This table was created to provide freely available data sets for employees in production environment who aim to acquire first experiences and know-how in hands-on application of Machine Learning (ML).
In addition, the table contains data sets from multiple fields, providing a wide range of use-cases. There are categories (A-G) to assign data sets into ML- and AI-application areas.

A - Process Design
B - Optimization of Routing and Scheduling
C - Predictive Process Control
D - Self Learning Machines and Assets
E - Anomaly Detection
F - Predictive Maintenance
G - Product Design

Use-Case	Description	Publishing Date	Learning Task	Number of Instances	Number of Attributes	Instances in Minor Class	A	B	C	D	E	F	G	Link
3D Printer	The aim of the study is to determine how much of the adjustment parameters in 3D printers affect the print quality, accuracy, and strength. There are nine setting parameters and three measured output parameters.	22.09.2018	Regression	50	12	25	◒	○	●	○	○	○	○	Kaggle / 3D-Printer
Mercedes-Benz Greener Manufacturing	In this competition, Daimler challenged Kagglers to tackle the curse of dimensionality and reduce the time that cars spend on the test bench. This data set contains an anonymized set of variables, each representing a custom feature in a Mercedes car.	2016	Regression	4,210	378	23	◒	◒	◒	○	◒	○	○	Kaggle / Mercedes-benz-manufact.
APS Failure at Scania Trucks	This set contains data from heavy Scania trucks in daily usage. The system in focus is the Air Pressure system (APS), which generates pressurized air used in various functions, such as braking and gear shifting.	01.02.2018	Classification	60,000	171	1,000	◒	○	●	○	◒	●	○	UCI / Scania-trucks
SECOM	The data was collected from a semiconductor manufacturing process. It represents a selection of features, in which each example represents a single production entity with associated measured features.	19.11.2008	Classification	1,567	591	104	◒	○	●	○	●	○	○	UCI / SECOM
Cylinder Bands	Process delays known as cylinder banding in rotogravure printing were substantially mitigated using control rules discovered by decision tree induction. ML shows to be promising for knowledge acquisition.	01.08.1995	Classification	512	40	200	◒	○	●	○	○	○	○	UCI / Cylinder-Bands
Bosch Production Line Performance	The data for this competition represents measurements of parts as they move through Bosch's production lines. Each part has a unique ID. The goal is to predict which parts will fail in quality control.	2016	Classification	1,183,747	2		◒	○	●	○	◒	○	○	Kaggle / Bosch-production-line
Quality Prediction Mining Process	Data from a mining plant. The goal is to predict how much impurity is in the ore concentrate that is measured every hour.	2017	Regression	734,000	24	12,269	○	○	●	○	○	○	○	Kaggle / Mining-Process
Energy Optimization	This data was collected from a demonstrator of a high storage system, which transports one package between two spots. The high storage system consists of 4 short conveyor belts and 2 rails.	01.07.2018	Classification, Regression	4 files; à 20,000	20	10,200	○	○	●	○	●	○	○	Kaggle / Energy-Optimizatio n
Production Plant Data for Condition	Data for 8 run-to-failure experiments were provided and 8 features related to the component were selected. Training and prediction data were selected using the leave-one-out method: data under test were selected as the target for the prediction.	01.09.2018	ClassificationRegression	8 files; à 20,000 inst.	26	15,800	○	○	●	○	◒	●	○	Kaggle / Monitoring
CNC Mill Tool Wear	Machining data was collected from a CNC machine for variations of tool condition, feed rate, and clamping pressure.	01.04.2018	Classification	18 files; à 500 inst.	48	2,304	○	○	●	○	●	◒	○	Kaggle / CNC-mill
Bolts	Data from an experiment, which analyzes the effects of machine adjustments on the time to count bolts. Bolts are dumped into a large metal dish. A plate that forms the bottom of the dish rotates counterclockwise.	04.10.2014	Classification, Regression	40	8	14	○	○	●	○	◒	○	○	OpenML / 857
Milling	The data was collected from experiments on a milling machine for different speeds, feeds, and depth of cut. Additionally, data from the wear of the milling process is acquired.	2007	Regression	167	13	59	○	○	◒	○	●	◒	○	Nasa / Prognostic-datarepository
Li-ion Battery Aging	This data set has been collected from a custom built battery prognostics testbed. The aim is to be able to manage this uncertainty of actual usage and make reliable predictions of Remaining Useful Life.	01.10.2008	Regression	2,167	12	636	○	○	●	○	○	●	○	Nasa / Resources133
Airfoil Self-Noise	The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The goal is to predict sound pressure levels.	04.03.2014	Regression	1,503	6	36	○	○	○	○	○	○	●	UCI / Airfoil-self-noise
CFRP Composites	Run-to-failure experiments were run on CFRP panels with periodic measurements to capture internal damage growth under tension-tension fatigue.	2008	Classification	3 files; à 4 Layouts; à 150 inst.	7	316	○	○	○	○	○	○	●	Nasa / Prognostic-data-repository
Mechanical Analysis	Fault diagnosis problems of electromechanical devices. Each instance contains many components, each one has eight attributes. Different instances in this database have different numbers of components.	01.06.1990	Classification	209	8	5	○	○	◒	○	●	○	○	UCI / Mechanical-Analysis
Versatile Production	Data from Versatile Production System (VPS) for a wide variety of tasks, including model learning, anomaly detection, and alarm management.	01.09.2018	Classification	8 files; à 10,000	6	65	○	○	◒	○	●	○	○	Kaggle / Versatile-Production
Steel Plates Faults	A data set of steel plates faults, classified into seven different types. The goal was to train machine learning for automatic pattern recognition.	01.11.2017	Classification	1,941	34	55	○	○	◒	○	●	○	○	Kaggle / Steel-plates
Bearing	Four bearings were installed on a shaft. The rotation speed was kept constant at 2,000 RPM by an AC motor coupled to the shaft via rub belts. Three data sets are included in the data packet. Each data set describes a test-to-failure experiment.	2007	Regression	3 files; à 2,156 / 984 / 4,448 inst.	8 / 4 / 4	984	○	○	◒	○	◒	○	○	Nasa.gov / Prognostic-data-repository
Plant Fault Detection	PHM Data Challenge 2015: Fault detection and prognostics, a common problem in industrial plant monitoring. The final aim is the ability to detect plant faults.	05.06.2015	Regression	70 files; à 127,691 inst.	10	700	○	○	◒	○	●	◒	○	PHM / Competition 15
Robot Execution Failures	This data set contains force and torque measurements on a robot after failure detection. All features are numeric although they are integer valued only.	23.04.1999	Classification	5 files; à 88 / 47 / 47 / 117 / 164 inst.	90	3	○	○	○	◒	●	◒	○	UCI / Robot-Execution
Turbofan Engine Degradation Simulation	The data was extracted from an engine, which is operating normally at the start of each time series until a fault occurs. The objective of the competition is to predict the number of remaining operational cycles before failure.	22.09.2010	Regression	4 files; à 20,000 inst.	26	76	○	○	○	○	●	◒	○	Nasa / Resources139
Gearbox Fault Detection	PHM Data Challenge 2009: Fault detection and magnitude estimation for a generic gearbox using accelerometer data and information about bearing geometry.	02.11.2017	Regression	560 files; à 133,000 inst.	3	65,000	○	○	○	○	●	◒	○	Nasa / Resources997
Anemometer Fault Detection	PHM Data Challenge 2011: Anemometer fault detection, a critical problem for the wind power industry, strongly affecting among other things the financing of a potential site.	03.05.2011	Regression	420 files; à 720 inst.	16	63,000	○	○	○	○	●	○	○	PHM / Competition11
Maintenance Action Recommendatio	PHM Data Challenge 2013: Maintenance action recommendation, which is a common problem in industrial remote monitoring and diagnostics.	2013	Regression	1,200,000	32	10,461	○	○	○	○	●	○	○	PHM / Competition13
Asset Health Condition	PHM Data Challenge 2014: Asset health calculation that is a common problem in industrial remote monitoring and diagnostics.	05.10.2014	Regression	270,831	4	9,200	○	○	○	○	●	◒	○	PHM / Competition 14
Genesis Demonstrator	The Genesis Demonstrator is a portable pick-and-place demonstrator, which uses an air tank to supply gripping and storage units. The data from the whole process is acquired.	01.07.2018	Regression	5 files; à (3x) 7,500 inst. (2x) 16,000 inst.	24	424	○	○	○	○	●	○	○	Kaggle / Genesis-Demonstrator
Maintenance of Naval Propulsion Plants	Data has been generated from a sophisticated simulator of Gas Turbines (GT), mounted on a Frigate characterized by a Combined Diesel Electric and Gas (CODLAG) propulsion plant.	11.09.2014	Regression	11,934	18	460	○	○	○	○	○	●	○	UCI / Naval-Plants
Azure Blob	Each machine includes a device, which stores data such as warnings, problems and errors generated by the machine over time.	13.06.2017	Classification	2,000,000	172	159,150	○	○	○	○	○	●	○	github / Azure-Predictive-Maintenance
Predictive Maintenance	The data set is in kind of time series, consisting of the log message and failure records of 984 days. The goal is to predict machine failure in advance.	01.09.2018	Classification, Regression	984	2	98	○	○	○	○	○	●	○	DeepLearning / Predictive-Maintenance
Aircraft Engine	The engine is operating normally at the start of each time series, and starts to degrade at some point during the series.	2008	Classification, Regression	3 files; à 45,000 inst.	26	105	○	○	○	○	○	●	○	Nasa / PHM 08 Aircraft Engine
Semiconductor CMP	PHM Data Challenge 2016: the challenge is focused on tracking the health state of components within a wafer chemical-mechanical planarization (polishing) system.	2016	Regression	2 folders; à 184 files; à 1,300 inst.	26	815	○	○	○	○	○	●	○	PHM / Semiconductor CMP
Condition Monitoring of Hydraulic	The data set addresses the condition assessment of a hydraulic test rig based on multi-sensor data. Four fault types are superimposed with several severity grades impeding selective quantification.	2018	Classification, Regression	2205	43680	756	○	○	○	○	○	●	○	UCI / Hydraulic Monitoring