

Problem 2:

You are given a data set containing various attributes that impact the result of the sanity test when a particular file in the given directory changes and is committed to GitHub. Build a classification model to predict if a sanity test case will fail on the platform for a given combination of attributes when a particular file is changed.

Enhance the model to identify the scenarios which have a very low probability of failure for the change in a particular set of files.

Data Attributes:

File Changed: Path of the file changed and committed in github

Component: Code Component area to which file belongs

Coder: Name of person coded the file change

Tester: Name of person who tested the file change

Category: Test category

Sanity Name: Test name which ran on that file

Platforms: Hardware platform on which the test was run

Sanity Result: Test case result (Pass/Fail/Aborted)