A hybrid machine learning process for anomalous satellite telemetry behaviour detection

Marcio Waldir Silva Junior1, Walter A. Dos Santos2

[1]Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil

Aluno de Mestrado do curso de Engenharia e Gerenciamento de Sistemas Espaciais – ETE/CSE.

[2] Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil – ETE/CSE

marciowaldir.sjr@outlook.com

Abstract. *In space missions, telemetry is a key source of systems health, and the lack of this may compromise the mission. Even with functional telemetry, there are some difficulties regarding the telemetry analysis. Some satellites have hundreds, even thousands of telemetries, and analyzing that to infer something about the system tends to be quite laborious. In this scenario, it can be difficult to perform in advance the detection, diagnosis, and prevention of anomalies and failures, decreasing the reliability and availability of space systems. Thus, shortening the system life and service continuity. This study proposes a data-driven approach, composed of a hybrid Machine Learning (ML) process capable of detecting anomalous behaviour on telemetry. Through statistics, data science processes, and ML algorithms, the proposed process was capable of classifying, with more than 90% of accuracy, four different circumstances for the telemetry system, sunlight, eclipse, twilight, and anomalous behaviour.*

Keywords: Machine Learning; KPCA; DBSCAN; KNN; Anomalous Behaviour;

1. Introduction

The usage of data science and machine learning algorithm to detect, diagnose, and prevent anomalous behaviour and therefore fails, has been applied to many different application fields (LI et al., 2018; SAFIZADEH, 2014; ZHANG, 2016), as to space systems as well, in special for satellites systems (GAO et al., 2012; YAIRI et al., 2017).

Artificial satellites usually provide important services in communication, remote sensing, scientific experiments, etc. Satellite damage entails not only a financial loss but also the loss of essential and sometimes strategic services. In this scenario, early detection, diagnosis, and prevention of anomalies and failures promote the reliability and availability of space systems, extending service life and long service continuity (AZEVEDO et al., 2012).

In space missions, telemetry plays an important role for systems status, and the lack of this may jeopardize a space mission since as it is usually the main source for the identification and prediction of anomalies.

Real-time and stored telemetry data is received during a satellite pass and analysed by specialists and operators, constituting the main source of identification and prediction of anomalies in artificial satellites (AZEVEDO et al., 2012). The complexity of these satellites with a big number of subsystems is reflected by the big number of TMs/TCs (WERTZ, 2011). Therefore, a large number of telemetries make its adequate data analysis an extremely complex task.

This study intends to present a data science approach for space systems telemetry data analysis with the main goal, at this first stage, of being capable of detecting anomalous behaviour, which are patterns in data that do not conform to a well define notion of normal behaviour (CHANDOLA et al., 2009), on the telemetry data from the power supply subsystem. Therefore, outliers and anomalies can be understand here as interchangeable terms This study also has a goal of being a backbone for a machine learning process capable of performing not only detection but also failure diagnosis.

2. Methodology

The machine learning process proposed and applied in this study is a fusion of different approaches found in the literature. Looking at it from a high abstraction level it can be understood as the data science process proposed in (SCHUTT, O´NEIL, 2013). Closely, the data-driven anomaly detection flow here is divided into four steps, being then, Data Preparation, Feature Reduction, Clustering, and Classification. Where the last two steps can be seen as a Machine Learning step in charge to generate and verify the model. This flow can be seen in Figure 1 and it follows the same proposed in (YAIRI et al., 2017; ZARE, 2018) and many others mentioned by (TABUROĞLU, 2019).

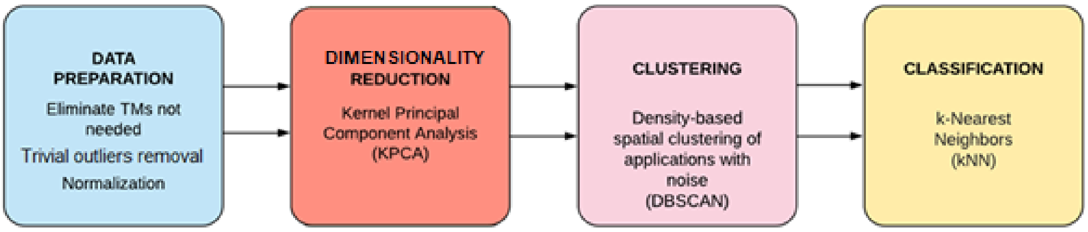

Figure 1. Data-driven anomaly detection flow.

The data extracted is from a TM CSV file and shaped as a pandas data frame which presents raw telemetry value that has to be processed to make it tidy before use (WICKHAM, 2014). Each telemetry channel presents a variable with different engineering magnitudes represented by different measurements units. It is common in these cases to normalize all the data so they will vary within a same range, and still carry their meaning. Before performing a normalization, some steps shall be earlier performed. Trivial outliers in this paper are understood as those that occur very rarely and abruptly, which neighbouring time steps differ from them. As demonstrated in (YAIRI et al., 2017),

The usage of the dimensionality reduction in this study was inspired by the lessons learned seen in (AZEVEDO et al., 2012) and supported by the idea that in machine learning, data with high dimensionality and multimodality, frequently are treated by dimensionality reduction as shown in (YAIRI at al., 2017). This process here is not only performing a feature extraction but also enabling the visualization of the data in three-dimensional space, key arguments for the usage of this technique as elicited in (MOHRI et al., 2018).

The PCA implementation by sci-kit learn used here mainly requires to set only one hyperparameter, the number of principal components (k) that shall be used to generate the new dataset. This was the procedure used in this article and its results can be in Figure 2.

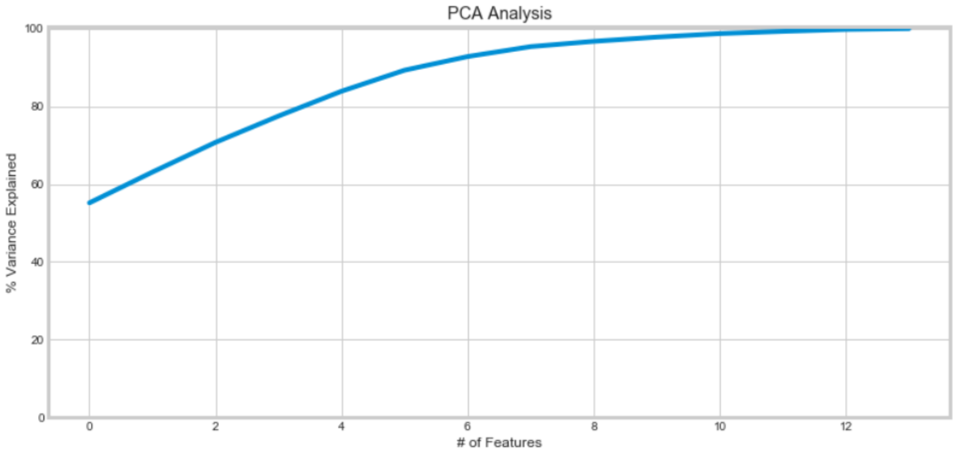

Figure 2. Percentage of variance explained by the number of features (principal components).

The choosing of the KPCA´s number of components, the kernel function, and the gamma hyperparameters was performed in a heuristic way, which can be considered a commonly used approach (ALAM, FUKUMIZU, 2014). However, a method was proposed by (ALAM, FUKUMIZU, 2014) to perform the hyperparameter selection in KPCA more deterministically, it hasn´t experimented here though.

To choose how many components would be used to generate the output data set, a similar approach to the PCA was used. Here, the eigenvalues from the eigenvectors were used as a data variance explaining variable. A plot similar to the PCA was generated and an assessment was made.

As this step of the machine learning process was being used as a data pre-processing step for the clusterization algorithm and this assessment was performed plotting the result of different kernel functions for the same number of components, and different values from gamma. This process tends to take some time as it can be seen as a manual fine-tuning. As an example, Figure 3 show the outcome of choosing the RBF kernel function.

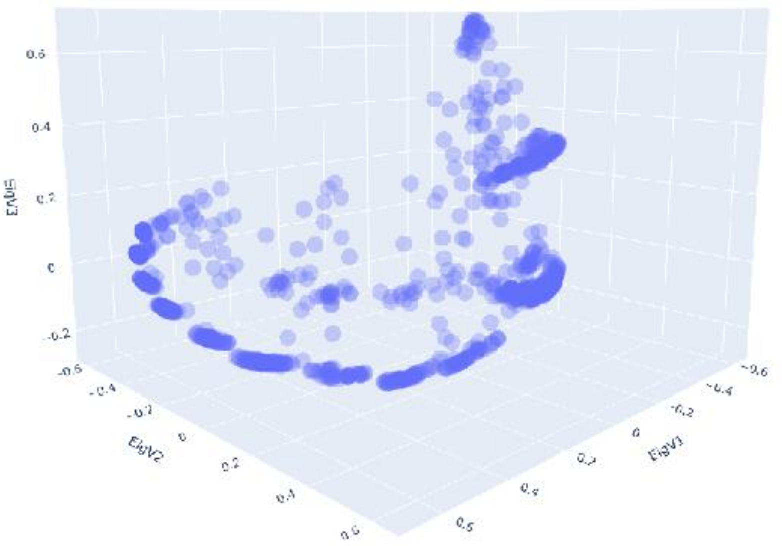

Figure 3. KPCA output for RBF as the kernel function.

The DBSCAN, different from the K-Means, doesn´t require a definition of how many clusters one would like to have but instead, through the hyperparameters epsilon (eps) and minimum of points (minPt), it returns how many clusters it was able to identify using theses hyperparameters and also returns those points which don´t fit any cluster. To find the most adequate values for both hyperparameters the outcome of the DBSCAN is evaluated by its Silhouette Coefficient (SC) value (KAUFMAN; ROUSSEEUW, 1990). The algorithm is executed with the eps varying from 0.1 to 1.0 in steps of 0.1. For each esp step, minPt values from 2 to 8 are tested. This generates a list of different combinations resulting in a different number of found clusters and their given SC which can be evaluated using the rule of thumb shown (KAUFMAN; ROUSSEEUW, 1990). This array then is appended to the original data frame as a column of labels, and the relationship between each telemetry and the found clusters is plotted through a series of histograms and a sample of it can be seen in Figure 4.

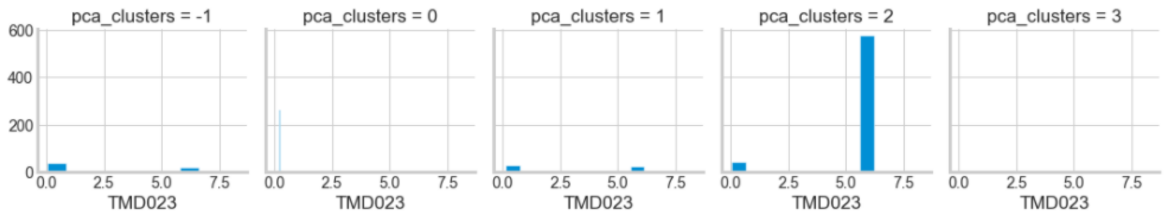

Figure 4. Model accuracy versus the number of K.

From this plot is possible to create the relationship of which cluster represents each circumstance of the satellite. Since TMD0023 represent one of the solar panels output current, it is possible to infer that cluster #2 is related to when the satellite has sun sight, by the other hand, cluster #0 represent when the satellite is in eclipse, while cluster #3 and #1 are related to the twilight situation and the clusters #-1 refers to the anomaly behaviour points.

The labeled data obtained thanks to the clusterization, now can be used to train a supervised classification algorithm as the KNN. This algorithm, besides simples, is easy to use and understand and its implementation on sci-kit learn requires only the number of neighbour's (k) definitions.

The method applied to determine the proper k value consists in training the algorithm and test it with different values within a range from 1 to 40. In the end, using metrics.accuracy function available in sci-kit learns we can find the accuracy of the trained model when comparing the classified data versus the real labels on the data frame. The outcome of this heuristic method is then used for created a plot that represents the accuracy of the model versus the value defined for k. The best accuracy value provided by a value of k bigger than 1 is then chosen. The outcome of this process can be seen in Figure 7.

3. Results and Discussion

3.1. KPCA+DBSCAN+kNN

Even though experiments were conducted with the usage of the PCA algorithm for dimensionality reduction as well, since the final choice was made for the usage of the KPCA algorithm, for the sake of space, the results regarding the PCA algorithm experiments will not be exposed neither discussed. Only the result regarding the chosen Machine learning process will be discussed

From the curve "Percentage of variance explained by the number of components" was obtained the values seen in Table 1.

Table 1. % Variance Explained by the number of principal components (PC).

Number of components	Explained Variance (%)
3	87,137
4	89,674
5	91,777
6	93,649

The usage of three components for the KPCA already produces a very good explained variance value, and as the data will be plot for visualization, chose three components among the other values turns to be the most adequate option

3.1.1. Without trivial outlier removal

The outcome dataset of the KPCA, seen in Figure 5, running a sigmoid kernel, with eigenvectors and gamma hyperparameters of 3 and 0.1 respectively, used as input for a DBSCAN with epsilon equals to 0.2 and a minimum number of samples of 4, provided a silhouette score of 0.537, with 5 clusters identified.

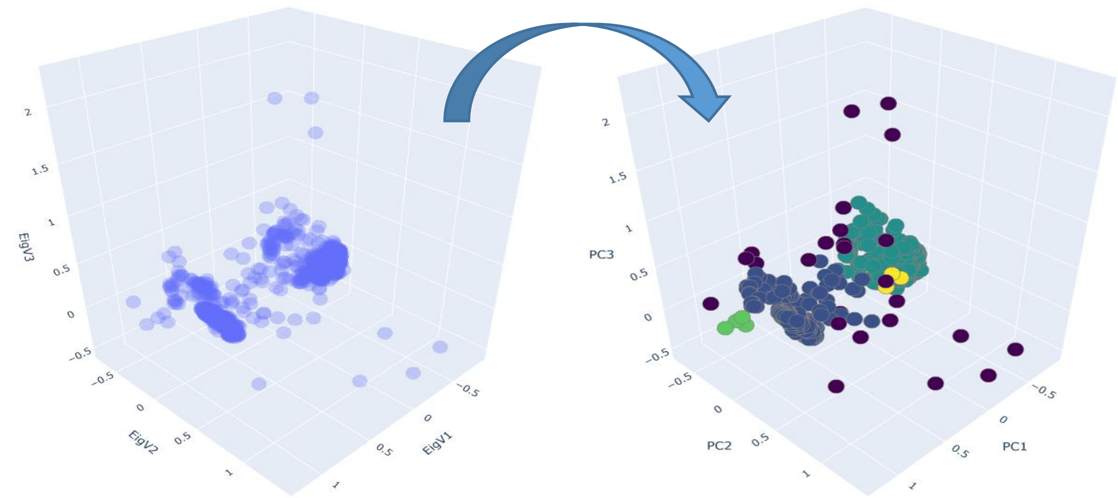

Figure 5. - KPCA output clustered

As seen, the clusters found through that performed tunning don´t meet the requirement of 4 well-defined clusters the represent the satellite's circumstances and one for the outliers/noise. Instead, two big clusters and two tiny clusters were assigned

This result was considered not adequate to proceed to the classification step, instead, as mentioned before, a trivial outlier removal technique was implemented to make this result more adequate.

3.1.2. With trivial outlier removal

The outcome dataset of the KPCA, seen in Figure 6, running a sigmoid kernel, with the values of eigenvector and gamma hyperparameters of 3 and 0.2 respectively, used as input for a DBSCAN with epsilon equals to 0.2 and a minimum number of samples of 7, provided a silhouette score of 0.516, with 5 clusters identified. However, it is important to draw the attention that after the trivial outliers were removed from the initial dataset, a new tunning of the KPCA hyperparameters had to be performed. In this case, the gamma had to be improved, and the hyperparameter "fit_inverse_transform", which learn the inverse transform for non-precomputed kernels

Different from the previous experiment, here, the outcome from both algorithms was quite satisfactory with 4 well define clusters and a cloud of sparse outliers assigned to the -1 cluster. This result is adequate and the indexes of the cluster can be used as labels to train a model with the kNN algorithm.

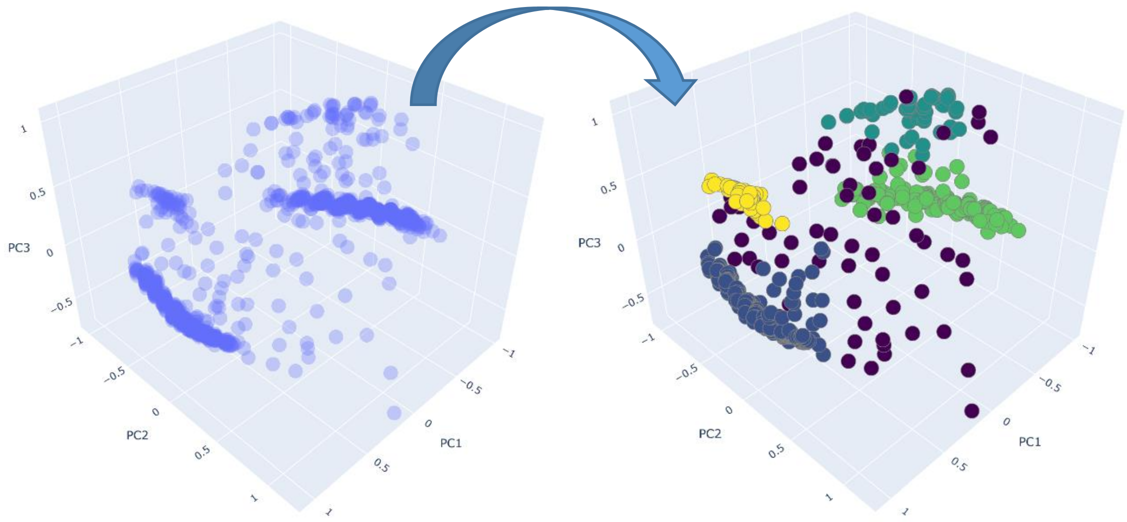

Figure 6. - KPCA output clustered with a trivial outlier cleared dataset

The process to choose the k parameters was performed having the above result as input for the kNN algorithm, and its result can be seen in Figure 7. The best outliers detection accuracy value achieved was 0.6875 using K equals 5. After training the model, an evaluation was run over a test dataset, the outcome of this test can be seen in Table 4

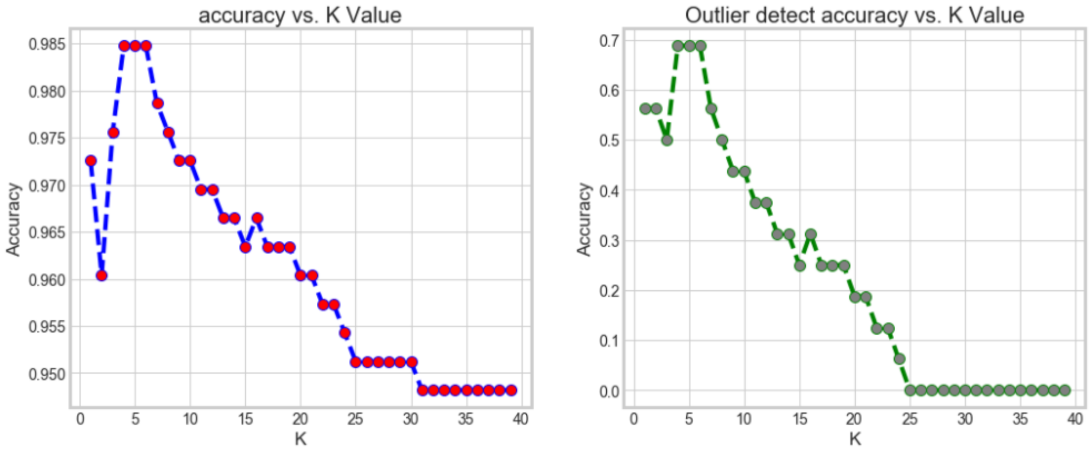

Figure 7. – KNN model accuracy versus number of k

Table 2. Confusion matrix for results obtained from KNN.

	Classified anomalous behavior	Classif. eclipsed	Classif. twilight	Classif. sun sight	Classif. twilight
Anomalous behavior	11	2	2	1	0
Eclipsed	0	135	0	0	0
Twilight	0	0	9	0	0
Sun sight	0	0	0	156	0
Twilight	0	0	0	0	12

The usage of the trivial outliers removal technique allied to the KPCA algorithm presented better result when comes to anomalous behaviour detection. Even though the KPCA output got improved by the removal of trivial outliers, the KPCA dataset generated at the end of the process presented clusters with considering the difference in shape and density, and due to that, the work of the DBSCAN algorithm got difficult. However, we were able to identify clusters with a silhouette score of 0.52. Not the best, but good enough. Eleven anomalies were correctly classified. Once more the general accuracy was a quite satisfactory value, around 0,979.

Among the kernel functions implemented in the KPCA sci-kit learn, the one which provided a more meaningful output was the kernel function sigmoid. Among the distance metrics implemented in the sci-kit learn for the kNN algorithm, two were tested, Euclidian and Manhattan, showing that Euclidian obtained a better accuracy when comes to anomalous behaviour detection in the test model. But for the sake of space, only the Ecludian related results were presented here.

4. Conclusion

The lack of a bigger and continuous dataset until a certain point harmed the classification skills of the model. Especially due to the fact it didn't contain identified anomalous behaviour from the telemetries that could be used to train the model over fails. This way, it is not possible to infer whether the classified anomalies from the test data are real anomalies and led to some problem or not. As mentioned, due to the limited availability of historical fault data, the accuracy of machine learning and other artificial intelligence techniques for fault diagnosis is greatly restricted as the samples required for adequate training of relevant data-driven methods are normally insufficient.

However, the samples used were enough to verify and validate the chosen machine learning procedure over the anomalous behaviour detection and generate the necessary for the next step on the research, the anomaly diagnosis.

The data processing stage where variables were removed depending on their relevance for the model associated with the appliance of a "trivial outliers" removal technique over the data proved to be not just good but also needed, to avoid a model biased by wrong data examples. And even though the clustering algorithm used is noise hardened, as its input data is the outcome from the dimensionality reduction stage, this process was required.

The choice of using the Kernel PCA over regular PCA to reduce the presented better result when comes to not just generate a clusterable data set, but also one which could enable the visualization of the satellite circumstances in orbit. Such aspects compensate that the hyperparameters and kernel function choosing was more time costly than simply choose the numbers of components as it is when using the PCA algorithm.

In the first moment, when the dataset presented clusters in which the size was more similar the DBSCAN showed better results when comes to the silhouette score and also the variability of hyperparameters selection. The difficulty to deal with different size clusters is a known back draw, and it was confirmed here when it generated cluster with a lower silhouette score after running in the dataset coming from the dimensionality reduction stage. Nevertheless, its usage was possible and its outcome was good enough when came to generate the labels for the dataset.

This way, it is possible to affirm that the DBSCAN algorithm served its purpose presenting good results, however, investigate the outcome of the appliance of different clustering algorithms in this context is not discarded, being a possible path to be branched out from this article

The appliance of the kNN algorithm with regards to hyperparameters tuning, proved to be easy and simple at a certain point. The reason behind why its accuracy to detect anomalous behaviour might come from how the algorithm works since it relies on the neighbours of the point under investigation to decide to which labels the given point belongs. And as anomalous behaviour points might not be as much in number as points from other labels, the classification of those anomalous behaviour points tends to have less accuracy than the overall classification. Therefore, the kNN can be considered as a good algorithm for studies purposes in this study case but maybe spent some effort looking for other solutions which could provide better accuracy, the SVM algorithm could be an option as a kickstart.

For now, what can be said about the proposed machine learning procedure is that, considering the proposed framework, the anomalous behaviour detection can be made but to assess whether the anomalous behaviour is a failure or a noise or even an unknown failure, a next step in the process has to be added. For future work we are planning to use a multi-level SVM.

Ackowlegments: *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, Instituto Nacional de Pesquisas Espaciais – INPE.*

References

AZEVEDO, Denise Rotondi et al. Applying Data Mining for Detecting Anomalies in Satellites. In: EUROPEAN DEPENDABLE COMPUTING CONFERENCE, 9., 2012, Sibiu. Proceeding. [S.L.]: Ieee, 2012. p. 212-217.

SCHUTT, R. & O´NEIL, C., Doing Data Science, O'Reilly Media, Inc, 2013.

WICKHAM, Hadley. Tidy Data. Journal Of Statistical Software. [S.L], p. 1-23. set. 2014.

WERTZ, J.R.; EVERETT, D. F.; PUSCHELL, J.J. Space Mission Engineering: The New SMAD, Microcosm Press, 2011.

ZHANG, Liangwei. Big Data Analytics for Fault Detection and its Application in Maintenance. 2016. 148 f. Tese (Doutorado) - Operation And Maintenance Engineering, Luleå University Of Technology, Luleå, 2016.

IBRAHIM, Sara K. et al. Machine Learning Techniques for Satellite Fault Diagnosis. Ain Shams Engineering Journal. [S.L.], p. 45-54. ago. 2019.

HASSANIEN, Aboul Ella; DARWISH, Ashraf; ABDELGHAFAR, ySara. Machine learning in telemetry data mining of space mission: basics, challenging and future directions. Artificial Intelligence Review. [S.L.], p. 3201-3230. set. 2019.

TABUROğLU, Seçil. A Survey on Anomaly Detection and Diagnosis Problem in the Space System Operation. Journal Of Intelligent Systems: Theory and Applications 2. [S.L], p. 13-17. jan. 2019.

GAO, Yu; YANG, Tianshe; XING, Nan; XU, Minqiang. Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines. In: 2012 7TH IEEE CONFERENCE ON INDUSTRIAL ELECTRONICS AND APPLICATIONS (ICIEA), 7., 2012, [S.L.]. 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA). [S.L.]: Ieee, 2012. p. 1983-1988.

JOLLIFFE, I. (2011). Principal Component Analysis. International Encyclopedia of Statistical Science, 1094–1096. doi:10.1007/978-3-642-04898-2_455.

SAFIZADEH, M.s.; LATIFI, S.K.. Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. Information Fusion, [S.L.], v. 18, p. 1-8, jul. 2014. Elsevier BV. http://dx.doi.org/10.1016/j.inffus.2013.10.002.

LI, Wei; PENG, Minjun; LIU, Yongkuo; JIANG, Nan; WANG, Hang; DUAN, Zhiyong. Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method. Annals Of Nuclear Energy, [S.L.], v. 113, p. 105-117, mar. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.anucene.2017.11.009.

YAIRI, Takehisa; TAKEISHI, Naoya; ODA, Tetsuo; NAKAJIMA, Yuta; NISHIMURA, Naoki; TAKATA, Noboru. A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Probabilistic Clustering and Dimensionality Reduction. Ieee Transactions On Aerospace And Electronic Systems, [S.L.], v. 53, n. 3, p. 1384-1401, jun. 2017. Institute of Electrical and Electronics Engineers (IEEE). http://dx.doi.org/10.1109/taes.2017.2671247.

GAO, Yu; YANG, Tianshe; XING, Nan; XU, Minqiang. Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines. 2012 7Th Ieee Conference On Industrial Electronics And Applications (Iciea), [S.L.], v. 7, n. 201150, p. 1984-1988, jul. 2012. IEEE. http://dx.doi.org/10.1109/iciea.2012.6361054.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection. Acm Computing Surveys, [S.L.], v. 41, n. 3, p. 1-58, jul. 2009. Association for Computing Machinery (ACM). http://dx.doi.org/10.1145/1541880.1541882.

Zare, Shokoofeh, "Fault Detection and Diagnosis of Electric Drives Using Intelligent Machine Learning Approaches" (2018). Electronic Theses and Dissertations. 7436. https://scholar.uwindsor.ca/etd/7436

ALAM. HYPERPARAMETER SELECTION IN KERNEL PRINCIPAL COMPONENT ANALYSIS. Journal Of Computer Science, [S.L.], v. 10, n. 7, p. 1139-1150, 1 jul. 2014. Science Publications. http://dx.doi.org/10.3844/jcssp.2014.1139.1150.

KAUFMAN, Leonard; ROUSSEEUW, Peter J.. FINDING GROUPS IN DATA: an introduction to cluster analysis. Brussels: Wiley-Interscience, 1990.

MOHRI, Mehryar; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. FOUNDATIONS OF MACHINE LEARNING. 2. ed. Cambridge: The Mit Press, 2018.

PAN, Dawei; LIU, Datong; ZHOU, Jun; ZHANG, Guoyong. Anomaly detection for satellite power subsystem with associated rules based on Kernel Principal Component Analysis. Microelectronics Reliability, [S.L.], v. 55, n. 9-10, p. 2082-2086, ago. 2015. Elsevier BV. http://dx.doi.org/10.1016/j.microrel.2015.07.010.

LOVRIC, Miodrag. International Encyclopedia of Statistical Science: principal component analysis. Berlin: Springer, 2011.

JOLLIFFE, I. T.. PRINCIPAL COMPONENT ANALYSIS. 2. ed. New York: Springer, 2002.