Project Proposal

Pablo Acereda

10 Oct 2019

1 Defining the Project

1.1 Detailed Research

Question: How complexity metrics are correlated to the outcome of supervise algorithms? Do complexity metrics affect the outcome of Feature Selection Algorithms? How complexity metrics and imbalance are related? Do complexity metrics tell us something about the quality of the datasets?

Aim: To explore complexity metrics on software defect datasets.

Objectives:

1. Literature search and literature review on complexity metrics, software defects, prediction techniques, etc.
2. Develop suitable models for Ho and Basu complexity metrics.
3. Identify and collect suitable data for analysis and evaluation.
4. Complete final report.

1.2 Keywords

Data Mining, Data Quality, Software Defect Prediction, Complexity Metrics, Imbalance.

1.3 Project Title

Analyse Complexity Metrics in Software Defect Prediction.

1.4 Client, Audience and Motivation

The target group of this paper is mainly academics. Those who would be more interested among that generalization, would be the academics in OBU and UAH; who could benefit from the new knowledge generated about how complexity metrics affect classification, attribute selection, etc.

1.5 Project Plan and Risk Management

1. Literature search - 1 week
2. Literature review - 3 weeks
3. Design and implementation - 2 weeks
4. Evaluation and Report Completion - 1 week

2 Initial Literature Review

2.1 Abstract

In this paper it is going to be analysed the complexity metrics affect classification, attribute selection, imbalance, etc., using supervised classification models on publicly available datasets. The obtained results are to be compared with the metrics from the dataset; regarding the attribute selection.

To achieve more accurate results, cross-validation methods are also implemented.

The experiment is going to be done using Python as programming language.

2.2 Initial Literature Review

2.3 Relevant Professional, Social, Ethical, Security and Legal Issues to the Project

2.4 Bibliography

References

[1] Tin Kam Ho and Mitra Basu. "Complexity Measures of Supervised Classification Problems". In: ().

[2] Sarah Nogueira, Konstantinos Sechidis, and Gavin Brown. "On the Stability of Feature Selection Algorithms". In: ().