Hackathon

10.02.2023

1 Informations

1.1 Data source / Industrial manufacturing plant

The source of the data is an industrial manufacturing facility that assembles printed circuit boards (PCB's) with components and then solder them within a furnace. Typical components are for example capacitors or resistors. Below is a picture of the plant.
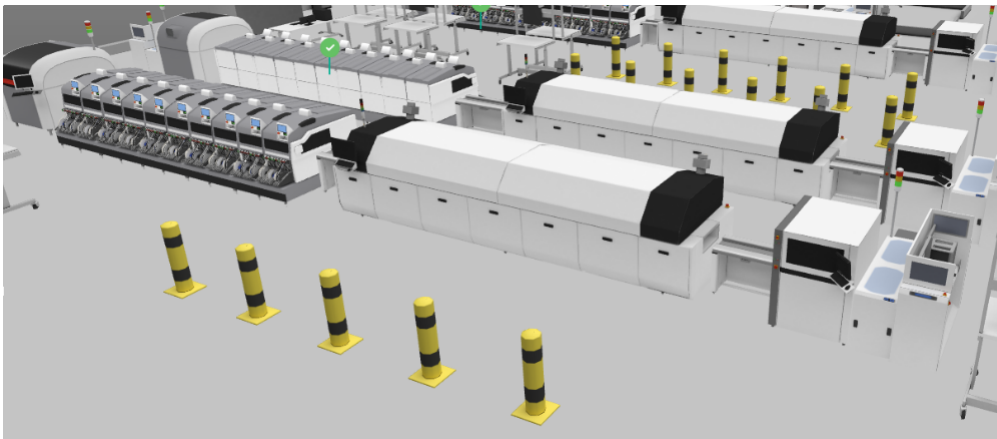

Figure 1: Schematic representation of the production line.

The machine on the left with the blue displays is the so-called placer. The pick-and-place machine consists of ten modules, which are arranged next to each other in a row and numbered analogously from one to ten. A PCB passes through each of these modules by means of a conveyor belt, but does not necessarily have to be assembled in each module. The machine to its right is the furnace, which is also passed by the pcbs via a conveyor belt.

1.2 Data description

There are two main data sources, one being the placer and the furnance. Both providing time series data.

FURNACE:
The data set of the furnace contains three columns for a period of **three months**.

The **first column** *time* contains timestamps and is clocked in 15-second intervals. Gaps may also occur.

The **second column** shows the average power consumption of the last 15 seconds in watts. It often happens that there is no value at all or a zero. In this case, this is a gap in the data set. The reasons for the gaps can be missing data, a change of work shift or maintenance of the machine. The work shifts are scheduled as follows: 06:00-14:00, 14:00-22:00 and 22:00-06:00.

The **third column** shows the average nitrogen consumption of the last 15 seconds in m^3/h. Again, it often happens that there is no value at all or a zero. This is also a gap in the data set for the same reasons as explained above.

time	energy	nitrogen_Nm3h
2022-10-11 19:50:30+00:00	9019.3995	19.440000
2022-10-11 19:50:45+00:00	6296.4840	19.428333
2022-10-11 19:51:00+00:00	15479.6640	19.585000
2022-10-11 19:51:15+00:00	11585.6500	20.127692
2022-10-11 19:51:30+00:00	6701.3010	19.685000
2022-10-11 19:51:45+00:00	9088.5830	20.792857
2022-10-11 19:52:00+00:00	16038.6930	21.813846
2022-10-11 19:52:15+00:00	9077.3420	21.088571
2022-10-11 19:52:30+00:00	7049.1325	21.680000
2022-10-11 19:52:45+00:00	11508.6340	22.217500

Figure 2: Extract from the furnace data set.

PLACER:

The placer record contains three columns for a period of **two months**.

The **first column** contains row numbering.

The **second column** *PCBID* contains the ID's, each of which is uniquely assigned to a PCB.

The **third column** *RECIPENAME* shows the names of the „blueprints" according to which the PCB's are assembled.

The **fourth column** *SIDE* shows the side to which the pcb is aligned (top/bottom).

The **fifth and sixth columns** *STARTTIME/ENDTIME* are containing the timestamps when a PCB assembly was startet by a module and when the assembling was completed.

The **sixth column** *MODULENO* shows in which module the PCB was assembled. Since there are ten Modules in the placer, this value is always between one and ten.

The **seventh column** *NUMCOMP* contains the number of components that were assembled to the PCB.

The **eighth column** is *NUMBLOCKS*. Strictly speaking, it is not individual PCBs that are assembled, but a large board consisting of a number of identical PCBs that are cut apart in a later processing step. Looking at the first row in Figure 3, the board consists of eight identical PCBs. Since there the board was assembled with 32 components, so each PCB was assembled with four components.

The **ninth** and last **column** is called *NUMERRORS*. This metric shows how many times a module has tried to place a component onto the PCB until it has succeeded.

	Unnamed: 0	PCBID	RECIPENAME	SIDE	STARTTIME	ENDTIME	MODULENO	NUMCOMP	NUMBLOCKS	NUMERRORS
0	0	9056093030018024_0048_0001_0001_20220831235024472	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:44	2022-09-01 00:00:17	8.0	32.0	8.0	0.0
1	1	9056093030018048_0048_0001_0001_20220831235430535	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:48	2022-09-01 00:00:18	5.0	72.0	8.0	0.0
2	2	9056093030017960_0048_0001_0001_20220831235851351	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:38	2022-09-01 00:00:20	2.0	168.0	8.0	0.0
3	3	9056093030017968_0048_0001_0001_20220831235947179	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:40	2022-09-01 00:00:31	1.0	208.0	8.0	0.0
4	4	9056093030017976_0048_0001_0001_20220831235553736	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:48	2022-09-01 00:00:57	4.0	200.0	8.0	0.0
5	5	9056093030018032_0048_0001_0001_20220831235145646	1097922-00_NXT3_TL4	0.0	2022-08-31 23:59:46	2022-09-01 00:01:03	7.0	160.0	8.0	0.0
6	6	9056093030017968_0048_0001_0001_20220831235947179	1097922-00_NXT3_TL4	0.0	2022-09-01 00:00:34	2022-09-01 00:01:16	2.0	168.0	8.0	0.0
7	7	9056093030018024_0048_0001_0001_20220831235024472	1097922-00_NXT3_TL4	0.0	2022-09-01 00:00:23	2022-09-01 00:01:19	9.0	56.0	8.0	0.0
8	8	9056093030017960_0048_0001_0001_20220831235851351	1097922-00_NXT3_TL4	0.0	2022-09-01 00:00:24	2022-09-01 00:01:34	3.0	184.0	8.0	0.0
9	9	9056093030018008_0048_0001_0001_20220901000053447	1097922-00_NXT3_TL4	0.0	2022-09-01 00:00:46	2022-09-01 00:01:38	1.0	208.0	8.0	0.0

Figure 3: Extract from the placer data set

2 Tasks to submit

In the file *furnace_blanked.csv* are excerpts, which were removed from the original furnace data set. In addition, gaps were inserted afterwards, which appear in regular intervals and are always two minutes long. The gaps are characterized by the fact that they have a value of -13.37. Additionally, zeros have been inserted after the gaps to prevent data from the future being used. The goal is to predict the original values of the gaps as best as possible. The source code as well as the completed *furnace_blanked.csv* file with the filled gaps must be submitted.

2.1 Evaluation criteria

The following criteria are included in the evaluation. They are ranked according to their weighting:

- The goal is to predict the original values of the gaps, marked by -13.37, as best as possible. For each of the two metrics, the distances will be determined by the mean squared error function.

- The presentation should cover the following content. Talk about the idea/your approach to the solution. Explain how you implemented your ideas: What libaries did you use. What was particularly difficult to implement.

The presentation will be significantly less weighted. In case of a tie, however, it can be the decisive point.