**Note:** For the complete list of datasets, including those that are divided into several subsets), refer to the end of this file.

**Dataset List Description:**

**Dataset extracted from PdM\_telemetry.csv:**

20 time series in total, 5 of each from 4 variables (volt, rotate, pressure, vibration), with length 200, 500, 2000, 5000, 8761, all from machine id=1’s data.

**Dataset from other sources, mainly DARTS:**

**(Datasets with \* are also used last semester)**

- 0.5 second:

1. HeartRateDataset – DARTS – 1800

*The series contains 1800 evenly spaced measurements of instantaneous heart rate from a single subject.*

*The measurements (in units of beats per minute) occur at 0.5 second intervals, so that the length of* *each series is exactly 15 minutes.*

*It uses an integer time index.*

- 10 min:

1. WeatherDataset – DARTS – 52695

*Includes 21 indicators of weather, such as air temperature, and humidity. The data was recorded every 10 min for 2020 in Germany.*

- 15 min:

1. ETTm1Dataset – DARTS – 69680

*The data of 1 Electricity Transformers at 1 station, including load, oil temperature.*

*The dataset ranges from 2016/07 to 2018/07 recorded every 15 minutes.*

*Field Descriptions:*

*\* date: The recorded date*

*\* HUFL: High UseFul Load*

*\* HULL: High UseLess Load*

*\* MUFL: Medium UseFul Load*

*\* MULL: Medium UseLess Load*

*\* LUFL: Low UseFul Load*

*\* LULL: Low UseLess Load*

*\* OT: Oil Temperature (Target)*

2. ETTm2Dataset – DARTS – 69680

*The data of 1 Electricity Transformers at 1 station, including load, oil temperature.*

*The dataset ranges from 2016/07 to 2018/07 recorded every 15 minutes.*

*Field Descriptions:*

*\* date: The recorded date*

*\* HUFL: High UseFul Load*

*\* HULL: High UseLess Load*

*\* MUFL: Medium UseFul Load*

*\* MULL: Medium UseLess Load*

*\* LUFL: Low UseFul Load*

*\* LULL: Low UseLess Load*

*\* OT: Oil Temperature (Target)*

- 30 min:

1. TaylorDataset – DARTS – 4032

*Half-hourly electricity demand in England and Wales from Monday 5 June 2000 to Sunday 27 August 2000.*

*Discussed in Taylor (2003) and kindly provided by James W Taylor. Units: Megawatts* *(Uses an integer time index).*

- 1 H:

1. ETTh1Dataset – DARTS – 17420

*The data of 1 Electricity Transformers at 1 station, including load, oil temperature.*

*The dataset ranges from 2016/07 to 2018/07 taken hourly.*

*Field Descriptions:*

*\* date: The recorded date*

*\* HUFL: High UseFul Load*

*\* HULL: High UseLess Load*

*\* MUFL: Medium UseFul Load*

*\* MULL: Medium UseLess Load*

*\* LUFL: Low UseFul Load*

*\* LULL: Low UseLess Load*

*\* OT: Oil Temperature (Target)*

2. ETTh2Dataset – DARTS – 17420

*The data of 1 Electricity Transformers at 1 station, including load, oil temperature.*

*The dataset ranges from 2016/07 to 2018/07 taken hourly.*

*Field Descriptions:*

*\* date: The recorded date*

*\* HUFL: High UseFul Load*

*\* HULL: High UseLess Load*

*\* MUFL: Medium UseFul Load*

*\* MULL: Medium UseLess Load*

*\* LUFL: Low UseFul Load*

*\* LULL: Low UseLess Load*

*\* OT: Oil Temperature (Target)*

3. EnergyDataset – DARTS – 35064

*Contains a time series with 28 hourly components between 2014-12-31 23:00:00 and 2018-12-31 22:00:00*

4. AirQuality – UCI ML Repo – 9357

*Responses of a gas multi-sensor device deployed on the field of a significantly polluted area in an Italian city. Hourly responses averages are recorded along with gas concentrations references from a certified analyzer.*

*Field Descriptions:*

*Date (DD/MM/YYYY) Time (HH.MM.SS)*

*True hourly averaged concentration CO in mg/m^3 (reference analyzer)*

*PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)*

*True hourly averaged overall Non Metanic HydroCarbons concentration in microg/m^3 (reference analyzer)*

*PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)*

*True hourly averaged NOx concentration in ppb (reference analyzer)*

*PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NOx targeted)*

*True hourly averaged NO2 concentration in microg/m^3 (reference analyzer)*

*PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO2 targeted)*

*True hourly averaged Benzene concentration in microg/m^3 (reference analyzer)*

*PT08.S5 (indium oxide) hourly averaged sensor response (nominally O3 targeted)*

*Temperature in Â°C*

*Relative Humidity (%)*

*AH Absolute Humidity*

- 1 D:

1. ExchangeRateDataset – DARTS – 7588

*The collection of the daily exchange rates of eight foreign countries, including Australia, British, Canada, Switzerland, China, Japan, New Zealand, and Singapore, from 1990 to 2016. Unfortunately,* *there were some inconsistencies concerning the dates, so the resulting Time Series is integer indexed.*

2. TemperatureDataset – DARTS – 3650

*Daily temperature in Melbourne between 1981 and 1990*

- 1 W:

1. ILINetDataset – DARTS – 1301 – Contain lots of missing values (0)

*ILI describes the number of patients seen with influenza like illness and the total number of patients. It includes* *weekly data from the Centers for Disease Control and Prevention of the United States from 1997 to 2022.*

*Components Descriptions:*

*\* % WEIGHTED ILI: Combined state-specific data of patients visit to healthcare providers for ILI reported each week* *weighted by state population*

*\* % UNWEIGHTED ILI: Combined state-specific data of patients visit to healthcare providers for ILI reported each* *week unweighted by state population*

*\* AGE 0-4: Number of patients between 0 and 4 years of age*

*\* AGE 25-49: Number of patients between 25 and 49 years of age*

*\* AGE 25-64: Number of patients between 25 and 64 years of age*

*\* AGE 5-24: Number of patients between 5 and 24 years of age*

*\* AGE 50-64: Number of patients between 50 and 64 years of age*

*\* AGE 65: Number of patients above (>=65) 65 years of age*

*\* ILITOTAL: Total number of ILI patients. For this system, ILI is defined as fever (temperature of 100°F [37.8°C]* *or greater) and a cough and/or a sore throat*

*\* NUM. OF PROVIDERS: Number of outpatient healthcare providers*

*\* TOTAL PATIENTS: Total number of patients*

2. USGasolineDataset – DARTS – 1578

*Weekly U.S. Product Supplied of Finished Motor Gasoline between 1991-02-08 and 2021-04-30*

- 1 M:

**1. MonthlyMilkDataset\* – DARTS** – **168**

*Monthly production of milk (in pounds per cow) between Jan 1962 and Dec 1975*

**2. AirPassengersDataset\* – DARTS** – **144**

*Monthly Air Passengers Dataset, from 1949 to 1960.*

**3. SunspotsDataset\* – DARTS** – **2820**

*Monthly mean relative sunspot numbers from 1749 to 1983.* *Collected at Swiss Federal Observatory, Zurich until 1960, then Tokyo Astronomical Observatory.*

**4. WineDataset\* – DARTS** – **176**

*Australian total wine sales by wine makers in bottles <= 1 litre. Monthly between Jan 1980 and Aug 1994.*

**5. GasRateCO2Dataset\* – DARTS** – **296**

*Two components, GasRate(ft3/min) & CO2%, length 296 (integer time index)*

6. IceCreamHeaterDataset – DARTS – 198

*Monthly sales of heaters and ice cream between January 2004 and June 2020.*

- 1 Q:

**1. WoolyDataset\* – DARTS** – **119**

*Quarterly production of woollen yarn in Australia: tonnes. Mar 1965 -- Sep 1994.*

**2. AusBeerDataset\* – DARTS** – **211**

*Total quarterly beer production in Australia (in megalitres) from 1956: Q1 to 2008: Q3*

**Datasets from above that can be divided into multiple subsets:**

WeatherDataset – DARTS – 52695 (**multi-variables**)

ETTm1Dataset – DARTS – 69680 (**multi-variables**)

ETTm2Dataset – DARTS – 69680 (**multi-variables**)

ETTh1Dataset – DARTS – 17420 (**multi-variables**)

ETTh2Dataset – DARTS– 17420 (**multi-variables**)

EnergyDataset – DARTS – 35064 (**multi-variables**)

Air Quality – UCI ML Repo – 9357 (**multi-variables**)

ExchangeRateDataset – DARTS – 7588 (**multi-variables**)

TemperatureDataset – DARTS – 3650

ILINetDataset – DARTS – 1301 (**multi-variables**)

GasRateCO2Dataset\* – DARTS – 296 **(2 variables)**

IceCreamHeaterDataset – DARTS – 198 **(2 variables)**

**Complete list of datasets:**

From PdM\_telemetry.csv:

1. volt\_200, 1 H, 200
2. volt\_500
3. volt\_2000
4. volt\_5000
5. volt\_8761
6. rotate\_200
7. rotate\_500
8. rotate\_2000
9. rotate\_5000
10. rotate\_8761
11. pressure\_200
12. pressure\_500
13. pressure\_2000
14. pressure\_5000
15. pressure\_8761
16. vibration\_200
17. vibration\_500
18. vibration\_2000
19. vibration\_5000
20. vibration\_8761

From other datasets listed above (name of variable extracted for that subset of time series is added to the dataset name):

1. **AirPassengersDataset\_DARTS\_144**, 1 M, 144
2. AirQuality\_UCI\_ML\_Repo\_CO\_1000, 1 H, 1000
3. AirQuality\_UCI\_ML\_Repo\_CO\_3000
4. AirQuality\_UCI\_ML\_Repo\_CO\_9357
5. AirQuality\_UCI\_ML\_Repo\_O3\_1000
6. AirQuality\_UCI\_ML\_Repo\_O3\_3000
7. AirQuality\_UCI\_ML\_Repo\_O3\_9357
8. AirQuality\_UCI\_ML\_Repo\_RH\_1000
9. AirQuality\_UCI\_ML\_Repo\_RH\_3000
10. AirQuality\_UCI\_ML\_Repo\_RH\_9357
11. **AusBeerDataset\_DARTS\_211**, 1 Q, 211
12. EnergyDataset\_DARTS\_fossialoil\_1000, 1 H, 1000
13. EnergyDataset\_DARTS\_fossialoil\_3000
14. EnergyDataset\_DARTS\_fossialoil\_10000
15. EnergyDataset\_DARTS\_fossilgas\_1000
16. EnergyDataset\_DARTS\_fossilgas\_3000
17. EnergyDataset\_DARTS\_fossilgas\_10000
18. EnergyDataset\_DARTS\_solar\_1000
19. EnergyDataset\_DARTS\_solar\_3000
20. EnergyDataset\_DARTS\_solar\_10000
21. ETTh1Dataset\_DARTS\_OT\_1000, 1H, 1000
22. ETTh1Dataset\_DARTS\_OT\_3000
23. ETTh1Dataset\_DARTS\_OT\_10000
24. ETTh1Dataset\_DARTS\_OT\_17420
25. ETTh2Dataset\_DARTS\_OT\_1000, 1H, 1000
26. ETTh2Dataset\_DARTS\_OT\_3000
27. ETTh2Dataset\_DARTS\_OT\_10000
28. ETTh2Dataset\_DARTS\_OT\_17420
29. ETTm1Dataset\_DARTS\_OT\_1000, 15 min, 1000
30. ETTm1Dataset\_DARTS\_OT\_3000
31. ETTm1Dataset\_DARTS\_OT\_10000
32. ETTm1Dataset\_DARTS\_OT\_20000
33. ETTm2Dataset\_DARTS\_OT\_1000, 15 min, 1000
34. ETTm2Dataset\_DARTS\_OT\_3000
35. ETTm2Dataset\_DARTS\_OT\_10000
36. ETTm2Dataset\_DARTS\_OT\_20000
37. ExchangeRateDataset\_DARTS\_aus\_1000, 1 D, 1000
38. ExchangeRateDataset\_DARTS\_aus\_3000
39. ExchangeRateDataset\_DARTS\_aus\_7588
40. ExchangeRateDataset\_DARTS\_brt\_1000
41. ExchangeRateDataset\_DARTS\_brt\_3000
42. ExchangeRateDataset\_DARTS\_brt\_7588
43. ExchangeRateDataset\_DARTS\_can\_1000
44. ExchangeRateDataset\_DARTS\_can\_3000
45. ExchangeRateDataset\_DARTS\_can\_7588
46. ExchangeRateDataset\_DARTS\_chn\_1000
47. ExchangeRateDataset\_DARTS\_chn\_3000
48. ExchangeRateDataset\_DARTS\_chn\_7588
49. **CO2Dataset\_DARTS\_296**, 1M, 296
50. **GasRateDataset\_DARTS\_296**
51. HeartRateDataset\_DARTS\_1800, 0.5 second, 1800
52. HeaterDataset\_DARTS\_198, 1 M, 198
53. IceCreamDataset\_DARTS\_198
54. ILINetDataset\_DARTS\_%WEIGHTED\_ILI\_1041, 1 W, 1041
55. ILINetDataset\_DARTS\_age0to4\_1041
56. ILINetDataset\_DARTS\_age5to24\_1041
57. ILINetDataset\_DARTS\_age25to49\_677
58. ILINetDataset\_DARTS\_age25to64\_364
59. ILINetDataset\_DARTS\_age50to64\_677
60. ILINetDataset\_DARTS\_age65\_1041
61. ILINetDataset\_DARTS\_illtotal\_1041
62. ILINetDataset\_DARTS\_num\_provider\_1041
63. ILINetDataset\_DARTS\_total\_patient\_1041
64. **MonthlyMilkDataset\_DARTS\_168**, 1 M, 168
65. **SunspotsDataset\_DARTS\_2820**, 1 M, 2820
66. SunspotsDataset\_DARTS\_500
67. TaylorDataset\_DARTS\_4032, 30 min, 4032
68. TemperatureDataset\_DARTS\_200, 1 D, 200
69. TemperatureDataset\_DARTS\_500
70. TemperatureDataset\_DARTS\_2000
71. TemperatureDataset\_DARTS\_3650
72. USGasolineDataset\_DARTS\_1578, 1 W, 1578
73. WeatherDataset\_DARTS\_pressure\_1000, 10 min, 1000
74. WeatherDataset\_DARTS\_pressure\_3000
75. WeatherDataset\_DARTS\_pressure\_10000
76. WeatherDataset\_DARTS\_windspeed\_1000
77. WeatherDataset\_DARTS\_windspeed\_3000
78. WeatherDataset\_DARTS\_windspeed\_10000
79. **WineDataset\_DARTS\_176**, 1 M, 176
80. **WoolyDataset\_DARTS\_119**, 1 Q, 119

**Regarding forecasting length:**

Data length<180: prediction length=24

180<Data length<300: prediction length = 36

300<Data length<1000: prediction length = 48

Data length>=1000: prediction length = 100