Bird Species Identification from an Image

Team Members: Sharavana Kumar H, Nishita Singh
Course: UE23CS352A – Machine Learning Mini Project
Date: October 2025

Problem Statement

Identifying bird species from images is challenging due to high intra-class variance (pose, lighting, background) and subtle inter-class differences. The goal of this project is to classify images into 200 bird species using the **CUB-200-2011 dataset**.

Approach

- **Data & Features:** Used 11,788 images of 200 bird species with 28 attribute groups and 312 binary visual features (color, shape, pattern).
- **Pre-processing:** Applied feature reduction (PCA) and feature selection (L1-based, variance threshold, univariate, tree-based).
- **Models Implemented:** Naive Bayes, KNN, Decision Trees, Random Forests, SVM, LDA, Logistic Regression.
- **Training & Evaluation:** Split dataset 70%/30%; evaluated models using testing accuracy and confusion matrices.

Implementation Overview

- Implemented models in Python using **Scikit-learn**.
- Experimented with PCA and feature selection to improve classifier performance.
- Optimized SVM kernels and Logistic Regression for high-dimensional data.
- Generated visualizations for training vs. testing accuracy and confusion matrices.

Conclusion & Challenges

- **Best Model:** Logistic Regression with PCA + feature selection achieved ~54% testing accuracy.
- Feature engineering significantly improved accuracy for the 200-class problem.
- Challenges included high dimensionality of features and handling class imbalance.
- The project demonstrates practical application of machine learning for fine-grained image classification.