**CS699\_Data\_Mining\_Project\_Proposal**

**(Kevin Rodrigues and Paritosh Shirodkar)**

**PC1 Software defect prediction**

One of the NASA Metrics Data Program defect data sets. Data from flight software for earth orbiting satellite. Data comes from McCabe and Halstead features extractors of source code. These features were defined in the 70s in an attempt to objectively characterize code features that are associated with software quality.

**Attribute Information**

1. loc : numeric % McCabe's line count of code

2. v(g) : numeric % McCabe "cyclomatic complexity"

3. ev(g) : numeric % McCabe "essential complexity"

4. iv(g) : numeric % McCabe "design complexity"

5. n : numeric % Halstead total operators + operands

6. v : numeric % Halstead "volume"

7. l : numeric % Halstead "program length"

8. d : numeric % Halstead "difficulty"

9. i : numeric % Halstead "intelligence"

10. e : numeric % Halstead "effort"

11. b : numeric % Halstead

12. t : numeric % Halstead's time estimator

13. lOCode : numeric % Halstead's line count

14. lOComment : numeric % Halstead's count of lines of comments

15. lOBlank : numeric % Halstead's count of blank lines

16. lOCodeAndComment: numeric

17. uniq\_Op : numeric % unique operators

18. uniq\_Opnd : numeric % unique operands

19. total\_Op : numeric % total operators

20. total\_Opnd : numeric % total operands

21. branchCount : numeric % of the flow graph

22. branchCount : numeric % of the flow graph

23. defects : {false,true} % module has/has not one or more reported defects

**Data Mining Goal:** To classify whether the module has one or more defect or not i.e true or false.

**Class Attribute:** The ‘defects’ attribute stores the class label once the model has classified whether a module has/has not one or more reported defects

**Citation:**

https://www.openml.org/d/1068