Zoo Animal Classification using Machine Learning

Presented by
A. Laxmi Prasanna
S. Sameeksha

Introduction

- Machine Learning can classify animals using biological traits
- ***Dataset:*** Zoo dataset with 101 animals, 17 features
- ***Goal:*** Predict the class type (Mammal, Bird, Reptile, Fish, Amphibian, Insect, Others)
- ***Importance:***

✓ Educational use in biology & AI

✓ Demonstrates ML on small categorical datasets

✓ Useful in zoology and data science training

Problem Statement

- ***Input:*** 17 binary/numeric features (hair, feathers, eggs, milk, aquatic, backbone, venomous, etc.)
- ***Output:*** Animal class (1–7)
- ***Challenges:***

➢ Very small dataset (101 rows)

➢ Class imbalance (some classes only 1 animal)

➢ Similar species (e.g., Alligator vs Salamander) confuse the model

Dataset Overview

Files Used:

- zoo.csv → animal features
- class.csv → mapping of class numbers → class labels

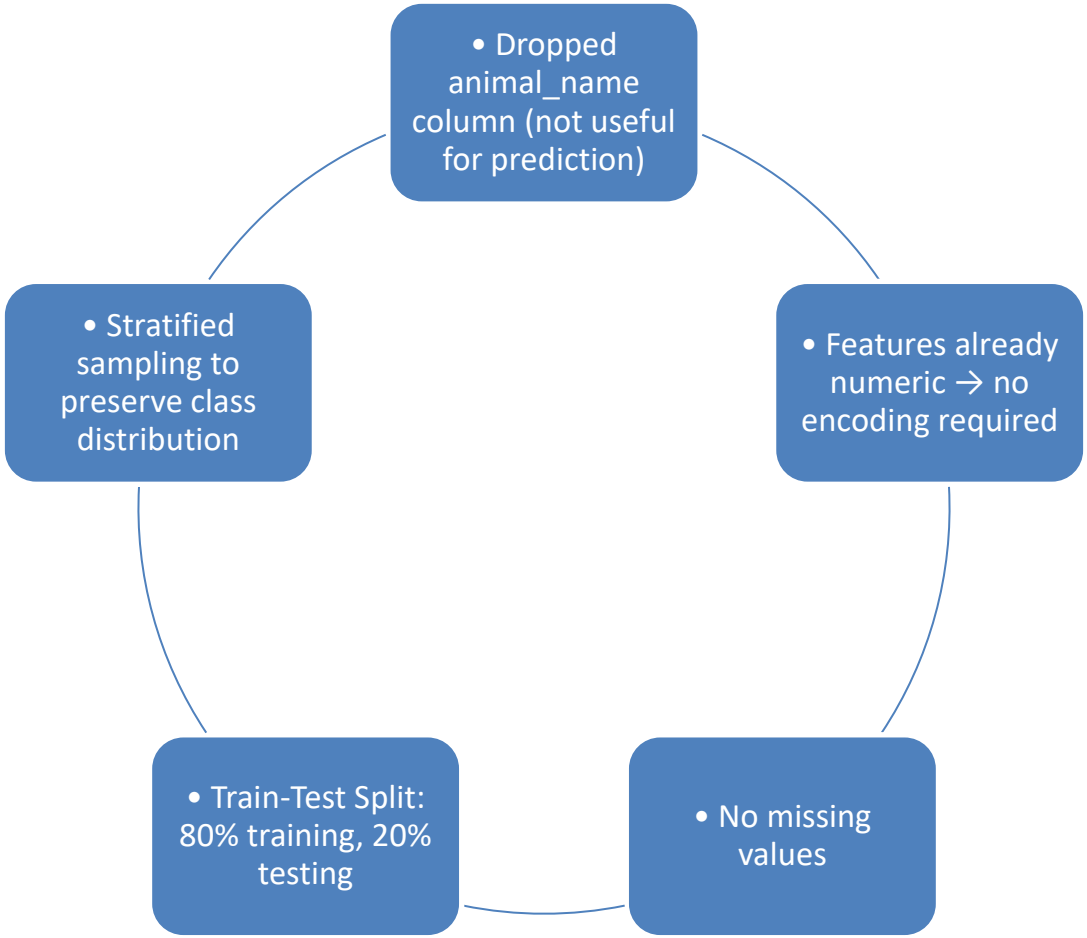
Structure:

- Features: 17 biological traits
- Target: class_type (1–7)

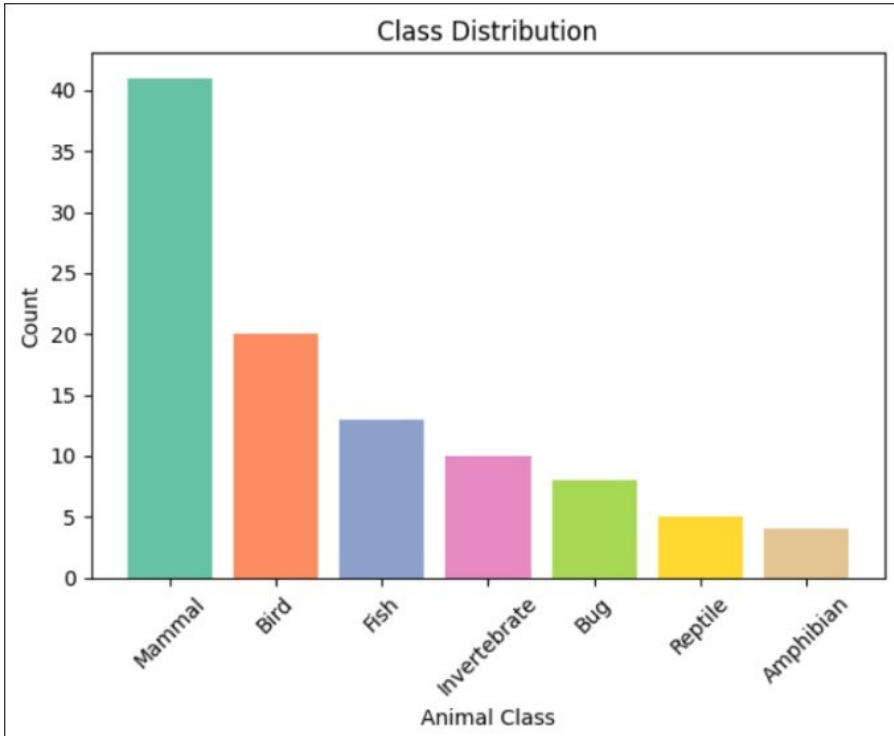
Class Distribution:

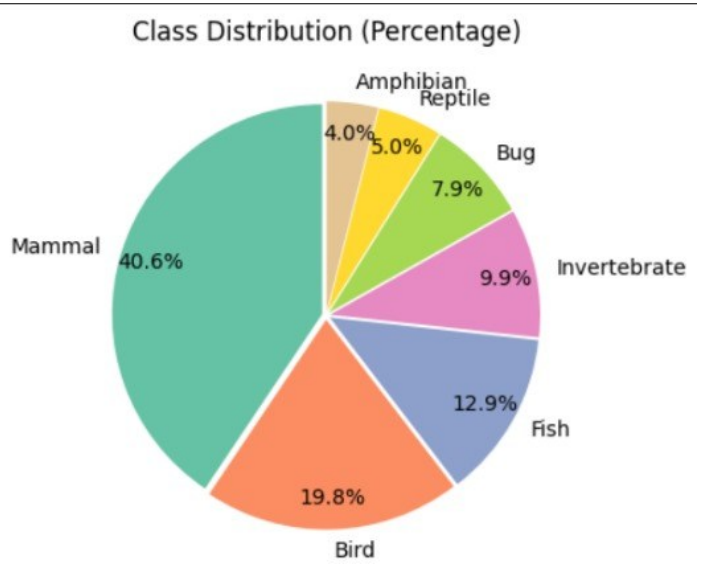
- Mammals dominate
- Some classes very rare
- challenge for ML

Data Preprocessing



Visual Exploration





Model Selection

Algorithms considered: Decision Tree, Logistic Regression

Chosen Model: XGBoost because:

Handles categorical-like binary/numeric features

Robust against overfitting

Performs well on small datasets

XGBoost Training & Tuning

- Hyperparameter tuning using GridSearchCV (5-fold CV)

Parameters tested:

- n_estimators: [50, 100, 200]
- max_depth: [3, 6]
- learning_rate: [0.1, 0.01]
- Best Model saved as best_xgb_model.pkl
- ***Evaluation Metric:*** Accuracy

Results (Initial)

- Test Accuracy: ~95%
- Most classes predicted correctly
- Misclassification:
- Alligator → Amphibian (wrong)
- Cause: Feature overlap with salamanders
- Display Confusion Matrix (visual placeholder)

Feature Engineering

Added new features based on biological knowledge:

cold_blooded → reptiles, amphibians, fish

scales → reptiles, fish

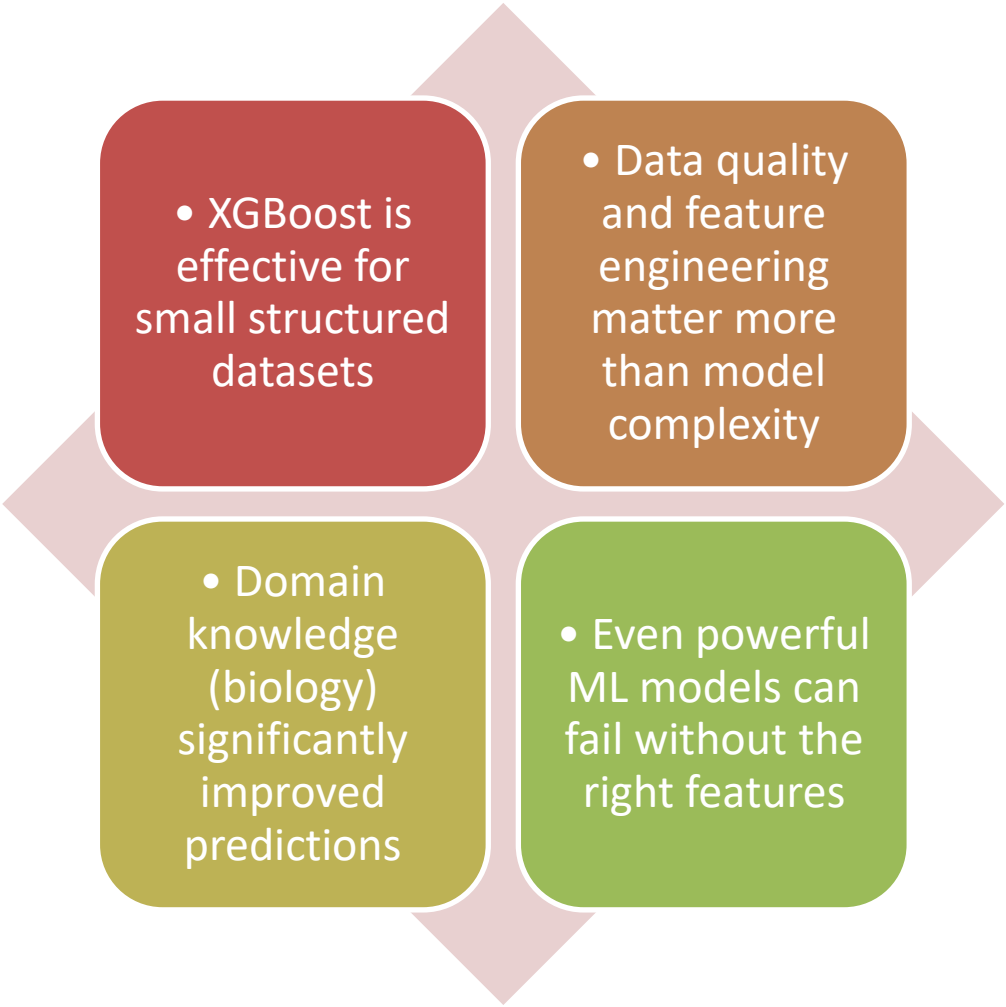
metamorphosis → amphibians

Purpose: Improve separation between reptiles & amphibians

Results (After Feature Engineering)

- Accuracy improved
- Alligator → correctly predicted as Reptile (3)
- Salamander → correctly predicted as Amphibian (5)
- Cobra → correctly predicted as Reptile (3)
- Updated Confusion Matrix shows fewer errors

Key Findings



Conclusion

Built a robust animal classifier with XGBoost + engineered features

Achieved high accuracy and fixed misclassification issues

THANK YOU

Presented by
A. Laxmi Prasanna
S. Sameeksha