

ISW2 – 2019/20

Specifiche Progetto Sezione Falessi

Machine Learning for Software Engineering

Deliverable 1

Misurare la stabilità di un attributo di progetto con modalità descritte nelle slide.

Deliverable 2

Eseguire uno studio empirico finalizzato a misurare l'effetto di tecniche di sampling e feature selection sull'accuratezza di modelli predittivi di localizzazione di bug nel codice di larghe applicazioni open-source.

In particolare è necessario usare:

- 2 progetti (come descritto nelle slide)
- walk forward come evaluation technique.
- No selection / best first come feature selection.
- No sampling / oversampling / undersampling / SMOTE come balancing
- RandomForest / NaiveBayes / lbk come classificatori.

La domanda a cui rispondere è se (e quali) tecniche di feature selection o balancing aumentano l'accuratezza dei classificatori:

- Per quali classificatori
- Per quali dataset
- Per quale EPV

Per rispondere alla domanda sopra si consiglia di creare, e poi analizzare, un file avente le seguenti colonne: dataset, #TrainingRelease, %training (data on training / total data), %Defective in training, %Defective in testing, EPVbeforeFeatureSelection, EPVafterFeatureSelection, classifier, balancing, Feature Selection, TP, FP, TN, FN, Precision, Recall, ROC Area, Kappa.

Report

I due deliverable devono essere descritti in un unico documento pdf. Ogni deliverable deve essere descritto in termini di Introduzione (e.g., contesto, scopo, etc.), Progettazione (e.g., Variabili (e.g., bugs (tipi, filtri), evaluation techniques), procedura di misurazione), Risultati (e.g, grafici, tabelle), Discussione (i.e., considerazione sui risultati), link sonarcloud del codice usato.