Unsupervised Learning Algorithms Applied to Anomaly Detection in Oil and Gas Wells

Hugo V. F. Azevedo[1], Lucas O. M. da Silva[2], Lucas G. O. Lopes[1], Eduardo T. de Lima Junior[1]

[1]*LCCV, Technology Center, Federal University of Alagoas*
Campus A. C. Simões, Tabuleiro dos Martins, 57072-900, Maceió, Alagoas, Brasil
hugo.azevedo@lccv.ufal.br, lucasomena@lccv.ufal.br, limajunior@lccv.ufal.br
[2]*Technology Center, Federal University of Alagoas*
Campus A. C. Simões, Tabuleiro dos Martins, 57072-900, Maceió, Alagoas, Brasil
lucas.mendes@ctec.ufal.br

Abstract. Monitoring through sensors is a powerful tool in the evaluation of vibrations, loads, deformations, among other problems in which gathering data allows to detect undesirable events that may arise in structures. Growing opportunities have been observed in companies offering sensing, monitoring, and digital transformation services, which offer cost reduction, increased operational safety and improved performance. Technologies for processing the data collected by sensors using machine learning (ML) methodologies have proven to be efficient tools in engineering processes. In the context of petroleum engineering, the prediction and detection of unexpected events stands out, by supporting decision-making processes and adding value to products and services. Thus, this paper aims to study and develop ML-based models for detecting anomalous states in oil wells, by applying classical techniques such as Support Vector Machines, Isolation Forest and Deep Neural Network. It is expected to compare the efficiency of these methodologies applied to time series datasets of pressure, temperature and flow rate, allowing to predict the anomaly occurrence and generate alerts to the production operator. It is observed the practical application and potential of the proposed methodologies for the intended product, being able to improve the fault detection process in oil wells, as well as ensure their integrity.

Keywords: Machine learning, Deep Learning, LSTM, Autoencoder.

1 Introduction

Oil and gas wells demand constant attention in the production phase to ensure that, in the extraction of the reservoir, the produced fluid reaches the surface efficiently and safely. To ensure that these wells are well maintained, monitoring equipment is installed in the construction phase so that it can be used in the future to track the behavior of the well. This is done with the use of pressure and temperature sensors installed along the well structure. The data provided by these sensors can then be used to verify whether the well is within its normal range or not.

As shown by Vargas et al. [1], the detection and classification of rare undesirable events are relevant tasks in various activities performed and/or monitored by humans. The goal is then to be able to timely identify and respond to these unexpected events, diagnosing the main causes of an anomalous behavior, to then take appropriate decisions and control actions to bring the operation back to a normal, safe, and operational state ([2]).

Technologies for processing the data collected by sensors using machine learning (ML) methodologies have proven to be efficient tools in anomaly detection, as exemplified by the works of Marchi et al. [3] and Malhotra et al. [4]. Therefore, this work aims to test machine learning techniques in the detection of anomalies in oil and gas wells. All models were tested using the Python computational language, the data used were obtained from sensors in wells in the production phase and can be accessed in the public dataset presented by Vargas et al. [1].

2 Methodology

The focus of this work is to test unsupervised learning models so that these models can be used in real-time production in the future. Unsupervised Learning techniques are used when one wants to learn information about the data set from unlabeled data. It is not known in advance what is an anomalous state and a normal state in the well. The focus is to test the entire dataset and try to take that information purely from the analysis done.

The pressure and temperature readings from the sensors are classified as a time series and are samples of time intervals in which the well was operating in a normal state and after some external influence began to operate in an anomalous state. The observations collected by the sensors used as an example in this paper can be seen in Figure 1.

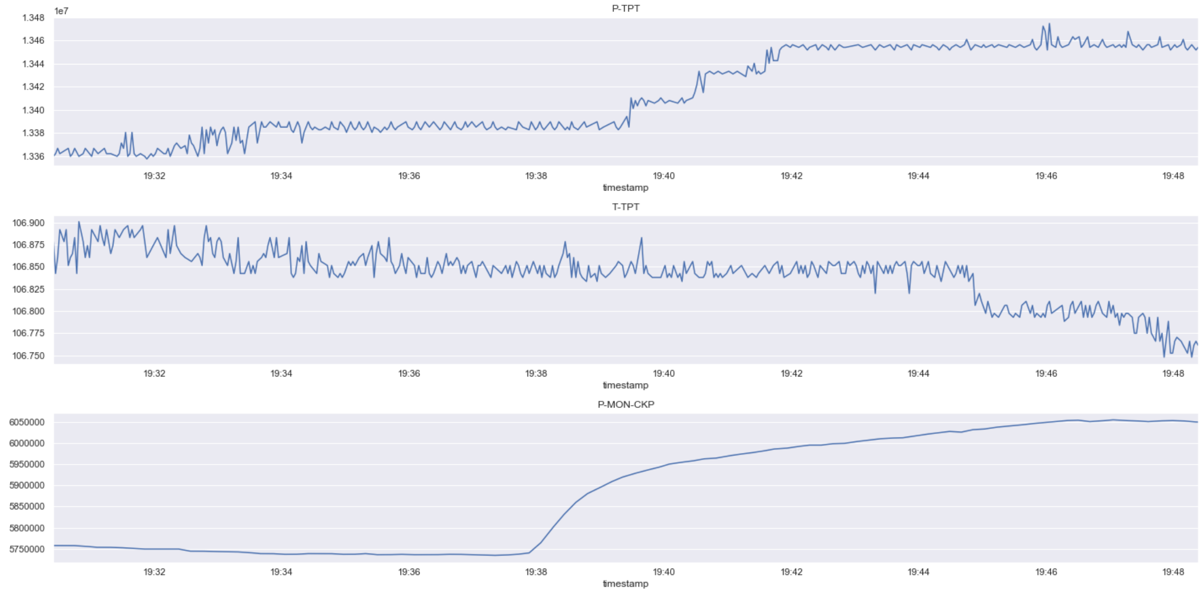

Figure 1. Example used to test the models

Two types of machine learning models were applied, as follows: One-Class Support Vector Machines (OCSVM) and Artificial Neural Networks (ANN) ([5]) of the type Long Short-Term Memory (LSTM) Autoencoder. Such models, when continuously trained, can be used for real-time anomaly detection. The use of OCSVM is already an established technique in the literature, as shown by Ma and Perkins [6] and Shawe-Taylor and Žličar [7].

The use of LSTM networks is justified in order to create more robust recurrent neural networks (RNNs) that avoid vanishing gradients. LSTM networks have a formulation that allows to selectively remember and forget information from the data to be analyzed. Thus a quantity of data to be considered can be determined in order to predict the next piece of data in the sequence. This type of machine learning model is also well known and used in the literature, for example in the works of Vavra and Hromada [8], Ma and Perkins [6] and Shawe-Taylor and Žličar [7], in which the networks were used to detect anomalous behavior in time series data, where the model was trained using data from the system in normal operation.

The autoencoder configuration type, used in the LSTM network, has as its main advantage the possibility of unsupervised training. Its goal is to learn the available information in the input data set in smaller dimensions, neglecting unnecessary information so that it is still possible to recover as much of the original as possible. All this compression of the original information is stored in the latent layer, so that, if necessary, the compressed information can be reconstructed to minimize errors.

3 Results and Discussions

The proposal for anomaly detection using Artificial Intelligence consists of a methodology capable of automatically learning what fits as a normal state and indicate when the data represents an anomalous state. From the input normal data set, the studied models enter the training phase and provide three outputs: 1) data normalization parameters; 2) neural network weights and 3) anomaly tolerance calculated during training. After this generation, when the outputs are saved, the process of testing and validating the proposed model enters, detecting abnormal behavior in new data acquired.

Among these three parameters obtained after network training, the first two are configuration parameters pertinent to the transformation of the input variables and weight coefficients of the neural network. Anomaly tolerance is calculated based on the Mean Absolute Error (MAE) score. Tolerance is based on samples that have an error greater than the mean of the error distribution plus three standard deviations, so that if error values follow a normal distribution, the tolerance value encompasses 99.73% of measured error values.

As a result, it was observed that the response provided by OCSVM was not as satisfactory compared to the responses of the LSTM Autoencoder network. Due to the simplicity of the OCSVM method, the hit percentage was lower than expected, the accuracy achieved was around 70%, consistent with the results presented by Vargas et al. [1]. Moreover, the use of this methodology is not very versatile, due to the fact that the adjustment of the degree of tolerance is linked to a new configuration of the model parameters. That is, whenever the degree of tolerance needs to be changed, the entire model must be changed and retrained to be operational.

LSTM networks, on the other hand, do not present this drawback, as the configuration can simply be automated for each set of data coming from different wells and valve states. Adjusting the network configuration and tolerance is more versatile in detecting anomalies after the training stage. The results obtained with this model were quite satisfactory, since the accuracy was around 90-95%, validating the model in the cases studied by dataset from the literature ([1]). It is therefore chosen to present only the answers provided by the LSTM model, since it was chosen as the best option to be used later in a real-time anomaly detection model. The 2 presents the normal data (in blue) and the data detected as anomalous (in red) of pressure in the CKP valve and the 3 the same classification for the pressure values of the TPT valve.

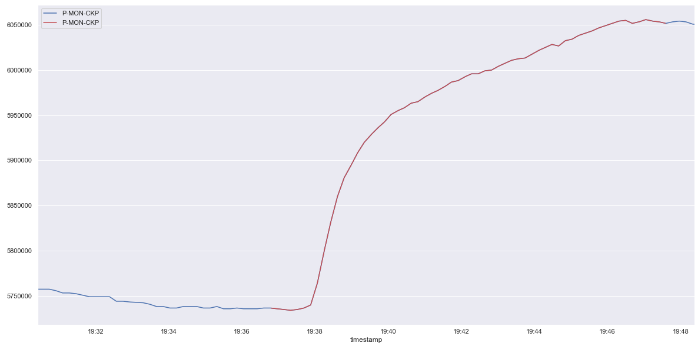

Figure 2. P-MON-CKP anomaly detection using LSTM Autoencoder Network

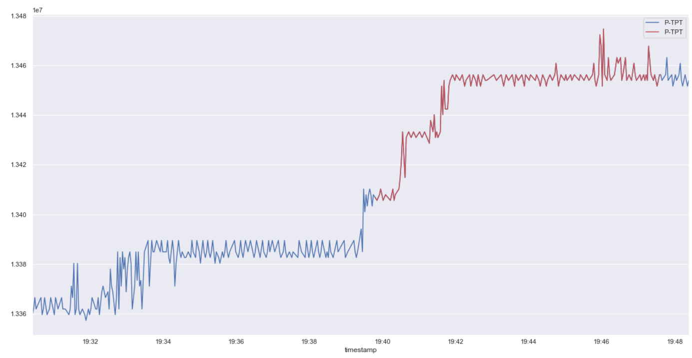

Figure 3. P-TPT anomaly detection using LSTM Autoencoder Network

4 Conclusions

From the results presented, the two methods tested were effective for the objective of detecting anomalies, but with the exception that the LSTM Autoencoder network had a significantly better result than that provided by

OCSVM. This can be explained by the very structure of these two models, the OCSVM is simpler in the way it is used, while the LSTM network is a robust network, more used in the literature and in the industry to detect anomalies in time series.

As future work, it is intended to implement these models in a real-time anomaly detection system, to be used by the industry to improve the monitoring process of oil and gas wells, in order to guarantee a good maintenance throughout the useful life of these wells. The desired system should be able to detect anomalies agilely, so that the operator can perform the well maintenance before causing possible damages and losses.

Acknowledgements. The authors would like to acknowledge Petrobras for technical and financial support.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] R. E. V. Vargas, C. J. Munaro, P. M. Ciarelli, A. G. Medeiros, do B. G. Amaral, D. C. Barrionuevo, de J. C. D. Araújo, J. L. Ribeiro, and L. P. Magalhães. A realistic and public dataset with rare undesirable real events in oil wells. *Journal of Petroleum Science and Engineering*, vol. 181, pp. 106223, 2019.
[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, vol. 27, n. 3, pp. 293–311, 2003.
[3] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller. A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional lstm neural networks. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 1996–2000. IEEE, 2015.
[4] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, volume 89, pp. 89–94, 2015.
[5] S. Haykin. *Redes neurais: princípios e prática*. Bookman Editora, 2007.
[6] J. Ma and S. Perkins. Time-series novelty detection using one-class support vector machines. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 3, pp. 1741–1745. IEEE, 2003.
[7] J. Shawe-Taylor and B. Žličar. Novelty detection with one-class support vector machines. In *Advances in Statistical Models for Data Analysis*, pp. 231–257. Springer, 2015.
[8] J. Vavra and M. Hromada. Optimization of the novelty detection model based on lstm autoencoder for ics environment. In *Proceedings of the Computational Methods in Systems and Software*, pp. 306–319. Springer, 2019.