| | |
|---|---|
| **Subject:** Artificial Intelligence | **Task** |
| **Instructor:** Imran Nawar | **Date:** 26 July 2025 |

# Project: Image Classification using ML + Feature Descriptors (ORB/SIFT)

Build an end-to-end image classification system using Machine Learning algorithms and feature descriptors (ORB or SIFT). This project will include dataset exploration, preprocessing, feature extraction, model training, evaluation, and real-time inference using Streamlit.

## Dataset:

- Intel Image Classification Dataset (Kaggle)
- Contains ~25,000 labeled images of 6 classes: buildings, forest, glacier, mountain, sea, and street.

## Steps:

1. **Load Dataset**
    - Open a Python Notebook (Jupyter notebook, Kaggle, Google Colab etc).
    - Download and unzip dataset from Kaggle.
    - Organize data into training and test folders by class labels.
    - Load images using cv2.
2. **Visualize Sample Images**
    - Randomly select and display 5–10 images per class using matplotlib.
3. **Preprocessing**
    - Resize images (e.g., 128x128).
    - Convert to grayscale (if needed).
    - Normalize pixel values (optional).
4. **Feature Extraction**
    - Use **ORB** or **SIFT** to extract local features from each image.
    - Convert features into a fixed-size vector using:
5. **Model Training**

Train and compare at least **three** of the following ML models using scikit-learn:

- Support Vector Machine (SVM)
- Random Forest
- k-Nearest Neighbors (k-NN)
- Logistic Regression

6. **Evaluation**
    - Use accuracy.
    - Plot confusion matrix using seaborn.
7. **Save Best Model**
    - Use joblib or pickle library to save the trained model with the best performance.

8. **Create Streamlit App**
    - Upload an image.
    - Perform preprocessing and feature extraction.
    - Load saved model and predict the class.
    - Display result and confidence score.
9. **GitHub Repository**
    - Create a public GitHub repo.
    - Push all project files: notebooks/scripts, Streamlit app, saved models, and README.
10. **Customize README.md**

Include:

- Project overview
- Dataset link and description
- Tools used (OpenCV, scikit-learn, Streamlit)
- How to run training & inference
- Screenshots of Streamlit app (if possible)

---

**Best of Luck!**