Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 25 JUNE 2025 |
| Team ID | SWTID1749974387 |
| Project Title | **Neural Networks Ahoy: Cutting-Edge Ship Classification For Maritime Mastery** |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Report:**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan:**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | To build an image classification model that can automatically identify the type of a ship — Cargo, Military, Carrier, Cruise, or Tanker — from a given image using deep learning and transfer learning techniques. |
| Data Collection Plan | * Search for image datasets related to **ship classification**, focusing on categories like Cargo, Military, Carrier, Cruise, and Tanker. * Prioritize datasets that include **labeled ship images** with clear category annotations. * Ensure a **balanced distribution** across classes to avoid model bias. * Where possible, collect data from **diverse environments** (different lighting, angles, sea/weather conditions) to improve generalization. * Use **image preprocessing** (resizing, normalization) and **data augmentation** techniques (rotation, flip, zoom) to enrich the training dataset and reduce overfitting. |
| Raw Data Sources Identified | * High-quality images of ships categorized into five types: *Cargo, Military, Carrier, Cruise,* and *Tanker*. * A labeled dataset suitable for training a **Convolutional Neural Network (CNN)** model, with sufficient samples in each category to ensure accurate classification. * Sample data has been preprocessed to a uniform size (224x224 pixels) and normalized, consistent with the input requirements of the **VGG16** architecture used in the model.   These datasets formed the basis for training the classification model used in the Flask web application, which allows users to upload a ship image and receive predictions with confidence scores and ship-type details. |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access**  **Permissions** |
| Kaggle Dataset | The dataset comprises applicant details (gender, marital status), financial metrics (income, loan amount), and loan approval outcomes. | [**https://www.kaggle.com/code/abdullahhaxsh/ship-classifier-using-cnn/data**](https://www.kaggle.com/code/abdullahhaxsh/ship-classifier-using-cnn/data) | CSV&Image | 83.85 MB | Public |
|  |  |  |  |  |  |