

I Encuentro de Estudiantes de **Doctorado** de Enpresagintza

Gorka Mendizabal Arrieta

gmendizabal@mondragon.edu

1. RESUMEN DE LA TESIS

Título: servicios posventa basados en datos industriales. Digitalización, monetización y ciclo de vida del dato.

Temas analizados:

1. IoT y Servicios digitales basados en datos
2. Ciclo de vida del dato
3. Monetización de los datos

Conferencias:

1. RESER-Coval (21/01/2021)
2. SSC 2021 (10/05/2021)
3. VI Laboratorio de Economía de Paz (24/05/2023)

Investigaciones realizadas en conjunto con:

1. Universidad de Florencia (Italia)
2. Universidad de Padua (Italia)
3. Universidad de Tampere (Finlandia)
4. Universidad de Deggendorf (Alemania)



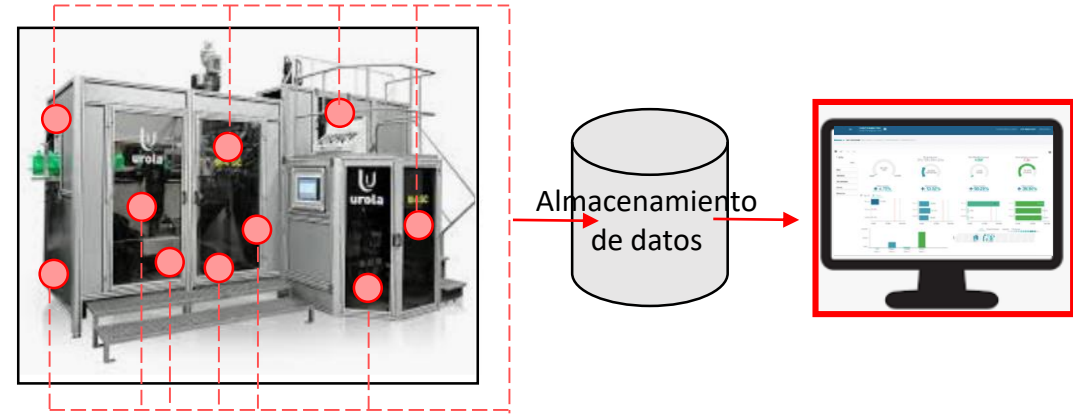
The screenshot shows the article page for "A pricing model to monetize your industrial data" by Gorka Mendizabal-Arrieta, Eduardo Castellano-Fernández, and Mario Rapaccini. The article is published in Frontiers in Manufacturing Technology, Volume 2, Article 1057557, on January 30, 2023. The abstract discusses data monetization (DM) as a relevant aspect of industrial manufacturing and proposes a mathematical model to price industrial data based on quality, entropy, and value. The introduction section is also visible, starting with "While data is a novel source of revenue and competitive advantage..."

2. ARTÍCULO CIENTÍFICO



Productor
Productor de información
Vendedor de los servicios digitales basados en datos

SENSORES INCORPORADOS A LA MÁQUINA



Comprador
Productor de datos
Cliente de los servicios Digitales basados en datos

2. ARTÍCULO CIENTÍFICO

Modelo de Shen et al. (2016) → datos personales

$$p_i = P_s \times \left(\frac{w_i}{w} \times \alpha + \frac{q_i}{q} \times \beta + \frac{r_j}{r} \times \gamma \right)$$

$$\alpha + \beta + \gamma = 1$$

- P_s : Supply Price of data
- $\frac{w_i}{w}$: Value weight
- $\frac{q_i}{q}$: Information entropy
- $\frac{r_j}{r}$: Data reference index



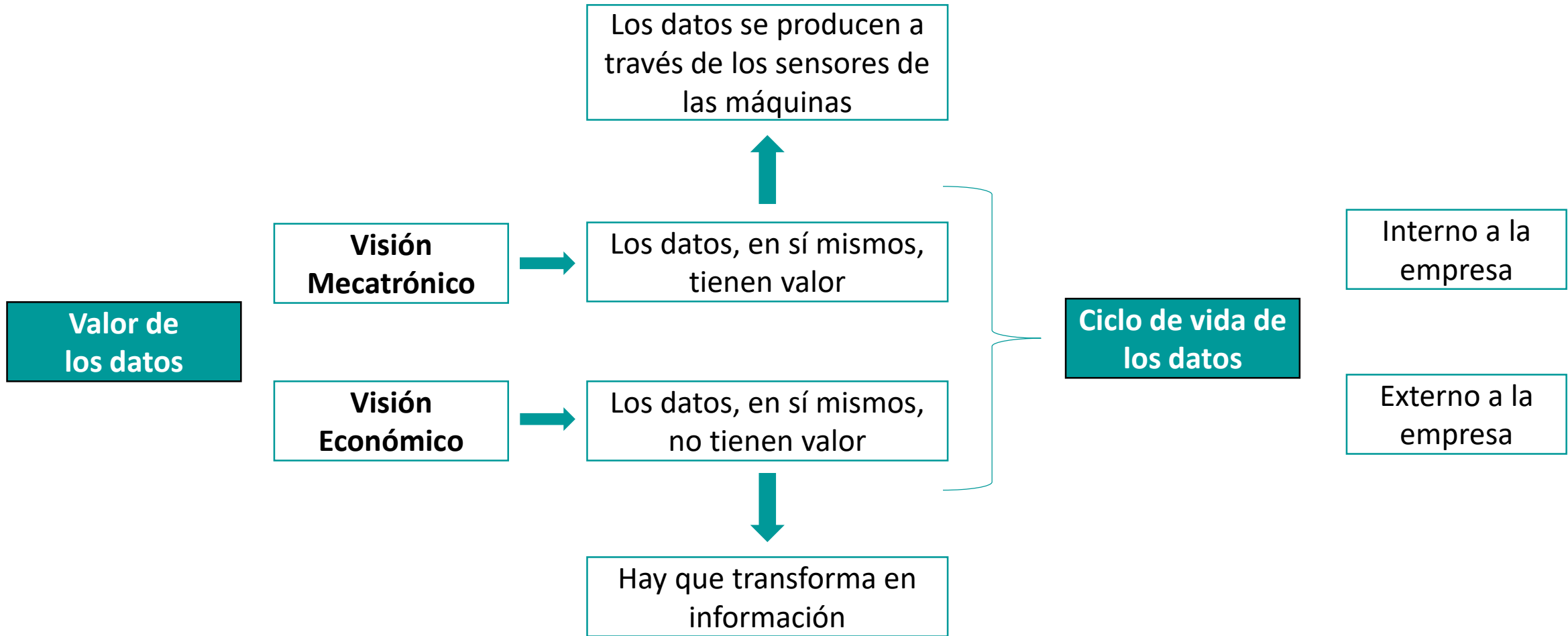
Modelo propuesto → datos industriales

$$p_i = P_s \times \left[\left(1 - \frac{DQ_i}{DQ} \right) \times \alpha + \frac{w_i}{w} \times \beta + \frac{q_i}{q} \times \gamma + \frac{CRI_j}{CRI} \times \delta \right]$$

$$\alpha + \beta + \gamma + \delta = 1$$

- P_s : Supply Price of data
- $1 - \frac{DQ_i}{DQ}$: Data quality
- $\frac{w_i}{w}$: Data (value) scores
- $\frac{q_i}{q}$: Information entropy
- $\frac{CRI_j}{CRI}$: Customer relevance index

2. ARTÍCULO CIENTÍFICO: valor y ciclo de vida de los datos



2. ARTÍCULO CIENTÍFICO: calidad del dato

$$1 - \frac{DQ_i}{DQ} : \text{Data quality}$$

	A
1	Date
2	2020-11-30T23:00:00.433Z
3	2020-11-30T23:00:01.433Z
4	2020-11-30T23:00:02.433Z
5	2020-11-30T23:00:03.433Z
6	2020-11-30T23:00:04.433Z
7	2020-11-30T23:00:05.437Z
8	2020-11-30T23:00:06.434Z
9	2020-11-30T23:00:07.433Z
10	2020-11-30T23:00:08.433Z
11	2020-11-30T23:00:09.433Z
12	2020-11-30T23:00:10.433Z
13	2020-11-30T23:00:11.433Z
14	2020-11-30T23:00:12.433Z
15	2020-11-30T23:00:13.433Z
16	2020-11-30T23:00:14.433Z
17	2020-11-30T23:00:15.433Z
18	2020-11-30T23:00:16.433Z
19	2020-11-30T23:00:17.433Z
20	2020-11-30T23:00:18.433Z
21	2020-11-30T23:00:19.433Z

01 H	02 H	03 H	04 H	05 H	06 H
3600	3580	3600	3575	3600	3600
07 H	08 H	09 H	10 H	11 H	12 H
3569	3552	3570	3600	3583	3300
13 H	14 H	15 H	16 H	17 H	18 H
3597	3600	3520	3600	3596	3589
19 H	20 H	21 H	22 H	23 H	24 H
3600	3585	3573	3600	3540	3500

Data tuple	Item	Pre-data quality lines (A)	Post-data quality lines (B)	DQ_i	DQ_i/DQ	$1 - \frac{DQ_i}{DQ}$
DT1	IT1	100	94	6	0,176	0,824
DT1	IT2	95	92	3	0,088	0,912
DT1	IT3	80	70	10	0,294	0,706
DT1	IT4	87	80	7	0,206	0,794
DT1	IT5	93	90	3	0,088	0,912
DT1	IT6	65	60	5	0,147	0,853
TOTAL		520	486	34	1	0

2. ARTÍCULO CIENTÍFICO: entropía

$\frac{q_i}{q}$: Information entropy

Data tuple	Item	RANGO				
Consumption	Total machine consumption	[0 - 20]	[20 - 40]	[40 - 60]	[60 - 80]	[80 - 100]
Temperature process	Room temperature	[13 – 14.5]	[14.5 -16]	[16 – 17.5]	[17.5 - 19]	[>= 19]
Feeding system process	Raw material consumption	[<= 240000]	[240000 - 480000]	[480000 - 720000]	[720000 - 960000]	[>= 960000]
Alarms	Types of registered alarms	[<=10]	[10 - 20]	[20 - 30]	[30 - 40]	[>=40]
OEE	Quality	[Buenas]	[Rechazo chorro]	[Rechazo trampilla]	[Rechazos tumbadas]	[Rechazos poros]
Electrical maintenance	Intensity	[<=40]	[40 - 80]	[80 - 120]	[120 - 160]	[>=160]

Descripción de los rangos seleccionados para cada una de las variables

```
intensidad_40 <- intensidad_TOTAL[intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 <= 40, ]
intensidad_80 <- intensidad_TOTAL[intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 >40 &
intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 <80, ]
intensidad_120 <- intensidad_TOTAL[intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 >=80 &
intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 <120, ]
intensidad_160 <- intensidad_TOTAL[intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 >=120 &
intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 <160, ]
intensidad_160_superior <- intensidad_TOTAL[intensidad_TOTAL$FINAL_CONSUMO_MATERIA.MAQUINA_INTENSIDAD_L1 >=160, ]
```

```
calidad_buenas <- sum(calidad_TOTAL$`Buenas (uds) [suma, 1 día]`)
calidad_Rchorro <- sum(calidad_TOTAL$`Rechazos chorro de aire (uds) [suma, 1 día]`)
calidad_Rtrampilla <- sum(calidad_TOTAL$`Rechazos trampilla antes Mae (uds) [suma, 1 día]`)
calidad_Rtumbadas <- sum(calidad_TOTAL$`Rechazos tumbadas y no cortadas (uds) [suma, 1 día]`)
calidad_Rporos <- sum(calidad_TOTAL$`Rechazos poros (uds) [suma, 1 día]`)
```

$$H(x_i) = - \sum_{j=1}^k p(x_{ij}) \log_2 p(x_{ij})$$

Probabilidad →

Log(pi;2) →

Entropía del Data tuple i →

```
p_temperatura13 <- nrow(temperatura_13) / nrow(temperatura_TOTAL)
p_temperatura14 <- nrow(temperatura_14) / nrow(temperatura_TOTAL)
p_temperatura16 <- nrow(temperatura_16) / nrow(temperatura_TOTAL)
p_temperatura17 <- nrow(temperatura_17) / nrow(temperatura_TOTAL)
p_temperatura19 <- nrow(temperatura_19) / nrow(temperatura_TOTAL)

LOG_temperatura13 <- log2(p_temperatura13)
LOG_temperatura14 <- log2(p_temperatura14)
LOG_temperatura16 <- log2(p_temperatura16)
LOG_temperatura17 <- log2(p_temperatura17)
LOG_temperatura19 <- log2(p_temperatura19)

HXi_temperatura <- -(p_temperatura13* LOG_temperatura13) -
(p_temperatura14* LOG_temperatura14) -
(p_temperatura16* LOG_temperatura16) -
(p_temperatura17* LOG_temperatura17) -
(p_temperatura19* LOG_temperatura19)
```

$$H(X) = \sum_{i=1}^n H(x_i) = - \sum_{i=1}^n \sum_{j=1}^k p(x_{ij}) \log_2 p(x_{ij})$$

Data tuple	Item	Information content	Ratio of information content
Feeding system process	Raw material consumption	0,896946	0,14498258
Temperature process	Room temperature	1,127666	0,18227615
Consumption	Total machine consumption	0,547874	0,08855838
OEE	Quality	0,701189	0,11334036
Electrical maintenance	Intensity	1,257334	0,20323571
Alarms	Types of registered alarms	1,655571	0,26760682
TOTAL		6,18658	1

2. ARTÍCULO CIENTÍFICO: CRI

CRI Questions	Explanation	CRI valuation
Se han vendido los servicios digitales?	Sí = 1 punto No = 0 punto	1
Nº de máquinas para las que se contratan los servicios digitales	1 punto por cada máquina	1
Peso relativo del cliente en la cuenta de explotación?	0,1 por cada punto porcentual	0,2
CRI		2.2
Ratio CRI		$2.2/(6*2,2) = 0,16666667$

3. CONCLUSIONES

1. No existe un consenso en la literatura científica sobre la definición de los datos.
2. Los servicios digitales basados en datos tienen un valor económico que se puede cuantificar.
3. La ecuación propuesta se basa en la investigación de Shen et al. (2016).
4. Se proponen 5 variables: precio del suministro, calidad del dato, valor de los datos, entropía, CRI.