AVAILABLE PROGNOSTIC DATASETS

Prepared by:

EL-Amrany Samir

AM2I

INS2I
INSIS

UNIVERSITÉ DE LORRAINE

Institut de Cancérologie de Lorraine
Alexis Vautrin
Ensemble, construisons l'avenir

PHM08

- The PHM08 dataset was provided by The NASA Prognostics Center of Excellence (PCoE).

- The PHM08 dataset is similar to (the Turbofan Engine Degradation Simulation data set known as CMAPSS) except the true Remaining Useful Life (RUL) values are not revealed. Users are expected to develop their algorithms using training and test sets provided in the package.

- The data is from the data challenge competition held at the 1st international conference on Prognostics and Health Management (PHM08)

ISET

Number of features	Number of training trajectories	Number of testing trajectories	Min length of training trajectories	Max length of training trajectories	Mean length of training trajectories
24	218	218	128	357	211

C **Data Set Citation**: A. Saxena and K. Goebel (2008). "PHM08 Challenge Data Set", NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA

ISET

MS AZURE PREDICTIVE MAINTENANCE DATA SET

- The AZURE PREDICTIVE MAINTENACE dataset was provided by a Kaggle user in the Kaggle plateform

- This dataset was available as a part of Azure AI Notebooks for Predictive Maintenance. But as of 15th Oct, 2020 the notebook (link) is no longer available..

- It consists of hourly average of voltage, rotation, pressure, and vibration collected from 100 machines for the year 2015.

ISET

Number of features	Number of tablets	Number of run-to-failure records	Min length of run-to-failure records	Max length of run-to-failure records	Mean length of run-to-failure records
20	100	761	96	8761	1113

○ **Data Set Link**:
https://www.kaggle.com/datasets/arnabbiswas1/microsoft-azure-predictive-maintenance

ISET

PREVENTIVE TO PREDICTIVE MAINTENANCE DATA

- This dataset was provided by a member of Essilingen University in the Kaggle platform.

- The degradation process that was observed in this dataset is related to the clogging of filters during the separation of solid particles from gas.

- the dataset contains training and test data, consisting both of 50 life tests respectively.

- We can use the test data in order to estimate the remaining useful life because this data contains randomly right-censored run-to-failure measurements and the respective RUL as ground truth to the prediction task.

ISET

Number of features	Number of run-to-failure records	Min length of run-to-failure records	Max length of run-to-failure records	Mean length of run-to-failure records
4	50	150	2581	788

○ **Data Set Citation**: Hagmeyer, S., Mauthe, F., & Zeiler, P. (2021). Creation of Publicly Available Data Sets for Prognostics and Diagnostics Addressing Data Scenarios Relevant to Industrial Applications. International Journal of Prognostics and Health Management, Volume 12, Issue 2, DOI: 10.36001/ijphm.2021.v12i2.3087.

ISET

WATER PUMP DATA

C This dataset was provided by a Kaggle user in the Kaggle platform.

C This is real scenario data which includes 7 system outages that occurred. These outages have caused huge problems for many people and have also caused serious life issues for some families.

C The data is from all available sensors and are all raw values. The total number of sensors is 52 units and contains the remaining useful life up to the next failure..

ISET

Number of features	Number of pumps	Min length of run-to-failure records	Max length of run-to-failure records	Mean length of run-to-failure records
52	7	7354	50249	24880

Data Set GitHub Link :
https://www.kaggle.com/datasets/anseldsouza/water-pump-rul-predictive-maintenance

ISET

LI-ION-BATTERY NASA DATASET

- The Li-ion Battery dataset was provided by The NASA Prognostics Center of Excellence (PCoE).

- It is a publicly available dataset containing experimental data on Li-Ion batteries, where the charging and discharging were conducted at different temperatures, and the impedance was recorded as the damage criterion.

- This dataset is often used as a benchmark for developing prognostics and Health Management Algorithms on Batteries.

ISET

Number of features	Number of Batteries	Min length of discharging cycles	Max length of discharging cycles	The mean length of discharging cycles
5	27	25	197	87

Data Set Citation: B. Saha and K. Goebel (2007). "Battery Data Set", NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA

ISET

LI-ION-BATTERY GIT-HUB REPOSITORY DATASET

C This dataset was by a Git-Hub user in the GitHub platform.

C it contains a total of 15064 data points for 14 Batteries including 7 features and the RUL information about each data point.

C To extract relevant information about this dataset I used a Python code with the link provided in the next slide.

ISET

Number of features	Number of charging-discharging Batteries	Min length of charging-descharging	Max length of charging-descharging	Mean length of charging-descharging
7	14	1051	1081	1076

C **Link to the data Set** : https://github.com/SuryaPrakash2/DATA-DRIVEN-APPROACH-FOR-PREDICTION-OF-REMAINING-USEFUL-LIFE-RUL-OF-LI-ION-BATTERY/blob/main/Battery_RUL.csv

C **Link to the code used to extract informations about the data**: https://colab.research.google.com/drive/1o6VM7_w-SHCqbTnUn4_ovV7eiReKT2Qs

ISET

ELEVATOR DATASET

- This dataset is from Huawei German Research Center and was provided by a Kaggle user in the Kaggle Platform.

- The dataset was collected from IoT sensors in the elevator industry for predictive maintenance of elevator doors.

- The dataset includes time-series data sampled at 4Hz during high-peak and evening elevator usage in a building, featuring data from various sensors such as electromechanical, ambiance, and physics sensors. The target of the dataset is to predict the absolute value of vibration.

ISET

Number of features	Number of elevators
7	112002

C **Link to the data Set** :
https://www.kaggle.com/datasets/shivamb/elevator-predictive-maintenance-dataset

ISET