Please first install python 2.7 and then install genism from this address:

https://radimrehurek.com/gensim/install.html

**PythonRun.jar:**

This jar file is responsible for running the machine learning part to find the textual similarity between bug description and corpus. This jar file needs three command line arguments as follows:

First argument is the directory which you stored these three files: (please save all these three files in one directory)

machinelearningwithlsi.py, Benchmark file, and Corpus file

Second argument is the name of benchmark file

Third argument is the name of corpus file. Please remember that you save benchmark and corpus files along with machinelearning.py in one directory.

Example:

java -jar PythonRun.jar "D:\research\codereview\developerrecom\MicrosoftPackageTest\testpythonrun" "EclipseBenchmark" "CorpusEclip"

**machinelearningwithlsi.py**

This file is responsible for running the LSI and finding the most similar files to the bug description. It creates the *output* folder inside the same directory which you already set as first command line argument for PythonRun.jar.

There are two different parameters called *num\_topics* and *chunksize* which are used by *machinelearningwithlsi.py*. If you faced memory issue please change the value of these two parameter to see if it helps.

*machinelearningwithlsi.py* creates the dictionary, corpus, tfidf model and lsi model and saves them in directory for later use.