ITU-ML5G-PS-015: Network failure detection and root cause analysis in 5GC by NFV-based test environment

Team Kaien
2021-12-01

Team member:

阿部 海燕 Kaien ABE
abe@people.kobe-u.ac.jp

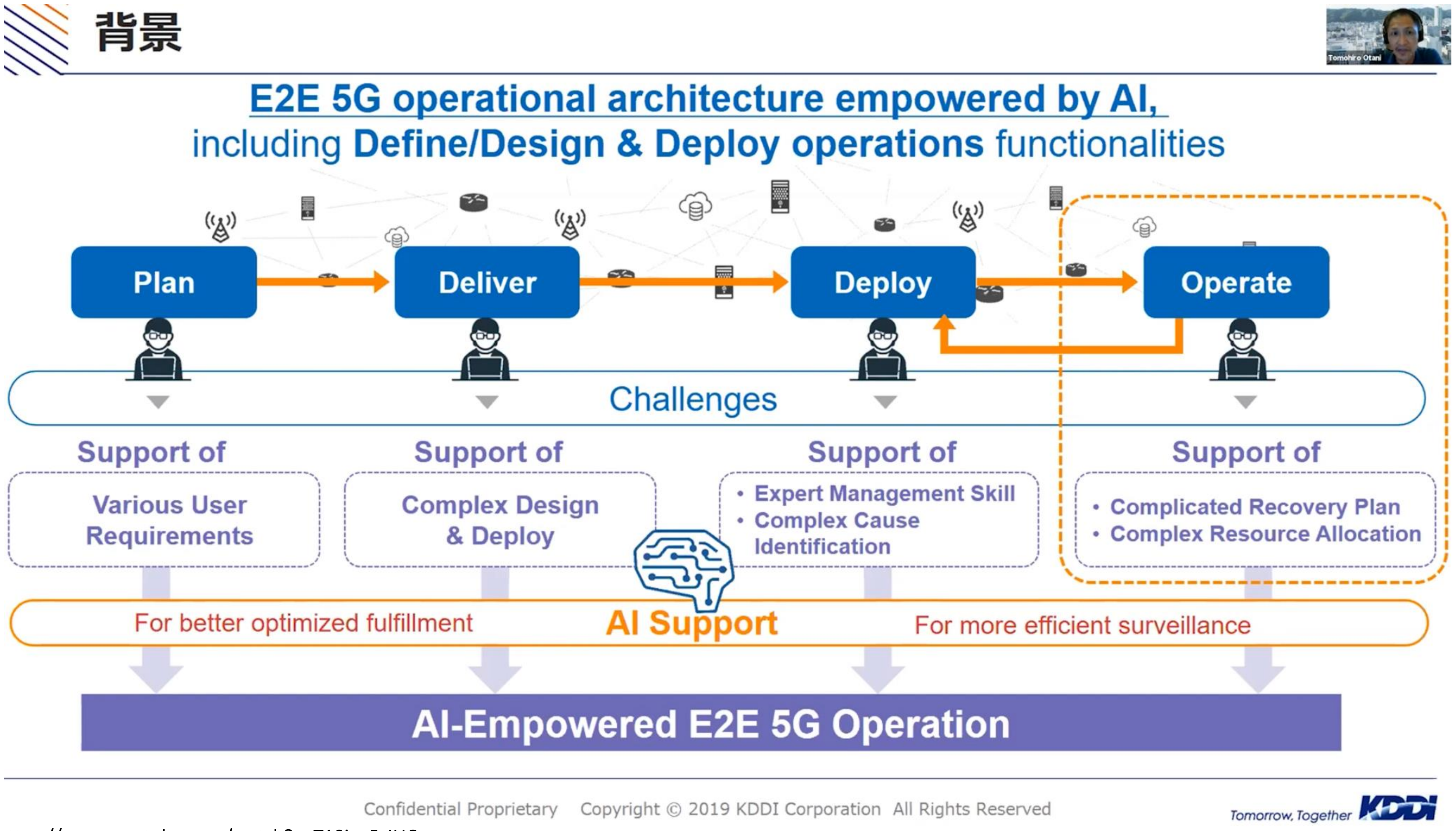
白石 善明 Yoshiaki SHIRAISHI
zenmei@port.kobe-u.ac.jp

KOBE UNIVERSITY

Content

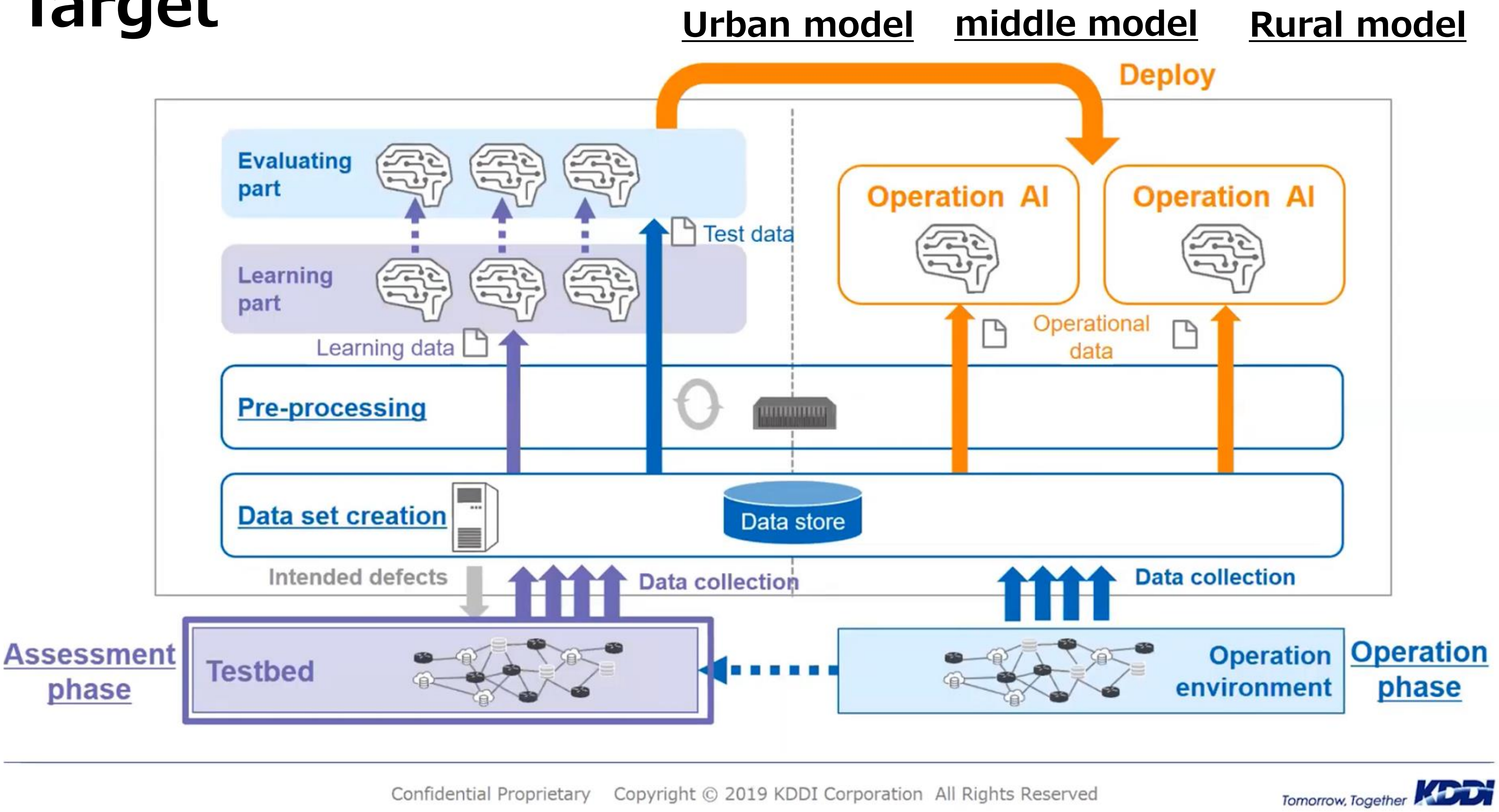
- Background
- Survey of Data Profile
- Feature Extraction Method
- Data clustering using unsupervised ML model
- Task #1 Abnormal detection model
- Task #2 Augmentation model
- Task #3 Failure cause analysis by SHAP value
- Future work

Project Introduction



https://www.youtube.com/watch?v=T18ivePrIUQ

Target



https://www.youtube.com/watch?v=T18ivePrIUQ

Data processing environment
計算機・ライブラ 環境

Computer configuration	
CPU/クロック周波数	Intel(R) Core(TM) i7-9750H/2.60GHz
メモリー（Memory）	32GB
OS	Ubuntu 18.04

Library	
Anacond3 (Environment)	Python3.8
json (process original json data)	csv （output data format）
Panadas (feature extraction)	sqlite （database for data post-processing
XGBoost (normal model)	XGBoost (increment model)

Original data is 77GB json file, it takes several days for feature extraction.

Dataset overview

Provided data: json files (77GB)	a Urban(都市)	c Rural(田舎)	b Middle(中間)
Train	98,533	98,533	12,360
Test	24,647	24,647	24,647
Failure label	2480(train)+620(test)	2480(train)+620(test)	310(train)+620(test)

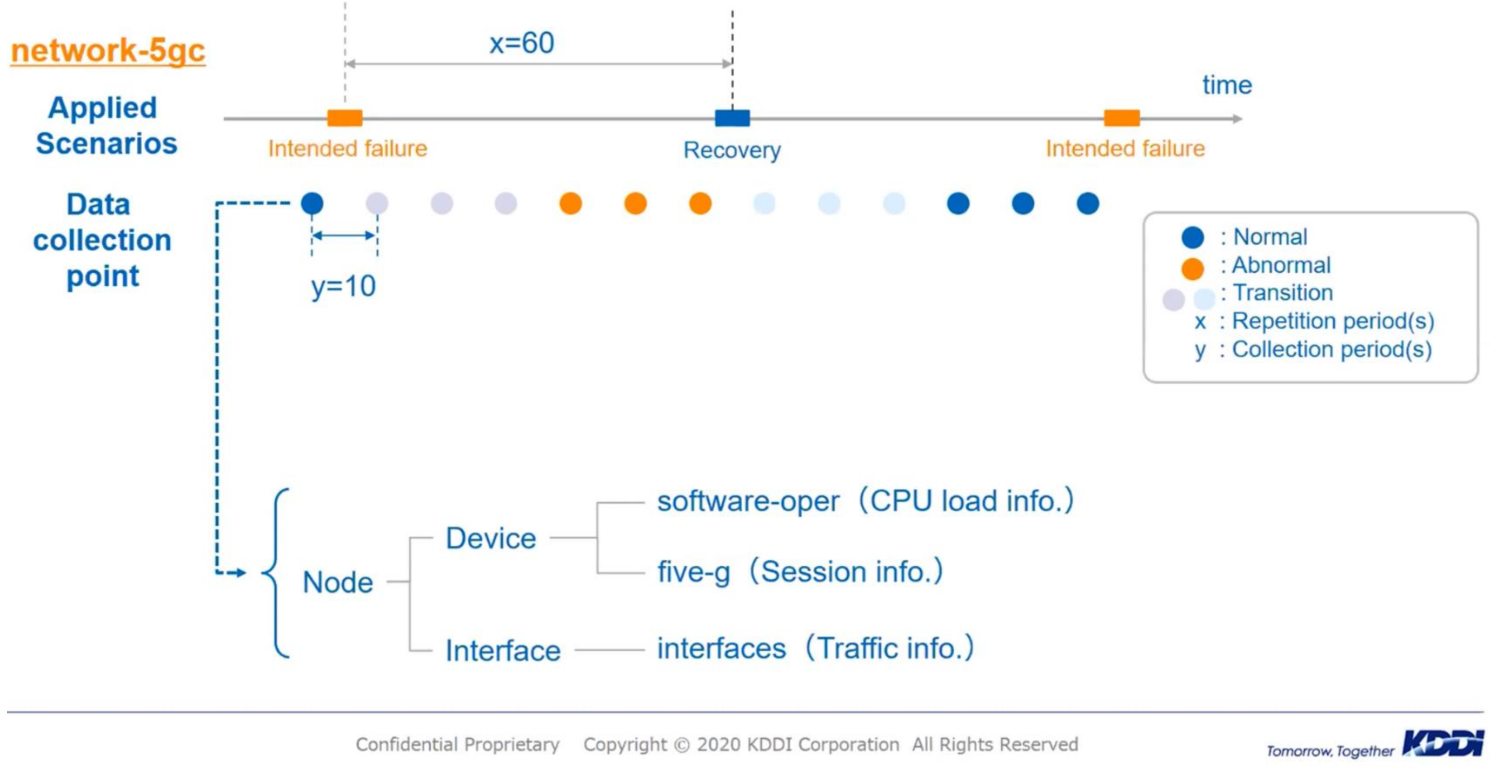
Extract features by python code

(One data is split to 3 data according to nodes number for more precise label)

Table data of features (101)

Train: (101 features) [nodes: amf, ausf, udm, respectively]	295,599 (=98,533×3) (×101 特徴量)	295,599 (=98,533×3) (×101 特徴量)	37,080 (=12,360×3) (×101 特徴量)
Test: (101 features) [nodes: amf, ausf, udm, respectively]	73,886 (≒24,647×3) (×101 特徴量)	73,889 (≒24,647×3) (×101 特徴量)	73,941 (=24,647×3) (×101 特徴量)
Abnormal data (from failure label)	4,645	4,011	1,212
Abnormal /Total (%)	13%	10%	10%

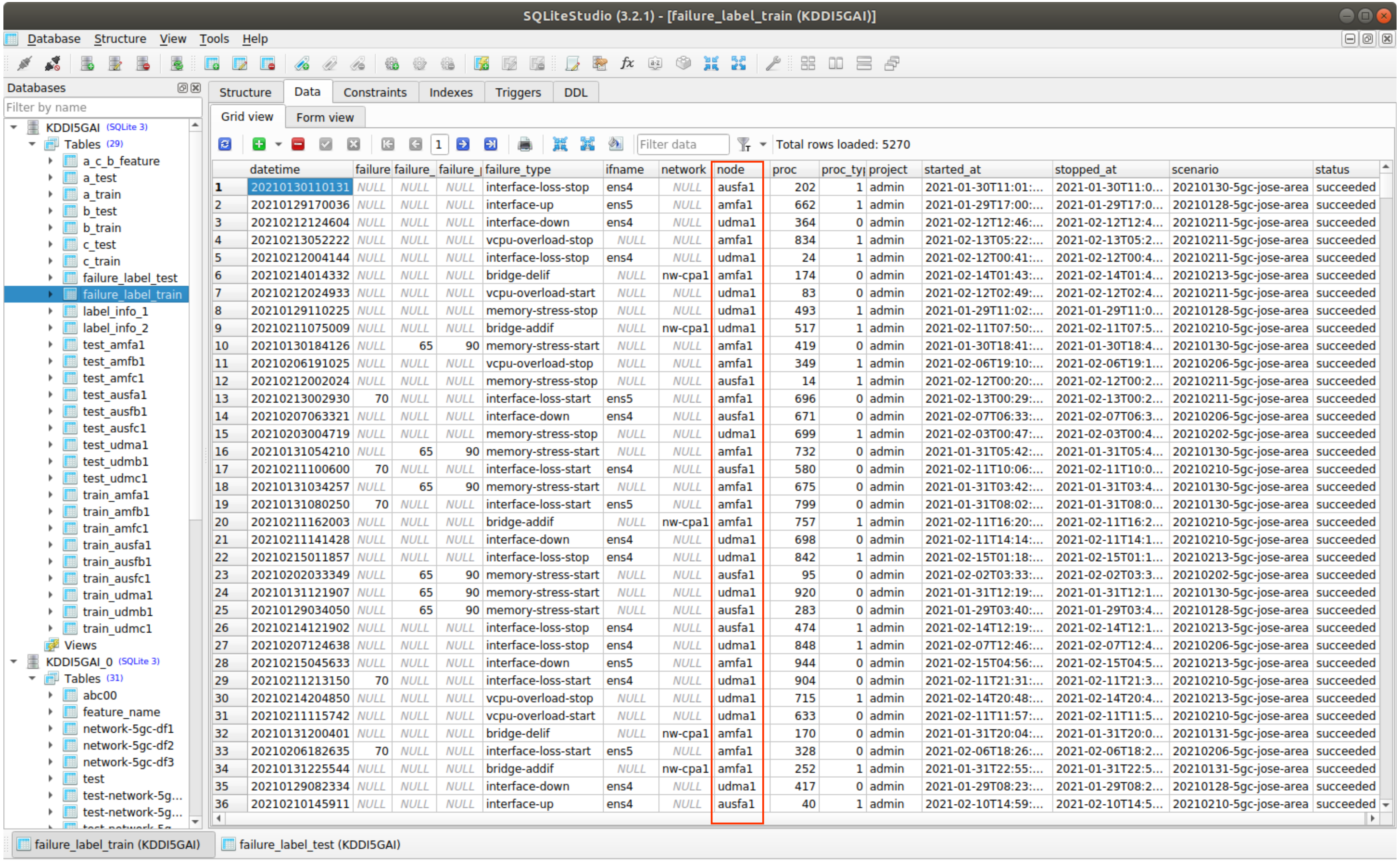
Data investigation



Nodes	Failure
amf	○ 障害発生
ausf	○ 障害発生
udm	○ 障害発生
gnb	× 障害なし
smf	× 障害なし
upf	× 障害なし
nrf	× 障害なし
dn	× 障害なし

8 types nodes are included in the dataset. but failures are happened in 3 types nodes according to Failure label. <u>In order to reduce calculation cost this time</u>, only these 3 types nodes are processed. In real scenario, all of nodes should be considered to process.

Failure label information



	datetime	failure	failure_	failure_	failure_type	ifname	network	node	proc	proc_ty	project	started_at	stopped_at	scenario	status
1	20210130110131	NULL	NULL	NULL	interface-loss-stop	ens4	NULL	ausfa1	202	1	admin	2021-01-30T11:01:...	2021-01-30T11:0...	20210130-5gc-jose-area	succeeded
2	20210129170036	NULL	NULL	NULL	interface-up	ens5	NULL	amfa1	662	1	admin	2021-01-29T17:00:...	2021-01-29T17:0...	20210128-5gc-jose-area	succeeded
3	20210212124604	NULL	NULL	NULL	interface-down	ens4	NULL	udma1	364	0	admin	2021-02-12T12:46:...	2021-02-12T12:4...	20210211-5gc-jose-area	succeeded
4	20210213052222	NULL	NULL	NULL	vcpu-overload-stop	NULL	NULL	amfa1	834	1	admin	2021-02-13T05:22:...	2021-02-13T05:2...	20210211-5gc-jose-area	succeeded
5	20210212004144	NULL	NULL	NULL	interface-loss-stop	ens4	NULL	udma1	24	1	admin	2021-02-12T00:41:...	2021-02-12T00:4...	20210211-5gc-jose-area	succeeded
6	20210214014332	NULL	NULL	NULL	bridge-delif	NULL	nw-cpa1	amfa1	174	0	admin	2021-02-14T01:43:...	2021-02-14T01:4...	20210213-5gc-jose-area	succeeded
7	20210212024933	NULL	NULL	NULL	vcpu-overload-start	NULL	NULL	udma1	83	0	admin	2021-02-12T02:49:...	2021-02-12T02:4...	20210211-5gc-jose-area	succeeded
8	20210129110225	NULL	NULL	NULL	memory-stress-stop	NULL	NULL	udma1	493	1	admin	2021-01-29T11:02:...	2021-01-29T11:0...	20210128-5gc-jose-area	succeeded
9	20210211075009	NULL	NULL	NULL	bridge-addif	NULL	nw-cpa1	udma1	517	1	admin	2021-02-11T07:50:...	2021-02-11T07:5...	20210210-5gc-jose-area	succeeded
10	20210130184126	NULL	65	90	memory-stress-start	NULL	NULL	amfa1	419	0	admin	2021-01-30T18:41:...	2021-01-30T18:4...	20210130-5gc-jose-area	succeeded
11	20210206191025	NULL	NULL	NULL	vcpu-overload-stop	NULL	NULL	amfa1	349	1	admin	2021-02-06T19:10:...	2021-02-06T19:1...	20210206-5gc-jose-area	succeeded
12	20210212002024	NULL	NULL	NULL	memory-stress-stop	NULL	NULL	ausfa1	14	1	admin	2021-02-12T00:20:...	2021-02-12T00:2...	20210211-5gc-jose-area	succeeded
13	20210213002930	70	NULL	NULL	interface-loss-start	ens5	NULL	amfa1	696	0	admin	2021-02-13T00:29:...	2021-02-13T00:2...	20210211-5gc-jose-area	succeeded
14	20210207063321	NULL	NULL	NULL	interface-down	ens4	NULL	ausfa1	671	0	admin	2021-02-07T06:33:...	2021-02-07T06:3...	20210206-5gc-jose-area	succeeded
15	20210203004719	NULL	NULL	NULL	memory-stress-stop	NULL	NULL	udma1	699	1	admin	2021-02-03T00:47:...	2021-02-03T00:4...	20210202-5gc-jose-area	succeeded
16	20210131054210	NULL	65	90	memory-stress-start	NULL	NULL	amfa1	732	0	admin	2021-01-31T05:42:...	2021-01-31T05:4...	20210130-5gc-jose-area	succeeded
17	20210211100600	70	NULL	NULL	interface-loss-start	ens4	NULL	ausfa1	580	0	admin	2021-02-11T10:06:...	2021-02-11T10:0...	20210210-5gc-jose-area	succeeded
18	20210131034257	NULL	65	90	memory-stress-start	NULL	NULL	amfa1	675	0	admin	2021-01-31T03:42:...	2021-01-31T03:4...	20210130-5gc-jose-area	succeeded
19	20210131080250	70	NULL	NULL	interface-loss-start	ens5	NULL	amfa1	799	0	admin	2021-01-31T08:02:...	2021-01-31T08:0...	20210130-5gc-jose-area	succeeded
20	20210211162003	NULL	NULL	NULL	bridge-addif	NULL	nw-cpa1	amfa1	757	1	admin	2021-02-11T16:20:...	2021-02-11T16:2...	20210210-5gc-jose-area	succeeded
21	20210211141428	NULL	NULL	NULL	interface-down	ens4	NULL	udma1	698	0	admin	2021-02-11T14:14:...	2021-02-11T14:1...	20210210-5gc-jose-area	succeeded
22	20210215011857	NULL	NULL	NULL	interface-loss-stop	ens4	NULL	udma1	842	1	admin	2021-02-15T01:18:...	2021-02-15T01:1...	20210213-5gc-jose-area	succeeded
23	20210202033349	NULL	65	90	memory-stress-start	NULL	NULL	ausfa1	95	0	admin	2021-02-02T03:33:...	2021-02-02T03:3...	20210202-5gc-jose-area	succeeded
24	20210131121907	NULL	65	90	memory-stress-start	NULL	NULL	udma1	920	0	admin	2021-01-31T12:19:...	2021-01-31T12:1...	20210130-5gc-jose-area	succeeded
25	20210129034050	NULL	65	90	memory-stress-start	NULL	NULL	ausfa1	283	0	admin	2021-01-29T03:40:...	2021-01-29T03:4...	20210128-5gc-jose-area	succeeded
26	20210214121902	NULL	NULL	NULL	interface-loss-stop	ens4	NULL	ausfa1	474	1	admin	2021-02-14T12:19:...	2021-02-14T12:1...	20210213-5gc-jose-area	succeeded
27	20210207124638	NULL	NULL	NULL	interface-loss-stop	ens4	NULL	udma1	848	1	admin	2021-02-07T12:46:...	2021-02-07T12:4...	20210206-5gc-jose-area	succeeded
28	20210215045633	NULL	NULL	NULL	interface-down	ens5	NULL	amfa1	944	0	admin	2021-02-15T04:56:...	2021-02-15T04:5...	20210213-5gc-jose-area	succeeded
29	20210211213150	70	NULL	NULL	interface-loss-start	ens4	NULL	udma1	904	0	admin	2021-02-11T21:31:...	2021-02-11T21:3...	20210210-5gc-jose-area	succeeded
30	20210214204850	NULL	NULL	NULL	vcpu-overload-stop	NULL	NULL	udma1	715	1	admin	2021-02-14T20:48:...	2021-02-14T20:4...	20210213-5gc-jose-area	succeeded
31	20210211115742	NULL	NULL	NULL	vcpu-overload-start	NULL	NULL	udma1	633	0	admin	2021-02-11T11:57:...	2021-02-11T11:5...	20210210-5gc-jose-area	succeeded
32	20210131200401	NULL	NULL	NULL	bridge-delif	NULL	nw-cpa1	amfa1	170	0	admin	2021-01-31T20:04:...	2021-01-31T20:0...	20210131-5gc-jose-area	succeeded
33	20210206182635	70	NULL	NULL	interface-loss-start	ens5	NULL	amfa1	328	0	admin	2021-02-06T18:26:...	2021-02-06T18:2...	20210206-5gc-jose-area	succeeded
34	20210131225544	NULL	NULL	NULL	bridge-addif	NULL	nw-cpa1	amfa1	252	1	admin	2021-01-31T22:55:...	2021-01-31T22:5...	20210131-5gc-jose-area	succeeded
35	20210129082334	NULL	NULL	NULL	interface-down	ens4	NULL	udma1	417	0	admin	2021-01-29T08:23:...	2021-01-29T08:2...	20210128-5gc-jose-area	succeeded
36	20210210145911	NULL	NULL	NULL	interface-up	ens4	NULL	ausfa1	40	1	admin	2021-02-10T14:59:...	2021-02-10T14:5...	20210210-5gc-jose-area	succeeded

Data processing (1) データ分割のアイデアとメリット

In the provided dataset, failures happened in three nodes (total nodes=8) so that we only processing 3 nodes information for decreasing data processing work.

	障害なし (No failure)	障害発生 (Failure)	障害なし (No Failure)	Positive
Data_1	amf	ausf	udm	Label=1
	101 Features	101 Features	101 Features	

One data is split to 3 data

Data_1_amf	amf	Label=0
Data_1_ausf	ausf 101 Features	Label=1
Data_1_udm	udm 101 Features	Label=0
	101 Features	

障害発生したノートを特定できる & 過学習を抑制できる

Failure nodes can be located precisely and feature number is decreased to prevent overfitting in ML

Data processing (2)-how to select features

特徴量の選抜基準

Nodes	障害	抽出項目数
amf	障害発生	204 (a,c) 154 (b)
ausf	障害発生	175(a,c,b)
udm	障害発生	177(a,c,b)
gnb	障害なし	
smf	障害なし	
upf	障害なし	
nrf	障害なし	
dn	障害なし	

■Delete category info.

{interface_type:softwareloopback,tunnel,healthy,unknown
interface_name,
interface_if_index,
Interface_if_name
platform-fru-rp,
phys-address,
Healthy, critical, unknow,up, down,
Null…
}

■Remain digital info.

{memory-states-,
per-core-states-,
interface-parameter…
}

■Common features among all of nodes

datetime	failure_type	ifname	network	node
20210209154552	interface-up	ens4	NULL	udma1
20210205234627	bridge-delif	NULL	nw-cpa1	ausfa1
20210205085657	interface-down	ens5	NULL	amfa1
20210209044653	memory-stress-stop	NULL	NULL	ausfa1
20210205122320	vcpu-overload-stop	NULL	NULL	amfa1
20210205141413	vcpu-overload-stop	NULL	NULL	amfa1
20210209115810	normal	NULL	NULL	NULL
20210205235154	vcpu-overload-stop	NULL	NULL	ausfa1
20210206041959	bridge-delif	NULL	nw-cpa1	ausfa1
20210205022336	interface-down	ens5	NULL	amfa1
20210205125623	interface-loss-start	ens5	NULL	amfa1
20210210014614	memory-stress-stop	NULL	NULL	amfa1
20210205205932	interface-loss-start	ens4	NULL	udma1
20210209055829	bridge-delif	NULL	nw-cpa1	amfa1
20210206020246	bridge-delif	NULL	nw-cpa1	udma1
20210205143538	interface-loss-stop	ens5	NULL	amfa1
20210205221325	memory-stress-stop	NULL	NULL	udma1
20210209225842	interface-loss-stop	ens4	NULL	udma1

get features (101)

This selection method is not perfect and presice. More deeply data observation to select more precise and informative features is needed.

Data processing (3)- how to label data

ラベルの付け方

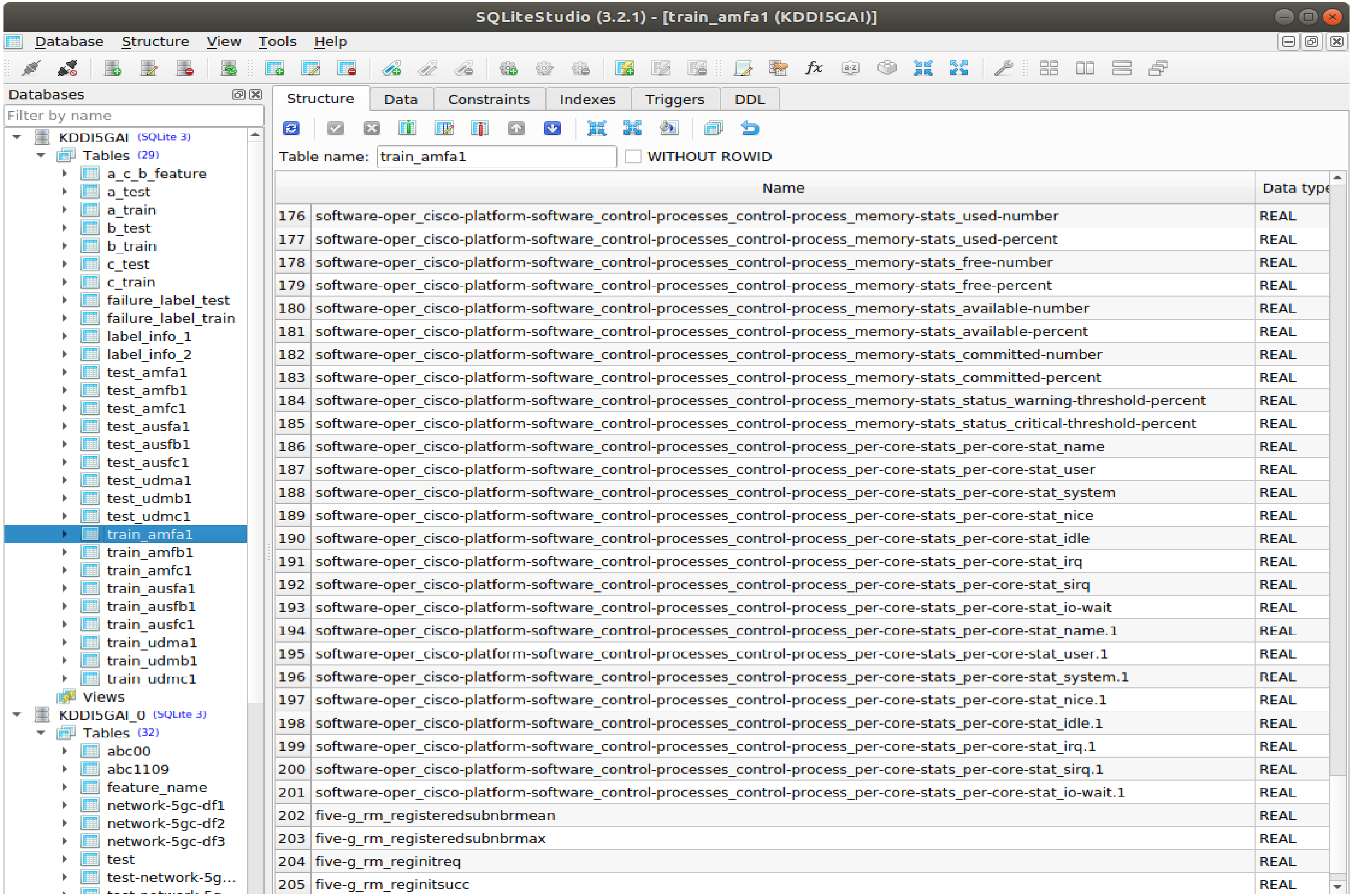
Form the failure information (failure start-time,failure type,node,ifname), we traced the train&test data, then found the following info. strongly related to failures, so that we label the training and test data according these info. There might be other info also related failures. Further more data investigation should be carried out.

Info & Features　特徴量	Abnormal　障害あり(label=1)	Normal　障害なし(label=0)
Memory_status： (148)(4512) software-oper_cisco-platform-software_control-processes_control-process_memory-stats_memory-status	'Critical' 'Warning'	'Healthy'
Minute_status_condition: 16 software-oper_cisco-platform-software_control-processes_control-process_load-avg-minutes_load-avg-minute_status_condition	'Critical’	'Healthy'
Operation_status: 1(203)2(4996) interfaces_interfaces-state_interface_oper-status	'down’ (start at, proc) (interface_down)	‘up’
memory-stats_used-percent： software-oper_cisco-platform-software_control-processes_control-process_memory-stats_used-percent	>90% (memory_stress_stop)	<=90
software-oper_cisco-platform-software_control-processes_control-process_memory-stats_free-percent	<65 (memory_stress_stop)	>=65
software-oper_cisco-platform-software_control-processes_control-process_memory-stats_free-number	<250000 (memory_stress_stop)	>=250000
software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_idle	0 (<0.5) (Bridge_delif)	>0.5
software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_nice	0 (<0.5) (vcpu_overload_stop)	>0.5

label

Verify by some features

Feature Profile 1 （features name）



	Name	Data type
176	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_used-number	REAL
177	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_used-percent	REAL
178	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_free-number	REAL
179	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_free-percent	REAL
180	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_available-number	REAL
181	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_available-percent	REAL
182	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_committed-number	REAL
183	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_committed-percent	REAL
184	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_status_warning-threshold-percent	REAL
185	software-oper_cisco-platform-software_control-processes_control-process_memory-stats_status_critical-threshold-percent	REAL
186	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_name	REAL
187	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_user	REAL
188	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_system	REAL
189	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_nice	REAL
190	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_idle	REAL
191	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_irq	REAL
192	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_sirq	REAL
193	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_io-wait	REAL
194	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_name.1	REAL
195	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_user.1	REAL
196	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_system.1	REAL
197	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_nice.1	REAL
198	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_idle.1	REAL
199	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_irq.1	REAL
200	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_sirq.1	REAL
201	software-oper_cisco-platform-software_control-processes_control-process_per-core-stats_per-core-stat_io-wait.1	REAL
202	five-g_rm_registeredsubnbrmean	REAL
203	five-g_rm_registeredsubnbrmax	REAL
204	five-g_rm_reginitreq	REAL
205	five-g_rm_reginitsucc	REAL

Feature Profile (2)(values)

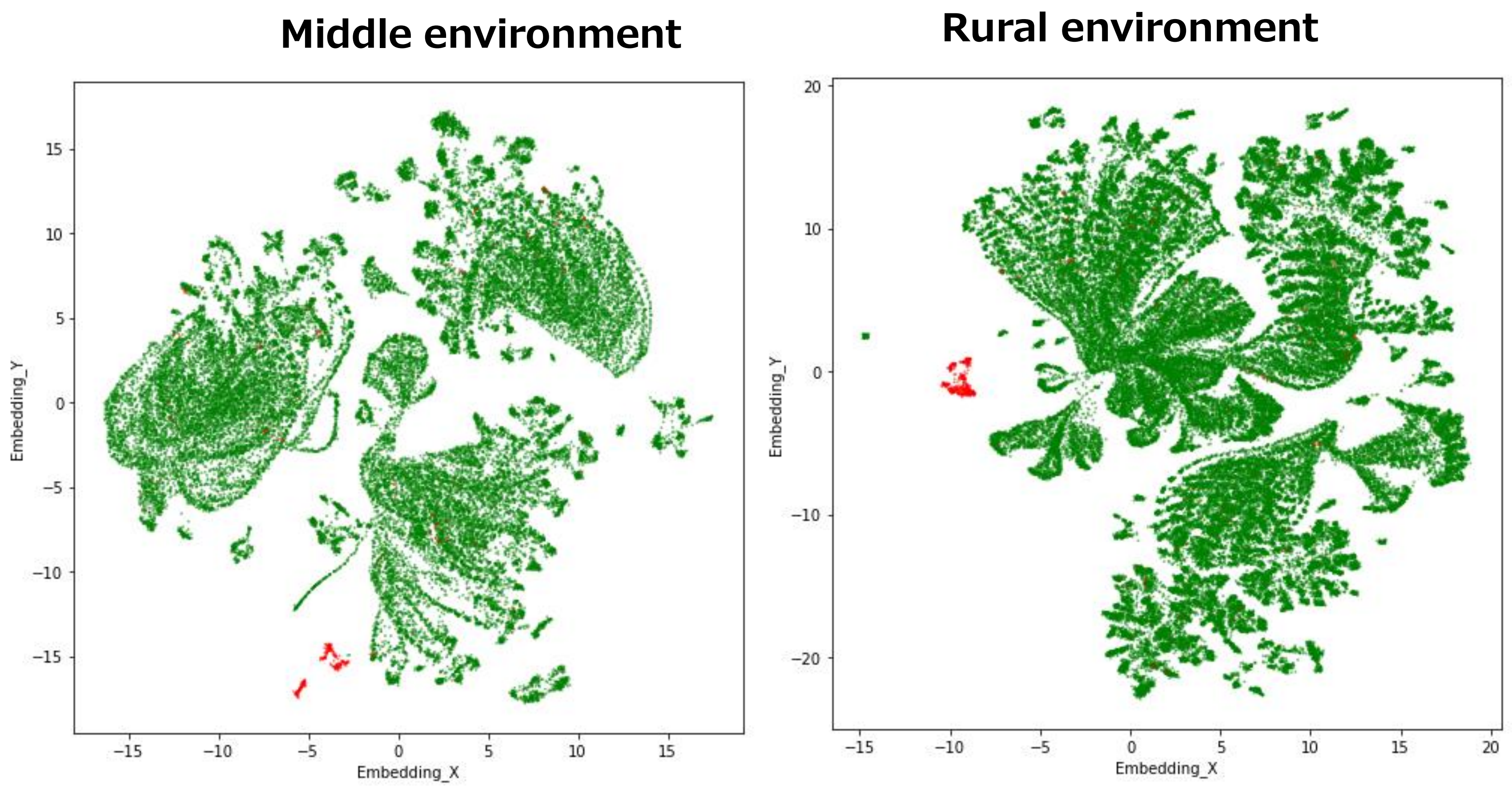
SQLiteStudio (3.2.1) - [train_amfa1 (KDDI5GAI)]

Total rows loaded: 98533

	ftware-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	software-	five-g_rm	five-g_rm	five-g_rm	five-g_rm	datetime
1	263	88	93	0	11.09	5.84	0	83.07	0	0	0	1	11.23	5.67	0	82.7	0	0.4	0	0	0	3488568	6301380	20210202100420
2	1151	88	93	0	0.7	0.3	0	98.49	0	0.1	0.4	1	0.9	0.5	0	98.49	0	0.1	0	0	0	3488568	6301380	20210202100740
3	756	88	93	0	3.29	2.2	0	94.41	0	0	0.1	1	5.93	3.82	0	89.85	0	0	0.4	0	0	3488568	6301380	20210202100440
4	1151	88	93	0	0.4	0.6	0	98.7	0	0	0.3	1	0.9	0.4	0	98.7	0	0	0	0	0	3488568	6301380	20210202100340
5	1087	88	93	0	0.5	0.6	0	98.5	0	0	0.4	1	1.2	0.4	0	98.4	0	0	0	0	0	3488568	6301380	20210202100410
6	1028	88	93	0	1.1	0.7	0	98.1	0	0.1	0	1	0.4	0.2	0	98.99	0	0	0.4	0	0	3488568	6301380	20210202100600
7	988	88	93	0	8.12	5.01	0	86.67	0	0	0.2	1	9.63	5.32	0	84.85	0	0.2	0	0	0	3488568	6301380	20210202100700
8	269	88	93	0	8.8	6.28	0	84.71	0	0	0.21	1	10.56	4.81	0	84.41	0	0.21	0	0	0	3488568	6301380	20210202100210
9	1654	88	93	0	2.6	1	0	96.1	0	0	0.3	1	2.52	0.91	0	96.58	0	0	0	0	0	3488568	6301380	20210202101030
10	208	88	93	0	3.03	1.31	0	95.56	0	0	0.1	1	3.13	1.21	0	95.45	0	0.1	0.1	0	0	3488568	6301380	20210202100850
11	777	88	93	0	16.09	4.74	0	78.87	0	0	0.3	1	15.7	9.28	0	74.23	0	0.69	0.1	0	0	3488568	6301380	20210202100540
12	1490	88	93	0	8.91	8.71	0	82.08	0	0	0.3	1	13.74	4.91	0	81.04	0	0.3	0	0	0	3488568	6301380	20210202100840
13	1216	88	93	0	1.67	0.79	0	97.54	0	0	0	1	2.36	0.98	0	96.36	0	0.1	0.2	0	0	3488568	6301380	20210202101210
14	1050	88	93	0	2.61	2.01	0	95.29	0	0	0.1	1	2.71	1.11	0	95.88	0	0	0.3	0	0	3488568	6301380	20210202100520
15	702	88	93	0	19.34	10.98	0	69.69	0	0	0	1	17.51	11.77	0	69.92	0	0.8	0	0	0	3488568	6301380	20210202100650
16	1660	88	93	0	9.34	2.98	0	87.38	0	0	0.3	1	6.86	5.17	0	87.18	0	0.4	0.4	0	0	3488568	6301380	20210202100830
17	1508	88	93	0	0.2	0.2	0	99.4	0	0.1	0.1	1	1.3	0.3	0	98.1	0	0	0.3	0	0	3488568	6301380	20210202100640
18	345	88	93	0	17.58	5.49	0	76.92	0	0	0	1	13.96	10.04	0	75.3	0	0.5	0.2	0	0	3488568	6301380	20210202101050
19	538	88	93	0	0.4	0.2	0	99	0	0	0.4	1	1.1	0.4	0	98.49	0	0	0	0	0	3488568	6301380	20210202100500
20	1698	88	93	0	0.6	0.3	0	98.99	0	0	0.1	1	1.01	0.3	0	98.39	0	0.2	0.1	0	0	3488568	6301380	20210202100240
21	934	88	93	0	0.4	0.4	0	99.2	0	0	0	1	1.1	0.5	0	98	0	0.1	0.3	0	0	3488568	6301380	20210202100810
22	169	88	93	0	0.69	0.3	0	99.01	0	0	0	1	1.57	0.79	0	97.25	0	0	0.39	0	0	3488568	6301380	20210202100930
23	1313	88	93	0	1.5	0.7	0	97.8	0	0	0	1	2.2	0.8	0	96.69	0	0	0.3	0	0	3488568	6301380	20210202100550
24	1397	88	93	0	1.09	0.64	0	98.18	0	0	0.09	1	0.36	0.09	0	99.54	0	0	0	0	0	3488568	6301380	20210202100800
25	1471	88	93	0	1.55	0.77	0	97.68	0	0	0	1	1.98	0.77	0	97.13	0	0	0.11	0	0	3488568	6301380	20210202100610
26	91	88	93	0	4.6	1.9	0	93.19	0	0.1	0.2	1	3.69	1.6	0	94.62	0	0	0.1	0	0	3488568	6301380	20210202100320
27	1645	88	93	0	1.5	1	0	97.5	0	0	0	1	2.79	1.39	0	95.32	0	0.1	0.4	0	0	3488568	6301380	20210202100400
28	325	88	93	0	6.36	3.74	0	89.9	0	0	0	1	7.75	3.12	0	88.62	0	0.2	0.3	0	0	3488568	6301380	20210202100910
29	233	88	93	0	0.46	0.18	0	99.27	0	0	0.09	1	0.73	0.36	0	98.63	0	0	0.27	0	0	3488568	6301380	20210202100530
30	1492	88	93	0	1.5	1.1	0	97.19	0	0	0.2	1	2.11	1.5	0	96.29	0	0	0.1	0	0	3488568	6301380	20210202100220
31	710	88	93	0	0.5	0.3	0	98.4	0	0	0.8	1	1.6	0.5	0	97.9	0	0	0	0	0	2980384	5395628	20210129052150
32	174	88	93	0	0.1	0	0	99.9	0	0	0	1	1.31	0.3	0	97.88	0	0	0.51	0	0	3097683	5611758	20210130034630
33	1679	88	93	0	0.6	0.5	0	98.7	0	0	0.2	1	1.4	0.6	0	97.9	0	0	0.1	0	0	3153315	5717056	20210130141240
34	882	88	93	0	0.4	0.2	0	99.2	0	0	0.2	1	1.2	0.4	0	98.4	0	0	0	0	0	3320819	6019960	20210131224040
35	399	88	93	0	1.2	0.7	0	98.09	0	0	0	1	1.91	0.8	0	96.99	0	0.1	0.2	0	0	3285112	5963452	20210131145050
36	485	88	93	0	8.31	3.9	0	87.49	0	0	0.3	1	7.04	3.52	0	89.03	0	0.4	0	4	4	3195458	5797354	20210130215800
37	411	88	93	0	0.8	0.7	0	98.29	0	0	0.2	1	0.6	0.3	0	99	0	0.1	0	0	0	3157215	5724656	20210130145850
38	787	88	93	0	0.7	0.7	0	98.3	0	0	0.3	1	1.1	0.8	0	98	0	0.1	0	0	0	3236808	5876814	20210131052500
39	73	88	93	0	0.2	0.1	0	99.3	0	0	0.4	1	0.9	0.4	0	98.7	0	0	0	0	0	3068964	5567700	20210129210750
40	1618	88	93	0	26.63	73.37	0	0	0	0	0	1	35.41	64.19	0	0	0	0.4	0	0	0	2936147	5313468	20210128210010
41	854	88	93	0	0.7	0.2	0	98.7	0	0.1	0.3	1	1.4	0.5	0	98.1	0	0	0	0	0	2928583	5301192	20210128192440
42	664	88	93	0	19.22	7.41	0	73.17	0	0	0.2	1	14.08	9.05	0	76.16	0	0.7	0	0	0	3143170	5700496	20210130120600

Data clustering by unsupervised model (UMAP)

- Normal
- Abnormal



Machine Learning by Xgboost 2-class model

Task-1 課程#1

Training by a (urban) : (Normal: 292,306; Abnormal: 3,293)
Test by c (rural) : (Normal: 73,082; Abnormal: 807)

	Precision	Recall	F-1
Xgboost	100%	100%	100%

Confusion Matrix		
	73,082	0(FP)
	0	807(TP)

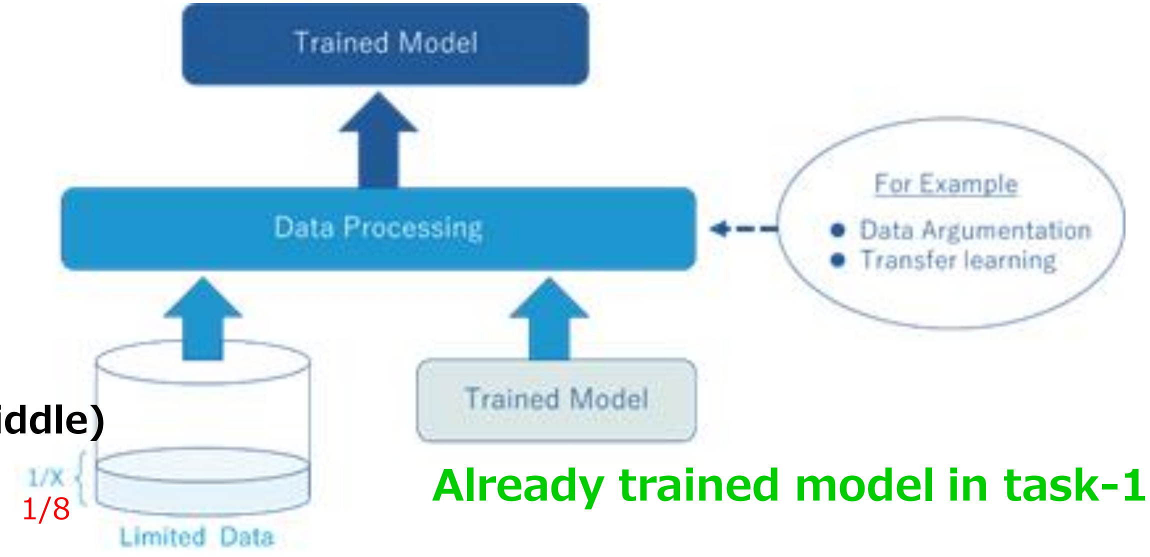
Training by c (rural) : (Normal:292,395; Abnormal: 3,204)
Test by a (urban) : (Normal: 72,534; Abnormal: 1,352)

	Precision	Recall	F-1
Xgboost	100%	100%	100%

Confusion Matrix		
	72,534	0(FP)
	0	1,352(TP)

Increment Learning by XGB Increment model

Task-2 課程#2



Training by argumentation new data b (middle)

Already trained model in task-1 by data a(urban)

- ◆ **Trained model by a (urban) (Normal: 292,306; Abnormal: 3,293) ＋ Training by b (middle)(Nomal:36,668; Abnormal: 412)**
- ■ **Test by b (middle) : (Normal: 73,129; Abnormal: 812)**

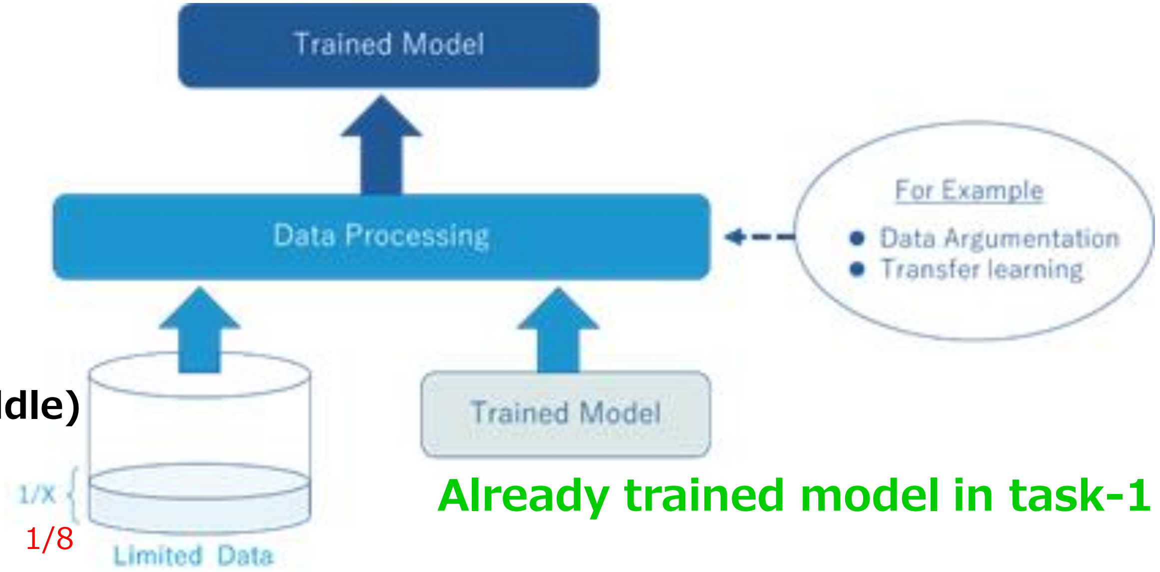
	Precision	Recall	F-1
Xgboost	100%	83%	91%

Confusion Matrix		
	73,129	0(FP)
	135	677(TP)

Increment Learning by XGB Increment model

Task 2 課程#2

Training by agumentation data b (middle)



Already trained model in task-1 by data c(rural)

- ◆ Trained model by c (urban) (Normal: 292,395; Abnormal: 3,204) + Training by b (middle)(Nomal:36,668; Abnormal: 412)
- ■ Test by b (middle) : (Normal: 73,129; Abnormal: 812)

	Precision	Recall	F-1
Xgboost	87%	100%	93%

Confusion Matrix		
	73,005	124(FP)
	0	812(TP)

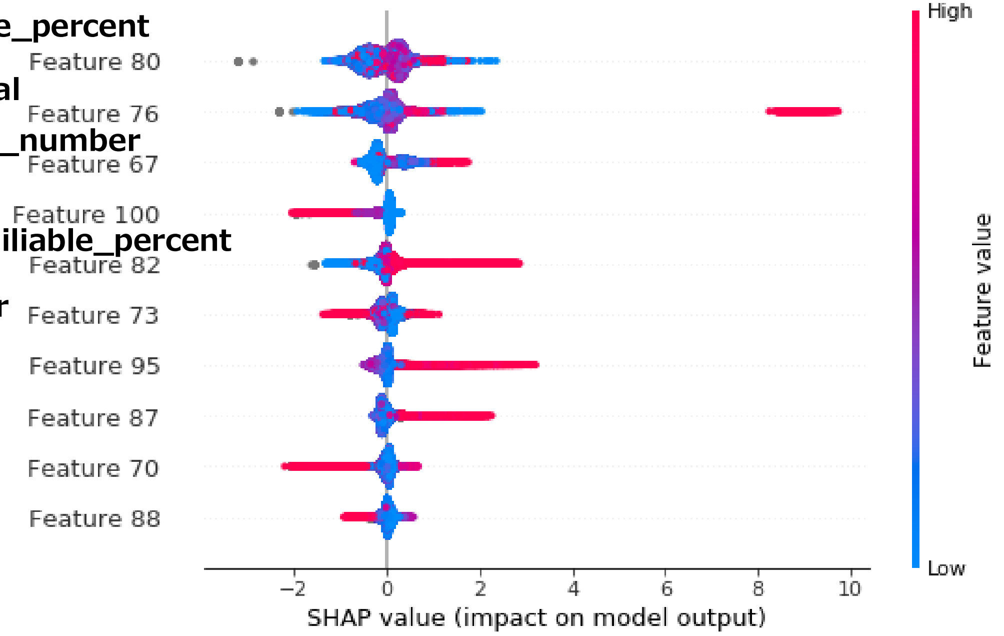
SHAP for Feature analysis
(SHapley Additive exPlanations)

障害原因分析

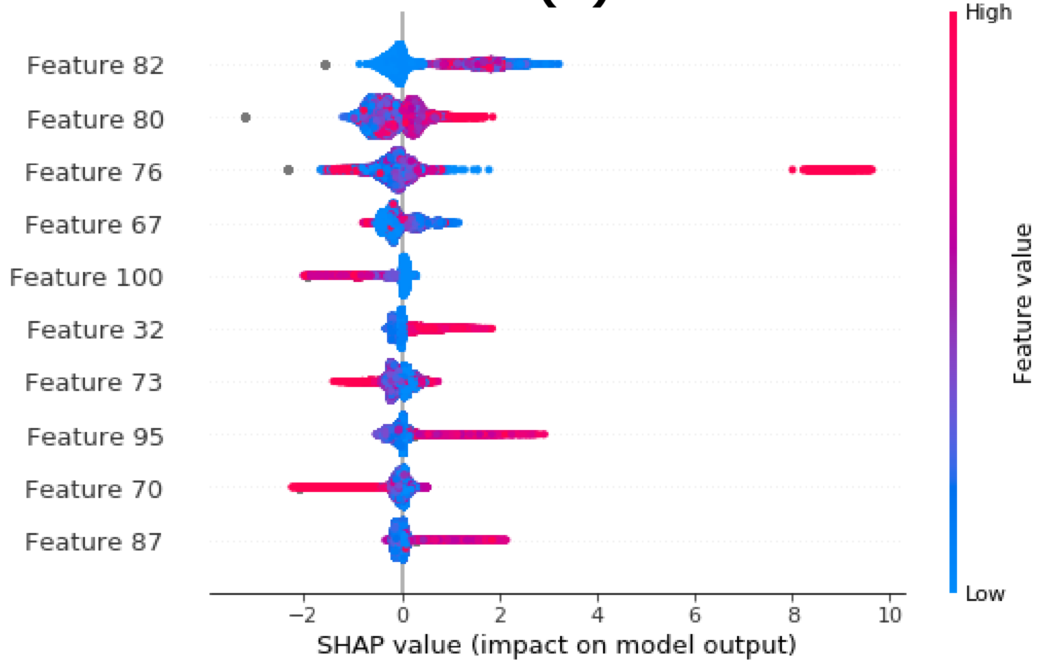
Machine learning based on tree models is comprehensive. When tree based models are applied in applications, we expect models can be interpretable, which means we can understand how the model uses input features to make predictions. ShapTreeExplainer bridges theory to practice based on classic game theoretic Shapley values. It makes possible to evaluate features impact when network failure happened.

Train in Rural (c)

Memory-stats_free_percent
Memory-stats_total
Load-avg-minutes_number
Per-core-stat_irq
Memory-stats_availiable_percent
Per-core-stat_user



Test in Urban (a)



By calculate the train or test data's SHAP values, we can know how the features influence the predictions results. So when the failure is detected, we can infer the most important failure reasons by SHAP value rank.

Improvement proposal 改善点など

- Data clustering or mapping is helpful to confirm the feature's effectiveness
- More deep analysis : how different information (CPU load info, traffic info, five-g session info) is related failure by training AI models respectively.
- How to choose features should be investigated more deeply.
- Comparing neural network models with tree models for more flexible deployment.