Anomaly Detection – Water Pump

AM20216801 YONG JUN KIM

Index

Anomaly Detection

- Anomaly detection refers to finding data showing different patterns in data. Such data is called anomaly, and anomaly detection is widely used in various fields including fraud detection, intrusion detection, and safety management.

Presentation Subject

Data

	Unnamed: 0	timestamp	sensor_00	sensor_01	sensor_02	sensor_03	sensor_04	sensor_05	sensor_06	sensor_07	...	sensor_43	sensor_44	sensor_45	sensor_46	sensor_47	sensor_48	sensor_49	sensor_50	sensor_51	machine_status
0	0	2018-04-01 00:00:00	2.465394	47.09201	53.2118	46.310760	634.3750	76.45975	13.41146	16.13136	...	41.92708	39.641200	65.68287	50.92593	38.194440	157.9861	67.70834	243.0556	201.3889	NORMAL
1	1	2018-04-01 00:01:00	2.465394	47.09201	53.2118	46.310760	634.3750	76.45975	13.41146	16.13136	...	41.92708	39.641200	65.68287	50.92593	38.194440	157.9861	67.70834	243.0556	201.3889	NORMAL
2	2	2018-04-01 00:02:00	2.444734	47.35243	53.2118	46.397570	638.8889	73.54598	13.32465	16.03733	...	41.66666	39.351852	65.39352	51.21528	38.194443	155.9606	67.12963	241.3194	203.7037	NORMAL
3	3	2018-04-01 00:03:00	2.460474	47.09201	53.1684	46.397568	628.1250	76.98898	13.31742	16.24711	...	40.88541	39.062500	64.81481	51.21528	38.194440	155.9606	66.84028	240.4514	203.1250	NORMAL
4	4	2018-04-01 00:04:00	2.445718	47.13541	53.2118	46.397568	636.4583	76.58897	13.35359	16.21094	...	41.40625	38.773150	65.10416	51.79398	38.773150	158.2755	66.55093	242.1875	201.3889	NORMAL

timestamp	sensor_00	sensor_01	sensor_02	sensor_03	sensor_04	sensor_05	sensor_06	sensor_07	sensor_08	sensor_09	...	sensor_51	machine_status
2018-04-12 21:55:00	0.000000	53.342010	52.821180	43.402775	202.526031	49.79289	3.219039	16.890910	16.869210	15.082470	...	324.652800	BROKEN
2018-04-18 00:30:00	1.093982	42.534720	47.699650	41.449650	206.038757	60.30106	12.304690	15.154800	14.185470	13.867190	...	183.738400	BROKEN
2018-05-19 03:18:00	2.258796	47.265630	52.734370	43.446178	200.115738	66.14643	13.592300	15.914350	15.147570	14.793110	...	257.523100	BROKEN
2018-05-25 00:30:00	2.321759	47.482640	51.475693	42.795135	612.152800	67.30158	14.062500	16.608800	15.943290	15.596060	...	267.361100	BROKEN
2018-06-28 22:00:00	0.364005	40.190970	45.225690	40.190971	201.368622	0.00000	11.335360	15.270540	15.183740	15.118630	...	202.699667	BROKEN
2018-07-08 00:11:00	0.001968	45.138890	52.907990	45.355900	500.000000	1.40131	0.028935	0.036169	0.036169	0.007234	...	174.768500	BROKEN
2018-07-25 14:00:00	2.318808	45.833332	52.994790	43.880210	420.503448	72.52040	14.185470	16.247110	15.697340	15.053530	...	205.729200	BROKEN

7 rows × 60 columns

Graph

- Sensor_04 from total Sensor_00 ~ Sensor_50
- 7 X of each graphs are presenting Broken part
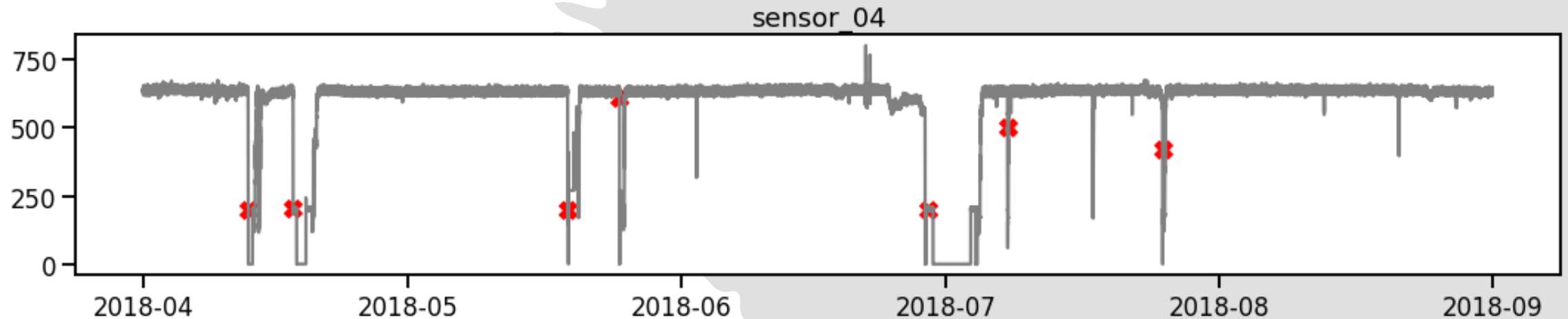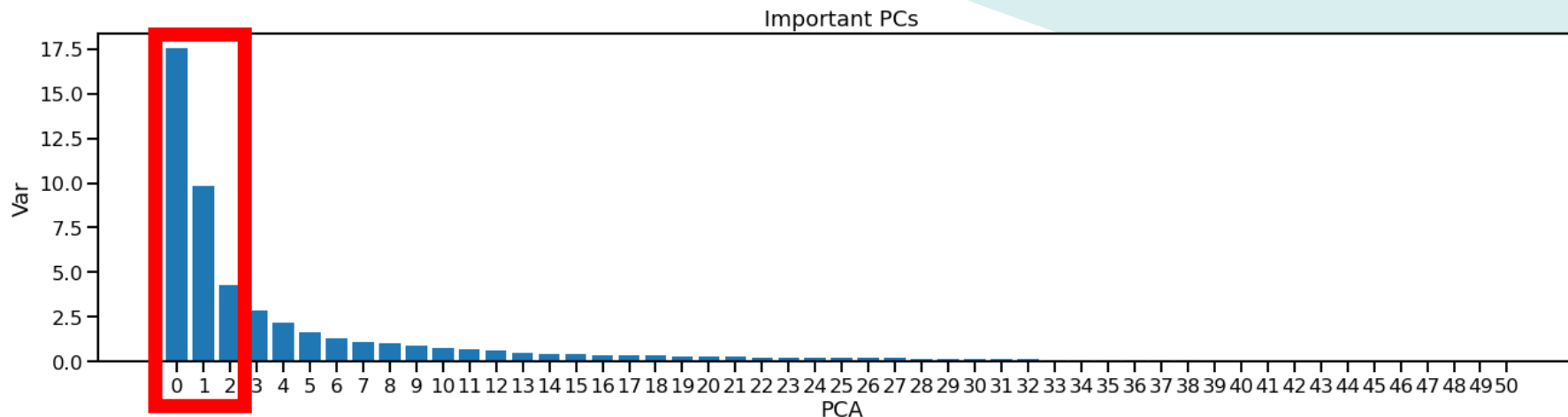
Isolation Forest

- Isolation forest basically expresses the dataset in the form of a decision tree.

- In order to separate the steady-state values, we have to climb deep down the decision tree.

- Conversely, outliers can be separated at the top of the decision tree.

- Using this characteristic, normal and abnormal are separated based on how many times it takes to climb down the decision tree to separate.
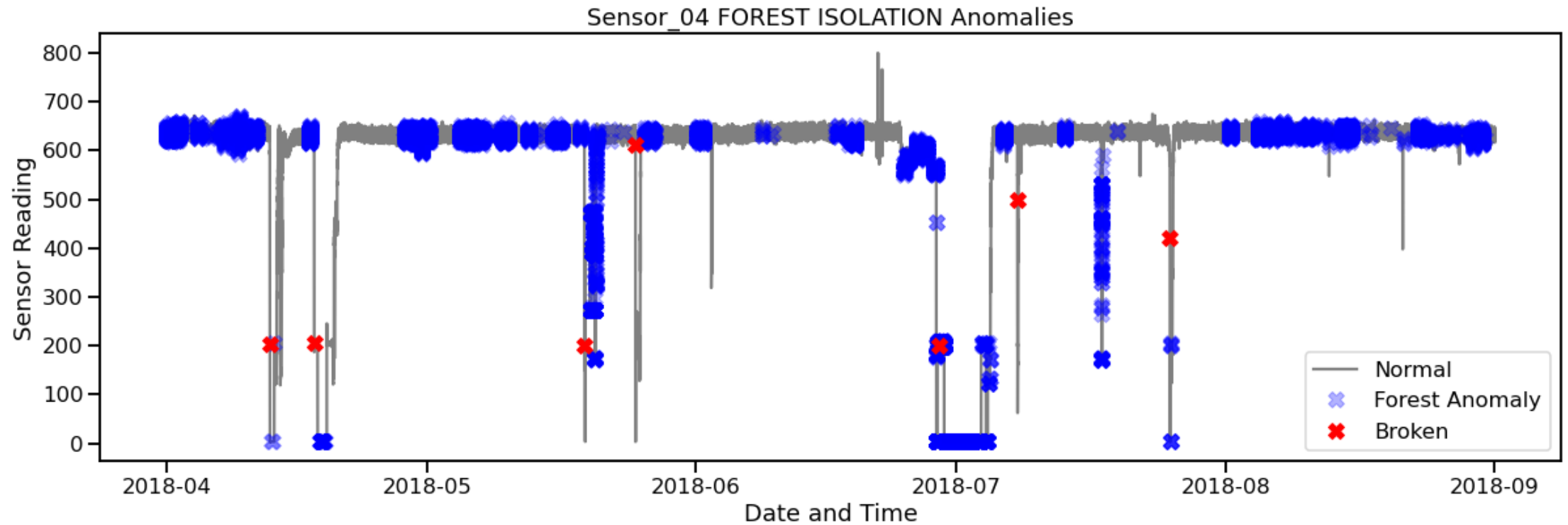
Advantages of Isolation Forest

- Compared to the cluster-based anomaly detection algorithm, the amount of computation is very low.

- A robust model can be created.

Important PCs and Number of records in cluster

- Principal Component Analysis (PCA) is a method to reduce the dimension by making a representative principal component by linear combination of the measured variables.

Important PCs

Anomaly after trained with Isolation Forest

Sensor_04 FOREST ISOLATION Anomalies

Thank you