Using AI in the fault management predictive model of the SKA TM Services: a preliminary study
Matteo Canzani (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 statisticalal 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.

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 devices 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.

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 known. 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.

Preliminary Conclusion and Future Work
In this paper a test has been performed on Artificial Intelligence algorithms 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 successfully 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