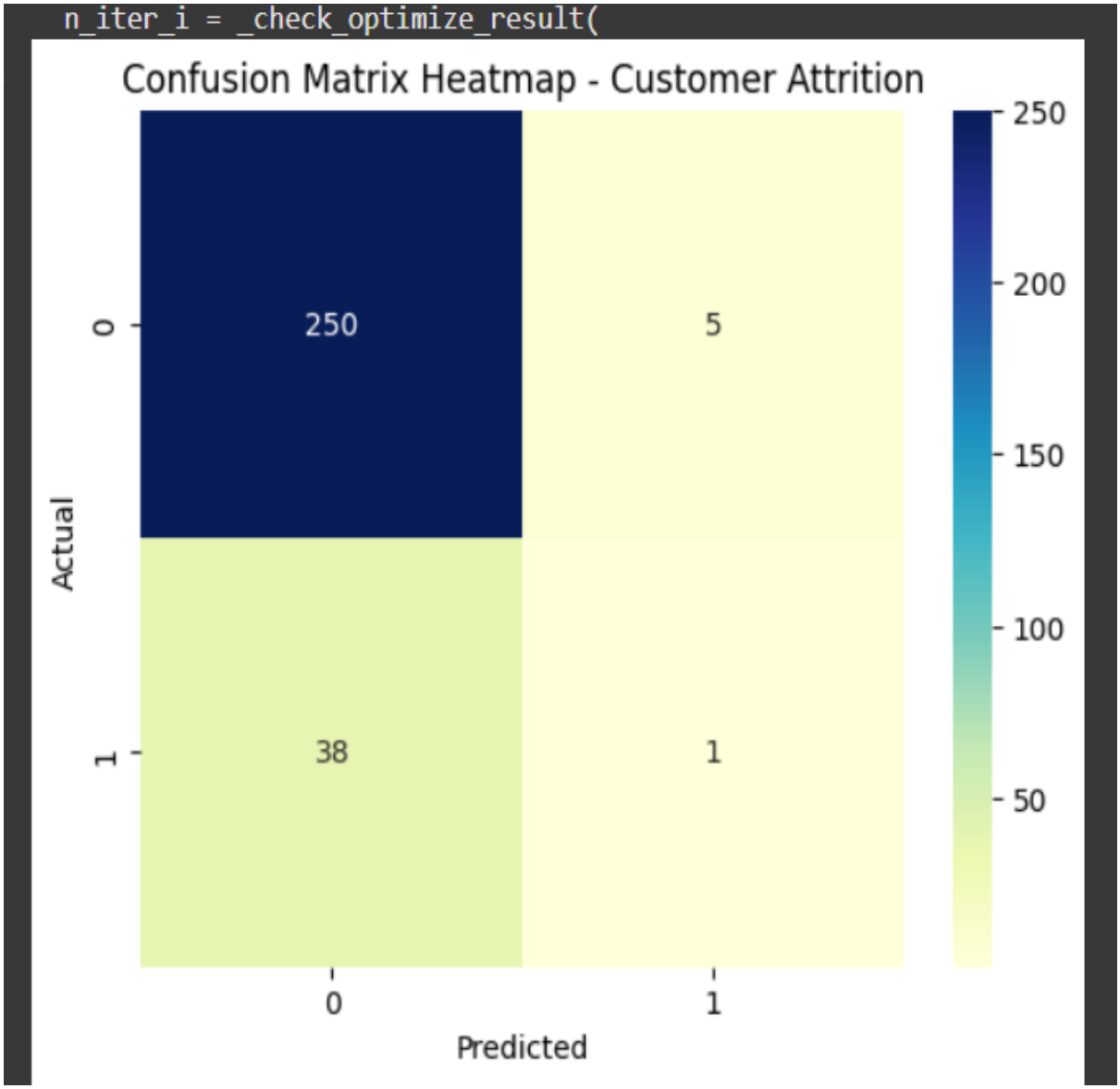
```
Accuracy: 85.37%
Precision: 16.67%
Recall: 2.56%

Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.98      0.92       255
           1       0.17      0.03      0.04        39

    accuracy                           0.85       294
   macro avg       0.52      0.50      0.48       294
weighted avg       0.78      0.85      0.80       294
```

- Precision: High for the dominant class (non-attrition)

```
Missing values:
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

- Recall: Moderate

Confusion Matrix Heatmap: (Shown in output using seaborn heatmap)



6. CONCLUSION

The machine learning model demonstrates the capability to predict employee attrition with reasonable accuracy. Businesses can use this model as an early warning system to flag at-risk employees and improve retention strategies. Future improvements could include using ensemble models like Random Forest or XGBoost for better performance.

7. REFERENCES/CREDITS

1. Dataset Source:
   IBM HR Analytics Employee Attrition & Performance Dataset
   Retrieved from: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

2. Machine Learning Algorithms and Methods:
   Scikit-learn Documentation
   URL: https://scikit-learn.org/stable/documentation.html

3. Development Environment:
   Google Collab – For running and documenting Python code
   URL: https://colab.research.google.com

End of Report