Evaluation Report

Contents

[Introduction 1](#_Toc82859709)

[Setup 2](#_Toc82859710)

[Analysis 2](#_Toc82859711)

[Appendix 3](#_Toc82859712)

Introduction

This report covers our analysis on the performance of the best algorithm from our proposed system against other software fault prediction algorithms devised from other research papers.

Setup

After evaluating the performances of each of our algorithms (which was done in a separate evaluation report), the best performing algorithm would then be selected for comparison with other algorithms from other research papers, such as the SDAEsSTE approach devised by Tong et al. The algorithms are compared based on the AUC and F1-scores results which were present on the research paper themselves, and the data is tabulated based on the various datasets which were used on the algorithms. The comparisons can be found in the table at the results section.

Results

The table in the following page shows our results when compared against other algorithms from other research papers. It is a comparison with the algorithms devised by Yucalar et al. and Tong et al. , and AUC and F1-scores are used to compare the performances between each of these algorithms. Algorithms that perform the best for a certain dataset will have their results bolded and highlighted.

**Table 1: AUC and F1-score results comparisons with other proposed algorithms**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Datasets used** | Results | | | | | |
| Our Algorithm | | VOT(ROF, RF, LB)  (Yucalar et al.) | | SDAEsSTE (Tong et al.) | |
| AUC | F1 | AUC | F1 | AUC | F1 |
| CM1 | 0.806 | **0.391** | N/A | N/A | **0.8373** | 0.2882 |
| JM1 | 0.654 | **0.325** | N/A | N/A | **0.7731** | 0.3174 |
| KC1 | 0.625 | **0.386** | N/A | N/A | **0.7426** | 0.3946 |
| MC1 | 0.958 | **0.457** | N/A | N/A | **0.9614** | 0.2217 |
| MC2 | **0.815** | **0.635** | N/A | N/A | 0.6846 | 0.5871 |
| MW1 | 0.73 | 0.08 | N/A | N/A | **0.8597** | **0.4073** |
| PC1 | 0.846 | 0.244 | N/A | N/A | **0.9062** | **0.3062** |
| PC2 | 0.913 | **0.25** | N/A | N/A | **0.9264** | 0.2154 |
| PC3 | **0.859** | 0.326 | N/A | N/A | 0.8381 | **0.3545** |
| PC4 | **0.889** | 0.527 | N/A | N/A | 0.8846 | **0.5544** |
| cm1 | 0.752 | 0.264 | **0.784** | N/A | N/A | N/A |
| jm1 | 0.633 | 0.285 | **0.753** | N/A | N/A | N/A |
| kc1 | 0.636 | 0.291 | **0.845** | N/A | N/A | N/A |
| pc1 | 0.865 | 0.354 | **0.876** | N/A | N/A | N/A |

Conclusion

From the analysis above, we can see that our algorithm outperforms Tong et al’s SDAEsSTE algorithm in certain datasets. On the other hand, our algorithm is not able to outperforms Yucalar et al’s algorithm which is a combination of ensemble predictors. To conclude, although our best algorithm is not the best in terms of performance, it is still a decently viable method to use when a dataset is found to be imbalanced, as parts of our algorithm are built to handle those situations.