

Workflow Log

Understanding & Collecting Data

- **Code Comment Generator:** Find datasets with code snippets and comments (GitHub, CodeSearchNet).
<https://www.kaggle.com/datasets/omduggineni/codesearchnet/data>
- **Bug Report Classifier:** Collect labeled bug reports (Bugzilla, JIRA).
<https://www.kaggle.com/datasets/vellyy/bug-hunter>

Preprocessing the Data

- **Code Comment Generator:** Extract function definitions, remove extra spaces, clean code.
- **Bug Report Classifier:** Remove stop words, special characters, and convert text into numerical form using TF-IDF.

Model Selection & Training

- **Code Comment Generator:** Fine-tune a pre-trained model like GPT-2 or T5.
- **Bug Report Classifier:** Train a classifier (Logistic Regression, Random Forest, or Neural Network) on cleaned bug reports.

Testing & Evaluation

- **Code Comment Generator:** Check if the generated comments are relevant.
- **Bug Report Classifier:** Measure accuracy using a confusion matrix and test on real bug reports.

Deployment

- Create a Flask API to accept code snippets and bug reports.
- Serve model predictions as JSON responses.