Instruction Manual for new Repo

Data Extraction:

- Github link - *Download the zip file*
- To extract bug report and store it in the csv file - *retrieveBug.py*

Data Preprocessing:

- Mapping the issue with its files changed (0/1) - *mapping.py*

Feature Extraction:

- The java codes are converted into various different representations install SrcML, Progex, Comex:
    - AST - use SrcML - run below command line in SrcML
        - *find /path/to/your/java/files -name "*.java" -exec srcml --position {} -o {}.xml*
    - CFG - use Progex - run the python file below command line in progex
        - *convertToCFG.py*
    - For the below 5 combination use comex and run the filename *dotfiles.sh* and in the *dotfiles.sh* write the combination you want:
        - DFG - use comex
        - AST + CFG (Comb1) - use comex
        - AST + DFG (Comb2) - use comex
        - CFG + DFG (Comb3) - use comex
        - AST + CFG + DFG (All) - use comex
- For vectorizing the source code representation and store it in json file -
    - For AST - run the python file - *vectorize_ast.py*
    - For CFG - run the python file - *vectorize_cfg.py*
    - For remaining combinations edit the file according to the combination you want and run the file - *comex_vect.py*

Training & Evaluating:

- For different source code representation run the file -
    - AST input - *Train_AST.py*
    - CFG input - *Train_CFG.py*
    - DFG input - *Train_DFG.py*
    - AST + CFG input - *Train_AST_CFG.py*

- AST + DFG input - *Train_AST_DFG.py*
- CFG + DFG input - *Train_CFG_DFG.py*
- AST + CFG + DFG input - *Train_ALL.py*

Testing :

To predict the top 10 files run the file - *Test.py*