Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 29 April 2025 |
| Skillwallet ID | **SWUID20250148853** |
| Project Title | Uncovering The Hidden Treasures Of The Mushroom Kingdom: A Classification Analysis |
| Maximum Marks | 4 marks |

**Data Quality Report Template**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | Image Variation | High | Collect images from diverse sources (different cameras, lighting conditions, angles). Implement data augmentation techniques (rotation, scaling, cropping) during preprocessing. |
| Dataset | Occlusion | Moderate | Include images with partial occlusion, and/or train the model to be robust to it. |
| Dataset | Insufficient Resolution | Moderate | Establish a minimum resolution threshold for images. Use super- resolution techniques, if feasible, to enhance the resolution of some |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | images. |
| Dataset | Unbalanced Classes | High | Employ stratified sampling to ensure proportional representation of each mushroom species. Use data augmentation for minority classes. Explore the use of weighted loss functions during training. |