

ABSTRACT

Provided is a computer-implemented method, system, and computer program product for automatic selection of tests for software system regression testing using machine learning including generating a test mapping including at least one test of a plurality of tests corresponding to a source file. The plurality of tests and the at least one source file are associated with a software repository. Further, determining a defective score for the at least one test based on historical test data of the at least one test, receiving a component criticality score and a defect definition corresponding to the source file, generating a key value corresponding to at least one test based on the defective score, component criticality score, and defect definition, determining a subset of tests of the plurality of tests based on the key value corresponding to the at least one test; and executing the subset of tests with the software repository.

<Figure 2>