**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 12 June 2025 |
| Team ID | SWTID1749710222 |
| Project Title | Unlocking Silent Signals: Decoding Body Language with Mediapipe |
| 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** |
| coords.csv (generated from Mediapipe) | Some key points may be missing or NaN due to poor detection or movement during recording | Moderate | Handle missing values by interpolation or forward fill, or drop rows with many NaNs |
| coords.csv (generated from Mediapipe) | Few samples may be labeled incorrectly due to human oversight during collection | Low | Manually review and relabel those samples or filter them if they are very few |
| coords.csv (generated from Mediapipe) | Presence of outliers or abrupt movements causing unusual joint positions | Low | Remove or filter outliers by Z-score or IQR method |
| coords.csv (generated from Mediapipe) | Imbalanced classes (happy, sad, victorious, fight) | Moderate | Apply resampling techniques (like SMOTE) or under-sampling to balance classes |
| coords.csv (generated from Mediapipe) | Few recordings may be blurry or unreliable due to poor lighting or movement | Low | Remove samples with low confidence scores provided by Mediapipe or filter manually |