

## Data Collection and Preprocessing Phase

|               |   |
|---------------|---|
| Date          | 29 November 2024  |
| Team ID       | 739984  |
| Project Title | AI-Generated vs. Real Image Classifier Using Deep Learning. |
| Maximum Marks | 2 Marks   |

### Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

| Data Source | Data Quality Issue    | Severity | Resolution Plan  |
|-------------|-----------------------|----------|--|
| Kaggle      | Class Imbalance       | High     | Use oversampling (e.g., SMOTE) or under sampling techniques to balance the dataset, or employ class-weighted loss functions to mitigate bias during model training.          |
| Kaggle      | Varying Image Quality | Moderate | Apply preprocessing techniques like image resizing, denoising, and contrast enhancement to standardize the quality, or use data augmentation to simulate different qualities |
| Kaggle      | Domain Shift          | High     | Use domain adaptation techniques or fine-tune models using transfer learning on domain-specific data to improve performance across varied sources.                           |