

[Open in app](#)[Get started](#)Published in Towards Data Science · [Follow](#)Guansong Pang · [Follow](#)

Oct 30, 2020 · 2 min read

# ADRepository: Anomaly Detection Datasets with Real Anomalies

## A continuously updated anomaly dataset collection

I have created a Github repository to provide a continuously updated collection of popular real-world datasets used for anomaly detection in the literature. Please find the link below to the repository.

### **GuansongPang/anomaly-detection-datasets**

In this repository, we provide a continuously updated collection of popular real-world datasets used for anomaly...

[github.com](#)

One main obstacle to the development of anomaly detection is the lack of real-world datasets with real anomalies. Although there are a number of relevant publicly available datasets at [UCI machine learning repository](#) and/or [Libsvm datasets](#), we may often need to devote a large amount of time to make the publicly available datasets ready for our anomaly detection models. This repository is created to tackle these issues. It will focus on ready-to-use datasets, i.e., without requiring any further data preprocessing, and continuously update it with datasets pre-processed by myself or other researchers.

The datasets that are now available in the repository are from diverse domains, such as credit card fraud detection, intrusion detection, disease detection, etc. Basic statistics of these datasets can be found below.



[Open in app](#)[Get started](#)

donors	619,326	10
census	299,285	500
fraud	284,807	29
celeba	202,599	39
backdoor	95,329	196
campaign	41,188	62
thyroid	7,200	21

Datasets used in Pang, G., Shen, C., & van den Hengel, A. (2019, July). Deep anomaly detection with deviation networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 353–362). See detailed introduction of these datasets in this paper and the source links therein.



[Open in app](#)[Get started](#)

bank	41,188	10	yes
census	299,285	33	50K+
AID362	4,279	114	active
w7a	49,749	300	yes
CMC	1,473	8	child>10
APAS	12,695	64	train
CelebA	202,599	39	bald
Chess	28,056	6	zero
AD	3,279	1,555	ad.
Solar-flare	1,066	11	F
Probe	64,759	6	attack
U2R	60,821	6	attack
R10	12,897	100	corn

## Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

[Get this newsletter](#)

### Deep Learning for Anomaly Detection: A Review

Anomaly detection, a.k.a. outlier detection, has been a lasting yet active research area in various research...

