Software Defect Prediction: Dự đoán lỗi phần mềm

Là 1 task nhỏ trong SE, input là các dự án bao gồm source code repo, output: dự đoán lỗi có thể có ở các level khác nhau: function, method.

Bài toán này hiện có 2 cách tiếp cận:

* Chỉ dựa trên độ phức tạp mã nguồn (software metrics)
* Dựa trên cả độ phức tạp mã nguồn + source code

Phân loại theo scope:

* Dự đoán trong 1 project: Dùng 1 version của source code để train, 1 version tiếp theo của source code để test của chính project đó. Benefit: thông thướng 2 version liên tiếp nhau thì sẽ có một số phần code chung, và các code change. Người ta dựa trên version trước đó để dự đoán lỗi version hiện tại thì rất hợp lý
* Dự đoán chéo project: Dùng 1 project A để train và test trên project B

Challenge Task:

* Chỉ dùng software metric, mỗi project sẽ gồm các source file, mỗi source file đặc trưng bởi 1 vector biểu diễn độ phức tạp mã nguồn
* Đọc các paper có liên quan - có khá nhiều paper, mỗi paper thể hiện các cách tiếp cận khác nhau. Recommend: chọn 2-3 bài thấy phù hợp để đọc. Đưa ra các idea của bài đó.
* Thử các cách tiếp cận khác nhau, mục đích đẩy cao performance của các thuật toán machine learning lên.
* Formulate bài toán thành 2 dạng:
  + Binary classification: error/non-error
  + Regression: number of errors
* Chỉ chạy thực nghiệm trên mô hình within-project SDP

**Thời gian thực hiện: 2 weeks. Week 1 - progress review, analysis, idea. Week 2 – Results**