Project Summary

Our project aims to classify images of snakes as venomous or non-venomous using machine learning using supervised classification with image-based color and texture features to train our baseline model

What Has Been Accomplished So Far

- The dataset was organized into two labeled folders — Venomous and Non-Venomous
- Each image was resized to 128×128 pixels, normalized, and converted to both RGB (for color histograms) and grayscale (for texture extraction)
- After experimenting with Logistic Regression (55% accuracy), we trained a Random Forest classifier with 150 estimators, achieving an accuracy of 64.2% on the test set (1 fold). Also tried out KNN classification yielding a test accuracy: 62.5%.
- A full data-processing pipeline was implemented, including automated feature extraction, training/testing splits, and performance reporting. A visualization of class distribution was also generated to confirm dataset balance.

What Has Fallen Behind

- Testing on 10 folds
- Implementation of Convolutional Neural Network (CNN) postponed to Milestone 2.
- Some dataset samples exhibit high variation in lighting, posture, and background, limiting performance of simple linear models.

Any Project Changes

We revised our baseline model from Logistic Regression to Random Forest after evaluating performance differences.

This change is motivated by several advantages over logistic regression:

- Handles nonlinear relationships between color/texture features and snake venom status.
- More robust to noise and irrelevant background data.
- No need for extensive feature scaling.
- Provides feature importance metrics for interpretability.

Another important change going forward, we are looking to achieve better feature extraction since model type does not seem to be the bottleneck.

We have completed dataset preparation, preprocessing, feature extraction, and baseline model training with measurable performance. The project is on track for Milestone 2, where we plan to integrate deep feature extraction (CNN-based) to further enhance accuracy.