Pokemon Mega Evolution Classification – Mini Machine Learning

Experiment

Internship Submission Project

1. Objective

To apply Machine Learning fundamentals on a small dataset and demonstrate understanding of data

preprocessing, model building, and evaluation. The goal: predict whether a Pokemon has a Mega Evolution

based on its stats.

2. Dataset

Source: Pokemon dataset (Kaggle / Custom)

Attributes Used: Total, HP, Attack, Defense, Sp. Atk, Sp. Def, Speed

Target: Mega_Evolution (1 Pokemon name contains 'Mega', 0 Otherwise)

Dataset Size: 800 entries * 12 columns

3. Methodology

- Preprocessing: Cleaned dataset, created binary target column, dropped irrelevant features.

- Model: Random Forest Classifier (Scikit-learn).

- Evaluation: Accuracy, Confusion Matrix, ROC Curve, Precision-Recall Curve.

- Visualization: Used Matplotlib and Seaborn for interpretation.

4. Results

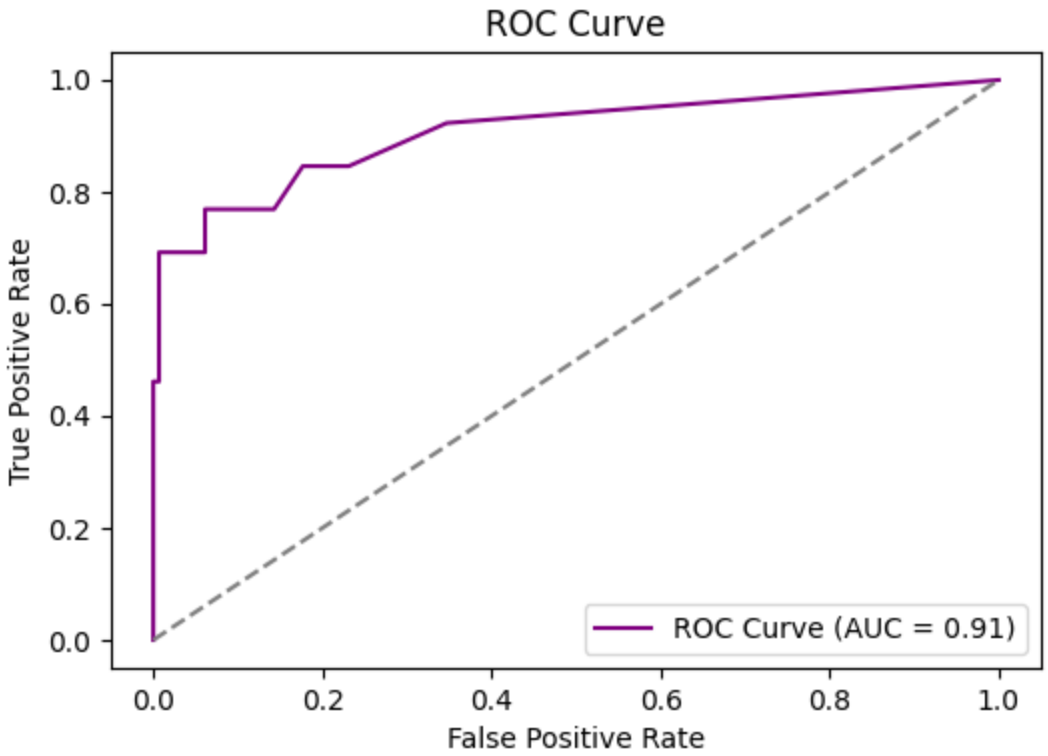
Metric Result Observation

Accuracy approx. 97% Model performed well even with minimal tuning

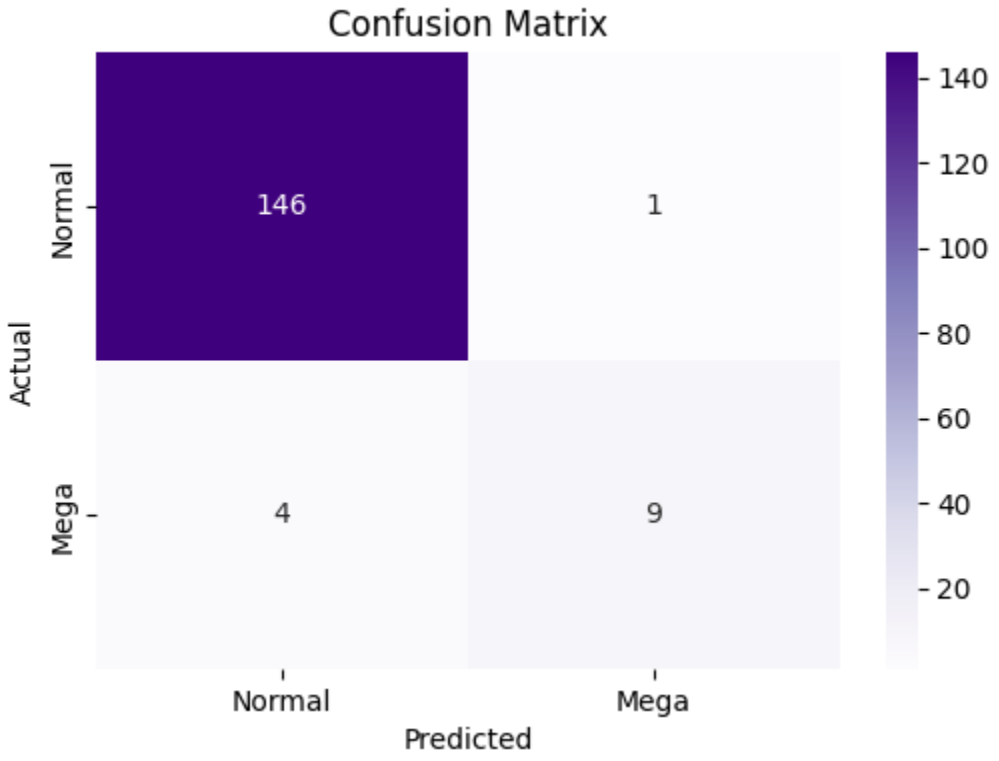
```
    61fa114..37304ab  main -> main
PS E:\PROJECTS\challenge\ML_Classification> venv\Scripts\activate
(venv) PS E:\PROJECTS\challenge\ML_Classification> python main.py

Model Accuracy: 0.97
```

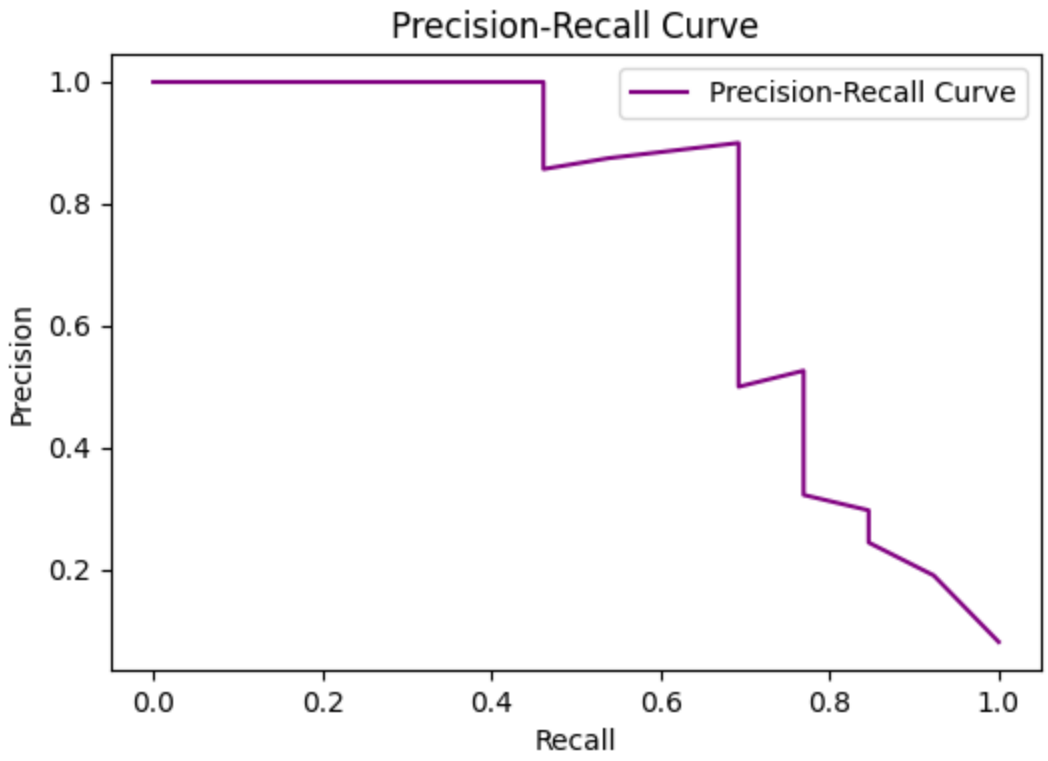
ROC AUC = 0.91 Indicates strong discrimination ability.



Confusion Matrix Visualized Balanced prediction



PR Curve Visualization



Key Insight: Class imbalance (few Mega Pokemon) slightly affected recall but didn't significantly degrade

overall performance.

5. Reflection

- What worked: Random Forest gave excellent results without heavy tuning. Visualizations made

performance more intuitive.

- What didn't: Dataset imbalance and limited feature scope restricted the model.

- What I learned: End-to-end ML workflow, importance of visual analysis, and modular reproducible code

practices.

6. Tools Used

Python

Pandas

Scikit-learn

Matplotlib

Seaborn

venv

git / github

7. Conclusion

A simple yet complete ML workflow was implemented — from data cleaning to visualization and reflection.

This project strengthened my understanding of ML pipelines and made me confident in applying these

concepts in real-world projects.