Lab Assignment #2: Image Spam Detection Using a Machine Learning Model

Objective

The primary aim of this lab assignment is to implement a machine-learning model for detecting image spam. Students are given the freedom to explore various image processing and data preprocessing methods, as well as different machine learning algorithms, to identify an optimal approach.

Note: You are allowed to use external libraries like scikit-learn for model implementation, and data preprocessing.

Dataset

For this assignment, you will be using the Image Spam Dataset: https://www.cs.jhu.edu/~mdredze/datasets/image_spam/. The program must include a feature that allows the grader to randomly set the ratio between the training and testing datasets upon execution.

Requirements

1. Image Processing & Preprocessing: You have the flexibility to experiment with various image processing and preprocessing methods. Your goal is to find the optimal combination.
2. Model Training: You are required to implement a machine-learning model. The use of external libraries like scikit-learn for model implementation is allowed.
3. Model Evaluation: Evaluate the trained model using the F1-score metric.
4. Result Analysis: Analyze the impact of the chosen image processing and preprocessing methods on the model's performance.
5. Documentation: Thoroughly document the entire process and your findings, including an explanation of the methods used for implementation.

Evaluation Criteria

- Code Completeness: Does the model function correctly?
- Model Performance: Performance as measured by the F1-score.
- Documentation: Is the entire process well-documented, including a clear explanation of the implementation?

Submission Items

1. Code File (.py)
2. The report containing implementation methods and result analysis (PDF)