ROBERT GORDON UNIVERSITY ABERDEEN
SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

Computing Application Software Development

A Desktop Application for Maintenance of Computer Systems

Darie-Dragos Mitoiu, Mr. Ian Harris

Introduction

The classical reliability theory that concerns the computer systems or the conventional approaches put in place for problem solving present a significant lack of consideration of the actual state of the computer systems as they are not capable to offer a reflection for the dynamic and the unforeseen runtime of a computer system or the failure of processes. These approaches are usually used in order to create a design that will allow for a long term or average functionality predictions to solve the technical problems that occur in computer systems in an optimal and efficient way when considering time and cost as the main factors of evaluation (Salfner, Lenk, Malek 2010).

Objective

The services and resources of a desktop or even a server type of computer system could be organized and scheduled in a manner which would reduce the impact caused by a possible system failure if maintenance predictions will be provided. In this specific case, the provision of computer systems failure predictions will be done with the help of techniques which imply the involvement of classification machine learning algorithms and the Hard-Disk component.

Methods

Due to the unique nature of the Hard-Disk S.M.A.R.T attributes as they are vendor specific, a series of steps were performed in order to solve the problem in cause, these steps are: the selection of the most relevant features when considering the influence on the failure event, the elimination of records which contain missing values from the dataset and the undersampling of the dataset.

Figures and Results

Algorithm	F-1 Score	ROC AUC	Accuracy (%)
Support Vector Machines	0.44	0.67	94.56
Random Forest Model	0.50	0.70	95.10
Logistic Regression	0.45	0.68	94.61
Multi-Layer Perceptron	0.46	0.69	94.52

Table 1: Best results for each algorithm used for the experiments

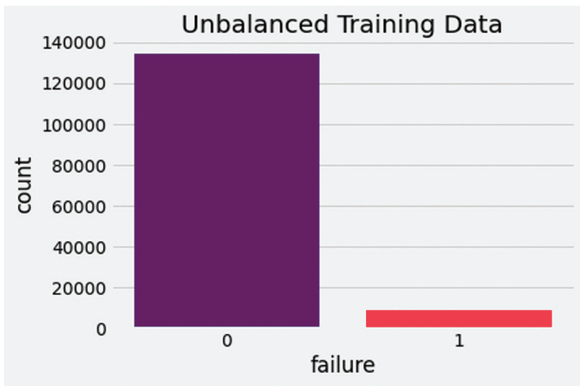

Figure 1: Initial class distribution

Figure 2: Balanced class distribution

According to Salfner, Lenk and Malek (2010) the core elements that contribute to the challenges and impediments of computer systems failure or downtime events predictions are factors such as dependability and resilience. These factors are represented by the continuous increasing of computer systems complexity, the continuous increasing of cyber security threats and last but not the least the dynamic and unforeseen runtime of computer systems. In addition to the core elements mentioned previosuly, according to Botezatu, Giurgiu, Bogojeska and Wiesmann (2016) when researching computer systems failure predictions a known fact is that most computer systems will present a storage system, this system being an important component and also the root cause of the most computer systems failures in data centres, in most cases this storage system it is a hard-disk and usually most hard disks present a mechanism called Self-Monitoring, Analysis and Reporting Technology (S.M.A.R.T), this mechanism contributes to the difficulty of trying to predict unforeseen failure events based on the storage system. The dynamicity of the hard-disk S.M.A.R.T indicators will be the main factor of difficulty as these indicators are manufacturer specific, the specialised encoding of the hard disk indicators will be different for each hard-disk model and their normalization values will also vary.
The above theory can be considered true, as it was necessary to train the machine learning models presented in table 1 using hard-disk S.M.A.R.T attributes from a single vendor in association with preliminary pre-processing operations to eliminate any errors that could arise from not doing so.

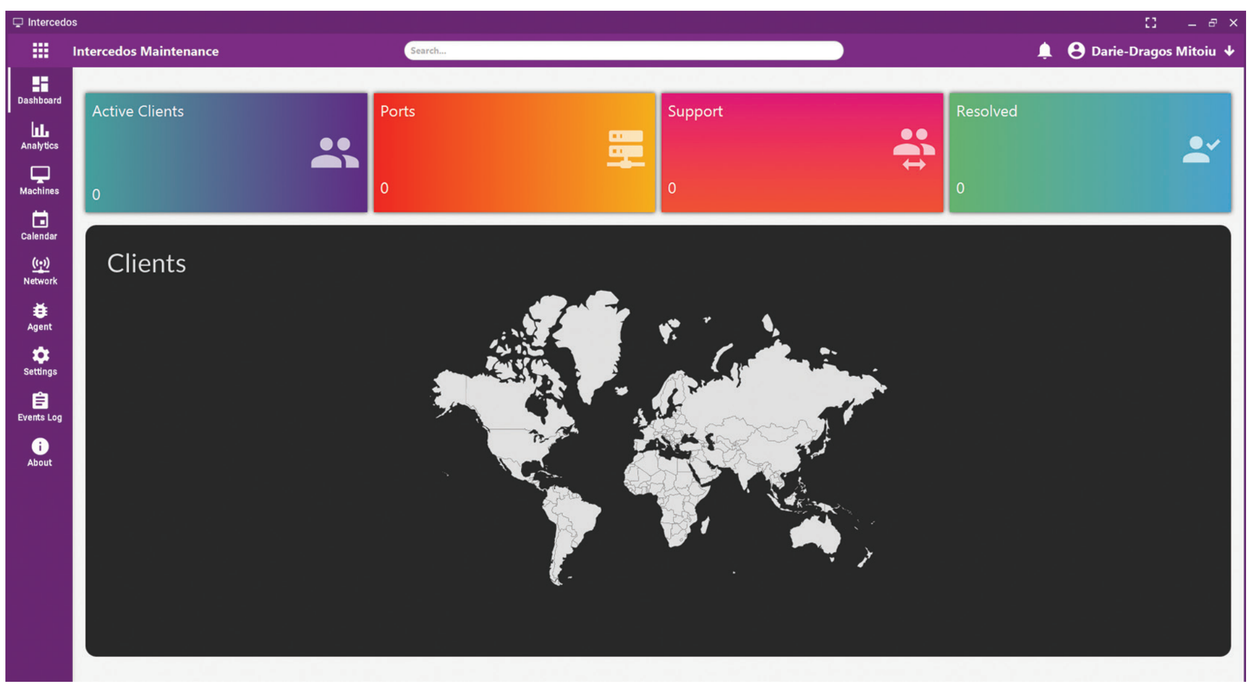

Figure 3: Dashboard of the real-world desktop application

Conclusion

While the metrics presented in table 1 may look promising, it must be considered that the analysis was performed using a single vendor of hard-disk components, this approach was necessary in order to ensure correct forecasting of computer systems failure or disk replacements events as different vendors use different hard-disk S.M.A.R.T attributes to evaluate the state of the storage system.
It is safe to say that the forecast of computer systems failure using the hard-disk component it is possible and it can be translated into a real-world application by training models using existing hard-disk data or collecting such data from a reasonable number of machines.

Acknowledgments

I would like to thank to my colleagues from Robert Gordon University Aberdeen who provided an insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations or conclusions presented in this paper.
I would also like to show my gratitude to Felix Salfner, Maren Lenk and Miroslaw Malek from University of Berlin for sharing their pearls of wisdom with us during the course of this research.

References

1. Salfner, F., Lenk, M. and Malek, M. (2010), A survey of online failure prediction methods, ACM Computing Surveys (CSUR) 42(3), 1–42.
2. Botezatu, M.M., Giurgiu, I., Bogojeska, J. and Wiesmann, D., 2016, August. Predicting disk replacement towards reliable data centers. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 39-48).