


Big Data Analytics

Aigües de Castellbisbal, Empresa Mixta S.L.

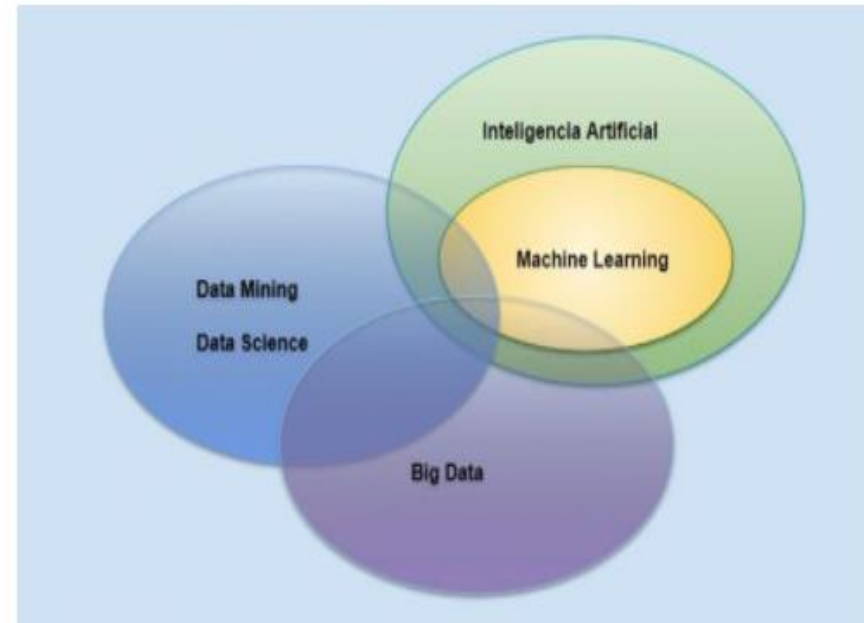
Aigües de Castellbisbal


ELEMENTS DEL BIG DATA ANALYTICS

Data Science, conjunt de principis, processos i tècniques que guien la extracció de coneixement des de les dades amb l'objectiu de millorar la presa de decisions.

Data Mining, es la extracció de coneixement des de les dades a través de tecnologies que incorporen els principis del Data Science.

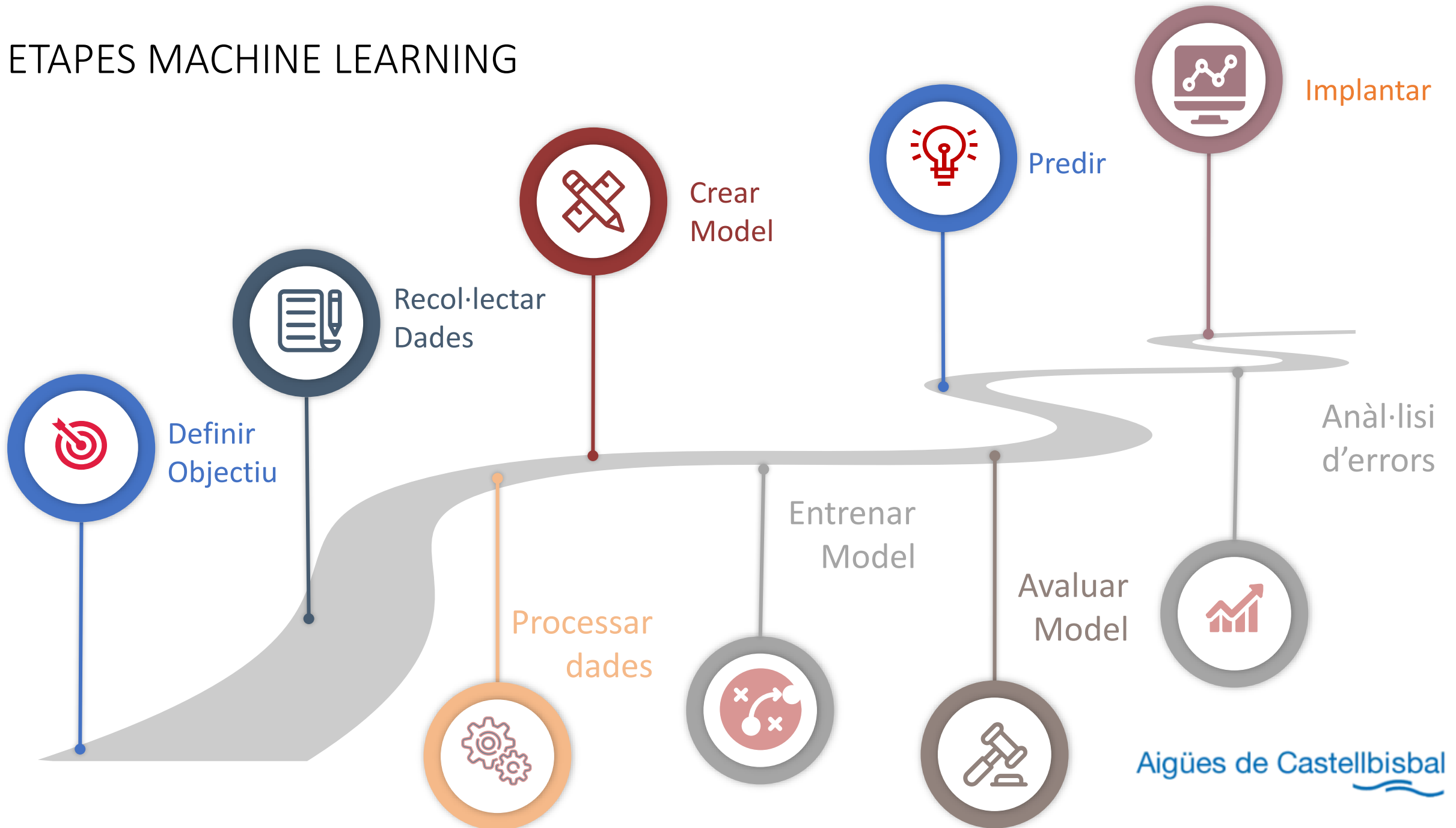
Big Data, conjunt de dades de molta volumetria per ser analitzades sota un sistema de processament tradicional. Requereix noves tecnologies de processament.



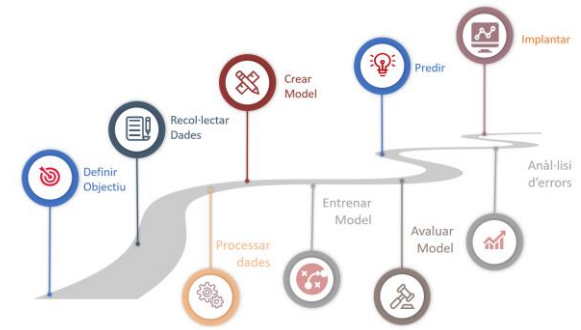
Intel·ligència artificial, John McCarthy va fer servir aquest nom per primer cop al 1956. I ho va definir com “la ciència i enginyer per fer màquines intel·ligents”. Es a dir, tenir computadores amb l'habilitat de raonar com essers humans.

Machine Learning, disciplina de la intel·ligència artificial que, mitjançant algorismes, dona als ordinadors la capacitat d'identificar patrons en dades massives i elaborar prediccions.

ETAPES MACHINE LEARNING

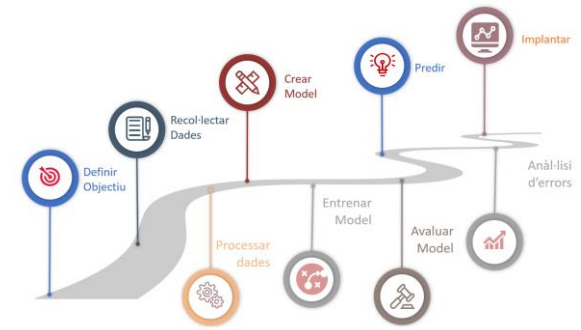


Etapa 1: Definir objectiu



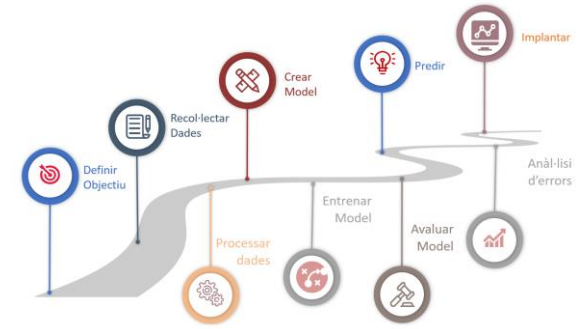
- Es vital entendre el problema a resoldre, ja que gran part de les decisions preses al llarg del projecte seran dependents d'aquesta etapa.
- Es necessari definir uns objectius **clars, específics, mesurables i assolibles** en un període determinat de temps.
- En aquesta etapa, son els experts del sector els que han de facilitar la informació.
- Les preguntes següents han de quedar respostes al finalitzar esta etapa:
 - Quin es el principal objectiu? Que es vol predir?
 - Com podem fer-ho?
 - Es pot assolir l'objectiu amb les dades existents?
 - Que es vol ensenyar al sistema?
 - Quin es l'estat actual de la variable objectiu?
 - Com es mesuraran els resultats?
 - Com es consumirà el model?

Etapa 2: Recol·lectar dades



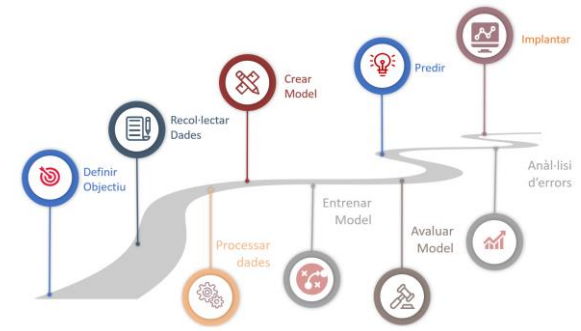
- Es defineix la **quantitat** i el tipus de dades necessàries, així com l'origen de les dades.
- La **qualitat** de les dades té un impacte directe al funcionament del model.
- Es més important la **qualitat** que la quantitat. No obstant, l'ideal es plantejar un equilibri, es a dir, com més i millors dades tinguem per començar, millor serà el rendiment del model.
- En alguns casos, no es disposa de totes les dades necessàries per resoldre el problema, pel que s'obtenen d'entitats externes.

Etapa 3: Processar Dades



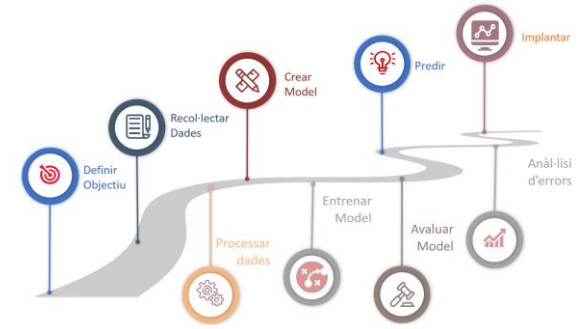
- L'objectiu principal es **visualitzar y analitzar** quines son les variables que representen millor allò que volem.
- Aquí s'observen algunes de las correlacions entre las variables independents i les dependents.
- Es probable que hi hagi dades que continguin irregularitats i afectin a la qualitat de la informació. Cal detectar-les i corregir-les. Les més habituals son:
 - Dades incompletes
 - Dades sorolloses
 - Dades incoherents
- Les dades necessiten un format determinat per ser processats per la màquina de la forma més senzilla possible.
- Les tasques de **processament i neteja** han d'estar ben realitzades abans de la posada en marxa del model.

Etapa 4: Crear model



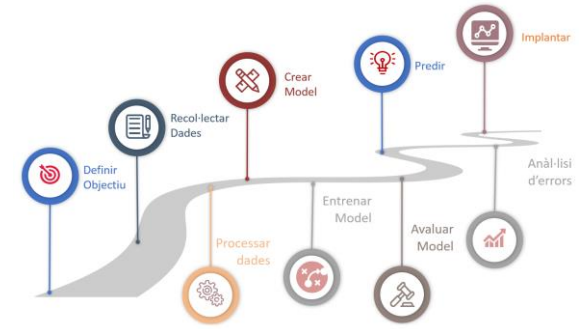
- En funció de l'objectiu i de les dades, cal decidir entre un algoritme de d'aprenentatge supervisat o aprenentatge no supervisat.
- I el tipus de model de Machine Learning:
 - **Classificació binaria:** prediu un resultat binari, es a dir, experiments estadístics amb només 2 resultats possibles. Per exemple, aquest correu electrònic es spam o no ho es?
 - **Model de classificació multiclasse:** proporciona prediccions para varies classes, es a dir, es capaç de predir un de més de dos resultats. Per exemple, ¿quina categoria de productes es més interessant per a aquest client?
 - **Model de regressió:** prediu un valor numèric. Per exemple, quina serà la temperatura de Barcelona demà?

Etapa 5: Entrenar model



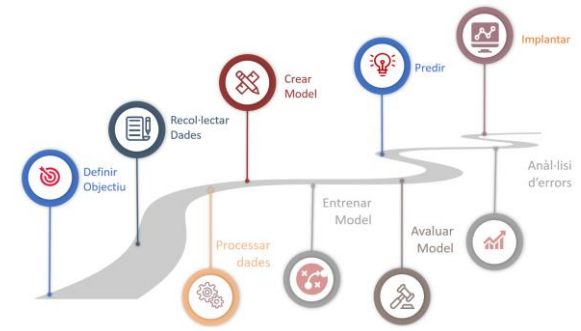
- L'entrenament consisteix en que el model aprengui a assolir l'objectiu plantejat inicialment a partir de les dades.
- L'algoritme es capaç de plantejar correlacions a les dades d'entrenament que no han estat assignats als atributs d'entrada i es proporciona un model que emmagatzema aquestes correlacions.
- Del conjunt de dades processades, es farà servir un 70% del total per realitzar l'entrenament.

Etapa 6: Avaluar el model



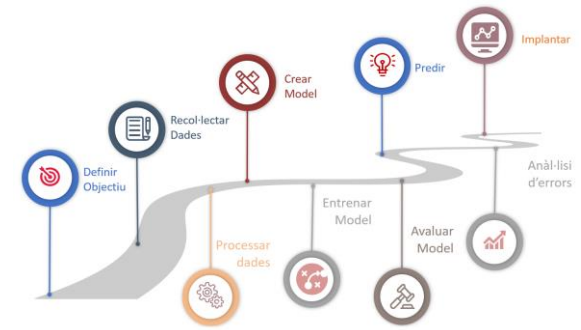
- L'avaluació consisteix en determinar el % d'encert sobre dades que el model encara no coneix. Del conjunt de dades processades, es farà servir un 30% del total per realitzar la validació.
- Un 50% de precisió es insuficient per validar un model, doncs indica que la meitat de vegades fallarà.
- Existeix la possibilitat que el model funcioni bé per les dades d'entrenament i no ho faci per les dades de validació (problema de overfitting). Si això passa, haurem d'ajustar els paràmetres del model i entrenar-lo de nou per millorar els resultats.

Etapa 7: Predicció



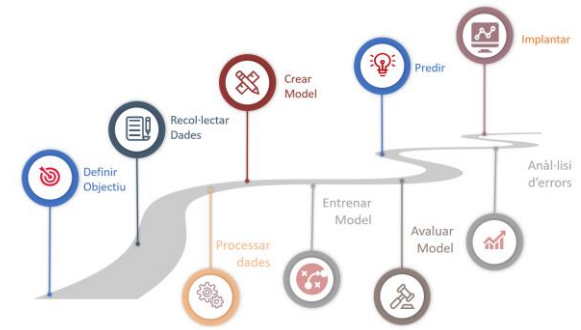
- El modelo haurà de realitzar prediccions amb dades els més semblants al món real.
- El test es pot realitzar integrant el model amb un sistema real amb el que es pugui comunicar o bé amb la realització d'execucions amb dades noves.
- Es determinarà si els paràmetres de configuració son els correctes o si se n'ha d'ajustar algun.

Etapa 8: Anàlisi d'errors.



- Aquesta fase permet modelar i canviar els aspectes no rellevants pel projecte.
- Analitzar errors es important para entendre que es el que hem de fer per millorar els resultats de Machine learning. En particular, les opcions podran ser:
 - fer servir un model més complex.
 - fer servir un model més senzill.
 - adonar-nos que necessitem més dades i / o característiques.
 - Desenvolupar una millor comprensió del problema i entendre millor quina es la següent passa.

Etapa 9: Implantació



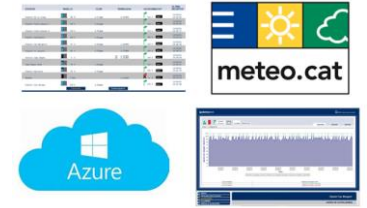
- El model s'ha d'integrar a un sistema.
- S'anirà avaluant cada cert temps per reajustar-lo, tornar-lo a entrenar o canviar en funció de les noves dades.

CAS PRÀCTIC: NIVELL DEL CLOR I RELACIÓ AMB ALTRES INDICADORS

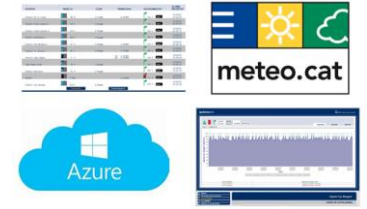
- Volíem analitzar si la temperatura es un factor que afecti al nivell de clor de la nostra xarxa, per tal de fer una previsió de la quantitat de clor a afegir.
- Per dur a terme l'anàlisi, les dades que hem fet servir han estat:
 - les dades del telecontrol dels darrers 30 dies, en un interval de 60 minuts.
 - la temperatura a Castellbisbal pel mateix període, en les mateixes hores.
- Aquestes dades les hem unificat a un fitxer i hem fet l'anàlisi amb les eines de Machine Learning d'Azure.

ESTACIÓ	NIVELLS	CLOR	TERBOLESA	ACCESSIBILITAT	ÚLTIMA RECEPCIÓ
Dipòsit de la Creixa	78 x	0,57 ppm	2,10 km/h	100 x	13/09/21 17:01:29
Dipòsit Orotó Blanca 1	130 x	0,61 ppm		100 x	13/09/21 13:01:35
Dipòsit Orotó Blanca 2	130 x	0,45 ppm		100 x	13/09/21 17:01:37
Dipòsit Santandreu	130 x	0,45 ppm		100 x	13/09/21 17:01:38
Dipòsit Can Margall	97 x	0,48 ppm	1,14 km/h	100 x	13/09/21 17:01:38
Dipòsit de Castell	83 x	0,78 ppm	2,75 km/h	100 x	13/09/21 13:01:32
Control Can Reguer	0 x			100 x	13/09/21 17:01:34
Can Reguer Gran	70 x	1,08 ppm		100 x	13/09/21 17:01:35
Dipòsit Agrícola	65 x	0,47 ppm		100 x	13/09/21 17:01:35
Resaca	2,30 x		2,00 km/h		09/09/20 09:00:00
Dipòsit Can Batlle	130 x	0,64 ppm		100 x	13/09/21 13:01:40



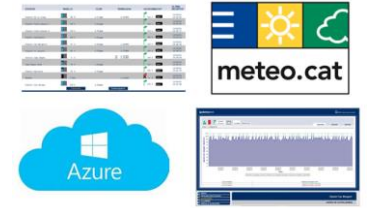


- Durant el procés, vam cometre errors en la recollida de dades que no ens permetien arribar a la fase de creació del model:
 - La **qualitat** de les dades no era bona. Fèiem servir massa indicadors del telecontrol, que no aportaven valor i no estaven influïent en la dada objectiu.
 - La **quantitat** de les dades era excessiva i cada iteració per començar el procés requeria massa temps de preparació.
- Un cop resultats, vam dividir els tests en 2 tipus:
 - Predictiu: preparació per predir la variable objectiu (nivell de clor) en funció de les dades que simulem.
 - Regressiu: ens mostra la relació entre les diferents variables i la variable objectiu (nivell de clor)



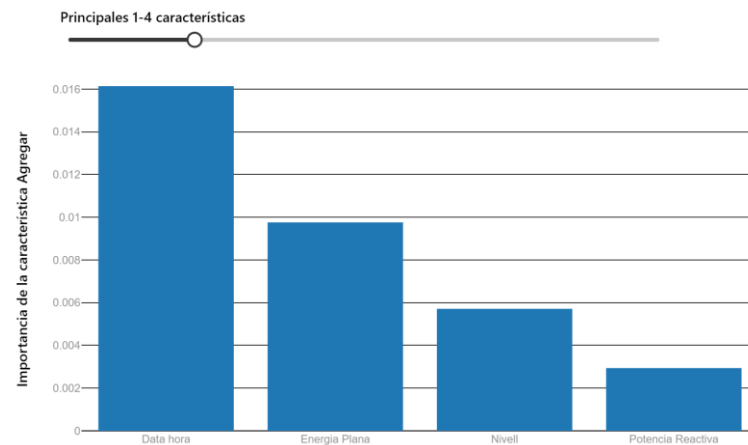
Resultats model predictiu:

- El model predictiu no va retornar valors fiables en la predicció del nivell de clor. Les característiques que ens presentava per fer una predicció no podien tenir una relació directa amb el nivell de clor.



Resultats model regressiu:

- El model regressiu ens va mostrar que la temperatura no era un factor determinant per al nivell de clor. Eren més importants variables com l'energia que es consumia al dipòsit, la data i hora en que es prenia el valor, o el nivell del dipòsit.



CAP A ON ENS ADRECEM?

Millors en la planificació i supervisió dels processos

The screenshot displays a comprehensive water utility management system interface. At the top, a table lists various tasks with columns for ID, date, user, type, category, priority, description, status, current phase, personnel, business line, planning date, variables, privacy, and budget. The table includes entries for commercial supply, city water meter reading, and a water leak repair in progress.

Below the table, a map shows the geographical location of the tasks, with a red line indicating a water main and a green line for a water service line. A sidebar on the left provides navigation options for different system components like 'Seguridad web', 'Seguridad App', and 'Parametrización'. A central panel shows a tree view of 'categorias' (categories) under 'Procesos' (processes).

The main visualization area features a line chart titled 'Diposit Can Margarit: Nivell (%)' showing water level fluctuations over time from August 2021 to September 2021. The chart includes a legend for 'Nivell (%)', 'Boia Mxim', and 'Boia Mm'. Below the chart, there are navigation buttons for 'MAPA', 'RESUM D'ESTACIONS', 'ESTACIONES', 'ALARMES', and 'MODULS AUXILIARIS'. The bottom right corner identifies the location as 'Diposit Can Margarit' and 'AIGÜES DE CASTELLBISBAL'.

Milliores en la gestión de clientes i recursos humanos

FECHAS ENTRE 2021-08-16 Y 2021-08-16

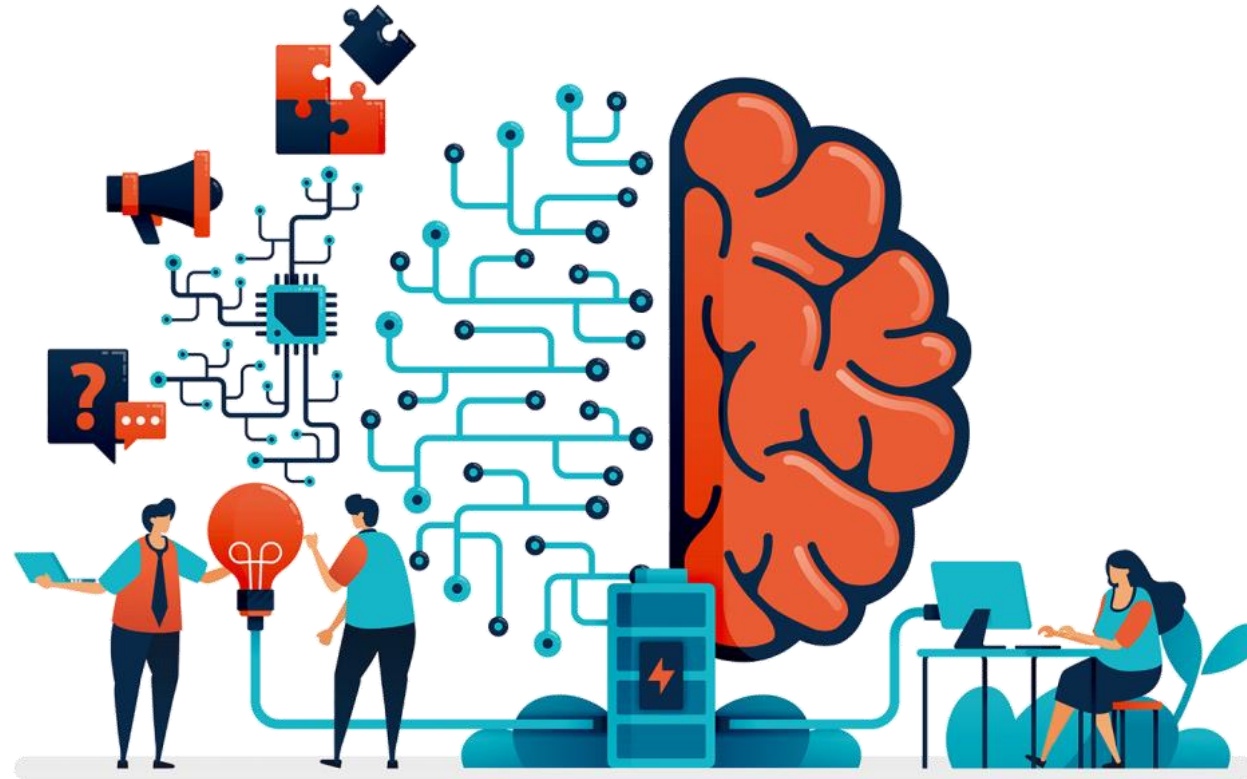
Concepto	Tarifa	Subconcepto	Nº Contratos	M3 Consumidos	M3 Facturados	Importe	% impuesto
NO FACTURADOS							
FACTURADOS SIN LINEAS DE CONSUMO							
CANON AGUA	1-DOMESTICA	CONSUMO					
TOTAL							
	2-INDUSTRIAL	CONSUMO					
TOTAL							
	3-MUNICIPAL	CONSUMO					
TOTAL							

Presupuesto

Empresa / Áreas de Castellbisbal Empresa / Presupuesto Empresa / Presupuesto

Nº	Descripción	Cliente	Año inicio	Año fin	Estado	Fecha Estado	% Margen	Estado Aceptación	Fecha estado aceptación	PVP Estudios	Coste directo estudios	Coefficiente medio administración	Coste Precios Horas	Coste Precios Compras	Coefficiente medio estudios	Total facturado
04F7_DAF0	04F7_Implementar en ACTUA la atención del presupuesto de la Empresa	N/A	N/A	N/A	FINALIZADO		0	4-EN EJECUCION								
04D5_DAF0	04D5_DAF0 Valorar e implantar sistema de satisfacción de clientes	N/A	N/A	N/A	FINALIZADO		0	5-FINALITZAT								
03P5_DAF0	03P5_DAF0_Proyecto imagen corporativa(nova web)	N/A	N/A	N/A	EN CURSO		0	1-PENDIENT								
01F2_DAF0	01F2_DAF0 Creación de la oficina virtual	N/A	N/A	N/A	FINALIZADO		0	5-FINALITZAT								
01F1_DAF0	01F1_DAF0 Ampliación horari oficina	N/A	N/A	N/A	FINALIZADO											
01D4(A3F1)_DAFO	01D4(A3F1)_DAFO Creación ota previa y atención personalizada con perfil de clientes VIP	N/A	N/A	N/A	EN CURSO											
01D1_DAF0	01D1_DAF0 Reorganización lugares trabajo y Contratación de los perfiles necesarios para el cambio	N/A	N/A	N/A	FINALIZADO											
APF2021	ACISA PLA FORMACIO_2021	N/A	N/A	N/A	EN CURSO											
APF2020	APF2020 ACISA PLA FORMACIO GC 2020	N/A	N/A	N/A	FINALIZADO											
AP40129	ESTUDI DE TARIFES 2021	ACISA	N/A	N/A	FINALIZADO											
AP40124	REUNIONS ESTRATEGIQUES_2021	ACISA	N/A	N/A	EN CURSO											
AP40123	TIQUETS VARIS_2021	ACISA	N/A	N/A	EN CURSO											
AP40106	REQUERIMENTS_2021	ACISA	N/A	N/A	EN CURSO											
AP40102	EVENTOS_2021	ACISA	N/A	N/A	EN CURSO											
AP40100	VENTILLODS_2021	ACISA	N/A	N/A	EN CURSO											
AP40099	ASBAR PAROL_EL CANVET_2021	AGBAR	N/A	N/A	EN CURSO											
AP40098	LECTURES TRIMESTRALS CUBETA_2021	CUACSA	N/A	N/A	FINALIZADO											
AP40095	SATISFACCIO CLIENTS_2021	ACISA	N/A	N/A	EN CURSO											
AP40094	NOV_ MEMORANDUMS_2021	ACISA	N/A	N/A	EN CURSO											
AP40093	SUBMINISTRAMENTS I CORREOS_2021	ACISA	N/A	N/A	EN CURSO											
AP40092	ISO9001_2021 (ADOS_DAF0)	ACISA	N/A	N/A	EN CURSO											
AP40091	HORES+MATERIALS SICCA_2021	ACISA	N/A	N/A	EN CURSO											

Eines de suport a la Direcció, per a una millor presa de decisions



Moltes gràcies per la vostra atenció

Aigües de Castellbisbal

