



**DITEN**

Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture  
Polytechnic School, University of Genoa

# Alla scoperta del Data Mining

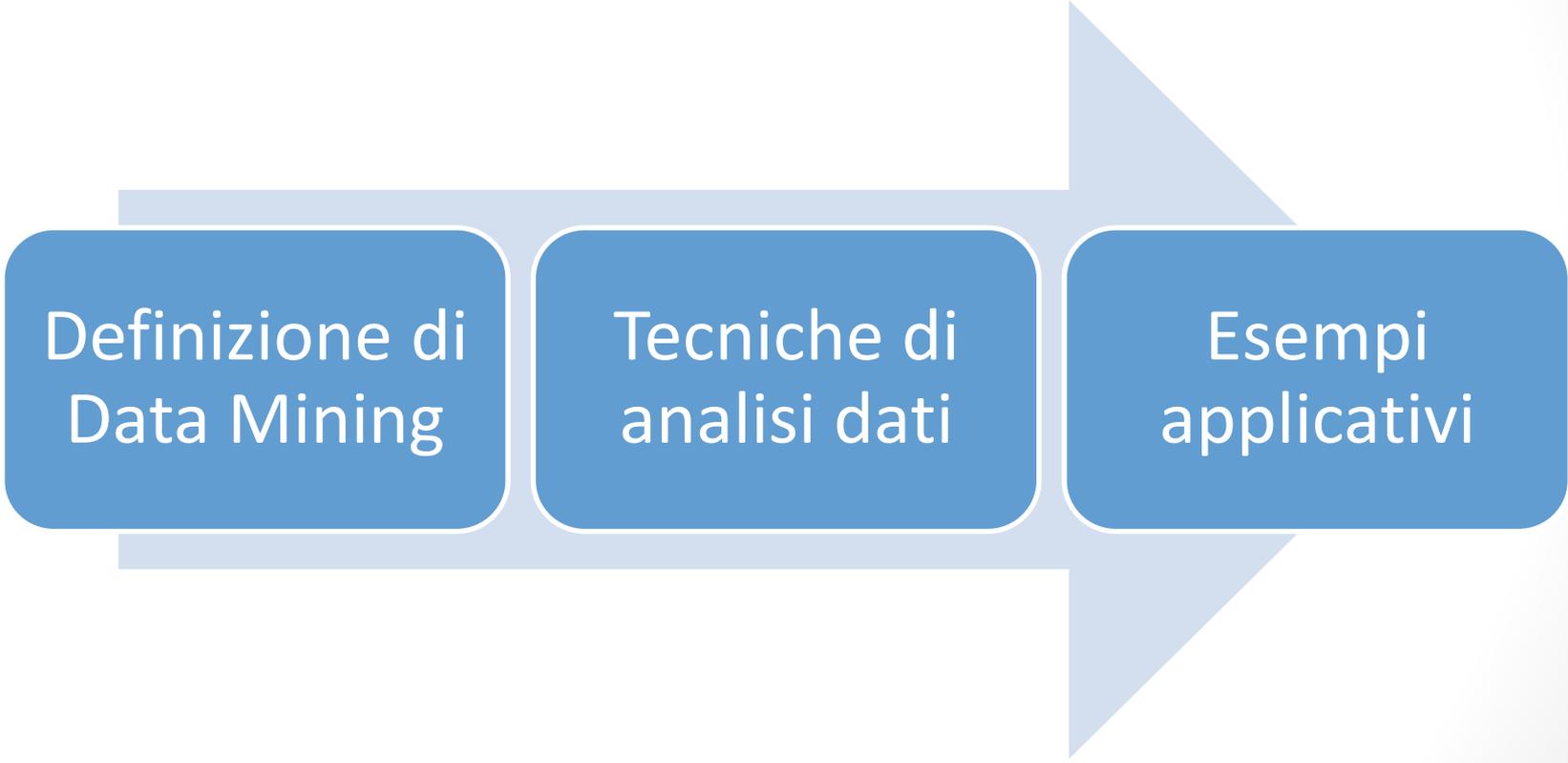
**Paolo Pinceti, Micaela Caserza Magro**  
**Università di Genova - Dipartimento DITEN**



*MAI Lab*

*Measurement & Automation for Industry*

# Outlines



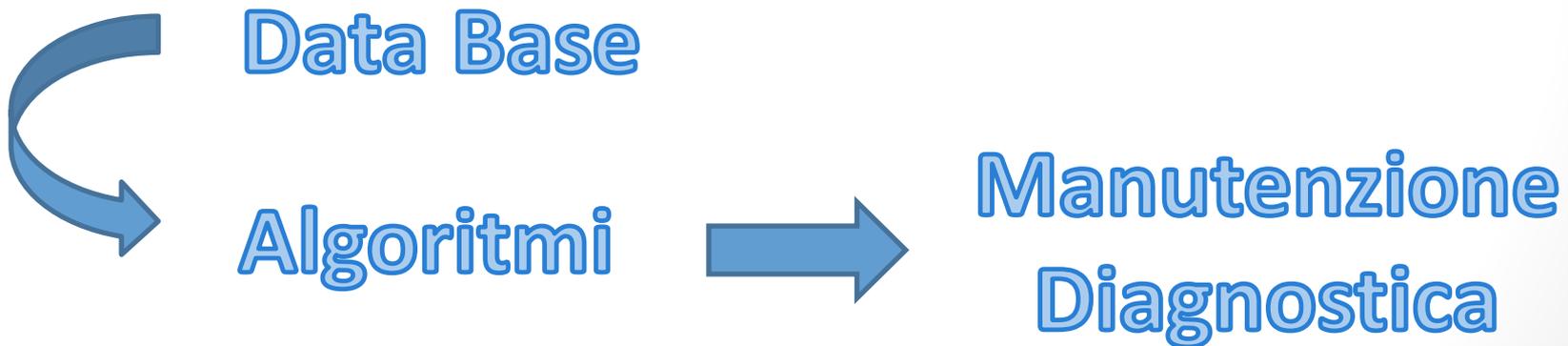
Definizione di  
Data Mining

Tecniche di  
analisi dati

Esempi  
applicativi

# Cosa è il Data Mining

Il Data Mining è un processo di estrazione di conoscenza da **banche dati** di grandi dimensioni tramite **l'applicazione di algoritmi** che individuano le **associazioni "nascoste" tra le informazioni e le rendono visibili**. Gli algoritmi di Data Mining sono stati sviluppati per far fronte all'esigenza di sfruttare il patrimonio informativo contenuto nelle grandi raccolte di dati che abbiamo a disposizione.



# Big Data

Strutture HW e tecniche SW per acquisizione, immagazzinamento elaborazione, ricerca di grandi quantità di dati.

$10^{11} \div 10^{15}$  byte

se 1 byte è 1 granello di sabbia:

$10^6$  byte = 1 megabyte = 1 pugno di sabbia

$10^9$  byte = 1 gigabyte = 1 scatola piena di sabbia

$10^{12}$  byte = 1 terabyte = 1 spiaggia grande come un campo da calcio

$10^{15}$  byte = 1 petabyte = molti chilometri di spiaggia

# Le «V» dei Big Data





chi usa  
BD ?



La domanda che sorge è:

Ma Big Data cosa ha a che fare con  
l'Automazione Industriale?

In senso stretto la risposta corretta è:

**NULLA !**

perchè....

Volume → kbyte - Mbyte

Velocità → non essenziale

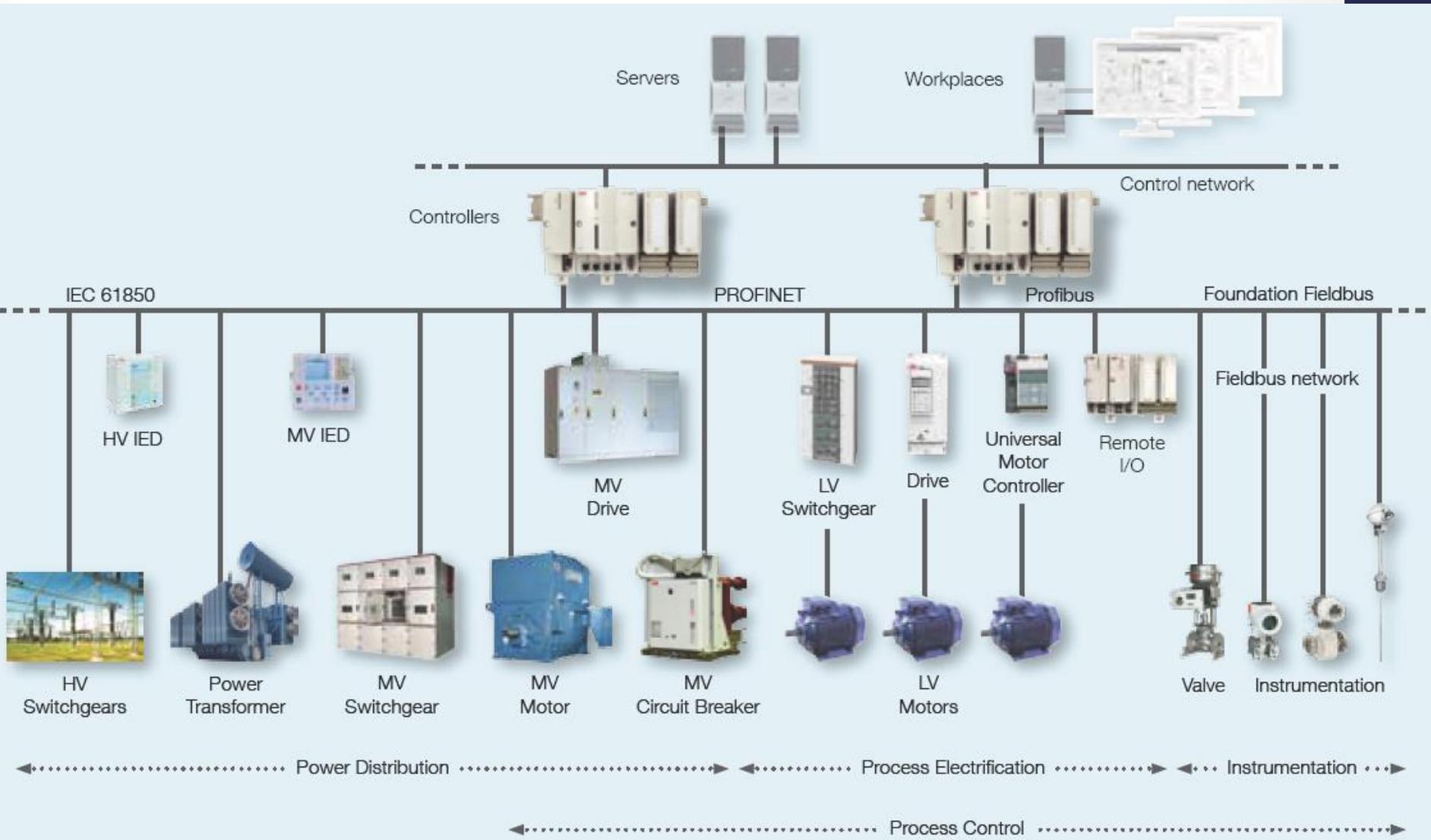
Varietà → dati omogenei

Visualizzazione → esterna

Validazione → a monte

però....

# Struttura database «processo»



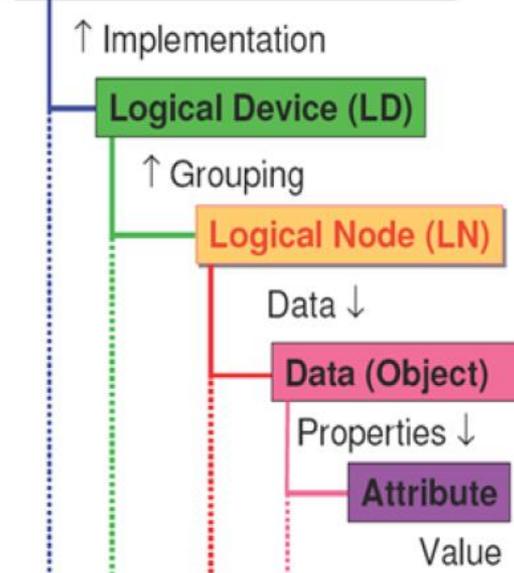
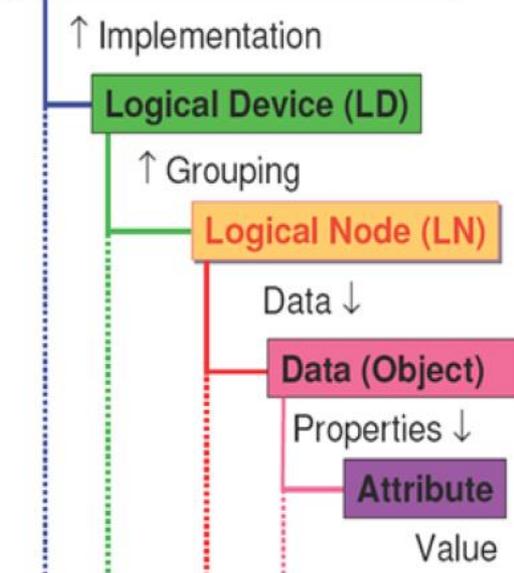
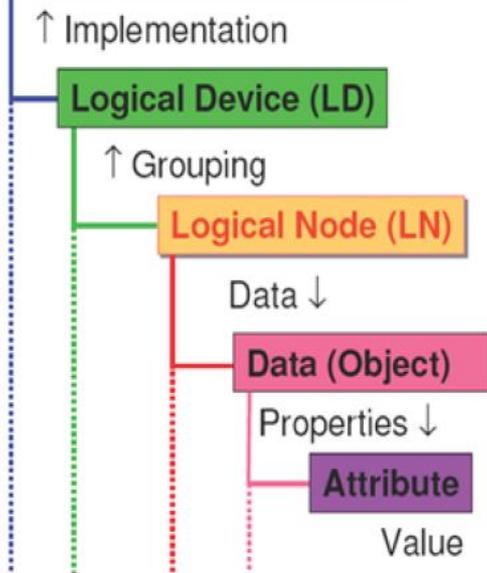
# Struttura database «elettrico»



Physical Device (IED)  
*defined as Server*

Physical Device (IED)  
*defined as Server*

Physical Device (IED)  
*defined as Server*



In entrambi i casi, il database di impianto contiene:

- dati di processo real-time
- dati storici
- allarmi e eventi
- dati diagnostici

sui quali è possibile applicare tecniche di analisi statistiche.

# Contenuti del DB di processo

Dati real-time

Dati storici

Allarmi ed eventi

Dati diagnostici

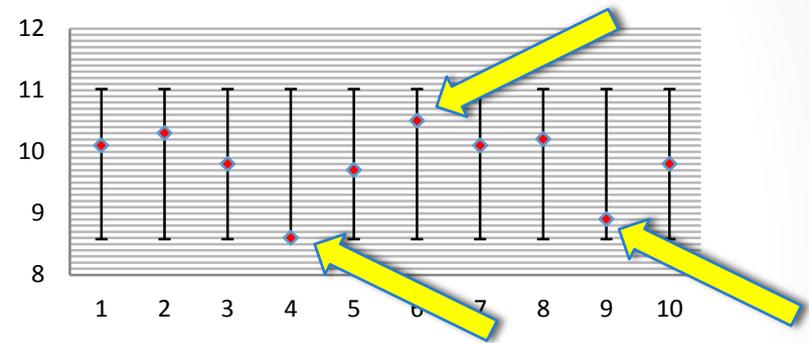
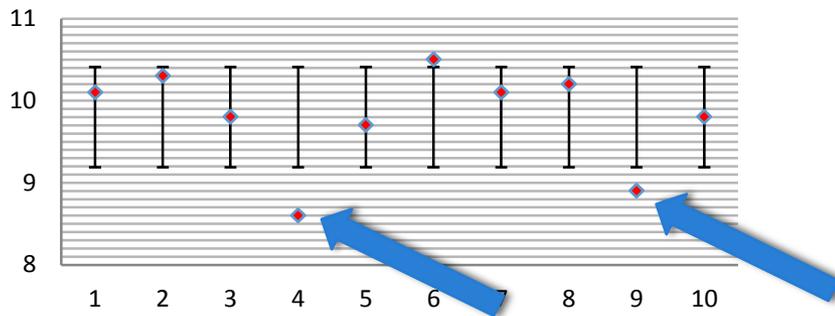
Analisi statistica

# Analisi dei dati

Macchina/Impianto	#1	#2	#3	...	#m
Parametro #1	$p_{11}$	$p_{12}$	$p_{13}$	...	$p_{1m}$
Parametro #2	$p_{21}$	$p_{22}$	$p_{23}$	...	$p_{2m}$
Parametro #3	$p_{31}$	$p_{32}$	$p_{33}$	...	$p_{3m}$
.....	...	...	...	...	...
Parametro #n	$p_{n1}$	$p_{n2}$	$p_{n3}$		$p_{nm}$

Orizzontale: analisi di proprietà uguali  
di apparati diversi

# Analisi dei valori medi



$$\sigma_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \langle x \rangle)^2}{N}}$$

varianza

numero di campioni

media aritmetica

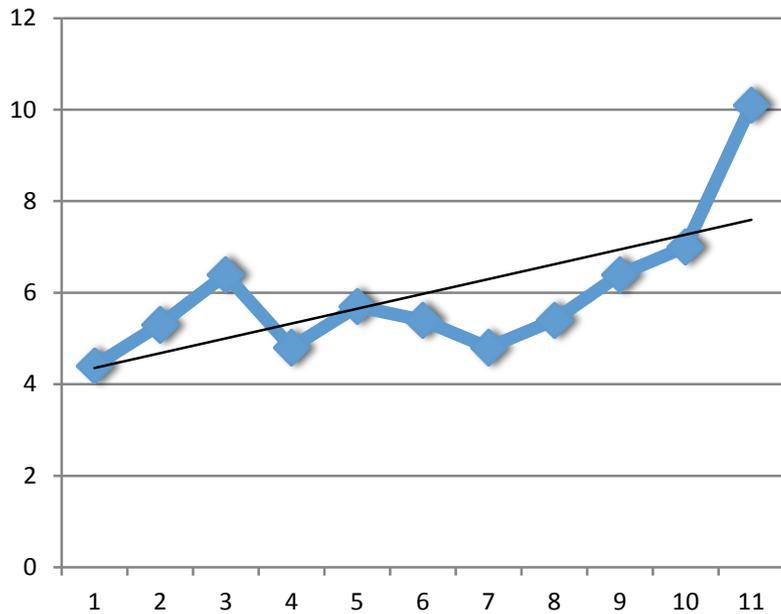
# Analisi dei dati

Macchina/Impianto	#1	#2	#3	...	#m
Parametro #1	$p_{11}$	$p_{12}$	$p_{13}$	...	$p_{1m}$
Parametro #2	$p_{21}$	$p_{22}$	$p_{23}$	...	$p_{2m}$
Parametro #3	$p_{31}$	$p_{32}$	$p_{33}$	...	$p_{3m}$
.....	...	...	...	...	...
Parametro #n	$p_{n1}$	$p_{n2}$	$p_{n3}$	...	$p_{nm}$

Verticale: analisi dell'andamento di una o più proprietà di un singolo apparato

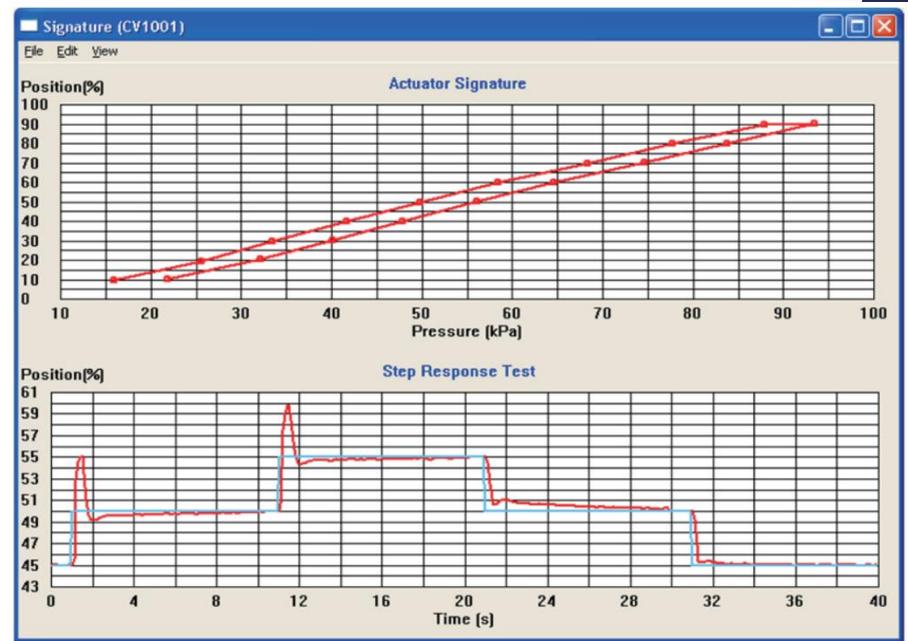
# Analisi verticale

## Analisi di trend



su parametri funzionali e/o su anomalie o guasti analizzati su un periodo sufficientemente lungo

es. ciclo di lavoro di un attuatore di valvola



## Analisi di firma

# Analisi dei dati

Macchina/Impianto	#1	#2	#3	...	#m
Parametro #1	$p_{11}$	$p_{12}$	$p_{13}$	...	$p_{1m}$
Parametro #2	$p_{21}$	$p_{22}$	$p_{23}$	...	$p_{2m}$
Parametro #3	$p_{31}$	$p_{32}$	$p_{33}$	...	$p_{3m}$
.....	...	...	...	...	...
Parametro #n	$p_{n1}$	$p_{n2}$	$p_{n3}$	...	$p_{nm}$

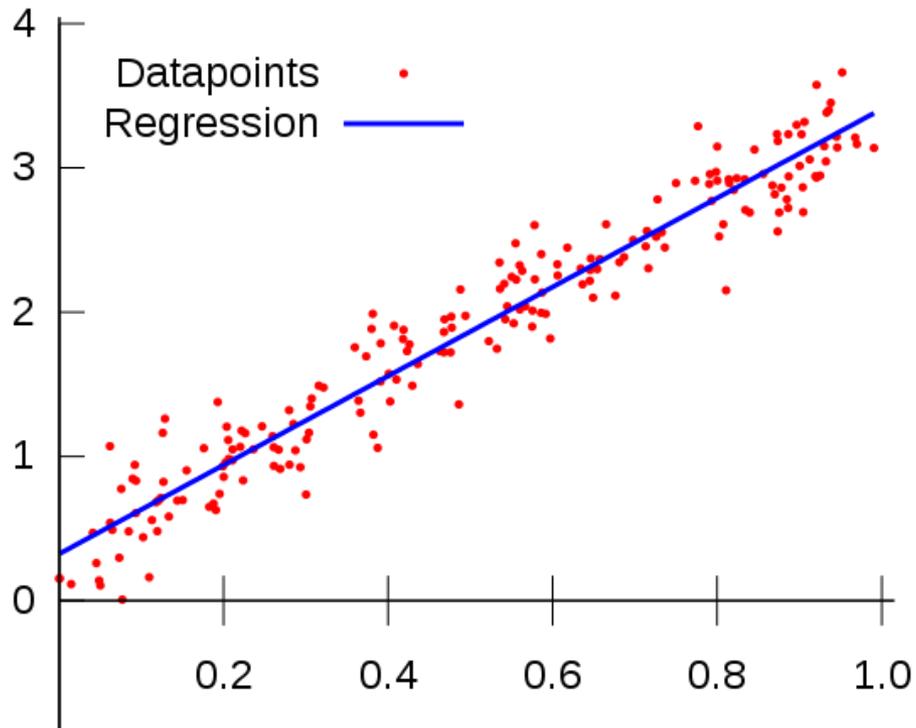
Mista (1): analisi incrociata di proprietà diverse dello stesso apparato

# Analisi dei dati

Macchina/Impianto	#1	#2	#3	...	#m
Parametro #1	$\rho_{11}$	$\rho_{12}$	$\rho_{13}$	...	$\rho_{1m}$
Parametro #2	$\rho_{21}$	$\rho_{22}$	$\rho_{23}$	...	$\rho_{2m}$
Parametro #3	$\rho_{31}$	$\rho_{32}$	$\rho_{33}$	...	$\rho_{3m}$
.....	...	...	...	...	...
Parametro #n	$\rho_{n1}$	$\rho_{n2}$	$\rho_{n3}$	...	$\rho_{nm}$

Mista (2): analisi incrociata di proprietà diverse di apparati diversi

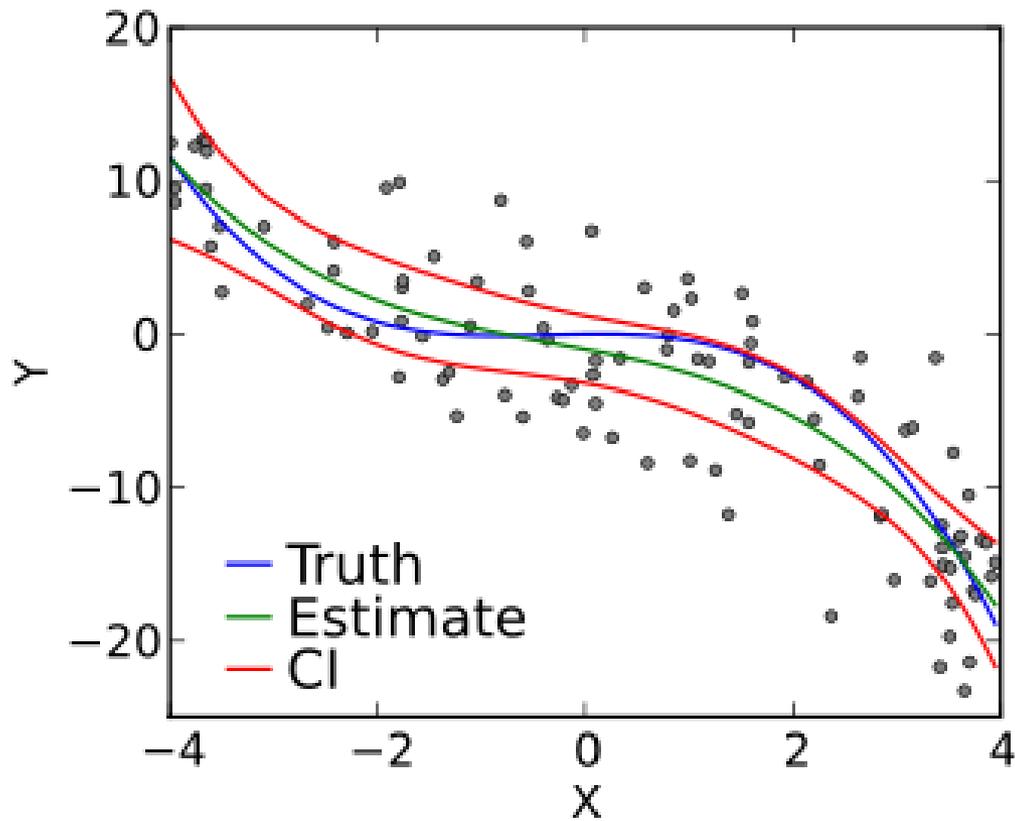
# Funzione di regressione



Regressione Lineare

E' necessaria un'analisi preliminare di ragionevolezza !!!

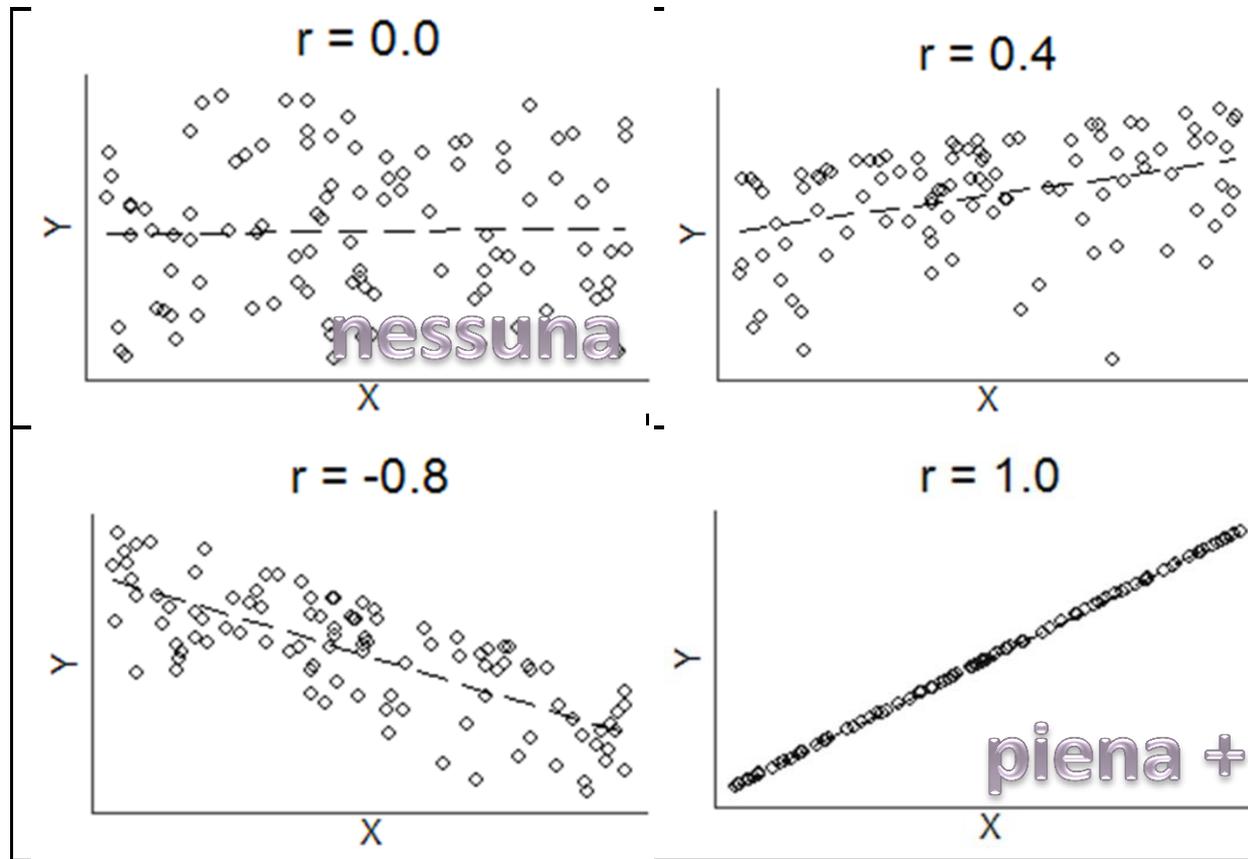
# Funzione di regressione



Regressione Polinomiale

# Coefficiente di correlazione

?



?

Attenzioni alle variabili nascoste!

X = vittime incendio  
Y = pompieri impegnati



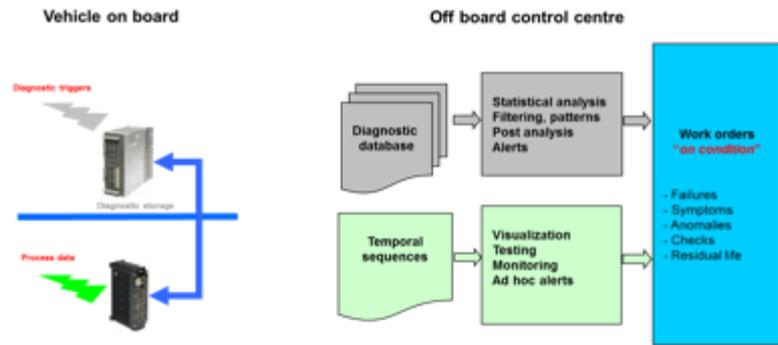
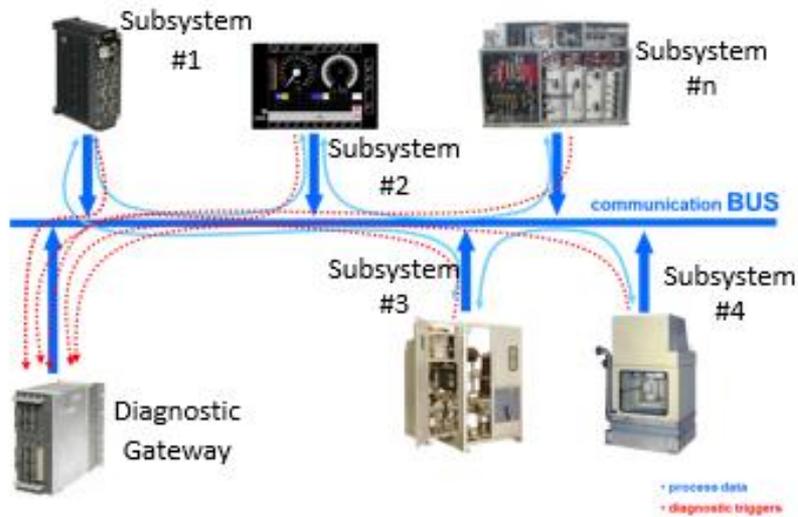
meno pompieri → meno vittime  
(Z = dimensione incendio)



**mct**

ESEMPIO DI APPLICAZIONE REALE

# Applicazione: flotta di locomotive



Raccolta dati e  
preparazione DB

2013\_06.csv [Sola lettura]

	A	B	C	D	E	F	G	H	I	J
1	Loc	version	DateON	HourON	DateOFF	HourOFF	SubSys	Proc	Message	LineVoltage
2	186111	7	01/06/2013	18:20:01	01/06/2013	18:20:33	PROT	ZSG	0607 - Lok	1,8
3	186111	7	01/06/2013	18:21:12	01/06/2013	18:22:15	PROT	ZSG	0607 - Lok	1,8
4	186111	7	01/06/2013	18:22:55	01/06/2013	18:25:04	PROT	ZSG	0607 - Lok	1,8
5	186111	7	01/06/2013	18:27:55	01/06/2013	18:29:26	PROT	ZSG	0607 - Lok	1,8
6	186111	7	01/06/2013	18:35:08	01/06/2013	19:12:55	PROT	ZSG	1087 - Strc	1,8
7	186111	7	01/06/2013	18:35:24	01/06/2013	19:12:55	PROT	ZSG	064F - Hau	1,8
8	186111	7	01/06/2013	18:41:29	01/06/2013	18:49:48	PROT	ZSG	0607 - Lok	1,8
9	186111	7	01/06/2013	18:55:01	01/06/2013	19:07:49	PROT	ZSG	0607 - Lok	1,8
10	186111	7	01/06/2013	19:09:20	01/06/2013	19:33:45	PROT	ZSG	0607 - Lok	1,8
11	186111	7	01/06/2013	19:13:41	01/06/2013	19:29:11	PROT	ZSG	1087 - Strc	0,9
12	186111	7	01/06/2013	19:13:55	01/06/2013	19:25:09	PROT	ZSG	064F - Hau	27,6
13	186111	7	01/06/2013	19:25:28	01/06/2013	19:29:10	PROT	ZSG	064F - Hau	27,6
14	186111	7	01/06/2013	19:30:19	01/06/2013	19:44:51	PROT	ZSG	1087 - Strc	1,8
15	186111	7	01/06/2013	19:30:43	01/06/2013	19:44:51	PROT	ZSG	064F - Hau	1,8
16	186111	7	01/06/2013	19:37:59	01/06/2013	20:02:59	PROT	ZSG	0607 - Lok	1,8
17	186111	7	01/06/2013	19:45:33	01/06/2013	19:53:44	PROT	ZSG	1087 - Strc	0,8
18	186111	7	01/06/2013	19:45:56	01/06/2013	19:52:15	PROT	ZSG	064F - Hau	28,2

# Applicazione: flotta di locomotive

Dati dal campo



Regole diagnostiche

Specifiche Produttore



# Dati diagnostici

No. Segnale	Significato	Utilità
064F	Numero manovre IR o IP	Indicatore usura interruttore meccanica
444D	Manovre di sicurezza	Indicatore usura interruttore elettrica/ malfunzionamento drive
0607	Velocità >3 km/h	
1087	Chiusura pantografo	Indicatore usura pantografo
21FF	Batteria attiva	



- Flotta 186: Multisistema
- Analisi di flotta
- Analisi da Gennaio a Novembre 2013

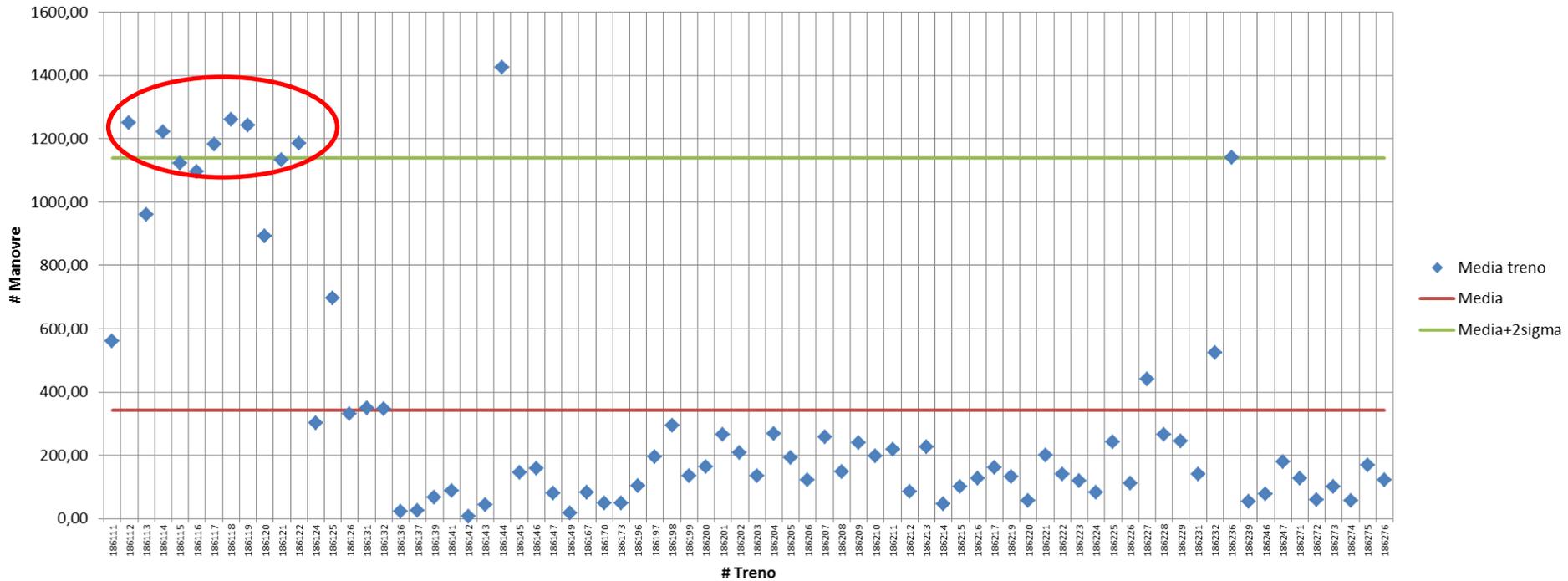


- KPI: numero manovre IR
- KPI: numero manovre IP
- KPI: numero manovre sicurezza

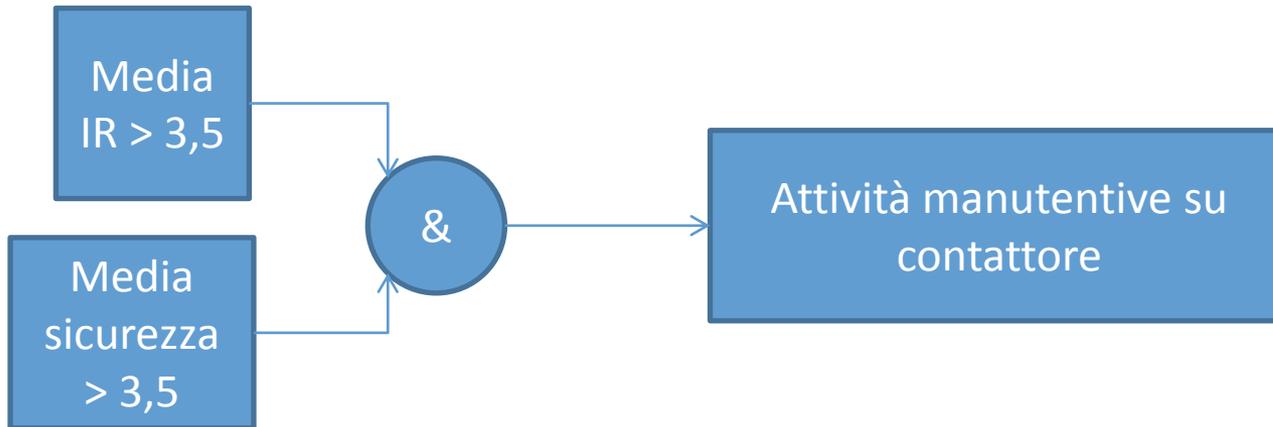


# Analisi orizzontale

## Media manovre IR



# Regole per la CBM



66% 

186112	186114	186118	186121	186122	186236
--------	--------	--------	--------	--------	--------

33% 

186117	186119	186144
--------	--------	--------

(falsi positivi)

( 30 )

(CBM = Condition Based Maintenance)



# Vantaggi ottenuti

Manutenzione  
ciclica

Manutenzione  
con Data Mining

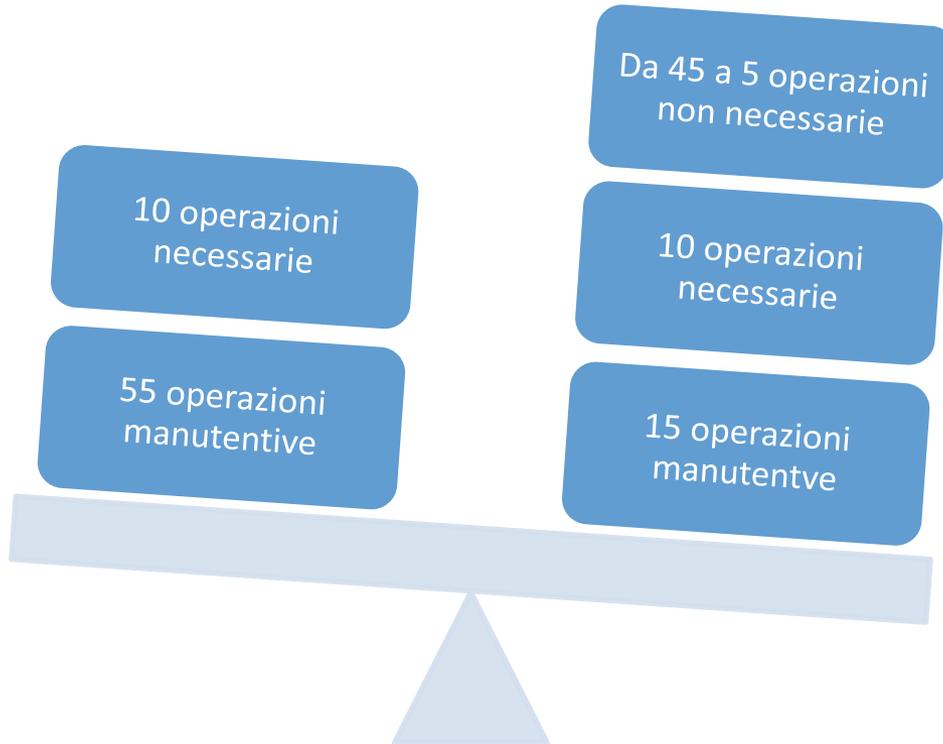
10 operazioni  
necessarie

55 operazioni  
manutentive

Da 45 a 5 operazioni  
non necessarie

10 operazioni  
necessarie

15 operazioni  
manutentive



# ESEMPIO APPLICATIVO: COMPRESSORE

# Dati diagnostici per compressori



Vibrazioni

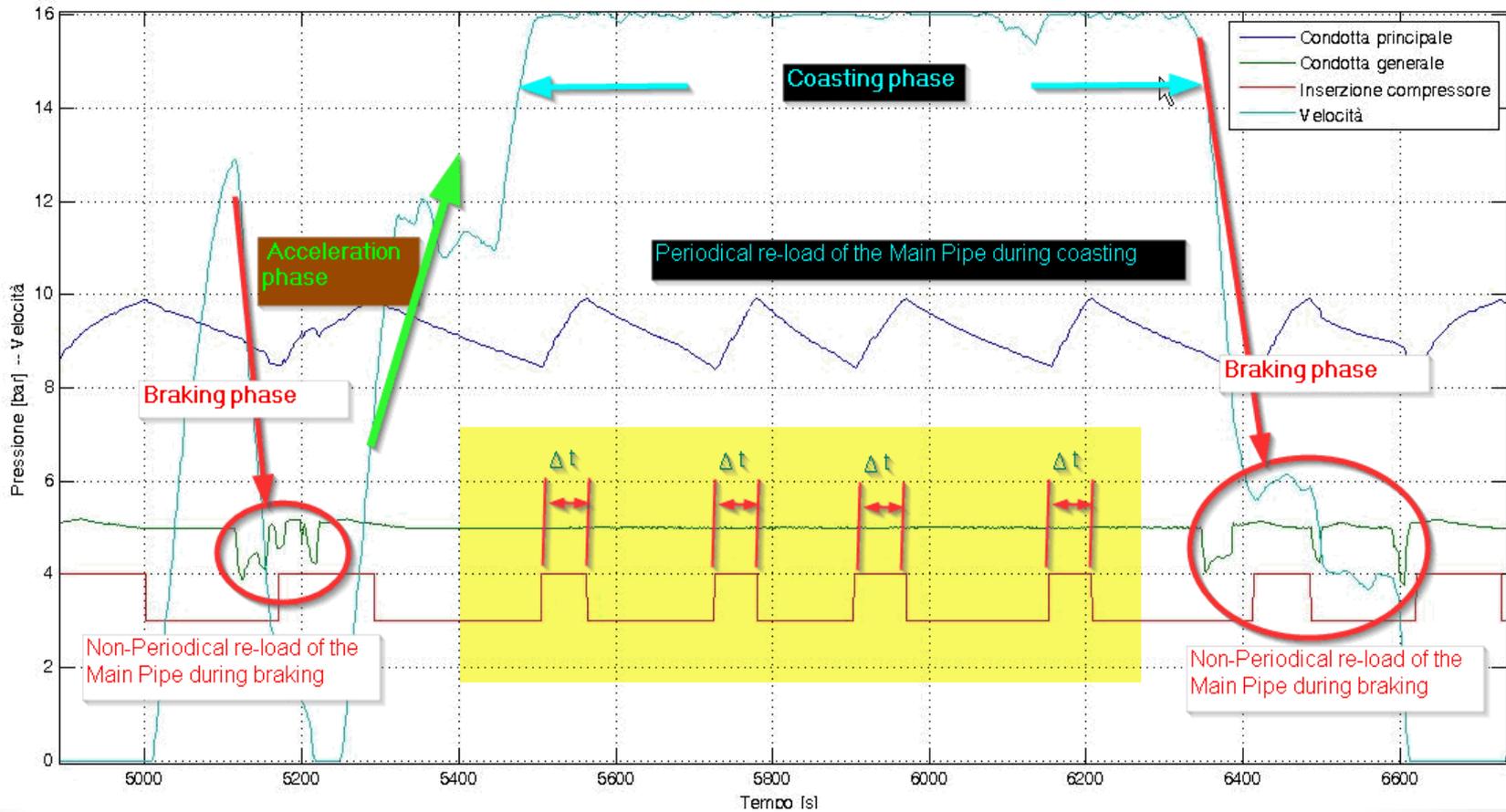
Temperatura

Ore di  
funzionamento

Tempi di  
ricarica

Numero  
avviamenti

# Firma per compressori



# Firma per compressori

Tempo di ricarica

- Ricarica delle perdite di carico
- Costante durante la vita utile

Tempo di funzionamento

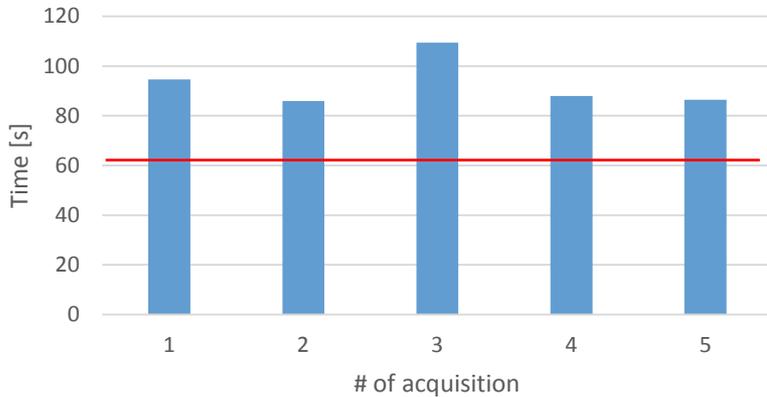
- Ore lavorate complessivamente dal compressore

Cicli di funzionamento

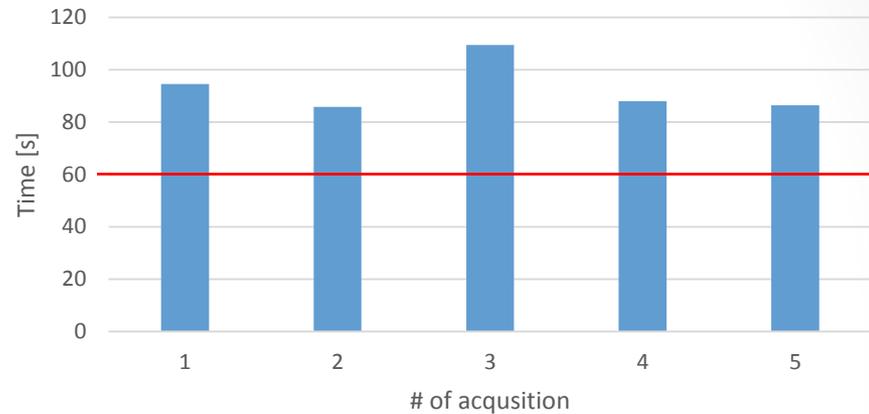
- Cicli di funzionamento normali

# Risultati

Compressor time ON in coasting

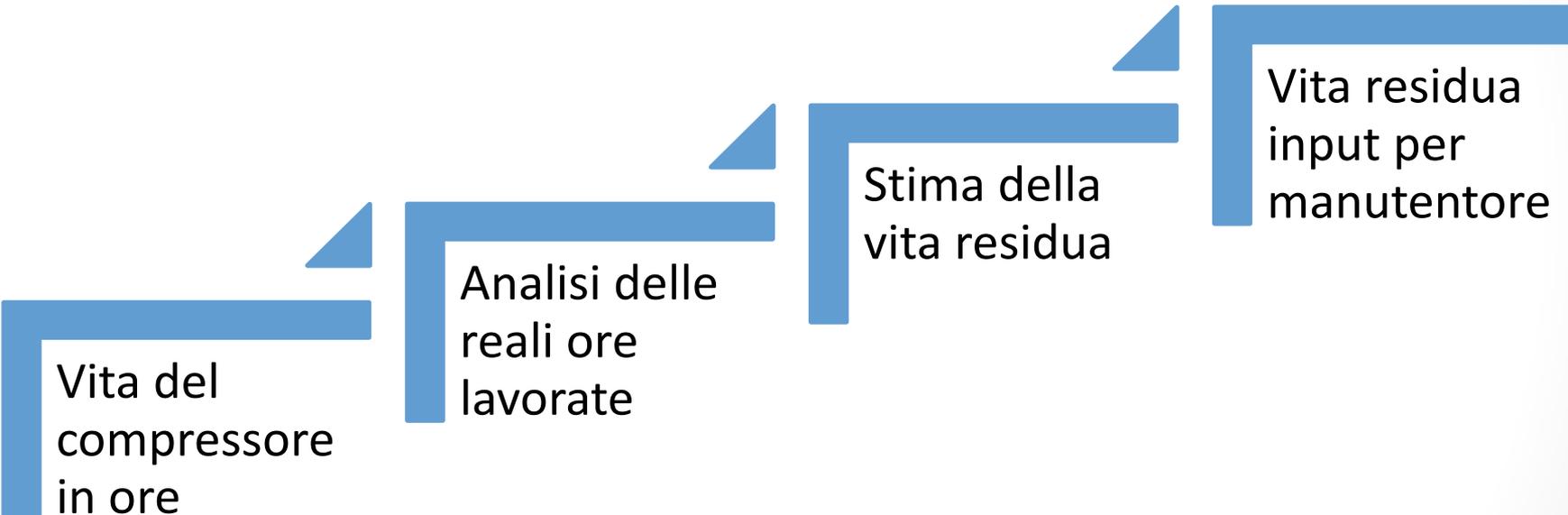


Compressor time ON in coasting



Loco ID	Average Time ON [s]	Average Time ON in coasting [s]	% compressor working
186111	90,83	74,48	25,30%
186112	122,91	73,98	30,05%
<b>186113</b>	<b>213,59</b>	<b>126,45</b>	<b>48,46%</b>
186114	73,81	64,47	18,58%
186115	75,49	59,74	16,98%
186116	68,27	54,39	16,13%
186117	88,11	71,74	25,36%
186118	75,79	60,09	22,77%
186119	76,03	60,17	18,64%
186120	77,69	64,45	22,42%
186121	90,03	75,18	26,03%
186122	89,81	67,35	26,89%
<b>186144</b>	<b>121,95</b>	<b>92,87</b>	<b>32,81%</b>
186236	76,05	63,22	21,18%

# Stima della vita residua



Vita del compressore in ore

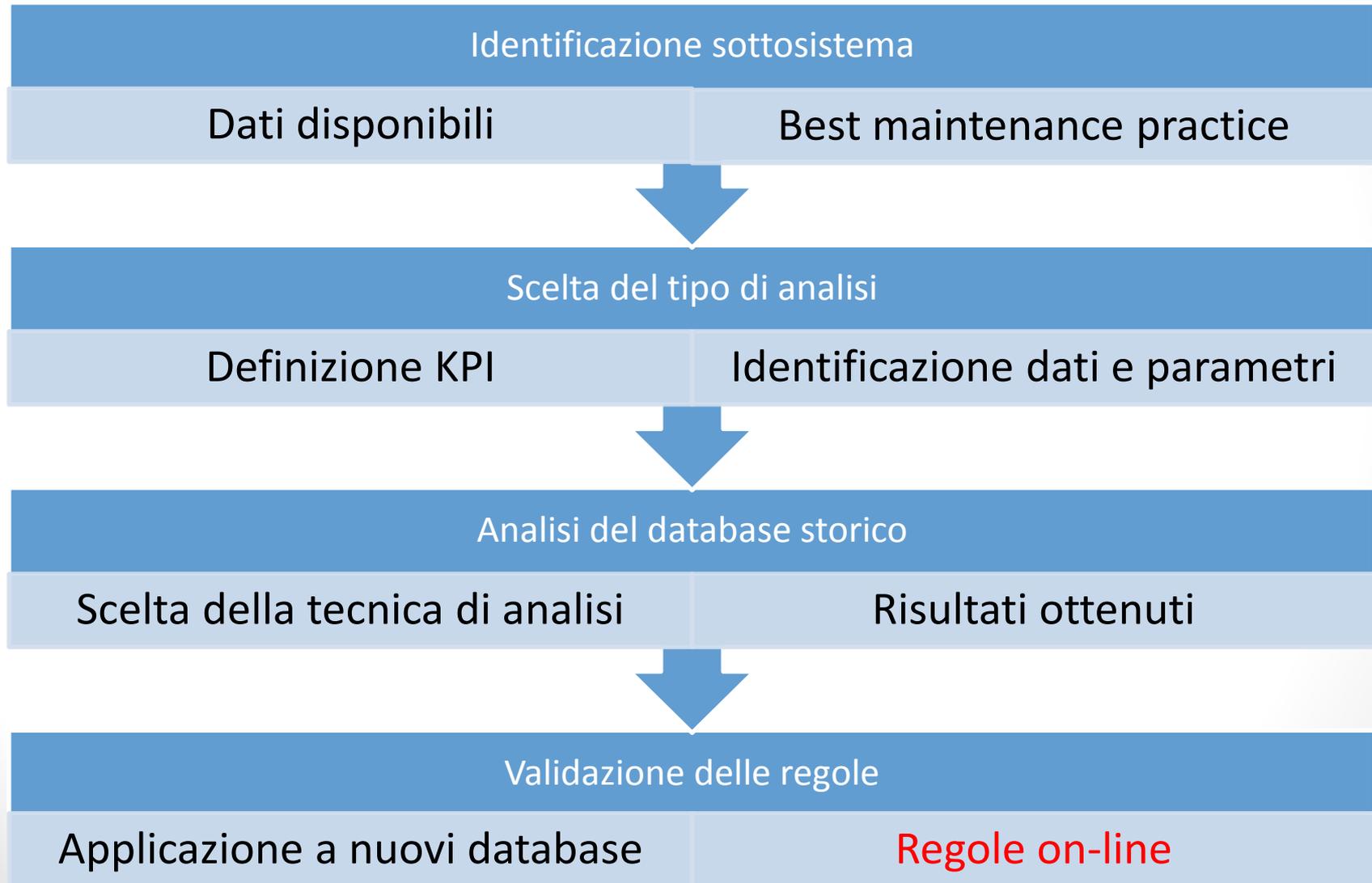
Analisi delle reali ore lavorate

Stima della vita residua

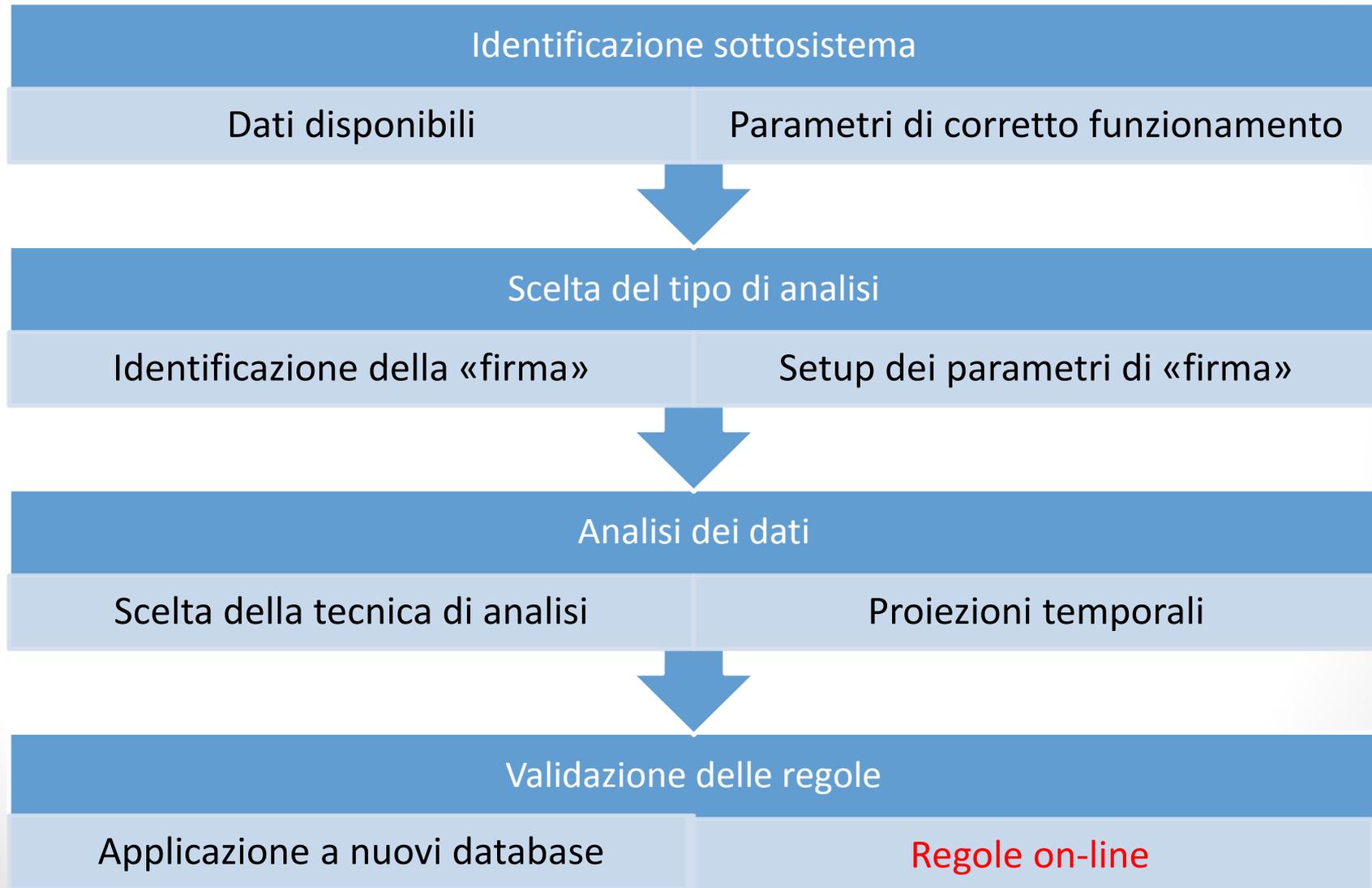
Vita residua input per manutentore

# REGOLE GENERALI PER L'APPROCCIO AL DATA MINING

# Flowchart data mining orizzontale



# Flowchart data mining verticale



# Conclusioni

L'analisi statistica dei dati può evidenziare comportamenti anomali e correlazioni tra parametri di processo e di ambiente utili ad ottimizzare, gli interventi manutentivi.



Si possono ottenere importanti benefici, e senza nemmeno essere Il Mago dei Numeri

