

ricardovargas /  
3w\_dataset

<> Code
Issues
Pull requests
Discussions
Actions
Security
Insights

The first realistic and public dataset with rare undesirable real events in oil wells.

MIT license

102 stars 54 forks 13 watching 1 Branch 0 Tags Activity

Public repository

master
1 Branch
0 Tags

Go to file
+
Add file
<> Code
⋮

File	Commit message	Time
<b>ricardovargas</b> Freeze this repository <span style="float: right;">2 years ago</span>		
data	Removed folder data/6	5 years ago
docs	Updating master_degree_dissertation_wan...	2 years ago
.gitignore	Initial commit	6 years ago
LICENSE	Initial commit	6 years ago
README.md	Freeze this repository	2 years ago
demo_1_benchmark_impact_of_usin...	Adding experiment for benchmark 1	5 years ago
demo_2_benchmark_anomaly_detec...	Adding experiment for benchmark 1	5 years ago
overview.ipynb	Demonstrations' removal	5 years ago

README
MIT license
✎

! **IMPORTANT!** This repository will remain available online, but without updates from May 31, 2022 onwards. All news about the 3W dataset will be published on <https://github.com/petrobras/3W>, the first Petrobras' repository on GitHub that supports the 3W project.

## 3W Dataset

To the best of its authors' knowledge, this is the first realistic and public dataset with rare undesirable real events in oil wells that can be readily used as a benchmark dataset for development of machine learning techniques related to inherent difficulties of actual data.

For more information about the theory behind this dataset, refer to the paper **A realistic and public dataset with rare undesirable real events in oil wells** published in the **Journal of Petroleum Science and Engineering** (link [here](#)). Specific challenges (benchmark that practitioners and researchers can use together with the 3W dataset are defined and proposed in this paper.

## Preamble

This is the supporting repository for the paper mentioned above. If you use the 3W dataset or the proposed benchmarks, please cite the paper and the 3W dataset:

```

@article{VARGAS2019106223,
  title = "A realistic and public dataset with rare undesirable real events in oil wells",
  journal = "Journal of Petroleum Science and Engineering",
  volume = "181",
  pages = "106223",
  year = "2019",
  issn = "0920-4105",
  doi = "https://doi.org/10.1016/j.petrol.2019.106223",
  url = "http://www.sciencedirect.com/science/article/pii/S0920410519306357",
  author = "Ricardo Emanuel Vaz Vargas and Celso José Munaro and Patrick Marques Ciarelli and André Gonçalves Medeiros and Bruno Guberfain do Amaral and Daniel Centurion Barrionuevo and Jean Carlos Dias de Araújo and Jorge Lins Ribeiro and Lucas Pierezan Magalhães",
}

```

```
keywords = "Fault detection and diagnosis, Oil well monitoring, Abnormal event management, Multivariate time series classification",
abstract = "Detection of undesirable events in oil and gas wells can help prevent production losses, environmental accidents, and human casualties and reduce maintenance costs. The scarcity of measurements in such processes is a drawback due to the low reliability of instrumentation in such hostile environments. Another issue is the absence of adequately structured data related to events that should be detected. To contribute to providing a priori knowledge about undesirable events for diagnostic algorithms in offshore naturally flowing wells, this work presents an original and valuable dataset with instances of eight types of undesirable events characterized by eight process variables. Many hours of expert work were required to validate historical instances and to produce simulated and hand-drawn instances that can be useful to distinguish normal and abnormal actual events under different operating conditions. The choices made during this dataset's preparation are described and justified, and specific benchmarks that practitioners and researchers can use together with the published dataset are defined. This work has resulted in two relevant contributions. A challenging public dataset that can be used as a benchmark for the development of (i) machine learning techniques related to inherent difficulties of actual data, and (ii) methods for specific tasks associated with detecting and diagnosing undesirable events in offshore naturally flowing oil and gas wells. The other contribution is the proposal of the defined benchmarks."
}
```

Vargas, Ricardo; Munaro, Celso; Ciarelli, Patrick; Medeiros, André; Amaral, Bruno; Barrionuevo, Daniel; Araújo, Jean; Ribeiro, Jorge; Magalhães, Lucas (2019), "Data for: A Realistic and Public Dataset with Rare Undesirable Real Events in Oil Wells", Mendeley Data, v1. <http://dx.doi.org/10.17632/r7774rwc7v.1>



If you have questions or want to contribute with this work, please drop me an email at ricardovargas at gmail dot com.

## 3W Dataset's Structure

The 3W dataset consists of 1,984 CSV files structured as follows. Due to the limitation of GitHub, this dataset is kept in 7z files splitted automatically and saved in the `data` directory. Before using 3W dataset, they must be decompressed. After that, the subdirectory names are the instances' labels. Each file represents one instance. The filename reveals its source. All files are standardized as follow. There are one observation per line and one series per column. Columns are separated by commas and decimals are separated by periods. The first column contains timestamps, the last one reveals the observations' labels, and the other columns are the Multivariate Time Series (MTS) (i.e. the instance itself).

## 3W Dataset's Overview

A 3W Dataset's general presentation with some quantities and statistics is available [here](#).

## Experiments for Proposed Benchmarks

The results of the following experiments can be used as baselines.

- Bechmark 1: Impact of Using Simulated and Hand-Drawn Instances (code and results [here](#)).
- Bechmark 2: Anomaly Detection (code and results [here](#)).

## Citations

If you know any other work that cites the 3W dataset, I will be grateful if you let me know by email (ricardovargas at gmail dot com).

1. R.E.V. Vargas, C.J. Munaro, P.M. Ciarelli. A methodology for generating datasets for development of anomaly detectors in oil wells based on Artificial Intelligence techniques. I Congresso Brasileiro em Engenharia de Sistemas em Processos. 2019. [https://www.ufrgs.br/psebr/wp-content/uploads/2019/04/Abstract\\_A019\\_Vargas.pdf](https://www.ufrgs.br/psebr/wp-content/uploads/2019/04/Abstract_A019_Vargas.pdf).
2. R.E.V. Vargas. Base de dados e benchmarks para prognóstico de anomalias em sistemas de elevação de petróleo. Universidade Federal do Espírito Santo. Doctoral thesis. 2019. [https://github.com/ricardovargas/3w\\_dataset/raw/master/docs/doctoral\\_thesis\\_ricardo\\_vargas.pdf](https://github.com/ricardovargas/3w_dataset/raw/master/docs/doctoral_thesis_ricardo_vargas.pdf).
3. Yan Li, Tingjian Ge, Cindy Chen. Data Stream Event Prediction Based on Timing Knowledge and State Transitions. PVLDB, 13(10): 1779-1792. 2020. <http://www.vldb.org/pvldb/vol13/p1779-li.pdf>.
4. Tao Lu, Wen Xia, Xiangyu Zou, Qianbin Xia. Adaptively Compressing IoT Data on the Resource-constrained Edge. 3rd {USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 20). 2020. [https://www.usenix.org/system/files/hotedge20\\_paper\\_lu.pdf](https://www.usenix.org/system/files/hotedge20_paper_lu.pdf).
5. Matheus A. Marins, Bettina D. Barros, Ismael H. Santos, Daniel C. Barrionuevo, Ricardo E.V. Vargas, Thiago de M. Prego, Amaro A. de Lima, Marcello L.R. de Campos, Eduardo A.B. da Silva, Sergio L. Netto. Fault detection and classification in oil wells and production/service lines using random forest. Journal of Petroleum Science and Engineering. 2020. <https://doi.org/10.1016/j.petrol.2020.107879>.

6. W. Fernandes Junior, R.E.V. Vargas, K.S. Komati, K.A. de Souza Gazolli. Detecção de anomalias em poços produtores de petróleo usando aprendizado de máquina. XXIII Congresso Brasileiro de Automática. 2020. [https://www.sba.org.br/open\\_journal\\_systems/index.php/cba/article/download/1405/1005](https://www.sba.org.br/open_journal_systems/index.php/cba/article/download/1405/1005).
7. Jianguo Liu, Jianli Gu, Huishu Li, Kenneth H. Carlson. Machine learning and transport simulations for groundwater anomaly detection, Journal of Computational and Applied Mathematics. 2020. <https://doi.org/10.1016/j.cam.2020.112982>.
8. Eduardo S.P. Sobrinho, Felipe L. Oliveira, Jorel L.R. Anjos, Clemente Gonçalves, Marcus V.D. Ferreira, Lucas G.O. Lopes, William W.A. Lira, João P.N. Araújo, Thiago B. Silva, Lucas P. Gouveia. Uma ferramenta para detectar anomalias de produção utilizando aprendizagem profunda e árvore de decisão. Rio Oil & Gas Expo and Conference 2020. 2020. [https://icongresso.ibp.itarget.com.br/arquivos/trabalhos\\_completos/ibp/3/final.IBP0938\\_20\\_27112020\\_085551.pdf](https://icongresso.ibp.itarget.com.br/arquivos/trabalhos_completos/ibp/3/final.IBP0938_20_27112020_085551.pdf).
9. I.M.N. Oliveira. Técnicas de inferência e previsão de dados como suporte à análise de integridade de revestimentos. Universidade Federal de Alagoas. Master's degree dissertation. 2020. [https://github.com/ricardovargas/3w\\_dataset/raw/master/docs/master\\_degree\\_dissertation\\_igor\\_oliveira.pdf](https://github.com/ricardovargas/3w_dataset/raw/master/docs/master_degree_dissertation_igor_oliveira.pdf).
10. Luiz Müller, Marcelo Ramos Martins. Proposition of Reliability-based Methodology for Well Integrity Management During Operational Phase. 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference. 2020. [https://doi.org/10.3850%2F978-981-14-8593-0\\_3682-cd](https://doi.org/10.3850%2F978-981-14-8593-0_3682-cd).
11. R.S.F. Nascimento, B.H.G. Barbosa, R.E.V. Vargas, I.H.F. Santos. Detecção de falhas com Stacked Autoencoders e técnicas de reconhecimento de padrões em poços de petróleo operados por gas lift. XXIII Congresso Brasileiro de Automática. 2020. [https://www.sba.org.br/open\\_journal\\_systems/index.php/cba/article/view/1462/1300](https://www.sba.org.br/open_journal_systems/index.php/cba/article/view/1462/1300).
12. R.S.F. Nascimento, B.H.G. Barbosa, R.E.V. Vargas, I.H.F. Santos. Detecção de anomalias em poços de petróleo surgentes com Stacked Autoencoders. Simpósio Brasileiro de Automação Inteligente. 2021.
13. R.S.F. Nascimento, B.H.G. Barbosa, R.E.V. Vargas, I.H.F. Santos. Fault detection with Stacked Autoencoders and pattern recognition techniques in gas lift operated oil wells. CILAMCE-PANACM. 2021.
14. R.S.F. Nascimento. Detecção de anomalias em poços de produção de petróleo offshore com a utilização de autoencoders e técnicas de reconhecimento de padrões. Universidade Federal de Lavras. Master's degree dissertation. 2021. [https://github.com/ricardovargas/3w\\_dataset/raw/master/docs/master\\_degree\\_dissertation\\_rodrigo\\_nascimento.pdf](https://github.com/ricardovargas/3w_dataset/raw/master/docs/master_degree_dissertation_rodrigo_nascimento.pdf).
15. Taimur Hafeez, Lina Xu, Gavin Mcardle. Edge Intelligence for Data Handling and Predictive Maintenance in IIOT. IEEE Access. 2021. <https://ieeexplore.ieee.org/document/9387301>.
16. Aurea Soriano-Vargas, Rafael Werneck, Renato Moura, Pedro Mendes Júnior, Raphael Prates, Manuel Castro, Maiara Gonçalves, Manzur Hossain, Marcelo Zampieri, Alexandre Ferreira, Alessandra Davólio, Bernd Hamann, Denis José Schiozer, Anderson Rocha. A visual analytics approach to anomaly detection in hydrocarbon reservoir time series data. Journal of Petroleum Science and Engineering. 2021. <https://doi.org/10.1016/j.petrol.2021.108988>.
17. Yan Li, Tingjian Ge. Imminence Monitoring of Critical Events: A Representation Learning Approach. International Conference on Management of Data. 2021. <https://doi.org/10.1145/3448016.3452804>.
18. B.G. Carvalho, R.E.V. Vargas, R.M. Salgado, C.J. Munaro, F.M. Varejão. Flow Instability Detection in Offshore Oil Wells with Multivariate Time Series Machine Learning Classifiers. 30th International Symposium on Industrial Electronics. 2021. <https://doi.org/10.1109/ISIE45552.2021.9576310>.
19. B.G. Carvalho, R.E.V. Vargas, R.M. Salgado, C.J. Munaro, F.M. Varejão. Hyperparameter Tuning and Feature Selection for Improving Flow Instability Detection in Offshore Oil Wells. IEEE 19th International Conference on Industrial Informatics (INDIN). 2021. <https://doi.org/10.1109/INDIN45523.2021.9557415>.
20. B.G. Carvalho. Evaluating machine learning techniques for detection of flow instability events in offshore oil wells. Universidade Federal do Espírito Santo. Master's degree dissertation. 2021. [https://github.com/ricardovargas/3w\\_dataset/raw/master/docs/master\\_degree\\_dissertation\\_bruno\\_carvalho.pdf](https://github.com/ricardovargas/3w_dataset/raw/master/docs/master_degree_dissertation_bruno_carvalho.pdf).
21. E. M. Turan, J. Jäschke. Classification of undesirable events in oil well operation. 23rd International Conference on Process Control (PC). 2021. <https://doi.org/10.1109/PC52310.2021.9447527>.
22. I.S. Figueirêdo, T.F. Carvalho, W.J.D Silva, L.L.N. Guarieiro, E.G.S. Nascimento. Detecting Interesting and Anomalous Patterns In Multivariate Time-Series Data in an Offshore Platform Using Unsupervised Learning. OTC Offshore Technology Conference. 2021. <https://doi.org/10.4043/31297-MS>.
23. R. Karl, J. Takeshita, T. Jung. Cryptonite: A Framework for Flexible Time-Series Secure Aggregation with Non-interactive Fault Recovery. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST. 2021. <https://eprint.iacr.org/2020/1561.pdf>.
24. A.O. de Salvo Castro, M. de Jesus Rocha Santos, F.R. Leta, C.B.C. Lima, G.B.A. Lima. Unsupervised Methods to Classify Real Data from Offshore Wells. American Journal of Operations Research. 2021. <https://doi.org/10.4236/ajor.2021.115014>.

25. W. Fernandes Junior. Comparação de classificadores para detecção de anomalias em poços produtores de petróleo. Instituto Federal do Espírito Santo. Master's degree dissertation. 2022. [https://github.com/ricardovargas/3w\\_dataset/raw/master/docs/master\\_degree\\_dissertation\\_wander\\_junior.pdf](https://github.com/ricardovargas/3w_dataset/raw/master/docs/master_degree_dissertation_wander_junior.pdf).
26. E.G.S. Nascimento, I.S. Figueirêdo, L.L.N. Guarieiro. A Novel Self Deep Learning Semi-Supervised Approach to Classify Unlabeled Multivariate Time Series Data. GPU Technology Conference Digital Spring. 2022. <https://www.nvidia.com/en-us/on-demand/session/gtcspring22-s41405>.
27. A.L. D'Almeida, N.C.R. Bergiante, G.S. Ferreira, F.R. Leta, C.B.C. Lima, G.B.A. Lima. Digital transformation: a review on artificial intelligence techniques in drilling and production applications. The International Journal of Advanced Manufacturing Technology. 2022. <https://doi.org/10.1007/s00170-021-08631-w>.

---

## Releases

No releases published

---

## Packages

No packages published

---

## Contributors 2



**ricardovargas** Ricardo Emanuel Vaz Vargas



**ricardoevargas** Ricardo Emanuel Vaz Vargas

---

## Languages

---

● **Jupyter Notebook** 100.0%