Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 04th JULY 2024 |
| Project Title | AI Feedback Generator For Development Teams Using Palm's Text-Bison-001 |

# 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 | The AI project aims to enhance the code review process by implementing an AI feedback generator using Palm's Text-Bison-001. By analyzing a dataset with features such as code snippets, code comments, and coding standards, the objective is to build a model that provides real-time, accurate feedback on code quality, facilitating efficient and informed decision-making in the development process. |
| Data Collection Plan |  **Search for datasets** related to code repositories, coding standards, and code reviews.   **Prioritize datasets** with diverse coding styles, languages, and complexity levels. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from public repositories such as GitHub and popular platforms like Kaggle & UCI. The provided sample data represents a subset of the collected information, encompassing variables such as code snippets, comments, and review feedback for machine learning analysis. |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| UCI | This data includes various code examples and programming assignments used in academic settings. | https://archive.ics.uci.edu/ml/datasets/Spambase | CSV | 13.6  kB | Public |