Using AI in the fault management predictive model of the SKA TM Services: a preliminary study

Matteo Canzari [(1)], Matteo Di Carlo [(1)], Mauro Dolci [(1)], Riccardo Smareglia [(2)]
[(1)] INAF – Osservatorio Astronomico d'Abruzzo – Italy, [(2)] INAF - Osservatorio Astronomico di Trieste – Italy

ABSTRACT

SKA (Square Kilometer Array) is a project aimed to build a very large radio-telescope, composed by thousands of antennae and related support systems. The overall orchestration is performed by the Telescope Manager (TM), a suite of software applications. In order to ensure the proper and uninterrupted operation of TM, a local monitoring and control system is developed, called TM Services. Fault Management (FM) [1] is one of these services, and is composed by processes and infrastructure associated with detecting, diagnosing and fixing faults, and finally returning to normal operations.
The aim of the study, introducing artificial intelligence algorithms during the detection phase, is to build a predictive model, based on the history and statistics of the system, in order to perform trend analysis and failure prediction. Based on monitoring data and health status detected by the software system monitor and on log files gathered by the ELK (Elasticsearch, Logstash, and Kibana) server, the predictive model ensures that the system is operating within its normal operating parameters and takes corrective actions in case of failure.

PREDICTIVE MODEL

A Predictive Model consists of the construction of a model based on data source useful to make predictions. It is primarily needed to prevent future events, but it can be applied also to past unknown events, regardless when they have occurred. There are a lot of methodologies that allow to create predictive models The current state-of-the-art of technologies, together with the increase of data volume and the processing power, makes it possible to apply Artificial Intelligence algorithms and machine learning to create powerful predictive models that learn certain properties from a training dataset in order to be able to make predictions.

Figure 1: A mind-map of Artificial Intelligence algorithms [2]

SUPERVISED MACHINE LEARNING ALGORITHMS

The technique of machine learning applied to the predictive model can be divided in two different areas: Regression and Pattern Classification[2]. The Regression consist in the study of the relations between a dependent variable and one or more independent variables (or predictors). At variance with it, the Pattern Classification is focused on the recognition of patterns and regularities in data and the assignment of discrete class labels to particular observations. In our study the attention has been focused on the Pattern Classification: in particular it can be further grouped in two subcategories: supervised and unsupervised. In the supervised learning, the class labels in the dataset, which are used to build the classification model, are knows. More generally, in a given set of input variables (X) and one output variable (Y), a supervised algorithm learns the mapping function from the input to the output Y = f(X) Once the function of mapping from the learning dataset, the aim of the algorithm is to compute values of (Y) from new given values of (X). It is called supervised learning because the process is supervised by a "teacher" (a human operator), who knows the real situation and can confirm, correct or reject the predictions the algorithm makes on the basis of what it has learnt. Learning stops when the algorithm achieves an acceptable level of performance. Most common algorithms of supervised learning are: linear regression for regression problems, Random forest for classification and regression problems, support vector machines for classification problems.

UNSUPERVISED MACHINE LEARNING ALGORITHMS

Unsupervised learning deals with unlabeled instances, and the classes have to be inferred from the unstructured dataset. More generally, it consists in knowing a set of input variables (X) without knowing output variables. The aim of unsupervised learning is modelling the structure or the data distribution in order to know them more. This learning is called unsupervised because there is no verification process by a "teacher": algorithms are left to their own devises to discover and present the interesting structure in the data. In particular, unsupervised learning can be further divided in: Clustering: used to discover groups of data Association: allows to identify rules that associate big portions of data Most common algorithms are: k-means for clustering problems. Apriori algorithm for association rule learning problems.

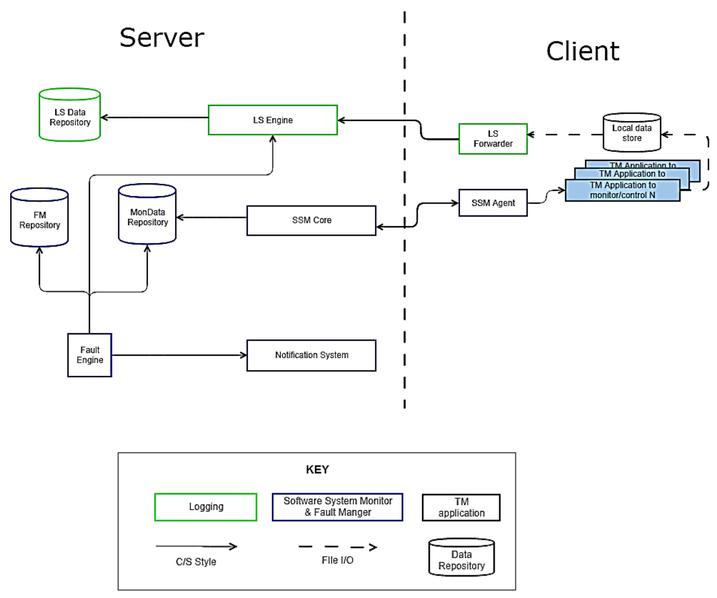

Figure 2: the architecture of SSM and Fault Manager. AI Algorithms are implemented in the Fault Engine Module

CASE STUDY

According to the study carried out, two different fields have been identified where the use of machine learning helps to improve the availability of the system. The first one is the validation of FMECA. In fact, FMECA analysis has the purpose of analyzing effects and the severity of failures on the whole system. So, by defining (X) as the possible inputs of the system that can cause failures and (Y) as the response of the system to the failure, FMECA identifies with a study the failure mode f(X), as the behaviour of the system when (X) occur For the validation of FMECA supervised learning algorithms, like Random forest for classification and regression problems, have been identified. The second one is the fault prediction of the system. Studied and established failure mode in the FMECA, it is possible to test f(X) using Random forest for classification and regression problems algorithm. to verify if the values of the (Y) obtained do not differ from (Y) estimated.
In order to perform fault prediction, unsupervised algorithms have been preferred. This because all the f(X) already known and already studied in FMECA have been validated with supervised methods.

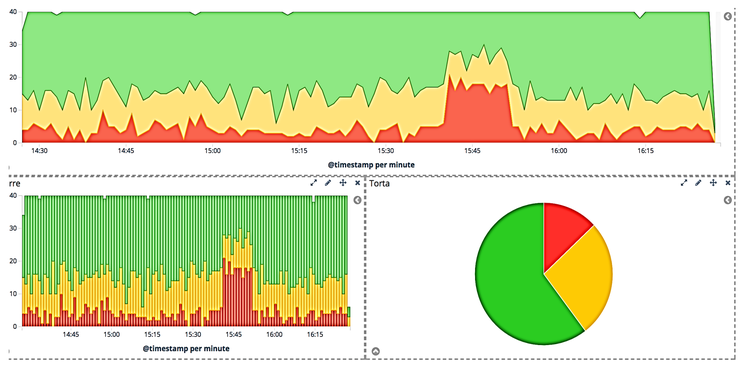

Figure 3: the dataset and the simulation of fault

FMECA does not cover all the possible cases in which the system can evolve. So, it is necessary to predict anomalous behaviours knowing only the values of (X). Among the most used tools the machine learning module of X-Pack (formerly Prelert) of Kibana [3][4] has been taken into consideration. This unsupervised learning patented software uses algorithms of Clustering based on Bayesian network. During the preliminary testing, has been simulated the functioning of a Tango Device Server which produced log based on his functioning and an error in the device server has been simulated. Previously there was a script that detected the error of the device according to parameters that would have been obtained in FMECA stage. For the learning stage the system has been working for 6 days, logging about 80.000 documents of logs. Machine learning algorithm, according to the learning already made, detected that the system was not working properly and reported the error.

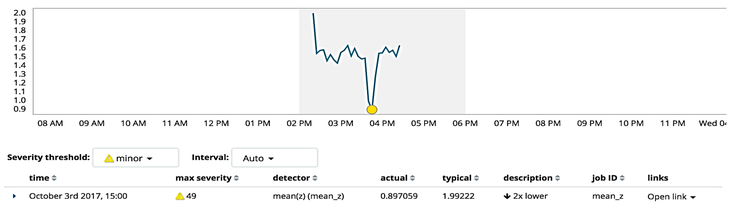

Figure 4: the AI algorithm detects the anomaly

PRELIMINARY CONCLUSION AND FUTURE WORK

In this paper a test has been performed on Artificial Intelligence algorhytms with the aim to automatically detect a failure or a fault in a system where these conditions had not been predicted by a preliminary FMECA. The test used a limited simulate dataset but succesfully proved the goodness of the adopted approach.
The results, even being very preliminary (the used dataset, made of a limited amount of simulated data, is not sufficient to describe the functioning of a complex system), are very encouraging and open the way to a more extensive work.
Next steps will use much larger datasets (at last one year of continuous operation) related to real telescopes, which will be processed by applying all the methods identified in this paper.

References

[1] Matteo Di Carlo, Matteo Canzari, Mauro Dolci (INAF - OA Teramo, Teramo), Domingos Barbosa, João Paulo Barraca, Jorge Bruno Morgado (GRIT, Aveiro), Riccardo Smareglia (INAF-OAT, Trieste): Tm Services: An Architecture for Monitoring and Con-Trolling the Square Kilometre Array (Ska) Telescope Manager (Tm) Proc. ICALEPCS 2017, TUPHA207;

[2] Python Machine Learning, Sebastian Raschka

[3] Anomaly Detection in Application Performance Monitoring Data, Thomas J. Veasey and Stephen J. Dodson, International Journal of Machine Learning and Computing, Vol. 4, No. 2, April 2014

[4] System and method for visualisation of behaviour within computer infrastructure, Patent EP 2645257 A2