



Aplicaciones de la IA al sector agropecuario



[CIDAI-PAI 02-2022-D07](#)

Aquest treball està disponible sota la llicència Creative Commons Attribution-NonCommercial-Share Alike 4.0 International (CC BY-NC-SA 4.0).

D'acord amb els termes d'aquesta llicència, podeu copiar, redistribuir i adaptar l'obra per a finalitats no comercials, sempre que se citi adequadament, tal com s'indica a continuació.

En qualsevol ús d'aquesta obra, no s'ha de suggerir que el CIDAI recolzi cap organització, producte o servei específic. No està permès l'ús del logotip del CIDAI.

Si adapteu l'obra, haureu de llicenciar la vostra obra sota la mateixa llicència Creative Commons o equivalent. Si creeu una traducció d'aquest treball, haureu d'afegir el següent avís d'exempció de responsabilitat juntament amb la citació suggerida: "Aquesta traducció no ha estat creada pel Centre d'Innovació per a la tecnologia de dades i la intel·ligència artificial (CIDAI). El CIDAI no es fa responsable ni del contingut ni de l'exactitud d'aquesta traducció. L'edició original en anglès serà l'edició vinculant i autèntica".

Qualsevol mediació relacionada amb les controvèrsies sorgides sota la llicència es durà a terme d'acord amb les normes de mediació de l'Organització Mundial de la Propietat Intel·lectual.

Citació recomanada: CIDAI-Projecte Alt Impacte: aplicacions de la IA en el sector agropecuari. Memòria final. Ref.: CIDAI-PAI 02-2022-D07.

Autors i col·laboradors

Aquest informe s'ha dut a terme amb la participació dels següents autors:

Nom	Entitat	Participació
Pol Torres	EURECAT	Coordinació, redacció i revisió
Joan Mas	CIDAI	Revisió
Fernando Vilariño, Coen Artens	CVC	Redacció del segon cas d'ús (AI-LAB i visió per computador) i revisió del document
Karina Gibert, Dante Conti	IDEAI-UPC	Redacció del primer cas d'ús
Mili Voltes, Ruth Gay, Alba Jiménez	Generalitat de Catalunya (Servei d'Alimentació Animal i Seguretat de la Producció Ramadera)	Revisió del primer cas d'ús
Iris Pallarol, David Avaro	BSC	Redacció del segon cas d'ús (model del governança i sostenibilitat)
Roberto García	UdL	Redacció del segon cas d'ús (espai de dades)

Els resultats presentats s'han obtingut amb la col·laboració dels següents experts:

Nom	Entitat	Participació
Pol Torres	EURECAT	Definició dels casos d'ús i coordinació del projecte
Fernando Vilariño, Coen Artens, Edgar García	CVC	Definició de AI-LAB i desenvolupament del segon cas d'ús
Karina Gibert, Dante Conti, Claudia Agüero	IDEAI-UPC	Implementació del primer cas d'ús
Mili Voltes, Ruth Gay, Alba Jiménez	Generalitat de Catalunya (Servei d'Alimentació Animal i Seguretat de la Producció Ramadera)	Definició i coneixement expert del primer cas d'ús
Iris Pallarol, David Avaro	BSC	Desenvolupament del model de governança i sostenibilitat
Roberto García, Rosa Gil, Lluís Miquel Pla, Pol Llagostera	UdL	Desenvolupament de l'espai de dades i suport en el segon cas d'ús
Daniel Babot	CEP	Suport logístic i coneixement expert en el segon cas d'ús.

Índex



1. RESUM EXECUTIU	5
2. INTRODUCCIÓ	7
3. SUPORT A LA DECISIÓ EN L'APLICACIÓ D'ANTIBIÒTICS EN EXPLOTACIONS RAMADERES	9
4. VISIÓ PER COMPUTADOR PEL SEGUIMENT INDIVIDUALITZAT DE PORCS EN GRANJA	27
5. ANNEX	50
6. REFERÈNCIES	67

Resum executiu



El projecte “Aplicacions de la IA al sector agropecuari” s’ha executat en el marc dels projectes d’alt impacte del CIDAI (*Centre of Innovation in Data Technologies and Artificial Intelligence*) amb el finançament de la Generalitat de Catalunya. En el desenvolupament d’aquest projecte, per part del CIDAI hi ha participat Eurecat, el Centre de Visió per Computador (CVC), el centre de recerca en ciència de dades i intel·ligència artificial de la Universitat Politècnica de Catalunya (IDEAI-UPC), la Generalitat de Catalunya i el Barcelona Supercomputing Center (BSC). Com a entitats externes també hi ha col·laborat la Universitat de Lleida (UdL) i el Centre d’Estudis Porcins (CEP).

Aquest projecte ha treballat en dos casos d’ús ben diferenciats dins del sector agropecuari. Per una banda la necessitat d’ampliar el coneixement en l’aplicació d’antibiòtics en explotacions ramaderes per part del Servei d’Alimentació Animal i Seguretat de la Producció Ramadera de la Generalitat de Catalunya va ser l’origen del primer cas d’ús enfocat a aplicar eines de suport a la decisió en l’aplicació d’antibiòtics en explotacions ramaderes. En segon lloc, un seguit de reunions amb agents sectorials de les terres de Lleida va desembocar en una col·laboració amb el Centre d’Estudis Porcins per desenvolupar algoritmes d’intel·ligència artificial i visió per computador pel seguiment del cycle de vida dels porcs per a la millora del benestar animal. Aquest cas d’ús va fer emergir també la necessitat d’utilitzar un espai per compartir les dades de manera segura. En aquest context, gràcies a l’expertesa de la UdL en aquest àmbit, degut a la seva participació en iniciatives com Gaia-X (iniciativa europea que te com a objectiu crear un ecosistema de dades obert, transparent i segur a Europa), es va optar per realitzar una prova de concepte d’un espai de dades del sector “agro”. De manera paral·lela al desenvolupament d’aquest segon cas d’ús, també es va redactar un informe de governança i sostenibilitat per tal d’establir una metodologia pel desenvolupament d’aquesta tipologia de projectes i identificar les necessitats per tal que s’implementin les tecnologies desenvolupades i tinguin impacte real en el territori. El marc metodològic que s’ha fet servir implementa l’instrument AI LAB, el qual integra les tres dimensions mencionades (identificació dels reptes de l’ecosistema, desenvolupament de solució tecnològica i infraestructura de dades, i model de governança, sostenibilitat i negoci) per avançar en l’escalabilitat i permetre vies d’oportunitat i negoci més enllà del pilot.

El primer cas d’ús sobre eines de suport a la decisió en l’aplicació d’antibiòtics en explotacions ramaderes utilitza una metodologia KDD (*Knowledge Discovery in Databases*), incloent recollida de dades, preprocessament i mineria de dades per avaluar i interpretar les dades dels antibiòtics aplicats en les explotacions ramaderes de Catalunya. L’anàlisi realitzat permet concloure la categoria zootècnica és un dels factors més discriminants sobre el consum d’antibiòtics mentre que l’indicador base conegut com a mgUR (mg Unitat de Referència) no està alineat amb l’indicador global nacional de referència IR (Índex de Referència) a les explotacions d’engreix.

El segon cas d’ús sobre seguiment del cycle de vida dels porcs per a la millora del benestar animal s’han utilitzat eines de visió per computador i xarxes neuronals per identificar de manera unívoca els diferents porcs del corral i fer-ne un seguiment en el temps. S’han extret mètriques de moviment/repòs i localització en el corral que serviran com a base per desenvolupar noves mètriques per extraure indicadors rellevants per al benestar animal. Per altra banda, la prova de concepte de l’espai de dades ha permès explorar una metodologia sòlida i segura d’intercanvi de dades i d’algoritmes, així com explorar-ne les seves vies de monetització. Finalment, l’informe de governança i sostenibilitat generat a partir del *background*, *foreground* i reunions amb agents sectorials proporciona una primera aproximació a la possible valorització d’aquestes tecnologies.

Tots els participants al projecte van signar un Acord de Confidencialitat per garantir la confidencialitat durant el transcurs del projecte i protegir l’intercanvi de dades. El Servei d’Alimentació Animal i Seguretat de la Producció Ramadera del Departament d’Acció Climàtica, Alimentació i Agenda Rural de la Generalitat de Catalunya va proporcionar dades anonimitzades relacionades amb l’aplicació d’antibiòtics a les explotacions ramaderes catalanes. El CEP va proporcionar gravacions d’una granja de porcs ubicada a la seva explotació experimental.

Introducció



Aquest document recull els resultats obtinguts del desenvolupament del projecte d'alt impacte en l'àmbit agropecuari (PAI-AGRO) desenvolupat durant el 2022-2023, en el qual hi ha participat els membres del CIDAI Eurecat, CVC, IDEAI-UPC, BSC i la Generalitat de Catalunya. També hi han col·laborat la Universitat de Lleida i el Centre d'Estudis Porcins.

El PAI-AGRO és un projecte multidisciplinari desenvolupat sobre l'eix central de l'agricultura i la ramaderia a Catalunya on es proposen dos casos d'ús alineats amb reptes rellevants d'aquests sectors: eines de suport a la decisió per a la identificació de patrons en l'aplicació d'antibiòtics a les explotacions ramaderes i anàlisi del cicle de vida dels porcs per a la millora del benestar animal. L'objectiu del projecte és desenvolupar eines intel·ligents que permetin resoldre els reptes plantejats en els diferents casos d'ús i establir el full de ruta per a la continuïtat de solucions tecnològiques treballades.

El primer cas d'ús neix d'una necessitat expressa del Servei d'Alimentació Animal i Seguretat de la Producció Ramadera (Departament d'Acció Climàtica, Alimentació i Agenda Rural, Generalitat de Catalunya) per tal d'identificar patrons i reduir l'aplicació d'antibiòtics a les explotacions ramaderes del país.

Per tal de potenciar la interacció amb el territori, per al segon cas d'ús s'ha desenvolupat un AI LAB centrat a les terres de Lleida. Aquest AI LAB proporciona una solució als reptes a llarg termini en el context de les agendes compartides de la RIS3CAT 2030 (Catalunya, s.f.).

L'AI LAB proporciona una estructura àgil per tractar reptes de llarg termini, adoptant els principis de les agendes compartides i del *test-before-invest* (o validar abans d'invertir). De manera explícita s'identifiquen els reptes i solucions al voltant de TRL 6, les quals, acompanyades d'un model de negoci i sostenibilitat, permeten prendre decisions d'inversió (*invest*) després de la validació de la prova (*test*). Aquest test no s'ha de veure només com una validació purament tecnològica, sinó també com a prova a l'ecosistema a partir del qual serà possible definir factors de mercat, estructurals i barreres reguladores que permetran definir un model de negoci factible.

Per una banda, és imprescindible que els reptes que solucioni el CIDAI estiguin lligats a reptes reals de territori. En aquest sentit, la co-creació amb les comunitats d'interès és clau, i el fet que els reptes triats estiguin alineats amb reptes a mig i llarg termini permet superar l'horitzó del projecte i garantir la futura incorporació del teixit empresarial del país als reptes. Per altra banda, la implicació dels socis CIDAI en el procés de co-creació i definició de governança permet injectar la tecnologia IA en reptes amb translació directa a mercat fent servir l'estat de l'art de les solucions tecnològiques basades en IA, proporcionant d'aquesta manera un lloc de partida capdavanter per als sectors empresarials catalans. En aquest sentit, s'ha desenvolupat una proposta de sostenibilitat per afavorir la inversió en les solucions proposades, a partir de models clars de la gestió de propietat intel·lectual (*background i foreground*) que afavoreixin la implantació d'un model de negoci factible a mig termini.

Suport a la decisió en l'aplicació d'antibiòtics en explotacions ramaderes



3.1. Introducció

Els antimicrobians, han estat fonamentals en la prevenció i lluita contra les malalties que afecten tant a les persones com als animals, no obstant en els darrers anys s'ha observat una pèrdua d'efectivitat d'aquests productes enfront organismes i malalties que abans neutralitzaven de forma ràpida i eficaç. Aquestes resistències als antimicrobians, suposen un greu problema, tan per a la salut pública, com per a la sanitat animal. Infeccions comunes que cursaven de manera lleu, poden perllongar-se, agreujar-se i inclús comportar la mort, a la vegada que tractaments que fa un temps eren vàlids, ara ja no són efectius.

De manera general, el terme de resistència als antimicrobians fa referència a la resistència de diferents microorganismes (cucs, paràsits, bacteris, virus, fongs...) als fàrmacs utilitzats per tractar les malalties que produeixen. En aquest sentit, els microorganismes multiresistents, són els que han adquirit resistència a la majoria d'antimicrobians.

En el cas concret dels bacteris, s'ha demostrat que han creat resistències als antibiòtics, que són els fàrmacs utilitzats per combatre'ls.

La resistència als antimicrobians és fruit de la capacitat natural dels bacteris a mutar i fer-se resistents. L'augment de les resistències antimicrobianes és conseqüència de diversos factors, un dels quals ha estat el seu elevat i erroni en la producció ramadera, de la mateixa manera que en salut humana.

Per lluitar contra aquesta problemàtica és important reduir l'ús dels antimicrobians i també la necessitat d'utilitzar-los. Es tracta d'un problema que afecta a nivell global, i per tant requereix d'una actuació conjunta (concepte *One Health*). En aquest sentit s'estan duent a terme accions coordinades a diferents nivells per a buscar una sortida al problema evidenciat, implicant professionals de tots els sectors afectats.

Un dels elements importants per lluitar contra les resistències és conèixer el consum d'antibiòtics a les explotacions ramaderes.

Des del Departament d'Acció Climàtica, Alimentació i Agenda Rural es realitza un càlcul de consums d'antibiòtics trimestralment per poder conèixer el valor real d'aquestes substàncies a les explotacions a partir de les prescripcions que realitzen el veterinaris, juntament amb les dades de capacitat, cens i moviments de les granges.

El consum s'obté a partir de la quantitat d'antibiòtic prescrita als animals (mg) i dels animals a tenir en compte de la granja (UR: unitats de referència).

Repte

Aquest estudi ha de permetre per una banda identificar quins són els factors que determinen un major/menor ús d'antibiòtics en funció del context de l'explotació i detectar desviacions de comportaments rellevants en explotacions similars.

Els resultats obtinguts han de permetre establir relacions entre els diferents paràmetres d'estudi i l'ús d'antibiòtics per tal de poder conèixer la relació dels diferents elements amb el consum d'antibiòtics i així posar el focus en aquells que són importants.

Objectius

L'objectiu principal és determinar la relació entre diferents elements de context (característiques de l'explotació i maneig) relacionats amb l'explotació ramadera i el consum d'antibiòtics (tipus i freqüència).

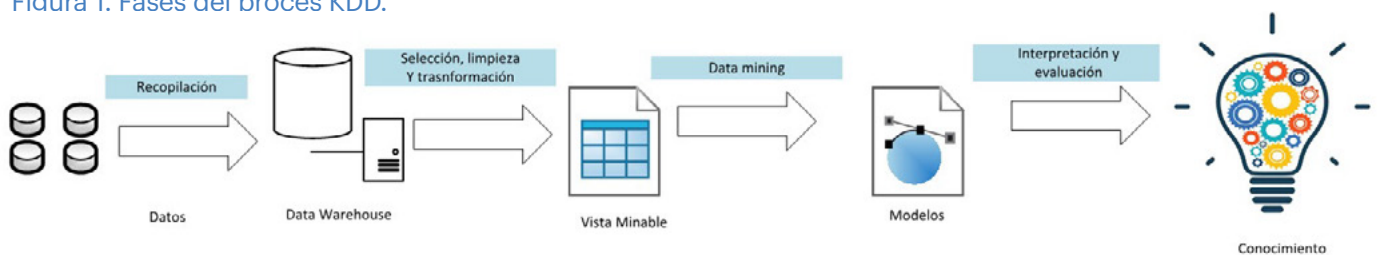
Per tal d'assolir aquest objectiu cal inicialment definir els estrats que agrupin explotacions amb característiques similars i posteriorment recollir les dades necessàries per cadascun dels àmbits.

El treball realitzat en aquest cas d'ús s'ha focalitzat sector porcí, però la metodologia i la plataforma intel·ligent és extrapolable a les altres tipologies d'animals per a les quals es disposa de la mateixa tipologia de dades.

3.2. Descripció general de la metodologia KDD

El tractament de les dades s'ha dut a terme amb la metodologia KDD (Knowledge Discovery in Databases). Podem definir el KDD com a "la identificació no trivial de patrons en les dades vàlides, noves, potencialment útils i comprensibles" (Fayyad, 1996). Aquest procés implica múltiples passos, com ara la preparació de dades, la cerca de patrons, l'avaluació del coneixement i el perfilat. No és un càlcul simple, sinó un procés complex de cerca i inferència. Un patró es considera coneixement si supera un llindar d'interès establert per l'usuari, encara que això no implica una definició filosòfica o àmplia del coneixement. El coneixement en aquest context és específic de l'usuari i del domini, determinat per les funcions i els llindars elegits per l'expert.

Figura 1. Fases del procés KDD.



Font: (<https://bit.ly/3CHqG1o>)

En concret, definirem les fases del procés KDD a:

- Recollida de dades
- Preprocessament: selecció, neteja i transformació de les dades
- Minería de dades
- Avaluació i interpretació

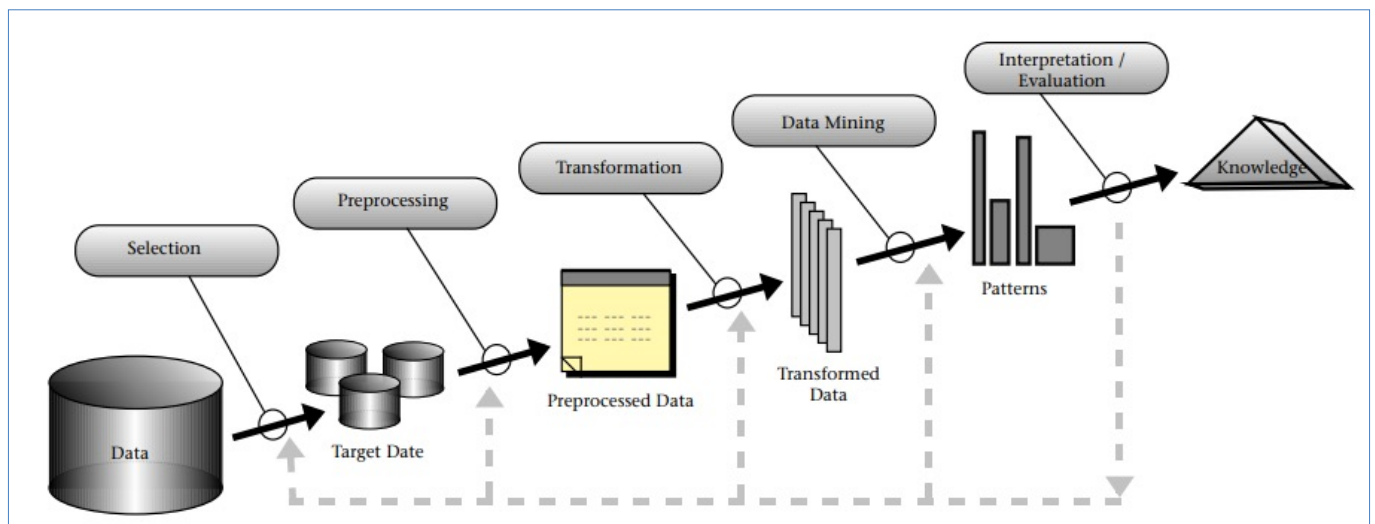


Figura 2. Descripció general dels passos de la metodologia KDD " (Fayyad, 1996)

Més detalls del procediment experimental es poden consultar a l'ANNEX (A1, A2).

3.3. Anàlisi de dades experimentals

Aquest capítol detalla l'experimentació final seguint les recomanacions dels experts enfocant l'anàlisi per kg d'animal (ús de les variables mgUR com a protagonistes).

Taula mestra final

En aquesta secció es mostren quatre registres de les variables afegides a la taula F1 que inclouen:

- Variables desagregades del consum i en percentatge, per tal de validar que coincideix amb les variable proporcionades en F1
- Quatre variables que els experts van considerar important afegir en l'anàlisi: es_premezcla, NO_premezcla, NOM ADS i esglaonada
- Altres variables provinents de les taules auxiliars per tal d'enriquir l'anàlisi

Figura 3. Variables de la taula mestra

Anio_Trimestre	mgConsumo_B	mgConsumo_C	mgConsumo_D	pct_mgConsumo_B	pct_mgConsumo_C	pct_mgConsumo_D
2021_2	250000	6300100	22311000	0.87	21.83	77.3
2021_2	44250	7020000	10112564	0.26	40.87	58.87
2021_2	1505000	19895150	38330000	2.52	33.31	64.17
2021_2	0	0	5566000	0	0	100

es_premezcla	NO_premezcla	CODI POSTAL EXPLOTACIÓ	SERVEI TERRITORIAL EXPLOTACIÓ	PROVINCIA EXPLOTACIÓ	COMARCA EXPLOTACIÓ	MUNICIPI EXPLOTACIÓ
NO	0.8181818181818181	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
NO	0.7575757575757575	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
NO	0.891304347826	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca
SI	0.4	25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca

NOM ADS	total_animals	prom_bloc1	prom_bloc2	prom_bloc3	prom_punt_total	esglaonada
1	NA	0	0	0	0	1
1	NA	0	0	0	0	0
1	NA	0	0	0	0	1
1	NA	0	0	0	0	0

- Es_premezcla: moda de las prescripcions discriminant entre sí/no premezcla (binària)
- No_premezcla: fracció de no premezcla reportat en les dades
- NOM ADS: 0 és sense descripció (assumint que no té ADS) i 1 en cas contrari (binària)
- esglaonada: si l'explotació és esglaonada (binària)

Aquesta taula mestra final pot fer-se servir per anàlisis interns i, sobre tot, per visualitzar els anàlisis gràfics en el panell de Power-BI per a la presa de decisions intel·ligent i amigable per l'usuari coneixedor en el camp.



3.4. Part 1. Anàlisi multivariant de variables

Anàlisi de correlacions amb totes les variables numèriques

Les següents figures mostren la correlació entre les variables que s’han analitzat.

Figura 4. Corrplot de totes les variables numèriques de la taula F1

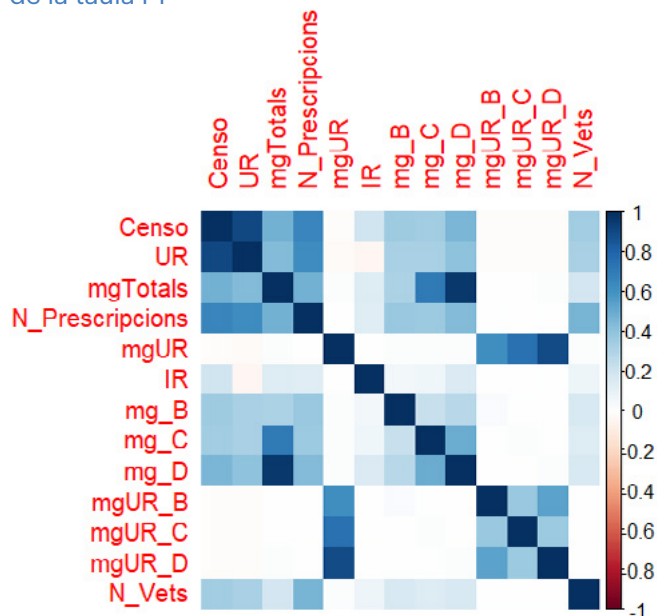


Figura 5. Biplot de totes les variables numèriques de la taula F1

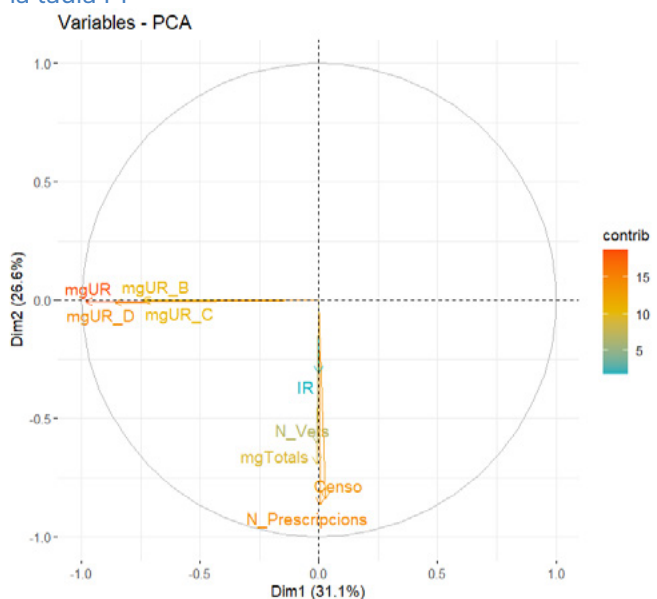
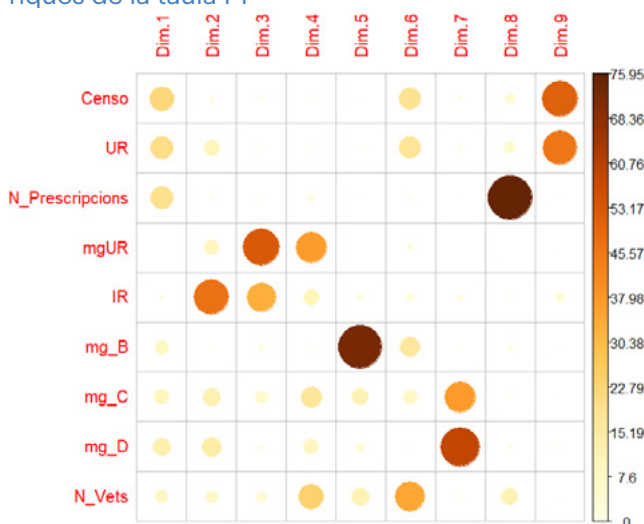


Figura 6. Mapa de calor de totes les variables numèriques de la taula F1



Es va fer un rànquing de les tres categories zootècniques més freqüents de les dades. La seva representació en tant per cent es troba en la tercera columna.

Taula 1. Top 3 categories zootècniques més freqüents (F1)

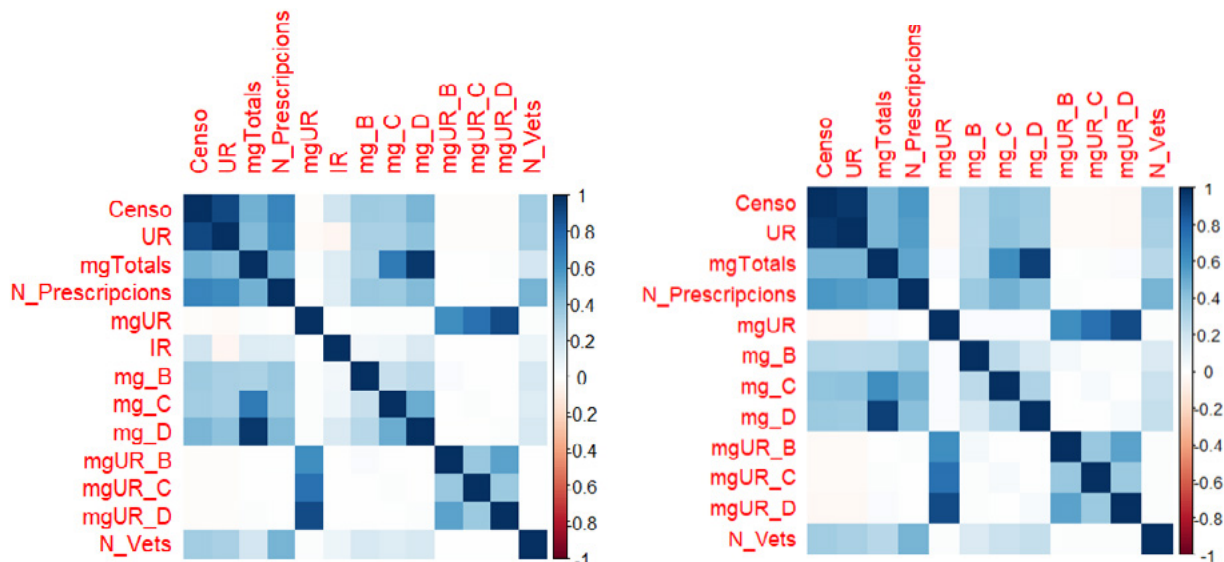
	number	%
Engreix	29871	72.505947
Producció cicle mixt	5346	12.976358
Transició de garrins	2574	6.247876

S’observa que les explotacions d’engreix són altament majoritàries (72,5%), i per tant, és una categoria dominant. Això indica que els anàlisis anteriors a aquest report són esbiaixats per la dominància de les explotacions d’engreix i no per la variable cens (la mida de l’explotació).

A continuació, es realitza un anàlisi multivariant discriminant per aquestes 3 categories zootècniques, ja que suposen el 91,6% del total d’explotacions, per ratificar o confirmar la hipòtesi dels experts.

1. Comparació de resultats globals amb les explotacions d'engreix

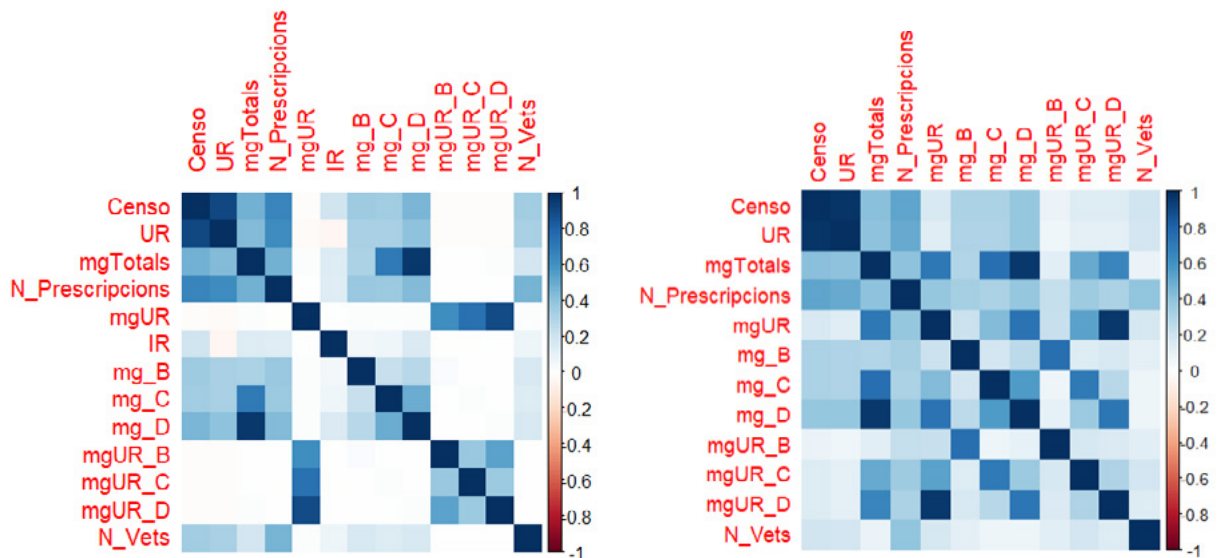
Figura 7. A l'esquerra: corplot de tots els registres. A la dreta: corplot dels registres només d'explotacions d'engreix



A les explotacions de categoria zootècnica Engreix s'observen pràcticament els mateixos patrons que a l'anàlisi del global de les explotacions.

2. Comparació de resultats globals amb les explotacions de producció de cycle mixt

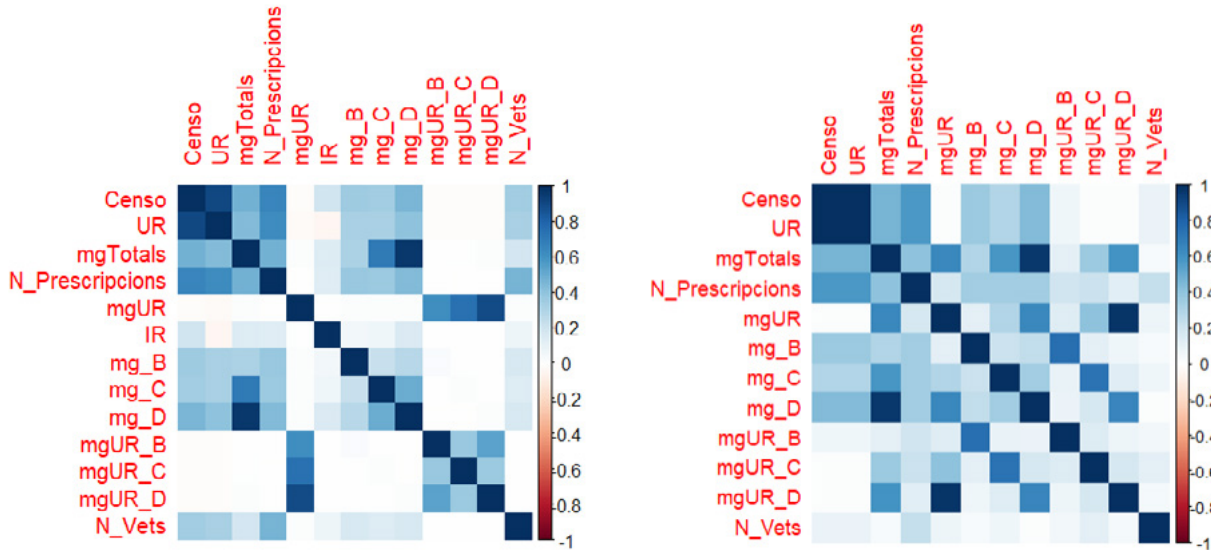
Figura 8. A l'esquerra: corplot del total de registres. A la dreta: corplot dels registres només d'explotacions de producció cycle mixt



En aquest cas l'associació per a les variables de consum canvia. S'observa que existeix certa relació entre la mida de la granja (cens), les UR i el nombre de prescripcions amb el consum (mg/UR).

3. Comparació de resultats globals amb les explotacions de transició de garrins

Figura 9. A l'esquerra: corplot dels registres d'F1. A la dreta: corplot dels registres només de transició de garrins



En aquest cas el corplot és similar a la situació del punt 2. Les explotacions d'aquestes categories zootècniques presenten característiques de consum d'antibiòtics diferents a les del conjunt total de les explotacions de porcí.

Com que les explotacions d'engreix representen un 72,5% del total d'explotacions, és va decidir realitzar dos últims ACP discriminant per categoria zootècnica en dos experiments:

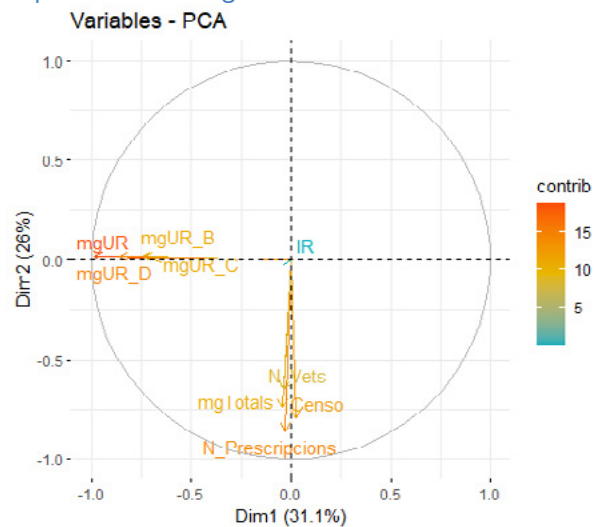
- Experiment 1: registres d'explotacions "Engreix"
- Experiment 2: registres d'explotacions de la resta de categories menys "Engreix"

Experiment 1

L'anàlisi de components principals indica que:

- Les variables relacionades amb el consum estan molt associades amb la grandària de l'explotació.
- Les variables de consum d'antibiòtic per categoria (B, C, D) i el mg/UR totals estan molt associades entre elles.
- El IR nacional no es correlaciona amb el consum d'antibiòtics de les explotacions de porcí d'engreix a Catalunya.

Figura 10. Biplot de l'anàlisi multivariant per l'experiment 1 – Engreix



Les figures següents mostren les correlacions entre les diferents variables. L'IR és una constant de manera que té una dimensió pròpia.

Figura 11. Heatmap de l'anàlisi multivariant per l'experiment 1 - Engreix

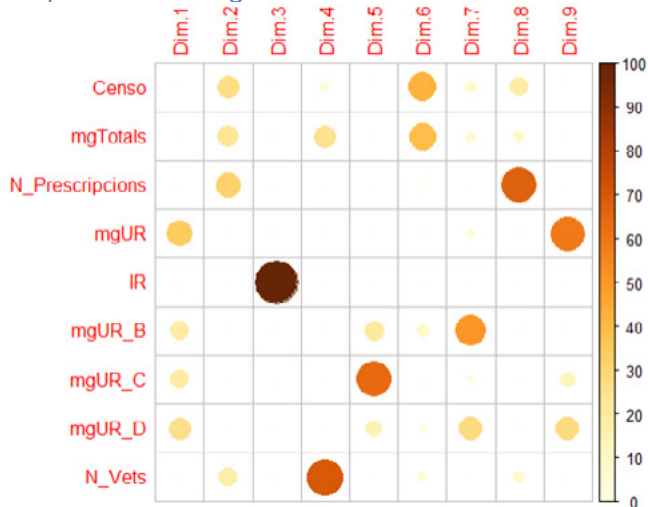
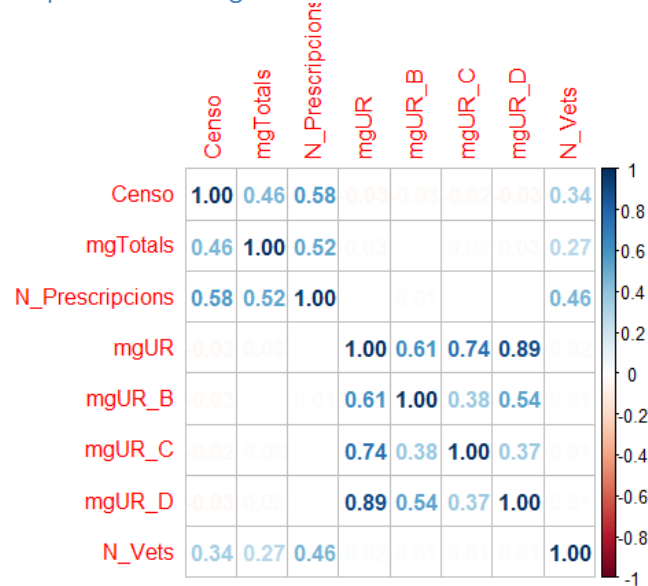


Figura 12. Corrplot de l'anàlisi multivariant per l'experiment 1 - Engreix



L'anàlisi de correlacions indica que:

- El consum d'antibiòtics (mgUR) i IR són independents.
- Existeixen relacions negatives però extremadament petites entre el cens i el consum d'antibiòtics de l'explotació. A mesura que augmenta l'explotació, disminueix el consum, i viceversa.
- El cens de l'explotació, consum d'antibiòtic i nombre de veterinaris estan molt associades.
- El consum d'antibiòtics (mgUR) està altament associat amb el consum de l'antibiòtic D (0,89), encara que també té una forta relació amb C i B (0,74 i 0,61).

Experiment 2

L'anàlisi de components principals indica que:

- Les variables que més influeixen en les característiques de consum d'antibiòtics de les explotacions que no són d'engreix són mgUR seguit de mgUR del tipus D. Tots dos estan molt associats.
- Els consums en mg per als antibiòtics D i C estan molt associats.
- El consum d'antibiòtic tipus B és independent dels anteriors.
- Les variables Cens, N_Prescripcions i N_Vets estan molt relacionades entre elles.

Les figures següents mostren les correlacions entre les diferents variables:

Figura 13. Biplot de l'anàlisi multivariant per l'experiment 2

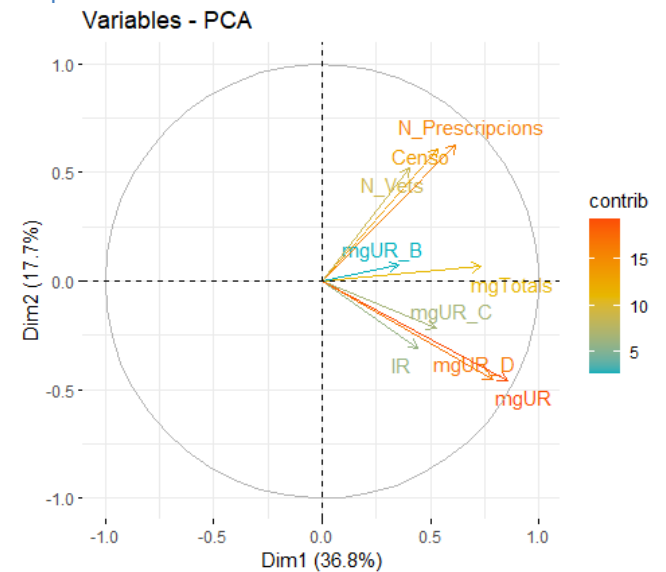


Figura 14. Heatmap de l'anàlisi multivariant per l'experiment 2

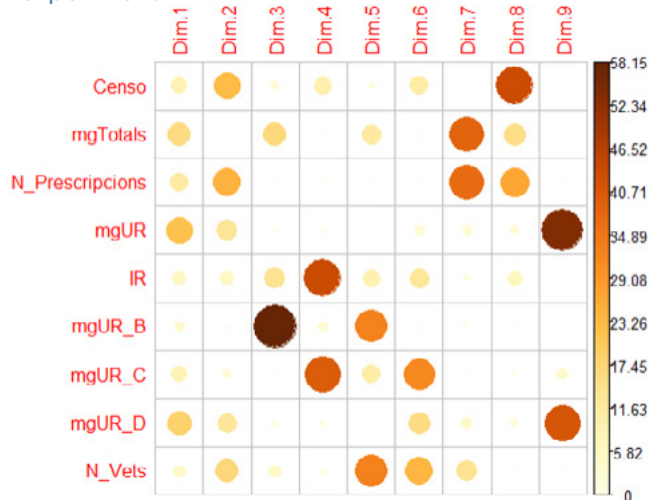
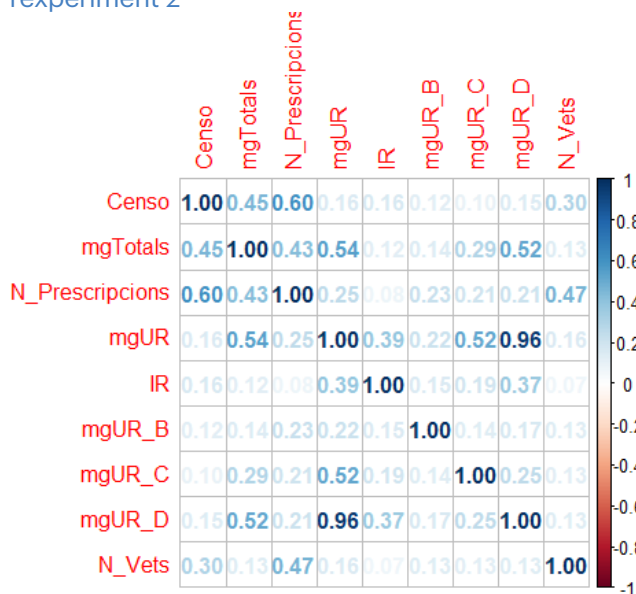


Figura 15. Corrplot de l'anàlisi multivariant per l'experiment 2



L'anàlisi de correlacions indica que:

- IR i el consum d'antibiòtic D estan molt relacionats.
- El consum del medicament B és independent de les altres dues tipologies.
- mgUR i mgUR_D estan molt relacionats.
- N_Vets està relacionat amb el cens de les explotacions i el nombre de prescripcions



3.5. Part 2. Clustering i profiling

El *clustering* jeràrquic té com a objectiu trobar patrons més discriminants per a diferenciar els codis d'explotació. Es recull una mostra de 15.000 registres de forma aleatòria. A continuació les variables seleccionades en l'anàlisi.

Taula 2. Clustering, escenari 1. Variables seleccionades

Variables
Censo
IdCZ
N_Prescripcions
mgUR
mgTotals
mgUR_B
mgUR_C
mgUR_D
N_Vets

Taula 3. Composició en nombre de les classes

Classes	1	2	3	4
Nombre d'ID explotacions	10.819	1.227	1.956	998

La classe més gran és la primera, i la més petita la quarta.

Taula 4. Classes per categoria zootècnica

	1	2	3	4
Altres	0	319	0	0
Centres d'inseminació artificial i centres de recollida de semen	0	89	0	0
Engreix	10818	0	0	0
Producció cycle mixt	0	0	1956	0
Producció de cycle tancat	0	291	0	0
Reproductors	0	186	0	0
Transició de garrins	0	0	0	998
Transició de reproductores nul·lipares i cria de reproductors	0	342	0	0
Varies Zoo	1	0	0	0

S'observa que les classes 3 i 4 són explotacions de producció de cycle mixt i transició de garrins respectivament i de forma exclusiva. Similarment passa amb el primer grup, el més gran, on tots els registres pertanyen a "Engreix" i una a "Varies Zoo". La classe 2 agrupa les categories zootècniques més variades, creant un patró de similitud.

Validació estadística

Es validen de forma estadística els resultats de l'anàlisi final anteriorment descrit. Aquí es mostren el test de les variables mgUR i IdCZ.

```
Kruskal-wallis rank sum test
data: samp$mgUR by samp$c4
Kruskal-wallis chi-squared = 4679.5, df = 3, p-value < 2.2e-16
```

```
Pearson's Chi-squared test
data: samp$IdCZ and samp$c4
X-squared = 123594, df = 24, p-value < 2.2e-16
```

Els dos tests són estadísticament significatius. Per tant, podem concloure que:

- Al menys en un dels grups, el consum de mg d'antibiòtic per UR és diferent que als altres.
- Al menys en un dels grups és diferent als altres segons la categoria zootècnica.

Es va procedir d'igual manera la validació estadística de la resta de variables numèriques i categòriques. Tots els test van ser estadísticament significatius.



Perfilat final

S'han assignat colors a les classes de forma orientativa o indicadora en base al consum d'antibiòtic per mgUR segons el criteri dels analistes. En vermell les classes més "greus" que requereixen la revisió de l'usuari final i en verd les que són més estables en el seu consum.

Taula 5. Guia final de la classificació de les explotacions

Classe	Id explotacions	Descriptiva de les classes
1	10819	Explotacions de mida, producció i consum mitjà. En mitjana, tenen 6 prescripcions i predomina el consum d'antibiòtic D. De mitjana 2 veterinaris. Són bàsicament explotacions d'engreix.
2	1227	Explotacions de mida, producció i consum mitjà-baix. Predomina el consum d'antibiòtic D i de mitjana tenen 1 veterinari. Gran variabilitat de la categoria zootècnica, la predominant és Transició de reproductores nul·líparas i cria de reproductors.
3	1956	Explotacions de mida gran, producció i consum mitjà. És la que té més prescripcions (30 de mitjana) i predomina el consum d'antibiòtic D. De mitjana 2 veterinaris. Correspon exclusivament a explotacions de cycle mixt.
4	998	Explotacions de mida, producció i consum més alt. De mitjana, tenen 27 prescripcions i altíssim consum d'antibiòtic D. De mitjana 2 veterinaris. Correspon íntegrament a explotacions de transició de garrins.

3.6. Part 3. Resultats finals amb tots els registres

Assignació de classes pel mètode CURE

Es va fer servir l'algorisme CURE per fer l'assignació de classes pels registres fora de la mostra, resolent el problema de la manipulació de grans bases de dades en R, tal i com es va explicar anteriorment.

Taula 6. Grandària de les classes

Classes	1	2	3	4
Nombre d'ID explotacions	29878	3400	5346	2574

Profiling final

Aquesta secció mostra els resultats gràfics de les variables numèriques en forma de Class Panel Graph (CPG) per tal de comparar les classes segons les variables (Gibert K. G.-R.-S., 2008).

Figura 16. CPG de les variables Censo, UR i N_Prescripcions

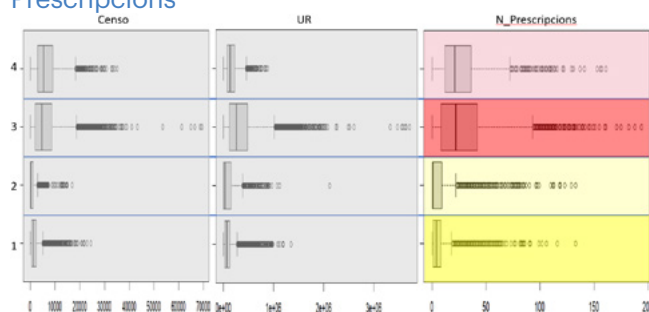


Figura 17. CPG de les variables mgTotals, mgUR, mgUR_B

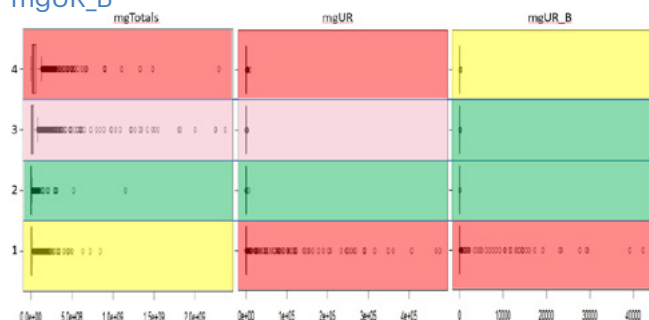
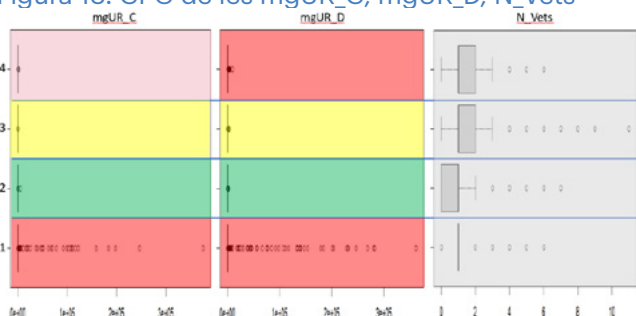


Figura 18. CPG de les mgUR_C, mgUR_D, N_Vets



El Class Panel Graph (CPG) és una eina visual que permet determinar en detall el comportament de les variables condicionades a cada classe. Les figures anteriors mostren aquest comportament en les variables numèriques utilitzades en el procés de cluserització. Els colors assignats dins del CPG és una altra eina de visualització per al perfilat, que mostra a l'usuari final d'una forma molt més amigable una interpretació directa amb l'associació d'aquests colors als colors d'un semàfor tradicional.

En el nostre cas aquests colors van ser assignats amb la comparativa dels centroides condicionats a les classes, tant per la mitjana com la distribució mostrada en els boxplot del CPG. Per tant, el verd representa valors de comportament positius i el vermell tot el contrari depenent de la mesura implícita de cada variable (mgUR vermell implica major consum d'antibiòtic). Finalment, fent un recompte dels colors presents a cada classe en el semàfor, s'estableix una tendència per al color definitiu de la classe o perfil com es mostra en la taula de perfilat descriptiu final (Taula 5).

A continuació es mostren les variables categòriques de cada classe. Amb aquests es finalitza el perfilat de les classes.

Figura 19. Barplot de la variable DesCZ

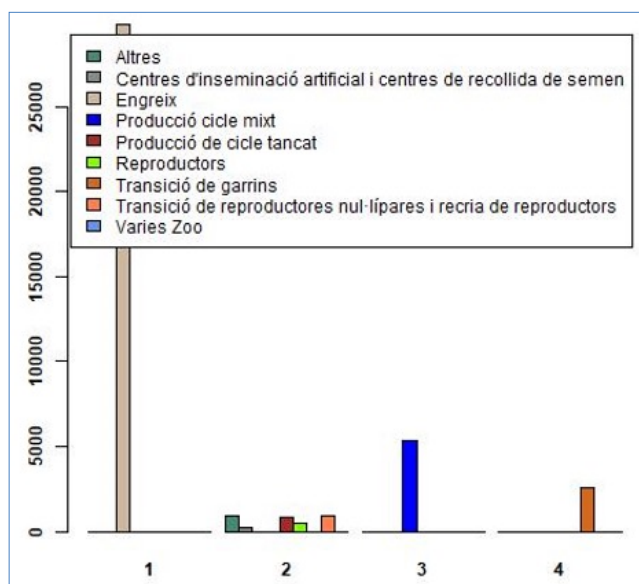
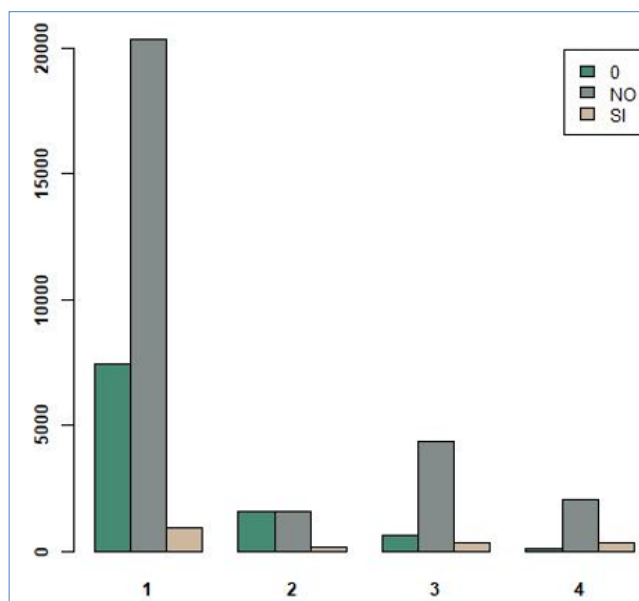


Figura 20. Barplot de la variable es_premezcla



Pel grup 2 existeix un equilibri entre les explotacions que administren tant antibiòtics via pinso (premescles) com altres formes farmacèutiques. En la resta de grups predomina les explotacions que mediquen via altres formes diferents als pinsos medicamentosos.

Figura 21. Barplot de la variable Servei Territorial

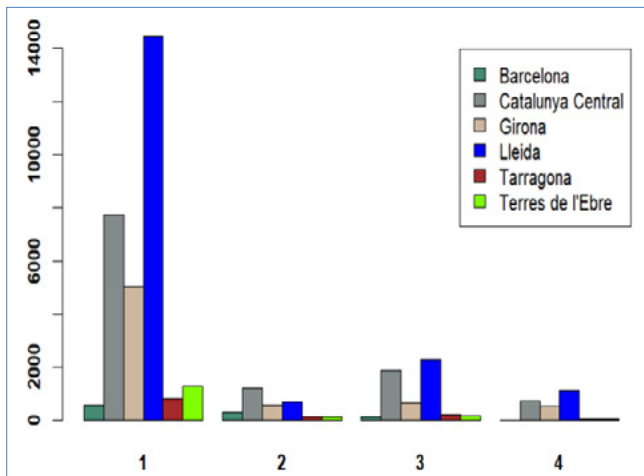


Figura 22. Barplot de la variable IR

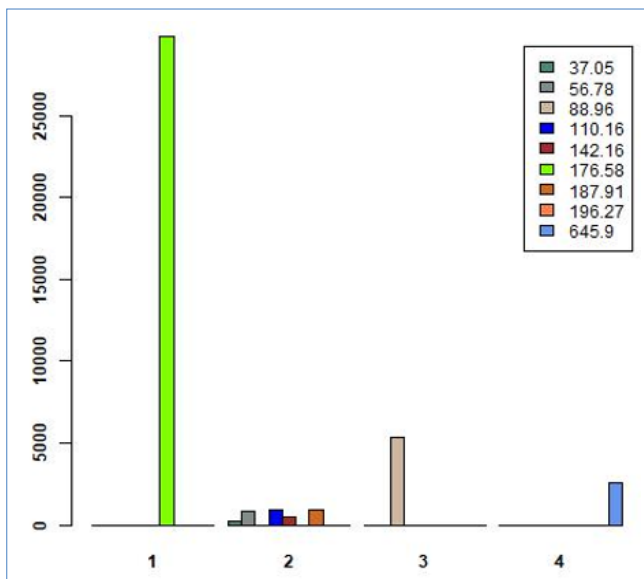


Figura 23. Barplot de la variable Província

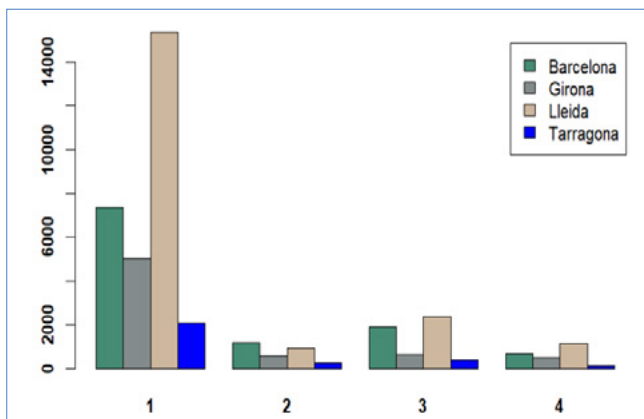


Figura 24. Barplot de la variable Nom.ADS

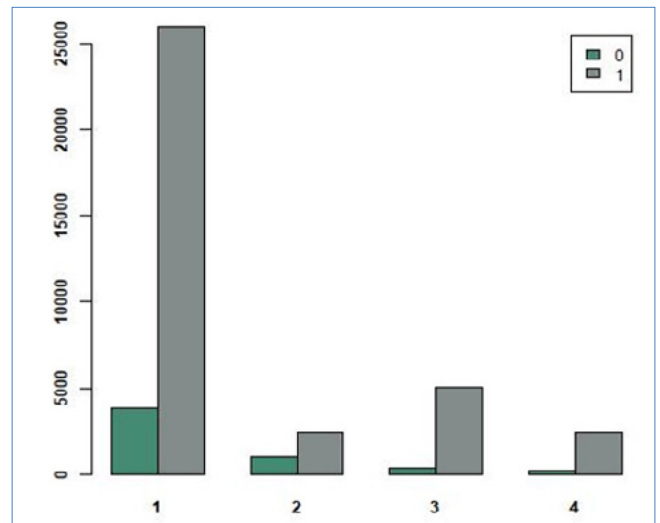
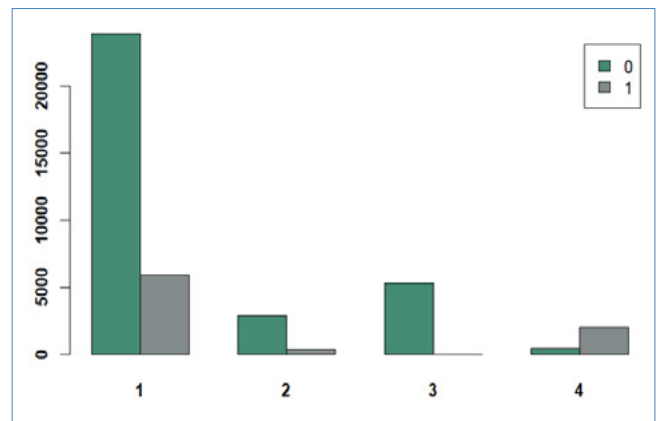


Figura 25. Barplot de la variable Esplaonades



3.7. Visualitzador – RanchingAnalytics.cat

Descripció del visualitzador

Amb les dades centralitzades i depurades s'ha procedit al disseny i implementació d'un visualitzador en plataforma Microsoft Power BI. Aquest visualitzador té com a objectiu complementar la informació obtinguda en l'anàlisi de les dades i els models explicats prèviament, per a així donar suport addicional d'una forma gràfica i amigable per a l'expert (usuari final) amb vista explicatives de consum d'antibiòtics en les explotacions de Catalunya resumides en un relat, o storytelling, que focalitza indicadors a nivell general sectoritzats per geografia i després a un zoom més detallat per a examinar cadascuna, i de forma individualitzada, de les explotacions ramaderes. Aquesta versió del visualitzador ([RanchingAnalytics.cat](#)) s'ha construït en conjunt amb els experts i ha estat fruit d'una interacció constant al llarg del desenvolupament del projecte.

Audiència

RanchingAnalytics.cat ha estat dissenyat per al públic en general. No obstant això, ha estat personalitzat per a manejar termes i conceptes propis de la ramaderia i l'ús d'indicadors associats al consum d'antibiòtics en aquest sector. Té un apartat al qual se li pot annexar glossari de termes, enllaços o informació addicional rellevant per a aquells usuaris amb menor coneixement de l'àrea.

Mitjà de divulgació

RanchingAnalytics.cat pot publicar-se en Internet. Aquesta versió té maneig intern. La decisió de la seva divulgació deu encara considerar-se i decidir-se per part dels propietaris de les dades (Generalitat de Catalunya).

Vistes principals (funcionalitats i Storytelling)

Ranchinganalytics.cat és un visualitzador que es divideix en tres grans seccions: (1) Una primera secció de presentació, crèdits i glossari de termes; (2) Una segona secció que detalla l'ús dels antibiòtics en les explotacions ramaderes (porcínes) de Catalunya amb Indicadors generals i posteriorment amb vista detallades i finalment, (3) Una tercera secció que compta amb el detall de la informació per cada codi d'explotació (explotació ramadera) incloent una vista comparativa final per a avaluar i explorar 2 explotacions al mateix temps (aquesta anàlisi també és viable a nivells de província, municipi, servei territorial d'explotació i comarca).

Figura 26. Pàgina principal del visualitzador

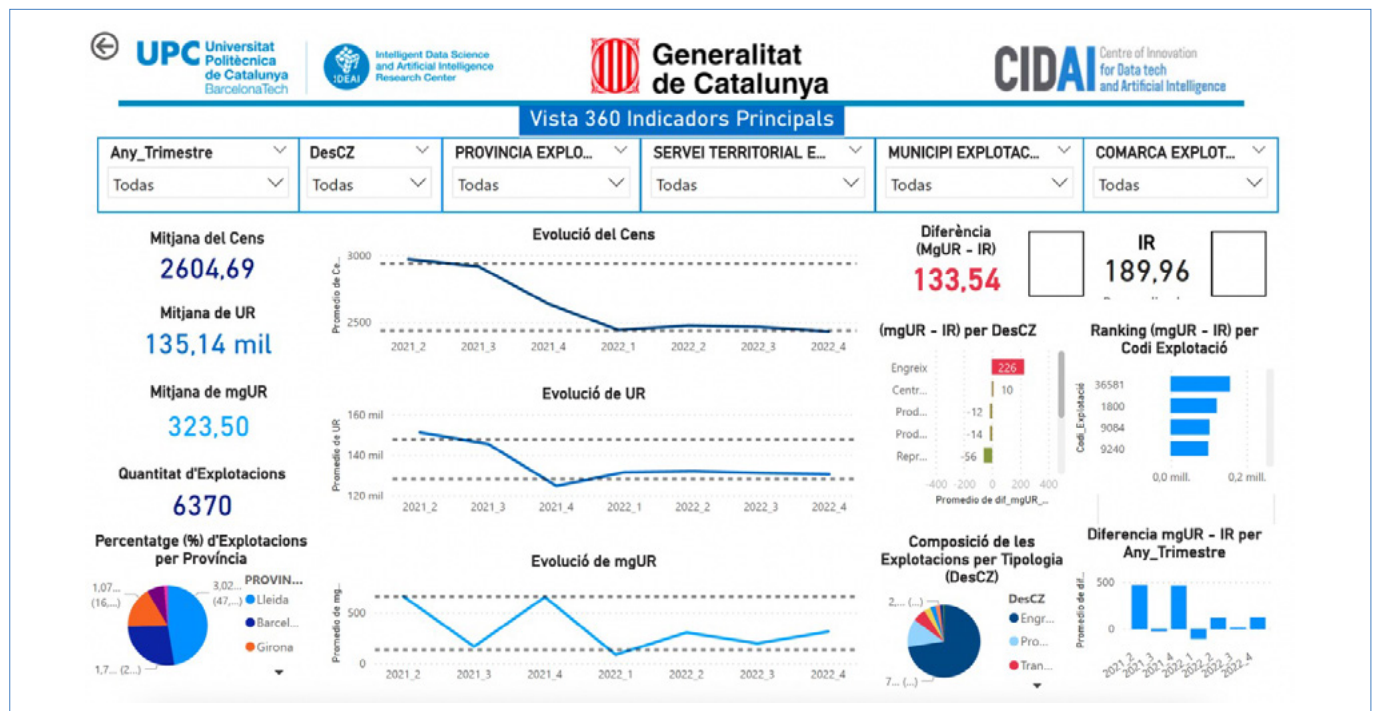


Secció 1 - Ús d'antibiòtics en explotacions porcínes de Catalunya

Vista 360 d'Indicadors Principals

Aquesta vista inclou indicadors principals sobre l'ús d'antibiòtics i descriptors del sector porcí a Catalunya. Posseeix filtres per a interactuar a diversos nivells: Trimestre, Descripció Zootècnica i ubicació geogràfiques. Es mostra a l'usuari evolucions del Cens, UR, mgUR, diferència entre el mgUR reportat a Catalunya i l'índex de referència nacional i altres gràfics descriptius. Posseeix botons d'ajuda (?) per a facilitar l'enteniment dels gràfics.

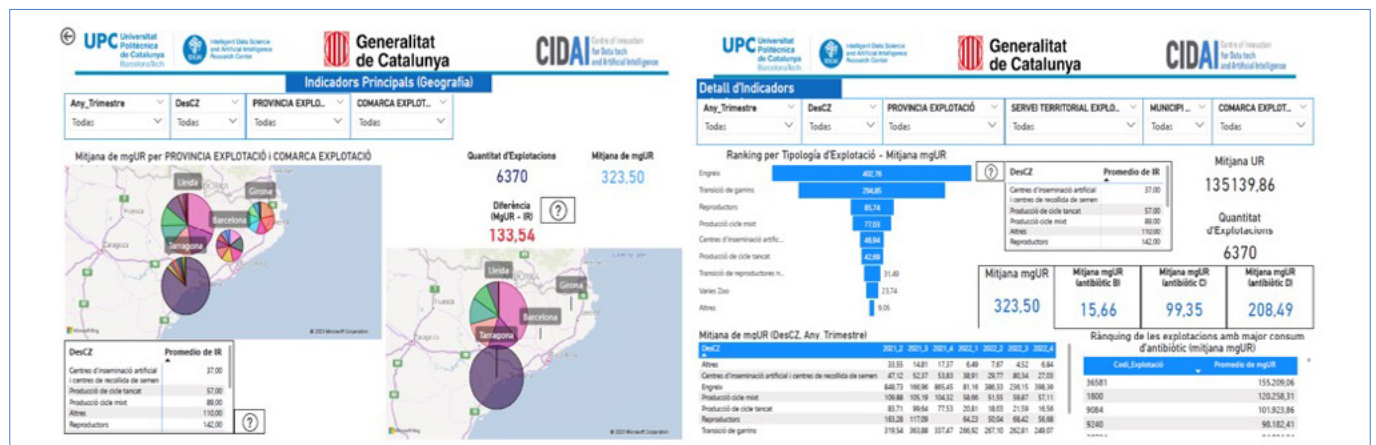
Figura 27. Vista d'indicadors principals



Detall d'Indicadors

Aquí s'aprofundeix l'anàlisi portant-lo a visualitzacions geogràfiques i als KPI de consum detallat d'antibiòtics per tipologia (B,C i D)

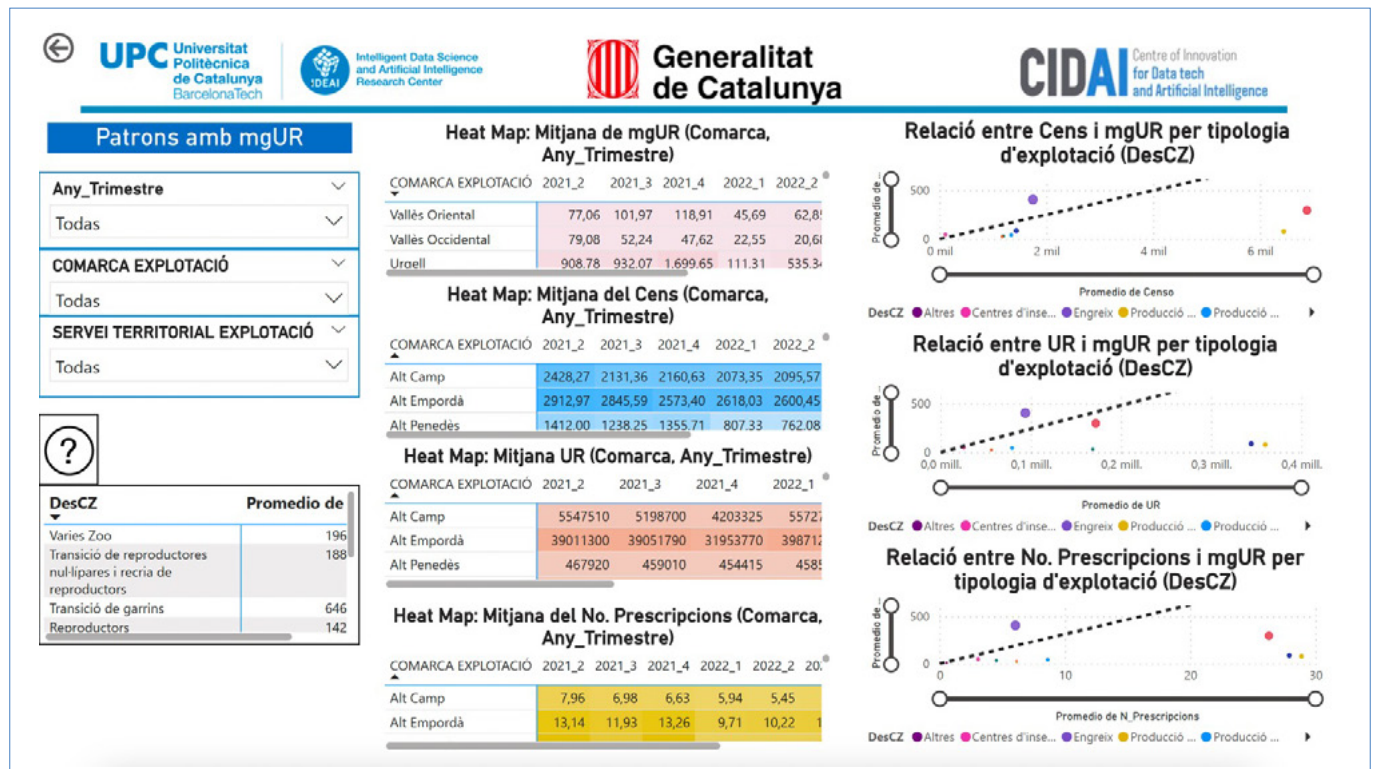
Figura 28. Detall d'indicadors



Vista de patrons i relacions

Aquesta vista mostra relacions explicatives entre el cens de l'explotació i el consum d'antibiòtics, entre les UR i el consum d'antibiòtics i finalment entre el nombre de prescripcions i l'índex de consum d'antibiòtics. Aquests quadres es mostren en base a la classificació zotècnica de les explotacions porcínes.

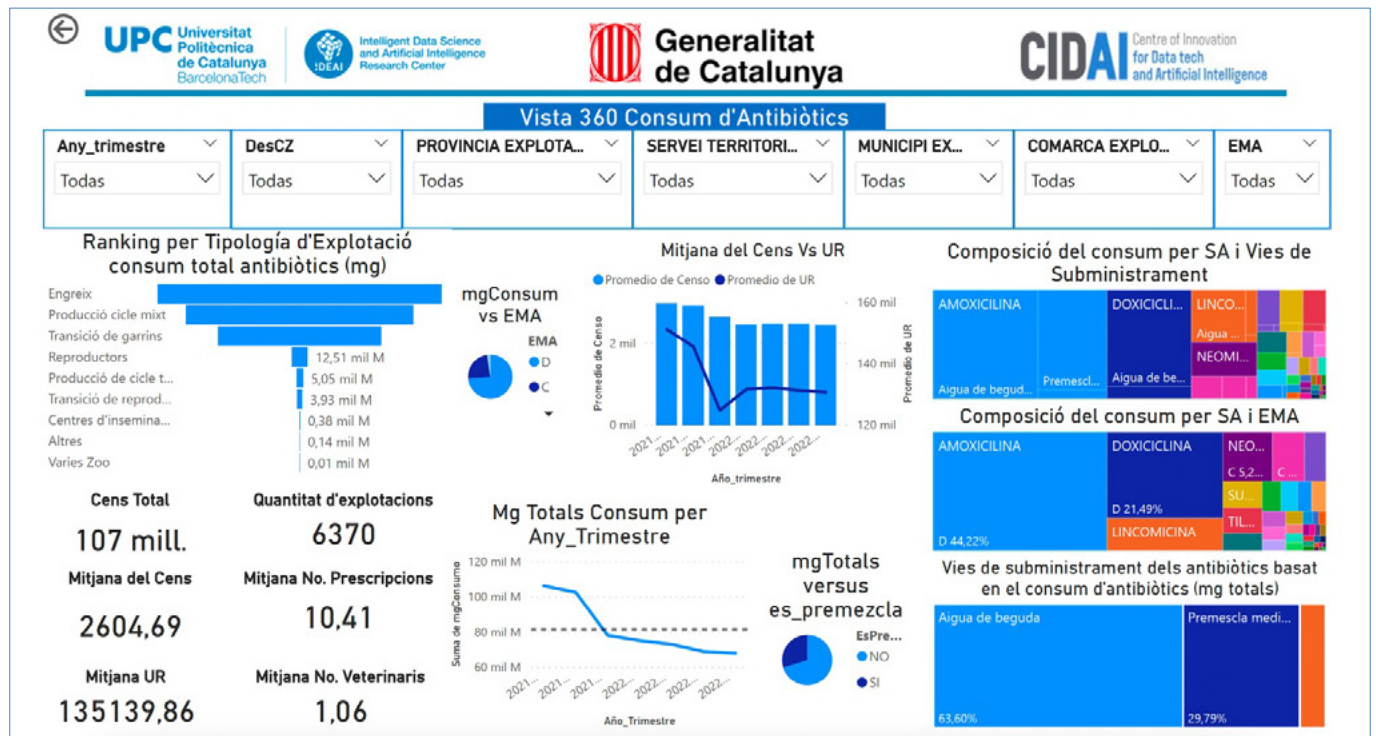
Figura 29. Vista de patrons i relacions



Vista detallada (ZOOM) del consum d'antibiòtics a Catalunya

Aquesta vista ofereix de manera detallada l'ús dels antibiòtics (vies d'administració, famílies d'antibiòtics més utilitzades, tendències i gràfics de composició). Els filtres faciliten explorar resultats per períodes, ubicació de les explotacions i classificació EMA de l'antibiòtic.

Figura 30. Vista detallada de consums



Secció 2 - Ús d'antibiòtics en explotacions porcínes de Catalunya

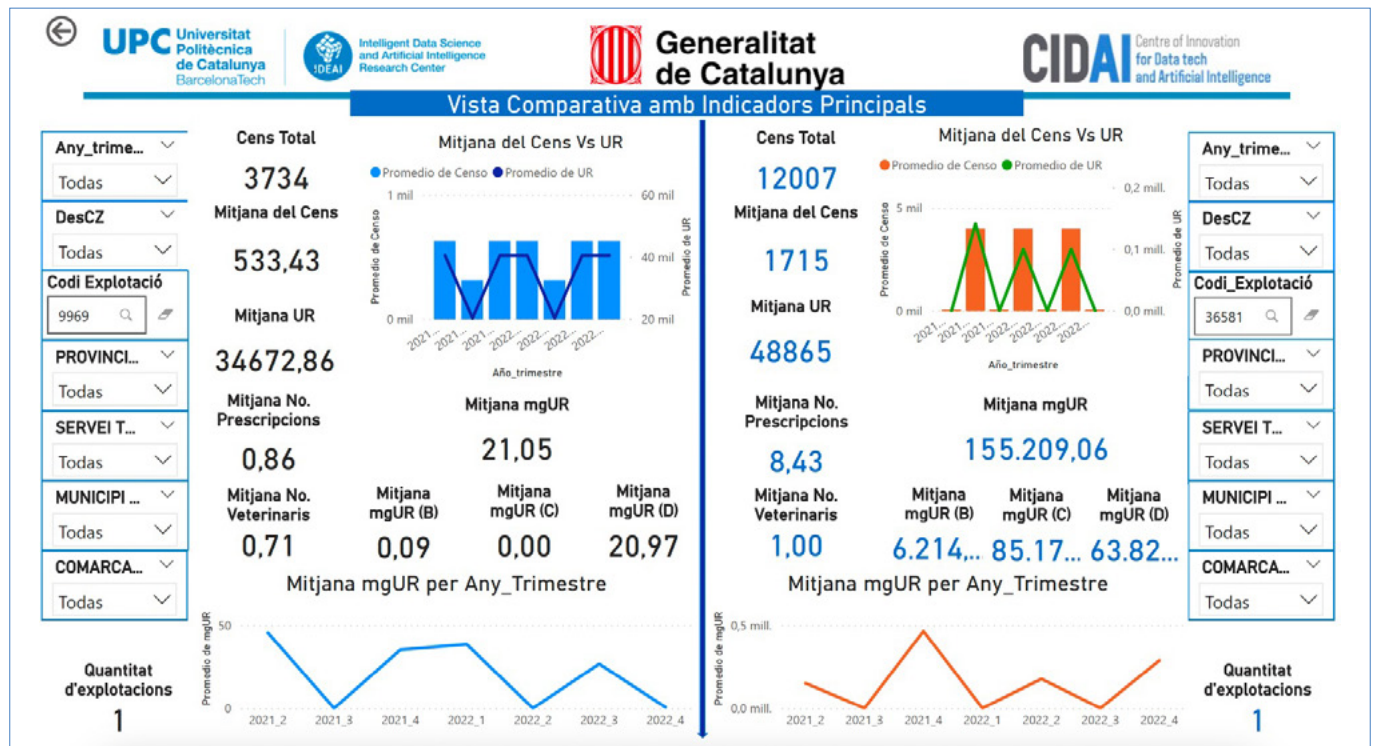
Detall a nivell de codi d'explotació: les vistes són similars a les mostrades anteriorment però amb informació individualitzada per a cada codi d'explotació present en la base de dades. A l'efecte de mantenir la privacitat de les dades, aquestes vistes no es mostren en aquest informe.



Vista Final

Aquesta vista permet comparar de manera simultània gran part de la informació descrita prèviament. És una eina potent per al procés de presa de decisions, ja que es permet comparar explotacions i períodes diferents, entre d'altres. (Nota: s'han anonimitzat les 2 explotacions comparades en la figures als fins de preservar la privacitat de les dades)

Figura 31. Vista comparativa amb indicadors principals



2.8. Conclusions

Durant el desenvolupament d'aquest primer cas d'ús s'han pogut complir els objectius principals de l'estudi. D'una banda, s'ha realitzat una anàlisi amb diferents escenaris i experiments per descriure les relacions entre les variables de consum d'antibiòtics i les característiques de les explotacions. S'ha determinat que la categoria zootècnica és un dels factors més discriminants sobre el consum d'antibiòtics. Per altra banda, s'han pogut detectar desviacions de comportaments rellevants sobre el consum d'antibiòtics al bestiar porcí en descobrir quatre grups ben diferenciats d'explotacions en el marc ramader porcí de Catalunya. Finalment, els experiments relacionats amb l' Anàlisi de Components Principals indiquen que l'indicador base subministrat pels experts conegut com a mgUR no està alineat amb l'indicador global nacional de referència IR a les explotacions d'engreix.

Visió per Computador pel seguiment individualitzat de porcs en granja



4.1. Introducció

Aquesta acció neix de la necessitat general de tots els PAIs d'anar més enllà del pilot, proporcionant instruments que permetin l'escalabilitat i la integració propera en mercat. Interpreta l'instrument AI LAB definit en l'Estratègia Catalana d'Intel·ligència Artificial¹ i li dona forma en un context específic aplicació. D'aquesta manera, el PAI es fa servir com a espai de prototipatge de l'instrument en sí mateix. Aquesta necessitat d'anar més enllà del pilot i proporcionar models de negoci a partir de la validació d'innovació és una demanda directa dels actors productius del territori a Catalunya, que precisen d'una visió clara del retorn de la inversió que suposa la validació de les tecnologies en espais experimentals reals.

A nivell metodològic, el AI LAB es desenvolupa cinc etapes diferenciades:

1. Mapeig i identificació de propietaris de reptes al territori.
2. Co-creació de la proposta de solució.
3. Implementació del prototip tècnic.
4. Disseny i implementació del prototip de l'espai de dades.
5. Creació del model de negoci i sostenibilitat.

A nivell de projecte, i a partir del mapeig fet, es proposa una acció a les Terres de Lleida, per donar suport a la comunitat d'explotacions agràries. Diferents reunions amb agents del territori, com són la Federació de Cooperatives Agràries de Catalunya, representats de la Diputació de Lleida i empreses del territori, ens ha portat a identificar el Centre d'Estudis Porcins com a agent clau amb qui intercanviar coneixement i desenvolupar de manera conjunta una solució tecnològica d'alt valor pel sector porcí i pel territori. El equip de treball per desenvolupar el projecte ha estat format per:

- Eurecat
- Centre de Visió per Computador (CVC)
- Centre d'Estudis Porcins (CEP)
- Universitat de Lleida
- Barcelona Supercomputing Center (BSC)

Repte

El repte principal d'aquest cas d'estudi ha estat doble:

1. Com a repte fonamental, identificar els actors principals del teixit agrícola y ramader de les terres de Lleida interessats en desenvolupar i implementar solucions tecnològiques per integrar-los en un procés de innovació de tecnologia.
2. Una vegada identificats els actors i reptes, s'ha desenvolupat una solució tecnològica, la qual ha estat desplegada en el Centre d'Estudis Porcins de Lleida, sota la coordinació de projecte d'Eurecat. Seguint l'aproximació AI LAB, aquesta solució consta de 3 resultats diferents: Prototip tecnològic, espai de dades, i model de governança, sostenibilitat i negoci.
 - **Prototip tecnològic:** S'ha implementat un sistema basat en Visió per Computador per seguir individualment el cicle de vida de cada porquet en una quadra, amb l'horitzó d'identificar tendències de comportament. Aquesta acció ha estat coordinada pel Centre de Visió per Computador.
 - **Espai de dades:** De manera paral·lela també s'ha identificat com a repte realitzar una prova de concepte de la implementació d'un espai de dades del sector agrícola, que ha estat liderat per la Universitat de Lleida.
 - **Model de governança, sostenibilitat i negoci:** Durant el desenvolupament del projecte també es marca el repte de desenvolupar un model de governança i sostenibilitat que permeti identificar el background i foreground dels agents implicats i definir el full de ruta que permeti garantir la continuïtat i explotació d'aquest projecte i definir les accions que facilitin desenvolupar altres iniciatives similars. Aquesta acció ha estat liderada pel Barcelona Supercomputing Center (BSC).

¹ "S'impulsarà el CATALONIA.AI LAB com a entorn de compartició i testatge que ha de ser una finestra oberta a les idees i les iniciatives d'arreu que quedin aplegades i catalogades: innovació social des de les persones i per a les persones. El seu objectiu és doble, impulsar el sector tecnològic oferint accés a infraestructures i equipament, i, d'altra banda, potenciar la innovació oberta amb empreses i administracions públiques mitjançant la participació en proves pilot." [CATALONIA.AI. L'estratègia d'intel·ligència Artificial de Catalunya](#). Generalitat de Catalunya Departament de Polítiques Digitals i Administració Pública Secretaria de Polítiques Digitals. 2020

Objectius

En el marc contextual definit, podem identificar els següents objectius del projecte:

- Extrapolar el coneixement i la tecnologia de seguiment de vaques al context d'una explotació porcina (CVC, 2022).
- Extraure informació d'alt valor per als responsables de les explotacions que permetin avaluar el comportament dels animals per tal de poder aplicar accions enfocades a garantir i millorar el seu benestar.
- Desenvolupar una prova de concepte d'un espai de dades per valoritzar tant les dades com els algoritmes.
- Establir un model de governança i sostenibilitat dels desenvolupaments tecnològics.



Impacte socioeconòmic

Els reptes atacats en aquesta acció estan específicament alineats amb els actors rellevants del territori i amb els actius reals dels socis de CIDAI. L'impacte del projecte està també alineat amb l'estratègia de les Agendes Compartides promoguda pel Govern, garantint així una contribució als objectius a llarg termini identificats dintre de l'Estratègia d'Especialització Intel·ligent de Catalunya. Finalment, el desenvolupament d'aquest projecte ha servit per reforçar la transformació digital del sector agroalimentari dins del marc de les accions previstes pel Digital Innovation Hub Agroalimentari i Forestal de Catalunya. Des de la perspectiva específica de les innovacions creades s'han creat els següents impactes:

- El projecte, des de la visió d'AI LAB, valida un instrument escalable que pot ser replicat no només en el sector "agro", sinó en els diversos verticals del teixit industrial català, enfortint el catàleg d'eines metodològiques estratègiques per la injecció de la IA en l'ecosistema industrial. La solució basada en visió per computador ha permès identificar una nova línia estratègica de treball que permet obtenir indicadors objectivament verificables correlacionats amb indicadors específics de benestar animal.
- El prototip d'espai de dades creat és el pioner en el sector "agro" a Catalunya, i serveix com a punt de partida per la consolidació d'un instrument fonamental per la posada en valor de les dades, implementant una nova estratègia de valorització per les explotacions porcines: "de bits a quilos".
- El model de governança, sostenibilitat i negoci, tot i amb una aproximació inicial, proporciona un instrument de referència fonamental per la indústria de cara a comprendre com valoritzar a curt termini la inversió feta en el procés de prototipatge, i dona a l'ecosistema un mapa de ruta aplicable a tots els verticals de la indústria.

Els impactes mencionats actuen com a llavor de potencials impactes futurs, que han cristal·litzat en col·laboracions estables dintre del marc de transferència de la innovació ².

² L'equip del CIDAI participant en aquest projecte ha creat diverses propostes com a spin-offs que han estat enviades a la convocatòria d'ajuts a les activitats de demostració de transferència de coneixements (PEPAC 2023-2027)

4.2 Seguiment del cicle de vida dels porcs per a la millora del benestar animal

Context de la necessitat de solució tecnològica

La ramaderia abasta el conjunt d'activitats relacionats amb la cria, l'explotació, i el comerç de bestiar. Aquest sector és encara molt convencional i tradicional. Tanmateix, als últims anys s'ha vist que les noves tecnologies com el Big Data, la Internet de les Coses i la Indústria 4.0 estan arribant al món de l'agricultura i la ramaderia. Tant és així que s'ha començat a introduir el concepte d'agricultura intel·ligent (*Smart Farming*), per referir-se a l'aplicació de Tecnologies de la Informació i Comunicació a aquests sectors.

Des del punt de vista de l'agricultor i ramader, l'agricultura intel·ligent hauria de proporcionar-li alt valor afegit, a través d'eines d'ajuda a la presa de decisions o per a una gestió de les seves explotacions més eficient. Això està estretament relacionat amb 3 camps tecnològics interconnectats:

1. **Sistemes de Gestió de la Informació:** Sistemes planificats per a la recollida, processament, emmagatzematge i disseminació de tota mena de dades necessàries per a gestionar les operacions i funcions de les explotacions agrícoles.
2. **Agricultura de Precisió:** Gestió de les variables espacials i temporals per a millorar els rendiments econòmics i l'impacte mediambiental. Inclou eines d'ajuda a la decisió per a la gestió integral d'explotacions amb l'objectiu d'optimitzar els inputs i preservar els recursos, facilitada a través de l'ús d'un sistema GPS, sistemes d'observació terrestre, imatges aèries obtingudes a través de drons i pels satèl·lits, que permeten el desenvolupament de mapes amb diferents variables espacials relacionades amb la producció, la tipologia de terreny, el contingut de matèria orgànica, nivells d'humitat i de nitrogen en sòl, etc.
3. **Automatització agrícola i robòtica:** L'aplicació de tecnologies de robòtica, el control automatitzat i la intel·ligència artificial en tots els nivells

de la producció agrícola i ramadera, incloent l'ús de robots i drons agrícoles.

Finalment, les aplicacions de l'agricultura intel·ligent no només es dirigeixen a explotacions agrícoles convencionals de gran grandària, sinó que també poden convertir-se en un impuls rellevant per a les granges familiars, cooperatives, i cooperatives federades, abastant tota la granularitat en la que s'estructura l'ecosistema.

Aquest cas d'ús es va definir per avaluar com aquest paradigma pot ajudar a identificar i automatitzar certs elements dels processos que es porta a terme al sector porcí. Des de la perspectiva d'infraestructures de producció, el sector del porc inclou: les granges on neixen els garrins, les granges on fan l'engreix del porc, els escorxadors on sacrifiquen els porcs, i les empreses de logística s'encarreguen de portar la carn als supermercats. A falta de les associacions de consumidors, aquest cicle comprèn tota la cadena de valor que va de la granja a la forquilla ³.

El sector porcí és un sector molt important a Catalunya. Segons l'Institut d'Estadística de Catalunya el sector porcí amb uns 6.500.000 caps de bestiar ocupa el segon lloc a la taula després d'aviram amb més de 43.000.000 caps de bestiar. El creixement dels porcs es porta a terme en granges d'engreix, i on gran part de la població porcina són, precisament, porcs d'engreix. L'engreix del porc dura uns 4 mesos, i quan el pes final dels porcs està al voltant de 110 o 120 kg, són portats a l'escorxador per al seu sacrifici.

En aquest projecte s'han estudiat varis aspectes de l'engreix dels porcs i s'han instal·lat un conjunt de sensors (càmeres) tant a la granja com a l'escorxador per obtenir dades que ajuden a avaluar el procés, fer seguiment i millores en aquest.

³ [From farm to Fork Strategy](#). For a fair, healthy and environmentally-friendly food System. European Union. 2020.

Figura 1. Esquema del cycle de vida del porc. Encerclat en blau, l'espai de prototipatge d'aquest projecte



Un factor molt important del procés d'engreix del porc és la seva alimentació. La manera clàssica per alimentar els porcs és amb pinso, però hi ha unes tendències que experimenten amb altres tipus d'alimentació com podria ser l'alimentació líquida, que és més fàcil de digerir pels animals. Fins i tot, hi ha iniciatives que contemplan la introducció d'aquest tipus d'alimentació a les granges. A la Figura 1 s'aprecia un esquema del cycle de vida del porc.

A les granges es porta a terme el procés de l'engreix del porc, el qual entra a la granja d'engreix quan té 10-11 setmanes i pesa uns 20-25 kg. Quan el porc pesa entre 110 i 120 kg, aproximadament al cap de 4 mesos, és traslladat a l'escorxador per al seu sacrifici. A les granges és molt important poder fer un seguiment del pes de cada animal ja que el pes del porc és el paràmetre per saber quan es pot portar el porc cap a l'escorxador. El pes del porc és important perquè a l'escorxador prefereixen treballar amb animals que tenen un pes homogeni per evitar ajustar les seves màquines. Tenint en compte aquesta informació aleshores seria idoni poder disposar d'un sistema per monitoritzar el pes dels porcs a l'engreix. La manera més fàcil evidentment seria utilitzar una bàscula que mesuri el pes del porc en temps real quan l'animal passa per la bàscula. Tanmateix, la bàscula per poder pesar correctament, l'animal ha d'estar un temps sense moure per poder obtenir un pes estable (la bàscula s'ha d'estabilitzar).

Finalment, a part de fer un seguiment del pes dels porcs a la quadra, cada vegada més hi ha un interès en monitoritzar els moviments i els comportaments dels animals. Aquesta idea té relació amb el concepte de benestar animal a les granges: els porcs han d'estar en les millors condicions durant el seu procés d'engreix, amb indicadors objectivament verificables.

Proposta de prototipatge

La solució proposada inclou els següents mòduls tecnològics:

1. Sistema de captació d'imatge amb càmeres instal·lades a una quadra experimental.
2. Sistema de segmentació dels porcs.
3. Sistema d'identificació individualitzada de cada porc.
4. Sistema de seguiment dels moviments del porc.

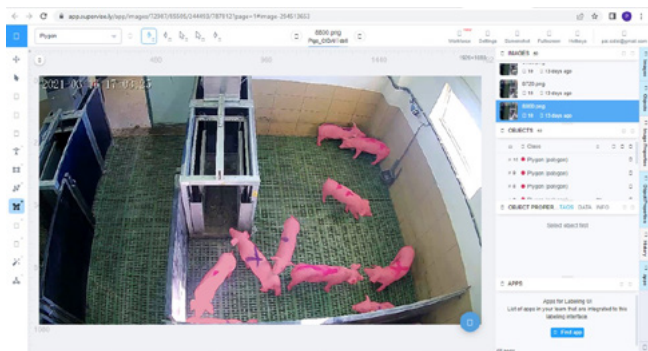
Una manera de monitoritzar els animals és mitjançant una càmera que s'instal·la a la quadra, tal i com es mostra a la Figura 2 es mostra una imatge d'una càmera que es va ubicar a les instal·lacions del CEP. En el context del projecte es va decidir aprofitar l'ús d'aquesta càmera per fer un seguiment dels porcs en el temps real i poder estudiar dels seus moviments.

Figura 2. Granja d'engreix de porcs del CEP. Quadra experimental amb els porcs marcats per la seva identificació. A dalt a l'esquerra és la bàscula. A la dreta, el recipient d'aigua.



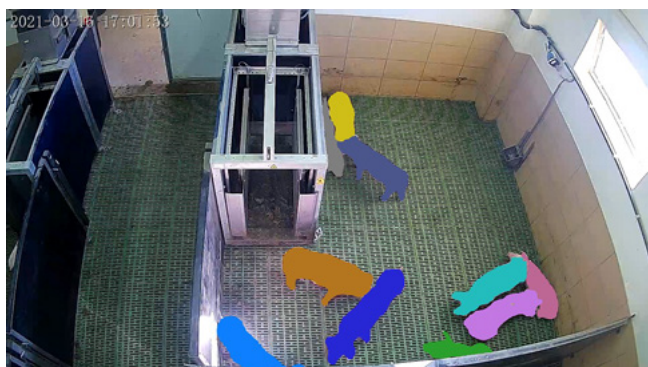
Per aquest motiu, es va decidir implementar un algorisme de detecció dels porcs a la imatge. Es va arribar a la conclusió que era important realitzar aquesta detecció a nivell de píxel per poder seguir els animals fins i tot quan s'apropen o quan es toquen. Aleshores es va optar per una xarxa neuronal de segmentació d'instància (*Instance Segmentation*), la particularment la Mask R-CNN. Per entrenar aquesta xarxa es va fer un conjunt de referència d'unes 100 imatges que es van anotar manualment (veure Figura 3).

Figura 3. Identificació dels porcs. Anotació manual per la creació del conjunt de referència per l'entrenament de la xarxa neuronal.



Up cop entrenada la xarxa, es va veure que la detecció dels animals funciona de manera molt eficient, tal com es pot veure a la Figura 4, on s'indiquen les deteccions dels porquets amb colors diferents.

Figura 4. Quadra de porcs amb identificació d'animals amb R-CNN



Tanmateix, per fer el seguiment dels porcs no és suficient poder detectar els porcs a la imatge amb precisió, sinó que s'ha de poder relacionar els porcs d'una imatge del vídeo amb els porcs d'una imatge següent. Aquest procés es coneix com a re-identificació, i es pot fer de diverses maneres:

1. Utilitzant un model per predir els moviments dels animals (per exemple amb un filtre de partícules o Kalman).
2. Utilitzant un algorisme per intentar generar un identificador únic (una mena d'empremta digital) per a cada porc. A continuació, amb aquests identificadors es poden posar en correspondència les deteccions que s'obtenen de dues imatges consecutives.

A la implementació final es fa servir una combinació de diferents tècniques, ja que hi ha prou casuístiques s'han de tenir en compte. Per exemple, donada la col·locació de les càmeres, a vegades els porcs desapareixen per un curt termini de temps, però després desapareix per oclusions d'altres porcs, i s'ha de recuperar el seguiment (*tracking*).

Un cop implementat l'algorisme de seguiment, es va aplicar l'analogia que la vida del porc en una quadra es podria estudiar com si fos un partit de futbol i que amb el seguiment dels animals es poden generar unes estadístiques com ara les que es generen en un partit (veure Figura 5).

Figura 5. Representació del moviment d'individus en un camp de futbol

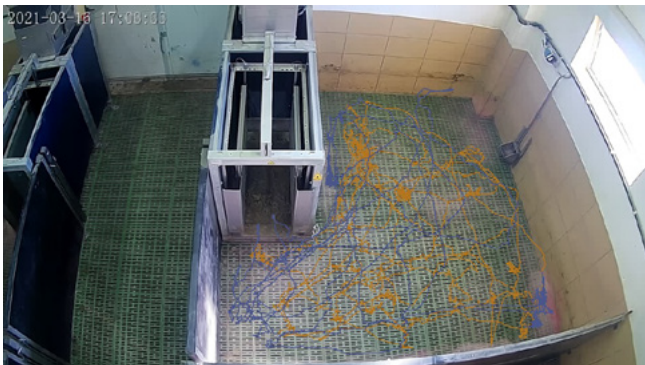


Hi ha un conjunt de paràmetres que val la pena estimar per tenir una primera idea del comportament dels animals a la quadra:

1. Camins dels porcs (distància, velocitat màxima).
2. Temps que passen a la zona de la bàscula.
3. Temps que passen aturats (potencialment dormint).
4. Temps que passen a la zona de l'aigua.
5. Mapes de calor per àrees de preferència.

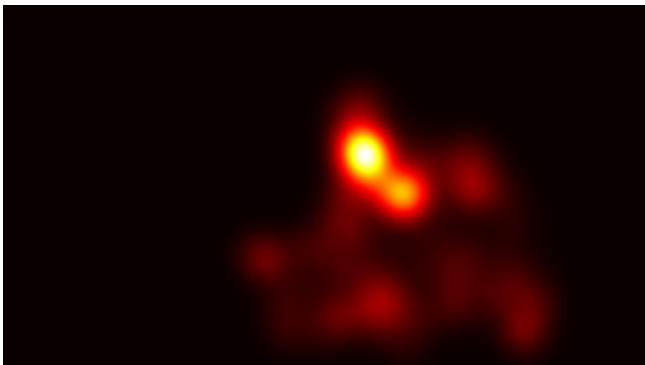
Per il·lustrar aquests càlculs, s'han generat imatges de referència on es resumeixen els camins que han realitzat els porcs durant el dia. A la Figura 6 es mostra els camins de dos porcs diferents.

Figura 6. Recorregut dels porcs dins de la quadra. En blau i carbassa dues la trajectòria que segueix el centre de massa de dos porcs diferents al llarg del dia. La imatge de la quadra buida proporciona la referència del moviment dels porcs.



També s'han generat uns mapes de calor per visualitzar on passa els porcs el seu temps. La Figura 7 mostra el resultat per a un porc concret. Aquest aproximació requereix contextualització amb respecte a la resta dels porcs, la qual cosa no ha estat objecte d'estudi en aquest projecte.

Figura 7. Mapa de calor d'un porc dins de la quadra



4.3. PoC d'espai de dades del sector agroalimentari

Actualment, ens trobem en les primeres etapes d'una nova revolució digital, comunament anomenada quarta revolució industrial, basada gran part en els recents avenços en el camp de la Intel·ligència Artificial (IA). Els models d'IA tenen la possibilitat de revolucionar la indústria i la societat en la que vivim gràcies a la seva capacitat per automatitzar tasques complexes i l'anàlisi de grans quantitats de dades.

Aquests models, però, necessiten grans bancs de dades per funcionar correctament i sovint trobem que a major quantitat de dades (i de més qualitat) obtenim models més precisos i capaços. La seva obtenció i processament és una tasca complexa i costosa però és el recurs que impulsa aquesta nova era digital. És per aquest motiu que actualment no és estrany escoltar l'expressió "les dades són el nou petroli".

Tot i això, avui dia és comú trobar "sitges de dades" o conjunts de dades guardats en sistemes i infraestructures totalment aïllades, impossibilitant que altres entitats en treguin i resultant en un desaprofitament del seu autèntic valor.

Els espais de dades (molt sovint anomenats amb la terminologia anglesa data spaces) són ecosistemes que busquen solucionar el problema degut al confinament de les dades. Permeten oferir i utilitzar recursos de manera segura entre diferents entitats per intercanviar i usar dades. No són només un emmagatzematge de dades sinó que incorporen funcionalitats per tal que dades de diferent tipologia i origen puguin ser compartides i utilitzades de forma segura. Creen un entorn ben organitzat, amb regles i directrius establertes i una seguretat sòlida.

D'una banda, els propietaris de conjunts de dades poden obtenir un benefici al compartir-les amb tercers, mentre mantenen el control del seu ús i privacitat. Per altre banda els espais de dades serveixen com a un catàleg on entitats i individus poden buscar i utilitzar una gran varietat de dades i algorismes que són executats dins el propi ecosistema.

A mode d'exemple, utilitzant el cas d'ús descrit anteriorment, considerem una granja porcina que disposa d'una càmera per a monitoritzar l'activitat dels animals. Per si soles, les imatges preses per aquesta càmera al llarg del temps no tenen gaire valor per

als propietaris, però no així per a altres individus que les poden trobar de gran utilitat. Per aquest motiu, un conjunt d'aquestes imatges és ofert en un espai de dades a canvi d'una certa retribució.

Les imatges estan ara a disposició de qualsevol organització o persona que les desitgi utilitzar, de manera que una segona entitat que es dedica a analitzar el comportament de bestiar les utilitza i les anota manualment (indica la posició exacta que ocupa cada animal dins de les imatges). Aquestes anotacions són dades noves que també tenen un gran valor i de fet, augmenten el valor original de les pròpies imatges ja que ara (junt amb les anotacions) poden ser utilitzades en a un rang més ampli d'aplicacions. Aquestes anotacions també són ofertes a l'espai de dades.

Per últim, una tercera entitat que disposa d'un model d'intel·ligència artificial enfocat a la detecció i seguiments de porcs, pot adquirir aquestes dades i utilitzar-les per a crear o millorar el seu model. Opcionalment també podrà oferir aquest model en l'espai de dades, de manera que altres parts interessades, incloent als propietaris inicials de la granja el puguin utilitzar en el seu benefici.

Així doncs, veiem com sense gran esforç podem rendibilitzar i donar valor a les dades que posseïm, mentre que contribuïm al desenvolupament de solucions més avançades beneficioses per a totes les parts.

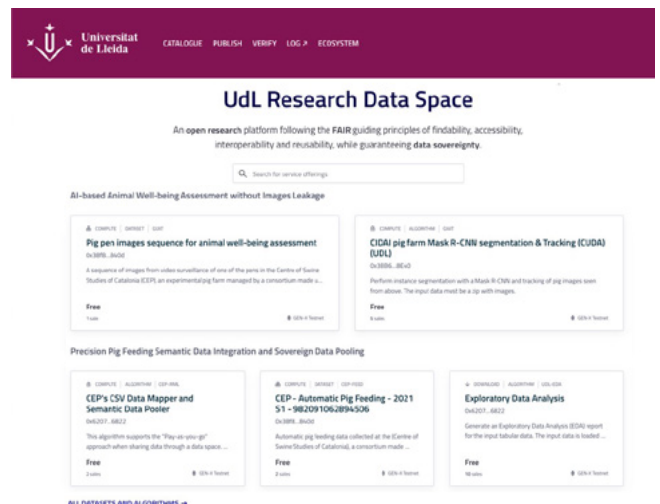
Pontus-x (Gaia-X, s.f.) és un bon exemple d'un espai de dades que ens permet oferir i utilitzar recursos i serveis. És un ecosistema descentralitzat, basat en Gaia-X, una recent iniciativa europea que te com a objectiu crear un ecosistema de dades obert, transparent i segur a Europa. Aquest espai utilitza el protocol Ocean per gestionar interaccions amb dades i serveis. Es tracta d'una tecnologia transparent i de codi obert que, gràcies a la seva descentralització, retorna el control de les dades i serveis als seus propietaris.

Aquest entorn utilitza la tecnologia Blockchain per a representar els recursos i gestionar les interaccions amb els usuaris. Quan publiquem un conjunt de dades o servei en aquest ecosistema, estem creant automàticament un NFT (criptovalor no fungible) que representa de forma única el nostre recurs i que inclou tot a la informació necessària per a utilitzar-lo de forma segura. Aquest NFT "viu" en la Blockchain, de manera que el seu accés és descentralitzat i és el propietari qui posseeix el control total sobre ell. Per altre banda, les transaccions per a adquirir i utilitzar les dades i serveis dins aquest ecosistema es realitzen per mitjà de monedes virtuals (tokens), què permeten una major transparència, seguretat i integració.

En el context d'aquest projecte, l'espai de dades inclou les següents funcionalitats:

1. Quan accedim al portal d'un espai de dades basat en Pontus-X, com per exemple el de la Universitat de Lleida mostrat en la Figura 8 ens trobem amb un catàleg de dades i algorismes a la nostre disposició, cada un amb un preu associat o ofert de manera gratuïta. En el cas dels conjunts de dades trobem dos modes d'accés que el propietari ha de triar. El primer mode "descàrrega", permet descarregar les dades un cop s'han comprat. El segon, anomenat "còmput", no permet accedir a les dades directament sinó que són utilitzades per mitja d'un algoritme dins del mateix ecosistema. Així, les dades mai surten d'aquest entorn, garantint així la seva privacitat i control. El propietari també pot decidir quins algorismes poden ser executats amb les seves dades.

Figura 8. Portal de l'espai de dades de la UdL

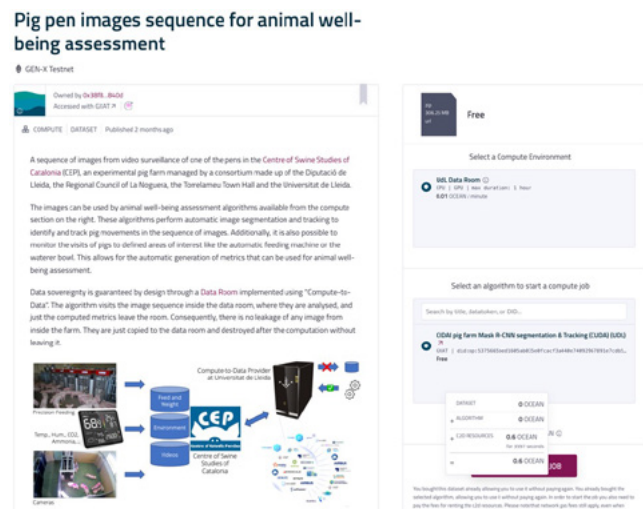


2. Quan volem utilitzar un algoritme amb unes dades en concret haurem d'abonar el preu associat a les dades, al algoritme i també al cost de la computació, en cada cas al propietari del recurs corresponent. El propietari de les dades, de l'algoritme o de la infraestructura de computació utilitzada. Aquest s'executarà dins la pròpia plataforma i podrem veure el seu estat i resultats en la secció de "tasques de còmput" en el nostre perfil (Figura 9).
3. Per altra banda, per a publicar dades o algorismes únicament haurem d'omplir la informació necessària. Haurem de seleccionar el tipus de recurs (dades o algoritme), informació bàsica com

el títol i la descripció, el preu i la ubicació de les dades o algorisme entre d'altres. Cal notar que Pontus-X no és un magatzem de dades, sinó que únicament gestiona el seu accés. Així doncs, nosaltres serem els responsables de pujar les dades a un servei web i a l'hora de publicar-les caldrà indicar com accedir a elles. Podrem utilitzar mètodes d'accés com ara una única URL que permeti descarregar les dades o protocols como ara IPFS que permeten un emmagatzematge de dades en una xarxa oberta i descentralitzada.

- En el cas d'oferir les dades per a descarregar, quan un usuari les adquireix, la pròpia plataforma generarà un enllaç segur que en cap cas exposarà la direcció original de les dades, mantenint així el seu control. A més a més les dades tindran un temps límit de validesa des de la seva compra, de manera que passat un temps establert el comprador no podrà accedir a les dades sense adquirir-les de nou.

Figura 9. Detall de l'espai de dades



- Per últim, els algorismes són executats utilitzant l'entorn de virtualització Docker. Podrem utilitzar qualsevol imatge Docker publicada en un repositori d'imatges privat o públic, com el Hub de Docker. A l'hora de publicar l'algorisme també haurem d'indicar la comanda i el fitxer de codi que s'haurà d'executar. l'execució tindrà lloc en les màquines proporcionades pel propi entorn tot hi que és possible utilitzar un altre proveïdor amb altres especificacions hardware.

4.4. Anàlisi de governança, sostenibilitat i explotació

Introducció

Una de les grans problemàtiques que tenen els pilots és que no aconsegueixen superar la fase de pilot, ja que no es consoliden dintre d'un model de negoci sostenible. Això té un efecte molt negatiu en els actors de la ciutadania i productors, ja que no veuen l'efecte de la seva participació en els processos d'innovació. El resultat és una pèrdua de legitimitat per part de les institucions que promouen accions d'innovació en el territori i una manca d'interès en els processos d'innovació oberta per part de la ciutadania.

De manera transversal a l'execució del projecte, el BSC ha desenvolupat una tasca específica per tal d'impulsar la transferència tecnològica de les diferents eines i solucions al mercat real, valorant-ne l'explotabilitat i l'impacte econòmic. Es proposen els següents objectius específics:

- Un anàlisi de la propietat intel·lectual dels projectes i la seva protecció
- El desenvolupament d'un pla de negoci
- L'avaluació del nivell de preparació de la tecnologia (TRL)
- L'aproximació de les propostes al mercat i a les institucions que formen part de l'ecosistema català de Deep Tech.

Per tal de desenvolupar aquests objectius es faran servir les dades obtingudes al desenvolupament tècnic dels treballs realitzats als diferents PAIs de CIDAI, així com l'experiència en gestió de la innovació i dels drets de propietat intel·lectual de l'equip de Transferència Tecnològica del Barcelona Supercomputing Center.

La metodologia utilitzada per elaborar aquest informe es pot trobar a l'ANNEX (A3).

Algoritme de traçabilitat i indicadors de benestar

Anàlisi de la protecció de la propietat intel·lectual

Drets de Propietat Intel·lectual (PI) preexistents

Universitat de Lleida			
COMPONENT (nom de la tecnologia preexistent/ coneixement/ software)	TIPUS (és una metodologia, un <i>software</i> , un <i>hardware</i> ?)	TITULARITAT (qui té drets de Propietat Intel·lectual sobre aquesta tecnologia?)	CONDICIONS D'ÚS I D'EXPLOTACIÓ (quines condicions s'apliquen a aquesta tecnologia? Se li ha assignat una llicència?)
UdL Research Data Space	Software	Universitat de Lleida	Apache License 2.0
Exploratory Data Analysis Data Space Algorithm	Software	Universitat de Lleida	Apache License 2.0
CEP's Mapper for Daily Pig Weight CSV Data Space Algorithm	Software	Universitat de Lleida	Apache License 2.0
CEP's Mapper for Automatic Pig Feeding CSV Data Space Algorithm	Software	Universitat de Lleida	Apache License 2.0

Centre de Visió per Computador			
COMPONENT (nom de la tecnologia preexistent/ coneixement/ software)	TIPUS (és una metodologia, un <i>software</i> , un <i>hardware</i> ?)	TITULARITAT (qui té drets de Propietat Intel·lectual sobre aquesta tecnologia?)	CONDICIONS D'ÚS I D'EXPLOTACIÓ (quines condicions s'apliquen a aquesta tecnologia? Se li ha assignat una llicència?)
Algorisme per fer seguiment d'un conjunt de vaques en un recinte basat en intel·ligència artificial.	Software	Compartida entre el CVC i Grupo Alcañiz	Conveni entre el CVC i l'empresa per l'explotació comercial per part de l'empresa

La següent taula mostra la PI preexistent utilitzada pels membres per implementar tasques amb propietat de tercers (coneixements de tercers/programari/ etc., incloent-hi codi obert).

COMPONENT (nom de la tecnologia preexistent/ coneixement/ software)	TIPUS (és una metodologia, un <i>software</i> , un <i>hardware</i> ?)	TITULARITAT (qui té drets de Propietat Intel·lectual sobre aquesta tecnologia?)	CONDICIONS D'ÚS I D'EXPLOTACIÓ (quines condicions s'apliquen a aquesta tecnologia? Se li ha assignat una llicència?)
Minimal Viable Gaia-X Portal	Software	deltaDAO	Apache License 2.0
Ocean Marketplace	Software	Ocean Protocol Foundation Ltd.	Apache License 2.0
ydata-profiling	Software	YData Labs Inc.	MIT License
rmlmapper-java	Software	RDF Mapping Language (RML)	MIT License
Mask R-CNN (Detectron2)	Software	Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick	Apache License 2.0
Deep-Person-REID	Software	Zhou, Kaiyang and Xiang, Tao	MIT License

PI resultant (IP foreground)

COMPONENT (nom de la tecnologia preexistent/ coneixement/ software)	TIPUS (és una metodologia, un software, un hardware?)	TITULARITAT (qui té drets de Propietat Intel·lectual sobre aquesta tecnologia?)	CONDICIONS D'ÚS I D'EXPLOTACIÓ (quines condicions s'apliquen a aquesta tecnologia? Se li ha assignat una llicència?)
Metodologia pel desplegament d'espais de dades sostenibles i amb sobirania de dades al sector agroalimentari	Metodologia	Compartida amb resta del participants del PAI	Apache License 2.0
Algorisme per fer seguiment d'un conjunt de porcs en un corral basat en intel·ligència artificial.	Software	CVC, CEP/UdL	Apache License 2.0

Context: el mercat

Característiques del mercat ramader

El mercat ramader es troba en evolució. Actualment conviuen explotacions ramaderes molt diverses, en quan a mida i nivell de tecnificació, des de granges molt tecnificades i que treballen amb eines informàtiques de control, a granges molt velles (en quant a instal·lacions, sense cobertura mòbil, etc.). Actualment utilitzen alguns programes de gestió com Porc-tec (de Agrotec soft) per paràmetres de productivitat o Bigdata per gestió del pinso, i participen puntualment en projectes d'innovació.

Les grans empreses integradores, mitjançant la compra de granges, estan creixent i augmentant la seva importància i poder, i generant una guerra de preus al mercat. Aquestes ja estan introduint eines per millorar la productivitat i controlar els costos (ex., alimentació de precisió per aconseguir estalvis en pinso i medicaments, control de dejeccions per aconseguir estalvis en pagament d'impostos, etc.).

A nivell d'explotacions individuals, s'observa que els ramaders estan menys predisposats a la innovació que altres agents del sector agro-alimentari (ex., empreses industrials més acostumades a controlar costos, productivitat, etc.).

Així doncs, la tendència serà a una tecnificació creixent de les granges, tot i que aquesta serà gradual, degut a que al sector ramader intervenen molts paràmetres que no sempre són fàcils de controlar.

Importància i paper de la innovació

Cada vegada és i serà més important i necessari monitoritzar els animals i obtenir paràmetres de mesura del seu comportament i benestar, així com de costos de producció de les granges / explotacions ramaderes. Principalment en termes de rendiment / productivitat i optimització (consums d'aigua, pinsos, etc.).

Aquest nivell creixent de control ve imposat per les administracions, les quals demanen més controls a nivell d'emissió de gasos i residus (CO2, amoníac, etc.), i també l'obtenció de registres. Això farà necessari adquirir i implementar eines per ajudar en aquestes tasques.

Per altra banda, el consumidor valora cada vegada més la carn de qualitat i conèixer l'origen dels animals, i està disposat a pagar més per això. Per la qual cosa, guanya importància conèixer la traçabilitat dels animals, informació que les noves eines informàtiques poden facilitar.

La innovació, la utilització de softwares i la monitorització marquen el camí del futur. Pel ramader és important que la innovació sigui clara a nivell cost-benefici. És crucial, doncs, afegir valor a les dades. No obstant, la realitat del dia a dia dels ramaders demostra que acostumen a tenir poc temps / recursos per analitzar les dades que actualment ja estan generant. Preval una mentalitat pràctica. Necessiten dades senzilles i útils, fàcils d'interpretar, i que serveixin per prendre decisions encaminades a incrementar la productivitat.

Programari

Actitud i valoració general

Davant de noves solucions tecnològiques, avui en dia detecten actituds receptives per part dels usuaris finals (ramaders / cooperatives). El plantejament d'un monitoratge individualitzat dels animals s'identifica com una tecnologia interessant, tot i que de moment no saben veure la utilitat final de les mètriques que s'obtenen actualment. No obstant, creuen que hi ha altres dades que són importants i que possiblement aquesta eina els hi podria facilitar, sempre pensant en millorar la productivitat. Tot i que tothom es mostra d'acord en millorar el benestar animal, l'objectiu principal és incrementar la productivitat, i no tenen com a prioritat la seva aplicació per certificacions. Alguns es mostren escèptics sobre el plantejament de l'eina, veuen difícil el seguiment d'un nombre elevat d'animals amb càmeres fixes, veuen una inversió elevada, per unes dades que no els hi semblen d'interès, a més de tenir altres prioritats on invertir (refrigeració de naus, làser per sitges, etc.).

Els consultors fan una valoració molt més positiva de l'eina i consideren que té molt potencial. Són més conscients de la dificultat tecnològica que suposa aconseguir monitoritzar el seguiment dels animals al corral i tenir-los identificats. Consideren que aquest tema està ben resolt. Així mateix, veuen de forma més clara tots els possibles paràmetres o indicadors clau (KPI) de comportament i benestar animal de segon nivell que es poden obtenir, a partir de les mesures obtingudes fins ara i que seran molt interessants i d'utilitat pels ramaders.

El conjunt dels agents sectorials entenen que els desenvolupaments actuals són el primer nivell o estadi -necessari- de tot un procés, essent els passos següents molt importants de cara a demostrar la utilitat de l'eina.

Valoració de la informació actual

A continuació s'indica la valoració per cadascuna de les mètriques calculades actualment:

- **Recorregut dels porcs** (distància). Cap interès a priori.
- **Temps que passen a la zona de la bàscula** (lloc on mengen, i on poden pesar els animals). No els hi serveix. Els interessen dades com: si mengen o deixen de menjar, quantitat del pinso que mengen, pes (és la informació més important i desitjable).

- **Temps que passen a la zona de l'aigua** (per tenir una idea de la quantitat d'aigua que beuen). No els interessa aquesta dada, els animals podrien estar a la zona de l'aigua per refrescar-se. Interessaria el seu consum d'aigua.
- **Temps que passen "dormint"** (o sense fer res, temps d'inactivitat) i mapes de calor (per saber on el porc passa el temps). No interessen per se. Interessarien en la mesura que aquestes variables de comportaments anòmals (passivitat o alta activitat, comportaments agressius, mossegades de cues, etc.) es poguessin relacionar amb malalties o manca de benestar, i permetin prendre decisions sobre actuacions (separar els animals, donar medicació, canviar les condicions de temperatura o aire de les naus per reduir l'estrès, etc.).

Als usuaris finals els hi costa veure l'interès o utilitat de les mesures actuals que dona l'eina. Hi veuen potencial, si pot donar altres dades més interessants i/o si aconsegueix correlacionar les variables actuals (ex., inactivitat) amb paràmetres més concrets i útils per prendre decisions encaminades a l'optimització / millor organització / productivitat.

Els consultors fan una valoració molt més positiva, i pensen que a partir de la informació actual ja es podrien derivar KPIs interessants per identificar problemàtiques (de climatització, conflictes, malalties, etc.), prendre decisions o obtenir la traçabilitat en tot el cicle de vida.

Expectatives d'informació. Mesures més interessants / desitjables i utilitats

Els indicadors que s'han identificat com a rellevants són:

- **Pes.** El més interessant i motivador per utilitzar l'eina. Especialment per granges d'engreix i de transició/deslletament. Conèixer el pes dels animals en cada moment i tenir dades històriques (corbes de creixement) té moltes utilitats.
- **Veure si hi ha desviacions respecte l'estàndard,** i si no arriben al pes, identificar les causes i prendre mesures (manca o problemes d'aigua, de pinso, malalties, etc.).
- **Decidir quan cobrir els animals** (ex., a partir de 150 kg).
- **Prendre decisions de canvis de tipus de pinso,** buscant el més adient segons el seu pes

(pinsos: starter, de creixement, d'acabament), també es podrien posar a corrals diferents segons pes i adaptar el pinso a cada corral.

- Previsió i selecció/marcatge dels animals per anar a l'escorxador (actualment es fa visualment).
- Quantitat d'aigua que beuen i quantitat de pinso que mengen. Gestió de consums i detecció malalties.
- Si beuen aigua o no / Si mengen o no. Detecció d'estats febrils, malalties.
- Comportaments passius (si canvia la jerarquia del corral, si es deixen pegar i no reaccionen, si no es mouen, etc.). Detecció d'estats febrils, malalties.
- Comportaments hiperactius (nerviosisme, estrès). Gestió condicions de les naus (temperatura, corrents d'aire, etc.) i gestió malalties.
- Comportaments agressius (mosseguen cues). Gestió de medicacions (més probabilitat d'infeccions).
- Temperatura individualitzada. Detecció d'estats febrils i gestió de malalties.

En aquesta línia, estan interessats en altres indicadors relacionades amb comportaments anòmals / malalties encara que tampoc saben si es poden obtenir o no mitjançant l'eina, ni com es podrien tractar. Mitjançant imatges es podrien identificar coixeses, diarrees, colors dels animals (ex., si és més groc, pot tindre un problema al fetge), greix al coll (correlació amb el pes), condició corporal abans i després de parir (per granges de mares), estats anímics (agressivitat, baralles, canvis de jerarquies al corral, mossegades d'orelles o cues, passivitat), animals morts. Si hi hagués la possibilitat d'incorporar so, es podria identificar la tos / respiració (per valorar si cal vacunar, donar medicació, etc.), crits o baralles (per detectar estrès o malalties), o la comunicació mare – cadells.

Per altra banda, també interessaven mesures i registres que demana / demanarà l'Administració. En principi són dades generals dels corrals / naus i no dades individualitzades. Es demanarà al 2025:

- Nivells de CO₂.
- Nivells d'amoníac.
- Registres de temperatura ambient.
- Nivell de llum per corral (mínim 40 luxes a nivell del cap de l'animal).

També preveuen que en el futur l'Administració pugui demanar més registres de paràmetres de benestar animal. Fins ara es reporten a l'Administració dades de medicaments i vacunes, però no en saben la traçabilitat. També es reporten als portals de l'Administració les dejeccions, els nivells de CO₂ i NH₃, però no tenen forma de saber si són reals o no. D'aquesta manera es podria controlar.

Per a les empreses i cooperatives, l'objectiu final de les dades sempre ha de ser millorar l'índex de conversió de les explotacions i, per tant, la seva productivitat. La millora del benestar animal interessa en la mesura que ajudi al objectiu anterior (anticipar malalties / optimitzar recursos). Al final el més important és poder relacionar qualsevol paràmetre amb l'índex de conversió, ja que l'objectiu és portar com més aviat millor l'animal a l'escorxador amb el menor consum d'aigua i de pinso. Si el benestar o malestar animal es tradueix en una millora de l'índex de conversió, llavors és molt important, i val la pena aquesta informació. Cal veure com correlacionar-la.

En aquests moments, consideren que l'eina no els serà útil per certificacions, benestar, ni d'altres (les dades les obtenen els inspectors i les granges no tenen cap necessitat d'informació addicional). Tampoc saben si el seguiment individualitzat els hi pot servir (l'Administració demana dades i registres per nau/corral). És possible que per futures demandes de l'Administració, si l'eina aconsegueix mesurar paràmetres que es relacionin de forma directa amb el benestar animal, els pugui ser d'utilitat.

Els consultors pensen en un ventall més ampli d'utilitats, incloent benestar animal, sanitaris, detecció de problemes a granges, certificacions, etc. Assenyalen la importància de recollir el màxim de dades possible per facilitar que a la fase d'anàlisi estadístic multivariant es puguin trobar correlacions significatives (fins i tot amb variables que a priori no semblaven importants).

Els KPIs de comportament s'han de relacionar amb variables tècniques de segon nivell, quantes més variables millor. Els veterinaris són els que han de llistar totes les possibilitats de variables directes i indirectes, perquè això és el que afegeix valor a les dades, i permet fer un anàlisi molt complet i mirar de trobar totes les correlacions possibles.

Expectatives sobre dades / interfície / funcionament

En relació a les dades, es necessiten dades fiables, molt simples i fàcils d'interpretar; poques dades, rellevants, pràctiques, que permetin la presa de decisions, KPIs i mesures que es puguin escollir, dintre d'un catàleg. Consideren molt interessant la possibilitat de personalitzar les KPIs / informació a contractar i establir el preu segons les dades contractades.

En relació a la interfície / outputs / tractament de dades, es suggereix que sigui molt visual, com ara informació evolutiva per poder tenir un històric, comparar entre diferents paràmetres / granges, etc. Un cop obtingudes les dades en brut, es podria treballar amb elles, interrelacionar-les amb variables de negoci (ex., amb el programa power value), descarregar els KPIs / dades per fer taules de controls, gràfics, i seguiment de producció. També, a la mateixa aplicació, poder controlar algunes variables (p.ex. baixes de corral), i incloure un espai per anotar dades per part de les granges.

Valoració del pipeline / procés

Instal·lació de sensors / càmeres

Els agents sectorials més oberts i positius, no veuen massa problema en instal·lar aquests dispositius, encara que mencionen alguns aspectes a tenir en compte (poden ser considerats com a barreres):

- Si les naus/corrals són d'uralita, per normativa no es poden foradar.
- Les càmeres han d'estar protegides o ser resistents, a les neteges/desinfeccions que es fan amb pressió i als gasos de les naus. Això, implica que a part d'una inversió inicial, possiblement hi hauran despeses de manteniment dels equips.
- Hi ha granges ubicades a llocs remots, sense cobertura d'internet.
- Els més escèptics sobre l'eina veuen molt costosa la inversió (són naus grans), poc pràctic comptar amb càmeres fixes (seria millor una aplicació que donés informació a partir de fotografies o sons en el mòbil).

En ambdós casos, l'increment de productivitat/optimització per les decisions preses a partir de les dades ha de compensar el cost / inversió a realitzar. El benefici ha de ser clar, per tal de que els ramaders/granges o les Cooperatives estiguin disposats a invertir.

Pels usuaris finals la òptima relació inversió – retorn és fonamental per a prendre la decisió de participar en aquesta iniciativa. Els consultors consideren que aquesta possible barrera (cost de la inversió) es pot superar buscant solucions imaginatives, com ara instal·lar càmeres només a alguns corrals i no a tota la nau, oferir fórmules de servitització, que facilitin al ramader la presa de decisions (ex., lloguer de càmeres pel primer engreix i opció de compra una vegada demostrada la utilitat de la solució).

Obtenció de dades / timings

En general, que es trigui 1 mes per ajustar el model i començar a obtenir dades no suposa un problema pels ramaders, si després les dades que obtenen són realment fiables i útils.

No obstant, es genera el dubte de si el model necessitarà 1 mes per obtenir dades per cada engreixada, o només la primera vegada. En cas de que fos per cada engreixada, l'eina perd interès, perquè és molt important seguir els paràmetres a l'inici de l'engreix.

Els que tenen més coneixements tècnics són més conscients de que quan més gran sigui el volum de dades que s'incorporin al software, el model s'ajustarà de forma més ràpida. Aquesta possible barrera es superarà si s'aconsegueix que més usuaris es sumin al projecte.



Espai de dades

Valoració i expectatives dels usuaris finals

Barreres inicials

A priori, els ramaders i Cooperatives mostren reticència a compartir les dades de les seves explotacions. Les compartirien si les dades són anònimes i al servidor es presenten de forma agregada (no volen que cap competidor tingui accés a les seves dades de productivitat). La política de protecció de dades és molt transparent amb l'ús que es fa de les dades.

Nivell d'interès / utilitat

Els que han vist útil la informació que podria donar l'eina, consideren també d'utilitat poder accedir a un servidor/núvol amb dades agregades de les explotacions que utilitzen el software, que permetin fer anàlisi comparatius dels resultats propis versus els de granges de la zona / província, els de diferents tipus d'explotacions, els de granges amb tipus d'alimentació similar, etc. (dades com costos, nombre de garrins per truja, índex de conversió, KPIs definits, etc.).

Així mateix, consideren que l'espai de dades també podria servir per prevenir malalties. En aquest cas, però, identifiquen que les granges són reticents a compartir aquest tipus de dades (ex., virus o problemes sanitaris), però són conscients de que compartir dades pot fer l'algoritme més precís / estable.

Les persones que no creuen els beneficis de l'eina tampoc valoren la possibilitat d'accedir a imatges o dades de l'espai de dades.

Funcionament

Es valora positivament el funcionament amb una subscripció a un preu raonable.

Valoració i expectatives dels consultors

Actitud general

Als *stakeholders* els agrada el plantejament i consideren que té molt potencial. Incorpora criteris democràtics, basant-se en l'economia de tokens.

Sobirania i compartició de dades

És un tema cabdal a definir, doncs per la seva experiència, la compartició de dades és "LA BARRERA" fonamental. Tot i que no és fàcil de resoldre, s'ha de crear un entorn fiable per no perdre el control de les dades. S'han d'oferir sempre dades agregades.

L'aprenentatge federat (*Federated Learning*), el qual permet utilitzar en algoritmes dades de tercers sense necessitat de tenir-les físicament els hi sembla una bona solució, així les dades no s'exporten/ descarreguen/copien. Encara que complicat, comenten la possibilitat de donar una garantia o segell per part d'una empresa certificada, per superar la desconfiança dels ramaders. Així doncs s'ha de definir una política de governança de les dades (consentiments, vots, delegació de gestió de dades, negociació, etc.), seguint la normativa europea. Tot i això, pensen que els usuaris finals no compartiran dades internes (ex., de productivitat), però si altres dades que puguin beneficiar al sector (ex., de consums d'aigua i de com reduir-la, de comportaments anòmals, etc.).

Ecosistema

Són molt conscients de la importància de l'ecosistema per donar valor afegit a la solució (i també per evitar que sigui fàcil de copiar).

Podria funcionar com un mercat, on es trobarien productors de dades, analistes i interpretadors de dades, desenvolupadors d'AI o softwares, inversors, etc. Tothom es beneficiaria, els que ofereixen el servei i els usuaris finals.

Sistema d'incentius

S'identifica la importància de generar models d'incentius que siguin motivadors i no generin desconfiança. Si paguen per les dades de forma directa, generarien més desconfiança.

Si tenen el suport de l'Administració pública, podrien ser més col·laboradors, pensant que és pel benefici de tot el sector (però després veurien que el servei es comercialitza de forma privada). També l'Administració podria ajudar a desenvolupar els models inicials.

Una altra opció seria optar per un model mixt, donar dades per ajudar a avançar al sector, i com a retorn el ramader obté un descompte en el servei. D'esta forma es podria fomentar la seva participació.

Usuaris potencials

Ramaders

El ramader pot estar interessat amb la solució per reduir el temps de gestió del dia a dia, anticipar-se i prendre decisions, guanyar tranquil·litat, millorar la productivitat final i reduir costos (tot i que moltes despeses són de la integradora).

Aquests volen una eina simple i pràctica, i que sigui rendible. Demanen poques dades, fiables i útils, que es puguin traduir en decisions. Algunes alarmes poden ser suficients (ex., alarma que avisi 3 dies abans d'una possible malaltia d'un animal; una alarma si un porc té un bon nivell de creixement i pot ser de més qualitat, etc.). No tenen temps per analitzar moltes dades i tenen poc personal / poc format per fer-ho. Això fa que sigui important vendre la solució quan la tecnologia ja estigui madura / sigui robusta / aportï valor, encara que sigui petit. Estan cansats de tecnologies que no els han servit perquè no estaven madures (ex. drons).

Granges amb més potencial

Les granges més tecnificades podran aprofitar millor les solucions. En principi la tendència és que la tecnificació augmenti, ja que el consumidor valora i demana conèixer la traçabilitat dels animals. També serà útil per explotacions i granges més grans, on s'han de seguir molts animals, i on alarmes o petites millores en KPIs poden tenir una repercussió més gran en la productivitat final.

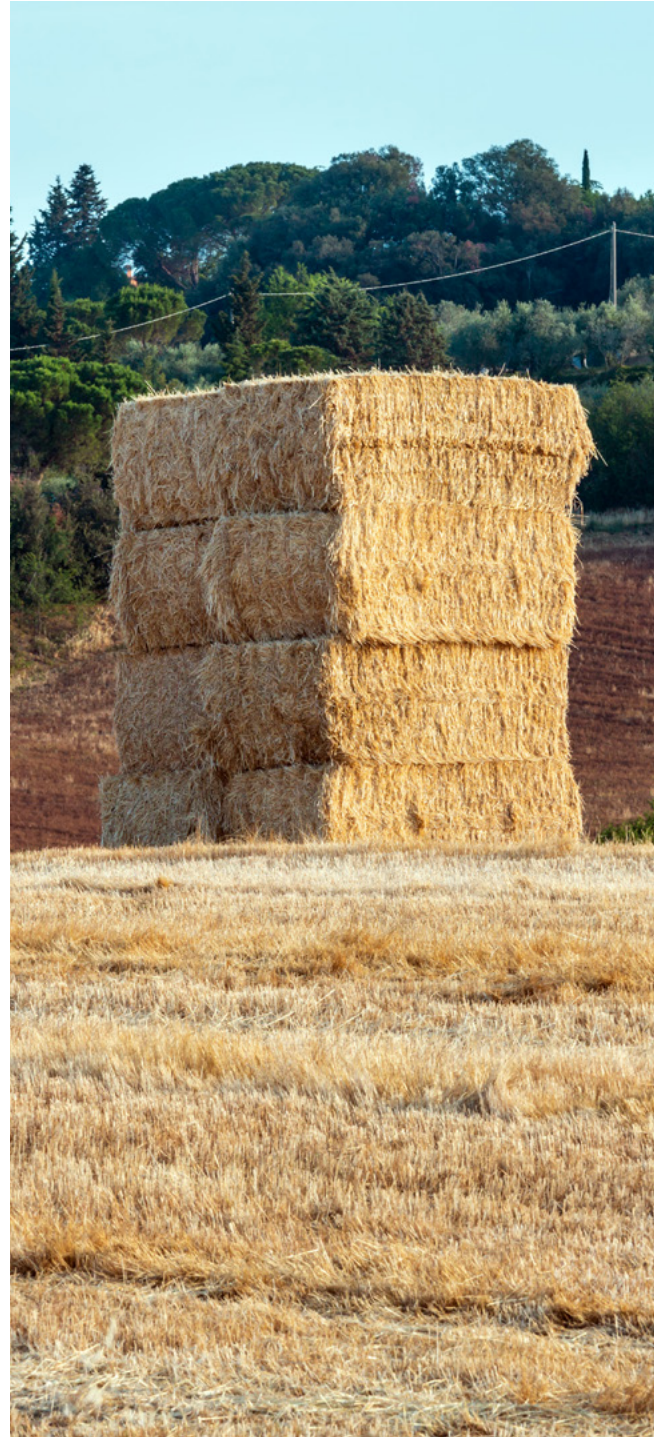
Actualment les granges més tecnificades són: granges d'engreix de porcs, alletament, granges de mares i de besàvies, granges avícoles i de boví.

Integradores

Teòricament poden ser un agent clau i amb potencial. Les Integradores estan acostumades a manegar i analitzar dades, fer comparatives entre les seves granges, saben quines funcionen i quines no, però no acostumen a conèixer les granges per dintre. La solució els hi pot ser molt útil per entendre el funcionament intern de cada granja, ajudant a identificar problemàtiques i/o els motius que justifiquen els seus millors o pitjors resultats, prendre mesures d'optimització/millora. A més, acostumen a tenir recursos disponibles. No obstant, al ser entitats rendibles, potser no tenen interès en invertir més recursos per millorar la productivitat i/o tenen altres prioritats. Al igual que els ramaders, prefereixen esperar a que l'eina estigui més madura.

Cooperatives

Les cooperatives també estan interessades en ajudar als seus ramaders oferint aquest servei al soci, sempre que hi vegin un retorn directe.



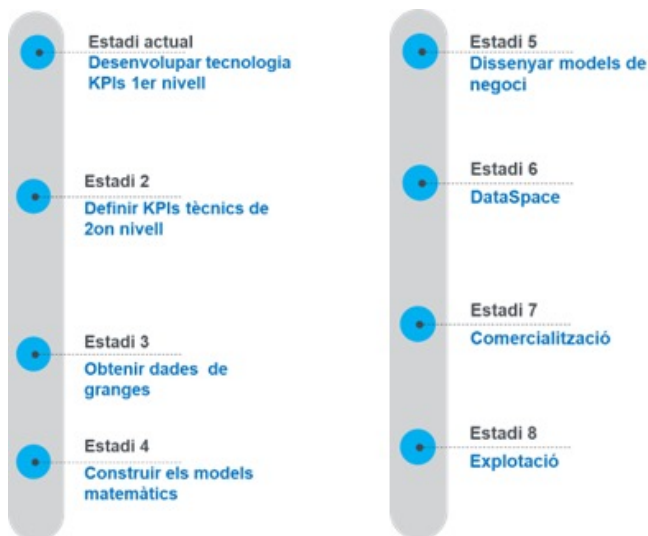
Valoració final

Els usuaris finals estan oberts a conèixer més l'eina. És interessant i els agrada poder disposar de mesures contínues de comportament i benestar animal, però cal seleccionar-les bé, perquè l'eina sigui rendible a nivell econòmic, portant a una millora dels KPIs, una reducció de costos, una optimització d'espais, una millora de l'índex de conversió, etc. La millora ha de ser clarament mesurable.

Cal tenir en compte que en la cria dels animals intervien moltes variables, que estan interrelacionades (també intervé el factor humà), i pot no ser fàcil aïllar paràmetres que permetin prendre decisions.

Els consultors estan convençuts del potencial del software i de l'espai de dades a mig termini, i li veuen moltes utilitats. Pensen que en una segona fase es podran trobar interrelacions entre variables que es traduiran en estalvis / productivitat per les explotacions. Pensen però, que el camí serà llarg i que s'hauran de superar barreres. Per altra banda, consideren molt important entendre al ramader i oferir-li l'eina quan estigui més madura i amb arguments que el motivin – poques dades i útils de cara a la gestió. Proposen desenvolupar models de comercialització imaginatius en tot el procés per trencar les barreres dels ramaders: models de servitització per que instal·lin les càmeres, sistemes d'incentius perquè comparteixin les seves dades, etc. I la possibilitat de comptar amb el suport i ajuts sectorials de l'Administració per desenvolupar els models i validar-los seguint l'esquema de següents passos representats a la Figura 10 i descrits a continuació:

Figura10. Esquema de next-steps



- **Estadi Actual / Etapa Tecnològica.** Aconseguir la tecnologia per obtenir els KPIs de comportament animal de primer nivell, a partir del seguiment individualitzat de cada porc identificat.
- **Estadi 2 / Etapa Tècnica.** Llistar de forma exhaustiva totes les possibles variables de benestar animal, comportaments anòmals, sanitàries, d'índex de conversió, productivitat (siguin directes o indirectes) que permetin trobar correlacions amb els KPIs de primer nivell i segmentar els models com es desitgi (ex., per etapes de creixement). Quantes més variables millor (consum d'antibiòtics, pinso, moviments, inactivitat, etc.). Quan menys òbvies millor (per evitar que es copii la solució). Els veterinaris poden pensar en dades més tècniques que el ramader i afegir més valor als models. També pot participar el ramader.
- **Estadi 3 / Etapa Recollida de Dades.** Obtenir dades de totes les granges / explotacions possibles (per enfortir el model) / validar les dades recollides.
- **Estadi 4 / Etapa Creació i Validació.** Realitzar l'anàlisi estadístic multivariant per trobar totes les correlacions significatives entre variables, i construir els models matemàtics segmentats (ex., mares, reproducció, engreix, moment de l'engreix), que permetin obtenir KPIs resumits / útils / amb incidència més directa pels ramaders.
- **Estadi 5 / Etapa Disseny Comercialització.** Dissenyar models de venda a ramaders / integradores / cooperatives, tenint en compte les seves necessitats i pensant en fórmules que motivin la contractació tant del software com del DataSpace (sistemes d'incentius).
- **Estadi 6 / Etapa Tecnològica + Legal.** Etapa paral·lela a l'anterior (no correlativa). Formular el plantejament tecnològic del DataSpace. Establir polítiques de protecció de dades, consentiments, etc.
- **Estadi 7 / Etapa Comercialització.** Aconseguir clients de l'eina / contractacions (ramaders, integradores i cooperatives).
- **Estadi 8 / Etapa Explotació.** Obtenció dels KPIs de cada explotació client (software). Amb les dades dels que accedeixin a compartir-les, alimentar les BBDD per enfortir / validar / sofisticar els models.

Model de governança de dades

Establir un model de governança és essencial en aquest projecte emmarcat al sector porcí, on es fan servir espais de dades, per diverses raons fonamentals. En primer lloc, assegura que les dades es recopilin, emmagatzemin i comparteixin de manera ètica i responsable, protegint la privadesa de les explotacions porcines i altres parts interessades. A més a més, contribueix a garantir el compliment de les regulacions de privadesa i protecció de dades, evitant possibles sancions legals. També estableix directrius clares per a la gestió de dades, promovent la transparència i la confiança de les parts interessades i facilitant una recerca ètica i responsable. En resum, el model de governança és fonamental per mantenir la integritat, la conformitat i l'èxit continuat del projecte en aquest sector.

Estructura organitzativa

L'establiment d'una estructura organitzativa clara i estable en aquest projecte del sector porcí, utilitzant espais de dades, és crític per diverses raons. En primer lloc, ajuda a definir rols, responsabilitats i jerarquies, assegurant que cada membre de l'equip conegui les seves funcions i com s'integra amb altres. Això promou una distribució eficient del treball i una comunicació efectiva, evitant confusions i duplicacions d'esforços. També facilita la presa de decisions i la resolució de problemes, ja que és clar qui és el responsable de què. A més, aporta estabilitat i cohesió a llarg termini, donant continuïtat al projecte.

A tal efecte, resultarà essencial establir:

- **Responsable de Dades.** Designar un responsable de dades, preferiblement un Chief Data Officer (CDO), qui supervisarà la governança de dades del projecte.
- **Comitè de Governança de Dades.** Establir un comitè que inclogui representants de la universitat, científics, experts en ètica i privadesa, i altres interessats rellevants.
- **Rols i Responsabilitats.** Definir rols i responsabilitats clares per a qui gestiona les dades, incloent científics, tècnics i personal de camp.

Polítiques i procediments

Política de recopilació de dades

És essencial establir una política de recopilació de dades en aquest projecte aplicat al sector porcí per garantir la coherència, la protecció de la privadesa i el compliment de les normatives en l'ús d'espais de dades. Això assegura que les dades es recopilin i es comparteixin de manera ètica i eficaç, protegint la privadesa de les parts interessades i mantenint la integritat de la recerca.

Instal·lació de Càmeres:

- Les càmeres seran instal·lades a les explotacions porcines per a fins de monitorització de l'activitat dels animals.
- La ubicació de les càmeres s'acordarà amb els propietaris de les explotacions i es garantirà que respectin la privadesa i la seguretat dels animals.

Consentiment informat:

- Les explotacions porcines han de donar el seu consentiment informat per a la recopilació d'imatges i dades relacionades amb els animals.
- El consentiment informat s'obtéindrà per escrit i es mantindrà com a registre.

Política d'accés i ús

Establir una política d'accés i ús és fonamental en aquest projecte porcí per regular l'ús de les dades, garantir la seguretat i la privadesa, i complir amb les normatives. Això ajuda a evitar abusos i assegura un ús ètic de les dades, establint directrius clares per als usuaris i protegint els interessos de les parts involucrades, incloent les explotacions porcines.

Autorització d'Accés:

- Les imatges recopilades seran propietat de les explotacions porcines i la Universitat de Lleida, i només podran ser utilitzades per les parts interessades segons els acords de l'espai de dades.

Finalitat de l'Ús:

- Les dades només es podran utilitzar per a l'objectiu específic de seguiment i anotació d'animals, i per a la millora de models d'intel·ligència artificial per al seguiment de porcs.

Restricció de Compartir Dades:

- Les dades no podran ser compartides fora de l'espai de dades sense l'autorització explícita de les explotacions porcines i la Universitat de Lleida.

Retenció de dades

Les directrius de retenció de dades en aquest projecte, asseguren una gestió eficient i segura de les dades, minimitzant els riscos de filtracions i complint amb les regulacions. A més, s'evita la retenció innecessària de dades, protegint la privadesa de les parts interessades.

Retenció de Dades:

- Les imatges originals i les anotacions relacionades es mantindran durant un període de temps específic, que ha de ser acordat amb les explotacions i les parts interessades.
- Després d'aquest període, les dades es suprimiran de manera segura, assegurant-se que no siguin recuperables.

Política de seguretat

L'acord sobre política de seguretat és crucial per garantir la continuïtat i la protecció del projecte porcí, protegint les dades, prevenint pèrdues i assegurant la confiança de les parts interessades. Això permet mantenir la integritat i l'èxit a llarg termini del projecte.

Seguretat de Dades:

- Les dades estaran protegides contra l'accés no autoritzat, l'ús indegut i la pèrdua.
- Les mesures de seguretat inclouran autenticació d'usuaris, xifrat i control d'accés.

Seguretat de les Càmeres:

- Les càmeres estaran protegides per evitar la manipulació o l'accés no autoritzat.
- S'establiran procediments per al manteniment regular de les càmeres per assegurar-se que estiguin operatives i recopilant dades correctament.

Protecció de Dades Anonimitzades:

- Les dades anonimitzades es protegiran per evitar la identificació d'animals o explotacions porcínes individuals.

Qualitat de dades

Les directrius sobre qualitat de dades asseguren que el projecte porcí mantingui la integritat i la fiabilitat de les dades, fonamentals per la seva continuïtat i èxit.

Estandardització de Dades:

- Caldrà definir estàndards per a la recopilació de dades, incloent la resolució de vídeo, el format de fitxer i la calibració de les càmeres.

Validació de Dades:

- Implementar processos de validació per garantir que les dades recopilades siguin precises i fiables.

Anotació de Dades:

- Les imatges han de ser anotades mitjançant l'ús de l'entorn Supervisely o eines similars, seguint les pràctiques d'anotació establertes.

Validació Visual:

- Les anotacions hauran de ser validades visualment per un equip d'experts per garantir que coincideixin amb la realitat observada a les imatges.
- Es revisarà l'exactitud de les anotacions en termes de la posició exacta dels animals i les màscares de segmentació.

Verificació de Coherència:

- Les dades han de ser subjectes a una verificació de coherència per assegurar que les anotacions siguin coherents en tot el conjunt de dades i que les etiquetes siguin utilitzades de manera consistent.

Proves de Fiabilitat del Model:

- Les dades anotades es faran servir per a provar i validar la fiabilitat del model de Machine Learning (Mask R-CNN).
- S'avaluarà el rendiment del model en termes de detecció i seguiment precís dels porcs a partir de les anotacions.

Retroalimentació i Correcció:

- Si es detecten errors o inexactituds en les anotacions durant la validació, es realitzarà la correcció i l'actualització de les anotacions pertinents.

Privadesa i ètica - Avaluació d'impacte de privacitat (AIPD)

L'Avaluació d'Impacte de Privacitat (AIPD) és essencial en aquest projecte per protegir la privadesa de les dades, assegurar el compliment legal, guanyar la confiança de les parts interessades, minimitzar riscos i garantir una recerca ètica. Mitjançant la identificació i la mitigació de riscos, l'AIPD contribueix a la protecció de la privadesa de les explotacions porcines, les persones i els individus involucrats i assegura el compliment de les lleis de privadesa aplicables.

Identificació de riscos

Risc d'Identificació d'Explotacions Porcines:

- Existeix el risc que les dades recopilades i compartides puguin utilitzar-se per identificar les explotacions porcines i la seva ubicació, la qual cosa podria suposar un risc de seguretat per als propietaris.

Risc d'Exposició de Dades Sensibles:

- Les dades de seguiment poden contenir informació sensible, com pràctiques de cria o salut animal, la qual cosa podria exposar detalls confidencials de les explotacions porcines.

Risc de Reidentificació d'Individus:

- Si no s'anonimitzen adequadament les imatges i les dades, existeix el risc que les persones puguin ser identificades en el context de les imatges, la qual cosa afectaria la seva privadesa.

Avaluació de l'impacte

Impacte en la Privadesa de les Explotacions Porcines:

- El risc d'identificació podria afectar la seguretat i la privadesa de les explotacions porcines, la qual cosa podria dur a la resistència a compartir dades.

Impacte en la Privadesa de les Persones:

- La manca d'anonimització adequada podria exposar individus a la reidentificació, la qual cosa comprometria la seva privadesa.

Mitigació de riscos

Anonimització de Dades:

- Garantir que les imatges i les dades s'anonimitzin de manera efectiva abans de ser compartides a l'espai de dades, de manera que no es pugui identificar les explotacions porcines ni individus.

Limitació de Dades Sensibles:

- Evitar recopilar o compartir dades sensibles que no siguin necessàries per als objectius del projecte.

Consentiment Informat:

- Continuar obtenint el consentiment informat de les explotacions porcines per a la recopilació i l'ús de dades i garantir que entenguin els riscos i beneficis.

Educació i Sensibilització:

- Proporcionar informació i formació a totes les parts interessades sobre la importància de la privadesa de dades i les pràctiques segures.

Supervisió Contínua:

- Implementar un procés de supervisió contínua de la privadesa de dades i revisar i actualitzar les mesures de privadesa segons sigui necessari.



Mesurament i monitoratge

D'igual manera, l'establiment de mètriques de rendiment i la realització d'auditories regulars són essencials per avaluar l'eficàcia, assegurar el compliment normatiu, millorar les pràctiques i protegir la privadesa de les dades en un projecte com aquest. Aquestes mesures permeten controlar la qualitat, demostrar el compliment de les regulacions i assegurar la transparència, la qual cosa és fonamental per una recerca ètica i la confiança de les parts interessades.

Mètriques de rendiment

Precisió del Model de Seguiment:

La precisió del model d'IA (Mask R-CNN) en la detecció i seguiment d'animals a les imatges.

Rendiment del Model:

- Velocitat de processament i temps de resposta del model durant l'ús pràctic.

Conformitat Legal:

- Grau de compliment de les regulacions de privadesa de dades i protecció animal.

Índex de Satisfacció de les Parts Interessades:

- Valoració de les parts interessades (explotacions, entitats d'anotació, etc.) sobre la gestió de les seves dades i el seu grau de satisfacció amb el projecte.

Compliment de Polítiques:

- Mesura de l'adhesió a les polítiques de governança de dades establertes.

Millores Implementades:

- Nombre i impacte de les millores realitzades en resposta a les auditories i les avaluacions de privadesa.

Nombre d'Usuaris Actius:

- La quantitat d'usuaris que accedeixen regularment a l'espai de dades per a utilitzar les dades i els recursos.

Taxa de Creixement d'Usuaris:

- L'augment en el nombre d'usuaris al llarg del temps, indicant l'atractiu de l'espai de dades.

Freqüència d'Ús:

- Amb quina freqüència els usuaris accedeixen i utilitzen les dades disponibles a l'espai.

Taxa de Retenció:

- La quantitat d'usuaris que continuen utilitzant l'espai de dades després del seu primer accés.

Descàrregues de Dades:

- La quantitat de vegades que les dades s'han descarregat des de l'espai.

Feedback de l'Usuari:

- Retroalimentació i valoració dels usuaris sobre la facilitat d'ús i la utilitat de les dades.

Projectes o Aplicacions Desenvolupades:

- Nombre de projectes o aplicacions que s'han creat utilitzant les dades de l'espai.

Impacte en la Recerca Científica o Aplicacions Pràctiques:

- Indicadors dels resultats i contribucions de la utilització de les dades en recerca científica o aplicacions pràctiques.

Auditories regulars

Les auditories regulars són crucials per assegurar la conformitat, identificar problemes i millorar la seguretat i la privadesa. Es recomana realitzar-les almenys una vegada a l'any per mantenir la integritat del sistema.



Comunicació

És fonamental establir protocols de consciència i formació per al personal i de comunicació externa en un projecte com aquest per diverses raons. Primer, assegura la comprensió i el compliment de les polítiques de privadesa i les normatives de protecció de dades per part de tots els implicats. A més, fomenta la consciència sobre l'ètica i la importància de la protecció de la privadesa. Per últim, una comunicació externa eficaç ajuda a establir una relació de confiança amb les parts interessades i el públic, demostrant compromís amb la privadesa i la responsabilitat ètica en la recerca científica. Així, s'eviten possibles problemes legals, es protegeix la imatge del projecte i es manté la integritat i la transparència en la gestió de dades.

Consciència i formació

Per a garantir la continuïtat al llarg del temps, el projecte haurà de proporcionar formació integral al personal i les parts interessades per garantir la comprensió i l'adhesió a les polítiques i procediments de governança de dades. Això inclouria sessions de formació periòdiques, materials didàctics i recursos en línia per a un aprenentatge continu i una correcta implementació de les polítiques i procediments de gestió de dades.

Comunicació externa

Per a una permanència superior al temps del PAI, serà crucial informar les parts interessades externes, com les explotacions agrícoles, mitjançant una comunicació clara i transparent. Es recomana la distribució de documentació explicativa que detalla com les dades es recopilaran, utilitzaran i protegiran per garantir la seva privadesa. A més a més, també poden ser útils les sessions d'educació i reunions informatives per abordar preguntes i preocupacions específiques. Així com l'establiment de canals de comunicació obert perquè les parts interessades puguin plantejar inquietuds o aclariments. Aquest enfocament assegura que les explotacions agrícoles estiguin ben informades i confiades en la gestió de les seves dades en el projecte.

Retroalimentació i millora

Els protocols de millora continua asseguren que el projecte porcí s'adapti a canvis, millori i mantingui la seva efectivitat al llarg del temps, garantint la continuïtat i l'èxit. Per a fer-ho, es recomana donar seguiment als següents passos.

1. Recopilació de Retroalimentació. Recopilar retroalimentació de les parts interessades mitjançant enquestes, reunions i comentaris.
2. Avaluació de Retroalimentació. Analitzar la retroalimentació per identificar tendències i àrees de millora.
3. Actualització de Polítiques i Procediments. Ajustar les polítiques i procediments segons la retroalimentació i les necessitats canviants.
4. Comunicació i Formació. Comunicar els canvis i oferir formació sobre les actualitzacions als membres de l'equip i altres implicats.
5. Implementació i Seguiment. Garantir que les noves polítiques es posin en pràctica i segueix el seu impacte.
6. Avaluació Continua. Mantenir la retroalimentació com una pràctica contínua, amb recopilació i ajustos periòdics.



4.5. Conclusions

El projecte desenvolupat proporciona les següents conclusions:

1. El model IA LAB apareix com una eina metodològica eficaç per resoldre reptes de territori, integrant l'ecosistema de manera eficient en procés d'injecció d'IA dintre de l'activitat sectorial específica.
2. La solució tecnològica de seguiment de l'activitat dels porcs en la quadra obre un nou escenari per la traçabilitat dels factors d'engreix de manera automàtica i individualitzada per cada animal. Això permet un augment de la qualitat dels processos i un potencial benefici per l'apertura de línies de mercat en les que es pot validar de manera objectiva la traçabilitat del benestar animal durant el procés d'engreix.
3. El model de desenvolupament de pilots requereix, a més de la implementació específica d'una solució tecnològica, un mecanisme que permeti crear valor de les dades generades. L'espai de dades dintre de l'àmbit agro creat proporciona el mínim l'esquelet d'infraestructura necessària per fer-ho possible.
4. El model de governança, sostenibilitat i negoci generat proporciona una primera aproximació a la valorització possible que la tecnologia emprada pot tenir.

L'explotació de les valors digitals generats en aquest projecte -dades i algorismes- permet, finalment, obrir una porta al sector per poder aprofitar la injecció d'IA dintre de les explotacions, avançant en la millora de qualitat dels processos d'engreix. En efecte, es tracta de un model d'implementació del camí "de bits a quilos" que l'ecosistema de Catalunya pot aprofitar a través del CIDAI i del diversos d'hubs d'innovació del país



Annex



A1. Detalls de l'experimentació – Base teòrica

Recollida de dades

En primer lloc cal desenvolupar una comprensió del domini de l'aplicació i el coneixement previ més rellevant identificant l'objectiu del procés KDD des del punt de vista del client. Per tant, alguns punts clau d'aquesta fase són la identificació de les fonts de dades rellevants per a l'anàlisi, el disseny del procés de recopilació, establint els mètodes i les tècniques adequats per obtenir les dades de manera eficient, seleccionant mostres representatives de la població per garantir resultats generalitzables; la recopilació de les dades, i la documentació dels procediments.

Finalment, es consideren aspectes ètics i legals, assegurant la privadesa, obtenint el consentiment informat i complint amb les regulacions i polítiques de protecció de dades.

Preprocessament: Selecció, neteja i transformació

Aquesta etapa del procés KDD és la part més rellevant i que consumeix més temps. Aquesta etapa es coneix com a preprocessament de dades, neteja de dades o preparació de dades, i es centra a preparar les dades per a la seva posterior anàlisi. Durant aquest procés, es fan servir tècniques estadístiques descriptives bàsiques i representacions gràfiques per comprendre i visualitzar les dades. De vegades també s'apliquen tècniques descriptives multivariants, com l'Anàlisi de Components Principals (ACP), per detectar valors atípics o patrons no lineals.

Les tècniques de visualització tenen un paper crucial en la identificació de problemes importants durant el preprocessament de dades. Aquestes tècniques poden variar en el nivell de sofisticació, però totes són útils per garantir la qualitat i l'adequació de les dades abans de fer l'anàlisi.

Tot i no haver-se formalitzat metodologies de preprocessament ben establertes, poden proporcionar-se algunes pautes. No obstant, per a aquest apartat de la metodologia se simplificarà el procés de preprocessament en tres punts: la selecció, la neteja i la transformació de les dades, tenint com a guia la metodologia proposada a (Gibert K, 2016) i mostrada a la Figura 1.

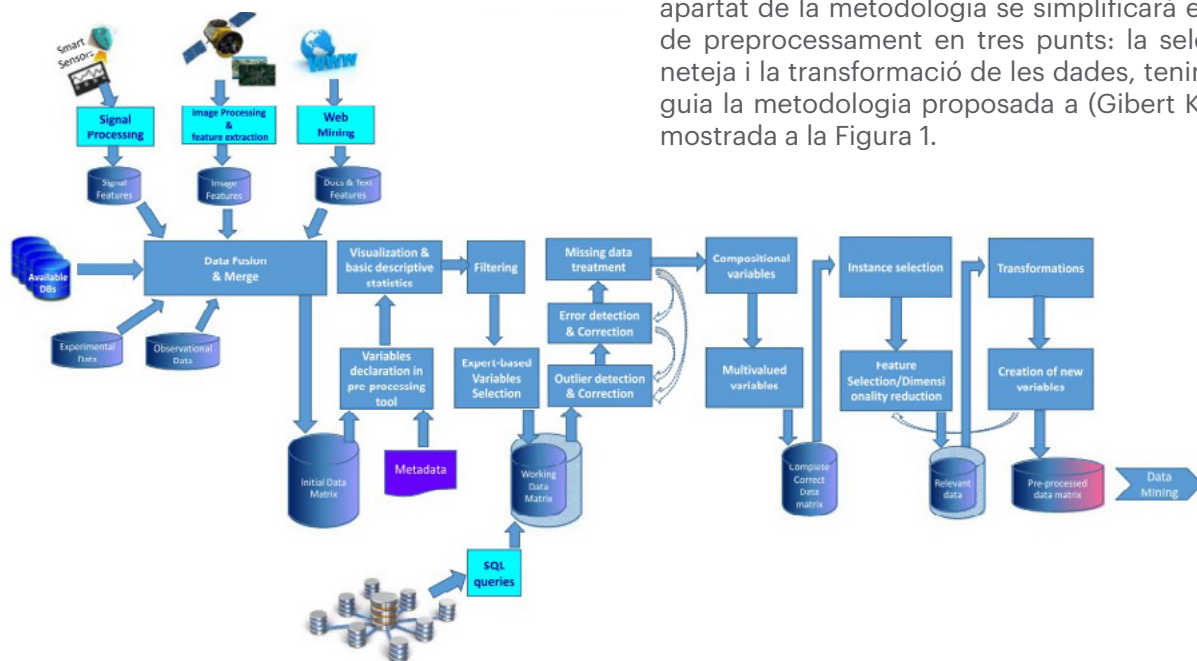
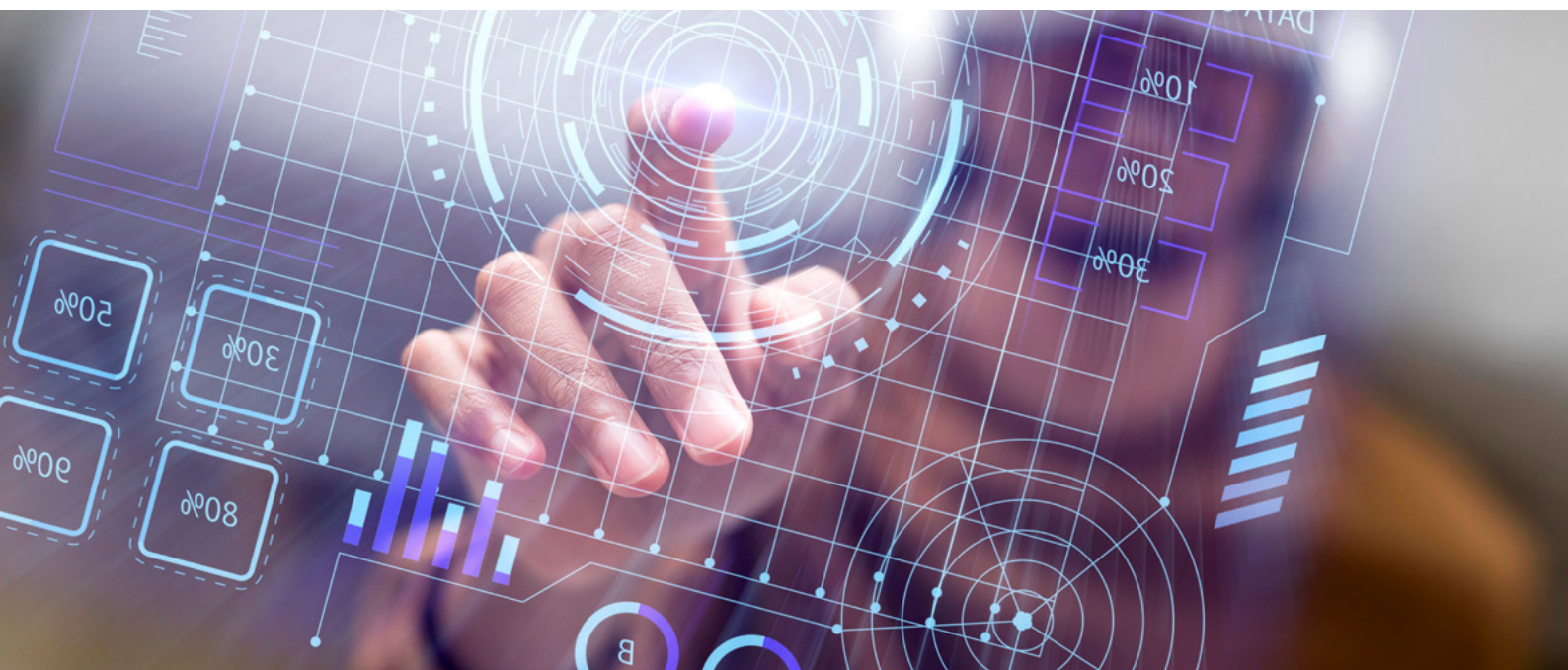


Figura 1. Procés de preprocessament (Gibert K, 2016)

Selecció de les dades

La selecció de dades implica identificar i recopilar els conjunts de dades rellevants que es faran servir en el procés de descobriment de coneixement. Aquesta etapa requereix una comprensió clara dels objectius i requisits del projecte, així com de les fonts de dades disponibles. A continuació s'anomenen alguns aspectes clau d'aquesta etapa:

1. **Definició dels objectius:** Abans de començar la selecció de dades, és fonamental tenir una comprensió clara dels objectius i requisits del projecte. Això implica definir quin tipus de coneixement es vol extreure de les dades i quines preguntes es pretenen respondre. Establir objectius clars ajuda a enfocar la selecció de dades a la informació rellevant per assolir aquests objectius.
2. **Identificació de les fonts de dades:** En aquesta etapa, cal identificar les fonts de dades disponibles que podrien contenir la informació necessària per a l'anàlisi. Aquestes fonts de dades poden incloure bases de dades, fitxers CSV, registres de transaccions, dades recopilades de sensors, dades en línia, entre d'altres. És important avaluar la qualitat i la integritat de les fonts de dades per garantir que siguin fiables i adequades per a l'anàlisi.
3. **Avaluació de la pertinència:** Un cop identificades les fonts de dades, cal avaluar la pertinència de cadascuna d'elles en relació amb els objectius del projecte. Això implica examinar les variables i atributs disponibles a cada conjunt de dades i determinar si proporcionen la informació necessària per respondre les preguntes plantejades. És possible que algunes fonts de dades no siguin rellevants i s'hagin de descartar, mentre que altres poden requerir transformacions o combinacions amb altres fonts per ser més útils.
4. **Grandària i representativitat de les dades:** En aquesta etapa, també s'ha de considerar la mida i la representativitat dels conjunts de dades. La mida adequada de les dades dependrà de la complexitat de l'anàlisi i dels recursos disponibles. És important assegurar-se que els conjunts de dades seleccionats siguin representatius de la població o el fenomen que s'està estudiant, per evitar biaixos o conclusions incorrectes.
5. **Consideracions ètiques i legals:** En seleccionar les dades, és important tenir en compte les consideracions ètiques i legals. Això implica garantir la privadesa i confidencialitat de les dades, així com complir amb les regulacions i polítiques de protecció de dades vigents. És possible que es requereixi obtenir el consentiment informat dels individus les dades dels quals s'utilitzaran, especialment en casos sensibles o en què es manegin dades personals.



Neteja de les dades

Una vegada s'han seleccionat les dades, es procedeix a la neteja. La neteja de dades és una etapa crítica en el procés de preparació de dades i té com a objectiu identificar i corregir problemes o inconsistències en aquests conjunts, que són els següents:

1. **Dades mancants.** Durant la recopilació de dades, és comú trobar valors mancants. Aquests poden ser causats per diversos motius, com ara errors a l'entrada de dades, problemes de transmissió o simplement la manca d'informació. En aquesta etapa, cal identificar les dades mancants i decidir com manejar-les. Això pot implicar eliminar les files o columnes que continguin dades mancants, imputar valors utilitzant tècniques com la mitjana o la interpolació, o utilitzar mètodes més avançats, com ara l'aprenentatge automàtic, per predir els valors mancants. En cas de dades mancants estructurals, els experts poden proporcionar regles per imputar-les manualment (knowledge-based logical rules) (Gibert K, 2016).
2. **Valors atípics.** Els valors atípics són observacions que difereixen significativament de la resta de dades. Aquests valors poden ser causats per errors de mesura, ingrés incorrecte de dades o fins i tot poden indicar situacions excepcionals o anòmales. És important identificar els valors atípics i determinar si s'han de corregir o eliminar. Això es pot fer utilitzant mètodes estadístics, com ara els límits de rang o criteris basats en la desviació estàndard, o mitjançant l'ús d'algorismes més sofisticats, com ara l'anàlisi de valors extrems
3. **Inconsistències i errors.** Els conjunts de dades sovint contenen errors o inconsistències, com ara valors incorrectes o contradictoris. Per exemple, un camp que registra l'edat d'una persona pot contenir valors negatius o més grans de 100 anys, cosa que és poc probable. En aquesta etapa, cal identificar i corregir aquest tipus d'errors, ja sigui eliminant les observacions amb problemes greus o corregint els errors coneguts utilitzant tècniques específiques.
4. **Dades duplicades.** Els conjunts de dades poden contenir duplicats, cosa que significa que hi ha observacions idèntiques o molt similars al conjunt de dades. La presència de dades duplicades pot esbiaixar els resultats de l'anàlisi i afectar la precisió dels models. Per tant, és important identificar i eliminar les dades duplicades per evitar biaixos innecessaris a l'anàlisi.

Transformació de les dades

La transformació de dades fa referència a l'aplicació d'operacions o tècniques per convertir les dades en una forma adequada per a l'anàlisi posterior. Això pot incloure la normalització de valors, la discretització de variables contínues, la codificació de variables categòriques, la reducció de la dimensionalitat o qualsevol altra transformació necessària per adaptar les dades a l'algorisme o mètode específic que s'utilitzarà a l'etapa següent del procés de descobriment de coneixement.



Mineria de dades

La mineria de dades és l'etapa en què s'utilitzen algorismes per analitzar i descobrir patrons a les dades. Aquests algorismes busquen de manera eficient i pràctica una llista específica de patrons o models, dins un espai de patrons que sol ser infinit. Tot i això, a causa de limitacions computacionals, s'imposen restriccions que limiten l'espai de cerca i el subconjunt de patrons que pot ser explorat pels algorismes de mineria de dades. Aquestes restriccions són necessàries per garantir un procés eficient i viable des del punt de vista computacional. En aquesta investigació es va fer una selecció del model de mineria de dades que cal aplicar al cas d'estudi. La literatura esmenta una tipologia variada de models, i la seva selecció depèn de la naturalesa de les dades, els objectius de l'estudi i l'ús que haguem de fer dels resultats d'acord amb el model seleccionat. Els autors Gibert et. al. a l'any 2018 (Gibert K. I.-M.-R., 2018) mostren un mapa conceptual que suporta la selecció adequada dels mètodes en mineria de dades. Aquí s'han utilitzat tècniques on no intervé una resposta variable i s'estudien les relacions entre variables (ACP) i individus (*clustering*).

Anàlisi de Components Principals (ACP)

L'Anàlisi de Components Principals (ACP) és un mètode estadístic que permet reduir la dimensionalitat de les dades mentre es conserva informació rellevant. Pertany a un conjunt de tècniques conegudes com a aprenentatge no supervisat, cosa que significa que no es basa en un model predictiu, sinó que busca extreure informació per identificar, per exemple, subgrups a les dades.

El mètode de l'ACP opera algebraicament amb els vectors propis (*eigenvectors*) i valors propis (*eigenvalues*) d'una matriu de dades numèriques i utilitza criteris de maximització de la variància per expressar-ne els components principals. Per això s'observa la inèrcia o variància explicada de les dimensions, les quals mostren si hi ha relacions fortes entre les variables d'aquests plans i suggereix el nombre d'eixos que cal estudiar. En general, se seleccionen les components que expliquen almenys el 80% de la variància total, és a dir, que recullen informació real.

El procés de l'ACP identifica les adreces en què la variància és més gran. Atès que la variància d'una variable es mesura a la mateixa escala elevada al quadrat, si no s'estandarditzen prèviament totes les variables perquè tinguin una mitjana de 0 i una desviació estàndard d'1, les variables amb una escala major dominaran les altres. Per tant, cal estandarditzar les dades abans de calcular les components principals, cosa que també ajuda a corregir possibles valors atípics (outliers), ja que aquestes poden tenir una influència directa en la variància de les variables.

Clustering

L'anàlisi de conglomerats clustering és un mètode estadístic emprat en el processament de dades, pel qual s'organitzen tots els elements de la base de dades en grups, o clústers, en funció de com estiguin estretament associats i les característiques que comparteixin. L'objectiu de l'anàlisi de conglomerats és trobar grups similars de subjectes, on la "semblança" entre cada parell de subjectes significa compartir certs trets sobre tot el conjunt de característiques.

El mètode utilitzat per realitzar el clúster és el mètode jeràrquic ascendent de Ward, que té com a objectiu trobar grups esfèrics compactes que minimitzin la variància total dins del grup. A cada etapa, s'uneixen els parells de grups que tenen la distància més petita entre ells i s'avalua la unió de tots els possibles parells de grups combinant aquells la fusió dels quals resulta en el menor augment de la pèrdua d'informació.

Seguint aquesta metodologia és que es creen els dendrogrames, diagrames d'arbre que mostren els conglomerats de les observacions i els nivells de similitud. El nombre de classes òptim es determina observant el dendrograma resultant i tenint en compte les característiques de les dades i els objectius de l'anàlisi. De manera que l'expert pot determinar el nombre òptim de classes observant el diagrama d'arbre o bé es poden fer servir altres algorismes com el mètode del colze. Val a dir que com més gran sigui la distància entre dos enllaços, la similitud, més gran serà la diferència en termes de característiques.

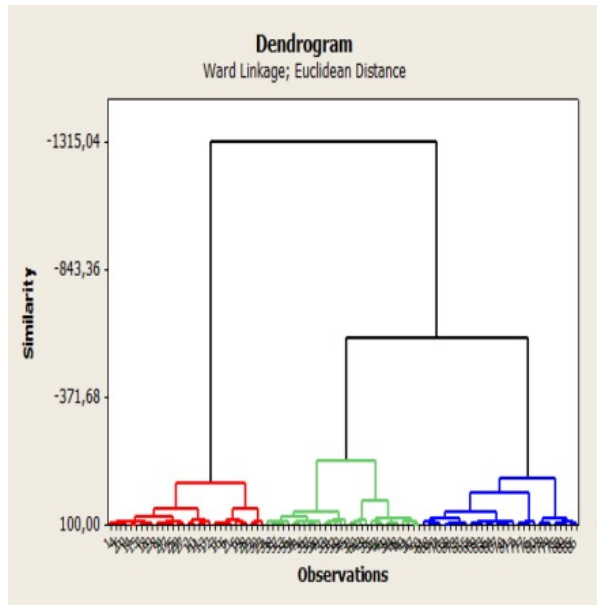


Figura 2. Dendrograma de les dades Iris Flower de R (<https://www-eio.upc.edu/teaching/DocenciaMulti-variant/>)

Per elaborar l'anàlisi de grups en dades mixtes (numèriques i categòriques) cal establir una mètrica que sigui adequada, i per això hi ha diversos algorismes. En aquest estudi es van emprar les mètriques de Gower i la distància euclidiana, que mesuren la distància entre observacions.

Interpretació i avaluació

El perfilat de les classes (clústers) es fa servir en el sentit de descriure les distribucions condicionades d'aquestes classes per poder identificar-ne les especificitats i poder-les interpretar en base a les variables incloses en el procés de clusterització. Entre les eines més utilitzades es té:

- **Càlcul de Centroides.** En general es fan amb la mitjana i la mitjana de les dades agregades (condicionades per cada classe) per a variables numèriques. En el cas de variables categòriques es calculen sobre la base de la moda o mitjançant taules que descriu la composició de les categories condicionat a cada classe.
- **Eines d'estadística descriptiva.** Boxplots múltiples, bicharts, pie charts i barcharts múltiples per descriure la composició de cada variable condicionada a les classes.

Validació

La validació es refereix a l'avaluació i la verificació dels resultats obtinguts durant les etapes de descobriment i modelatge. És un pas essencial per assegurar-se que els patrons, models o regles descoberts siguin fiables, precisos i útils en un context real.

La validació implica l'aplicació de mètriques i tècniques adequades per avaluar la qualitat i el rendiment dels resultats del KDD. Això pot incloure la divisió del conjunt de dades en subconjunts d'entrenament i prova, la comparació de les prediccions del model amb dades reals, el càlcul de mesures de precisió i exactitud i la realització de proves estadístiques per determinar la significació dels resultats.

Proves estadístiques de la relació entre variables:

- **Prova de Kruskal Wallis.** La prova de Kruskal-Wallis és una metodologia estadística no paramètrica utilitzada per avaluar si hi ha diferències significatives entre grups en una variable contínua. S'ordenen i assignen rangs a les dades, es calcula la suma dels rangs per a cada grup i s'obté una estadística de prova. Si l'estadística de prova és estadísticament significatiu, es conclou que hi ha diferències significatives entre els grups en la variable mesurada.
- **Prova de Chi-quadrat.** La prova Chi-quadrat és una prova estadística que s'utilitza per determinar si hi ha una associació significativa entre dues variables categòriques. Es basa a comparar la freqüència observada de cada categoria amb la freqüència esperada sota la suposició d'independència entre les variables. Primer es calcula un estadístic de prova txi-quadrat que es compara amb la distribució txi-quadrat teòrica. Segon, si el valor p associat amb l'estadística de prova és menor que el nivell de significació predefinit, es rebutja la hipòtesi nul·la i, finalment, es conclou que hi ha una associació significativa entre les variables.

A2. Detalls de l'experimentació – Aplicació al cas d'ús 1

Recopilació de dades

Per valorar la possible influència de diferents factors en el consum d'antibiòtics, tenint en compte l'orientació productiva, es va requerir la informació següent:

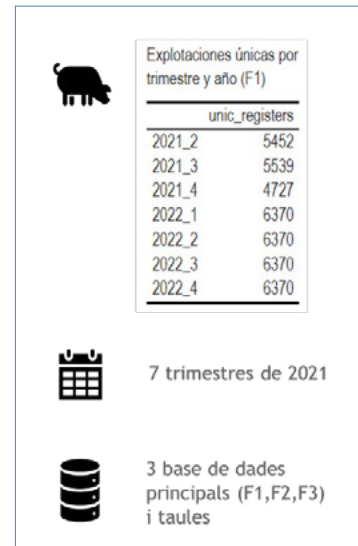
1. Grandària de l'explotació (mitjana d'ocupació de la granja)
2. Data de l'explotació (data origen o data darrera orientació productiva)
3. Si la granja forma part d'una empresa integradora
4. Nombre orígens dels animals
5. Distància des de l'origen dels animals a la granja de destinació
6. Edat dels animals a l'entrada
7. Entrades escalonades (sistema tot dins/tot fora)
8. Resultat de l'enquesta de bioseguretat

A partir de les dades de consum per explotació i espècie, es va valorar la possible influència de diferents factors relacionats amb el veterinari prescriptor en el consum d'antibiòtics a partir de dades de:

- Consum d'antibiòtics total i antibiòtics crítics / explotació/ veterinari /any
- Número prescripcions, nombre de granges, i dispersió granges per veterinari
- Nombre de veterinaris prescriptors per granja
- Identificació dels tractaments repetitius per tipus d'explotació i veterinari
- Tipus d'antibiòtic utilitzats

Es disposa de les dades de consum d'antibiòtics per explotació, espècie i prescripcions per veterinari per a l'any 2021. Aquestes dades ens les facilita el Servei d'Alimentació Animal i Seguretat de la Producció Ramadera de la Generalitat de Catalunya a través del MAPA (Ministerio de Agricultura Pesca y Alimentación). A més es disposa del veterinari prescriptor i de la categorització de l'antibiòtic (B, C i D).

Figura 3. Resum de la base de dades. Font de dades: CIDAI (6 de juliol 2023)



La base de dades original compta amb un total de 25 taules. Aquestes es divideixen en taules principals i taules auxiliars de tipus Excel (xlsx):

- F1, F2 i F3 són els fitxers principals que, alhora, cadascun es divideixen en 7 taules més agrupades per any i trimestre.
- Quatre taules auxiliars amb informació d'interès com ara la ubicació, els resultats de l'enquesta de bioseguretat, la mortalitat dels animals, etc.

Taula 1. Descripció dels fitxers originals

Tipus de taules	Nom del fitxer	Descripció
Principals	F1	Resum de la informació d’F2 i F3. Les primeres columnes són comunes als 3 fitxers i serveixen per relacionar dades.
	F2	Conté la informació del pes i l’UR de les explotacions.
	F3	Conté la informació de la dosi de cada prescripció i la informació del veterinari.
Auxiliars	Escalonades	Informació complementària de les explotacions
	Recollida de cadàvers	Dades dels cadàvers del bestiar per explotació del primer trimestre de l’any 2022
	Enquestes bioseguretat	Resultats de l’enquesta de bioseguretat en alguns quadrimestres a F1
	Registre d’explotacions	Informació geogràfica i dades jurídiques de les explotacions

Per a l’anàlisi de les dades, primer es van combinar les 7 taules d’F1, F2 i F3 en una de sola, resultant en una taula F1, F2 i F3 amb les dades trimestrals de 2021 i 2022 respectivament.

Preprocessament i transformació

Selecció, neteja i transformació

Abans de començar amb l’anàlisi dels valors que falten i la descriptiva de les dades, va tenir lloc la consolidació de totes les d’F1, F2 i F3. Per fer-ho, es va utilitzar Llenguatge de Consulta Estructurat (SQL) al programari de Rstudio amb la funció `sqldf()` del paquet `sqldf` (RStudio_Team, 2020). Per tant, es passarà d’un total de vint-i-un subtaules dividides per any i trimestre, a tres.

Qualitat de les dades

Es prossegueix amb l’estudi dels missings (valors que falten) de la base de dades. A les següents figures es mostren el percentatge de dades mancants per variable de cada fitxer de la base de dades.

Figura 4. Gràfic dels valors que falten del fitxer F1

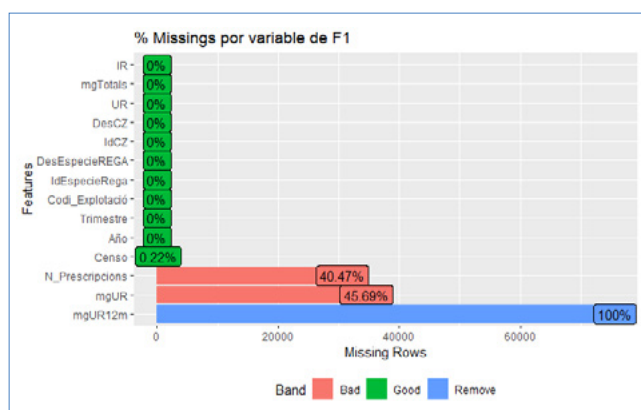


Figura 5. Gràfic dels valors que falten del fitxer F2

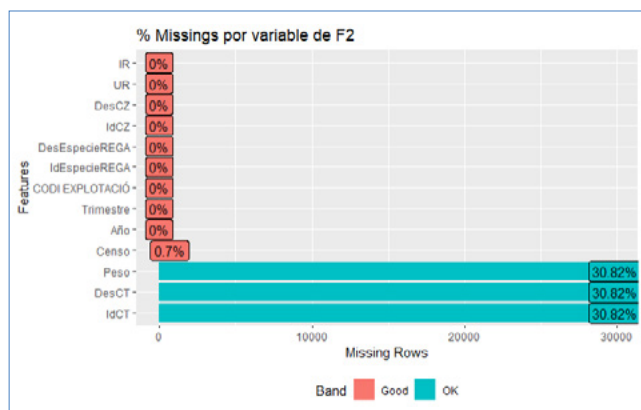


Figura 6. Gràfic dels valors que falten del fitxer F3

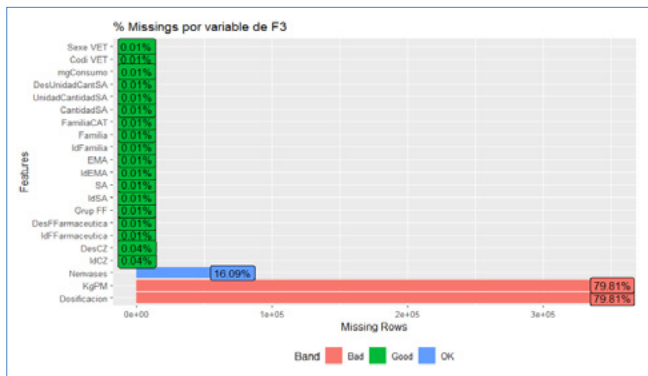


Figura 7. Gràfic dels valors que falten del fitxer Esllaonades

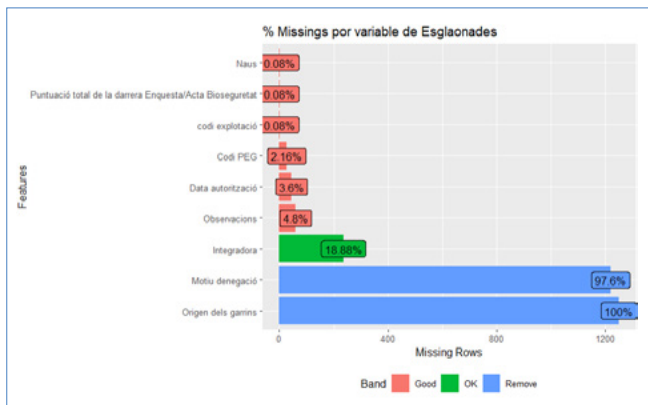


Figura 8. Gràfic dels valors que falten del fitxer Registre d'explotacions

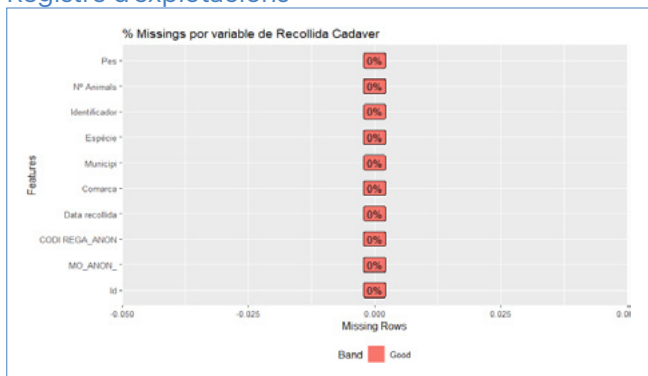


Figura 9. Gràfic dels valors que falten del fitxer Recogita de cadàvers

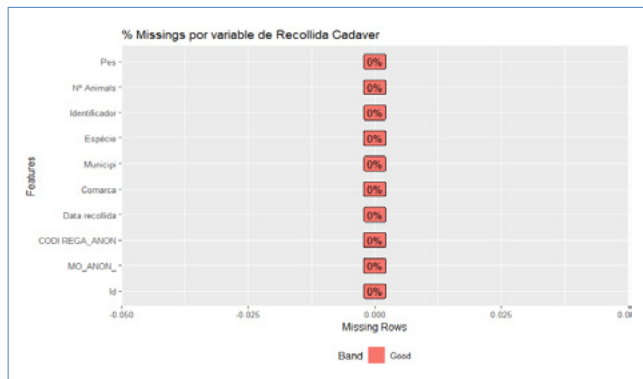
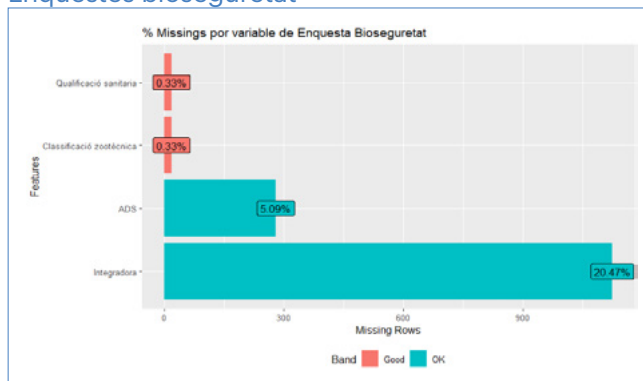


Figura 10. Gràfic dels valors que falten del fitxer Enquestes bioseguret



S'observa en els gràfics que en alguns casos hi ha variables amb molt poques dades mancants o cap, mentre que algunes concentren un percentatge important. Aquestes darreres no es van seleccionar per a l'anàlisi ni es van fer servir al llarg de l'estudi i, per tant, no es van tractar els missings que contenien.

Tractament dels valors mancants

A partir de reunions amb l'expert i l'usuari final es van definir regles lògiques basades en casos per imputar els valors que falten en la majoria dels casos reportats. Variables mgUR12m (F1), Nenvases, Dosificació, KgPM (F3), Origen dels garrins (Escalonades) són les variables que es van quedar fora de l'anàlisi pel seu alt percentatge de *missings*.

Cal recordar que no hi havia dades mancants a la taula auxiliar Recollida de cadàvers, així que no hi ha una taula amb regles lògiques per a aquest fitxer.

Taula 2. Regles lògiques del processament del fitxer F1

Variable	Nou valor	Regla lògica
N_Prescripcions	0	Si (N_Prescripcions =missing) y (Censo=0)
mgUR	0	Si (mgUR=missing) i (Censo=0)
Censo	0	Si (Censo=missing) i (N_Prescripcions=0)

Taula 3. Regles lògiques del processament del fitxer F2

Variable	Nou valor	Regla lògica
Peso	0	Si (Peso=missing) i (Censo=0)
Censo	0	Si (Censo=missing) i (UR=0)
IdCT	"99"	Si IdCT=missing
DesCT	"Sense categoria"	Si DesCT=missing

Taula 4. Regles lògiques del processament del fitxer F3

Variable	Nou valor	Regla lògica
IdCZ	"100"	Si IdCZ=missing
Codi VET	"9999"	Si Codi VET=missing
Sexe VET	"NA"	Si Sexe VET=missing

Finalment, es van eliminar 288 registres nuls, és a dir, que totes les columnes estaven buides (*missing*).

Taula 5. Regles lògiques del processament del fitxer Esclaonades

Variable	Nou valor	Regla lògica
Motiu denegació	"sense descripció"	Si Motiu denegació =missing
Integradora	"sense descripció"	Si Integradora =missing
Observacions	"sense descripció"	Si Observacions =missing
Data autorització	"01/01/1900"	Si Data autorització =missing
Codi PEG	"PEG999"	Si Codi PEG =missing
Naus	"sense descripció"	Si Naus=missing
codi explotació	"9999"	Si codi explotació=missing
Puntuació total de la darrera Bioseguretat	Mitjana tots els valors de la variable	Si Puntuació total de la darrera Enquesta/Acta Bioseguretat =missing

Taula 6. Regles lògiques del processament del fitxer Registre d'exploracions

Variable	Nou valor	Regla lògica
CODI ZOOTÈCNIC	"9999"	Si CODI ZOOTÈCNIC=missing
INTEGRADORA	"sense descripció"	Si INTEGRADORA=missing
NOM ADS	"sense descripció"	Si NOM ADS=missing
CAPACITAT PRODUCTIVA	"sense descripció"	Si CAPACITAT PRODUCTIVA=missing
FORMA DE CRIA	"sense descripció"	Si FORMA DE CRIA=missing
SUBESPÈCIE	"sense descripció"	Si SUBESPÈCIE=missing
DATA FI SUBEXPLORACIÓ	"sense descripció"	Si DATA FI SUBEXPLORACIÓ=missing
CLASSIFICACIÓ ZOOTÈCNICA	"9999"	Si CLASSIFICACIÓ ZOOTÈCNICA=missing
ESTAT SUBEXPLORACIÓ	"sense descripció"	Si ESTAT SUBEXPLORACIÓ=missing
CODI POSTAL EXPLORACIÓ	"sense descripció"	Si CODI POSTAL EXPLORACIÓ=missing
CRITERI DE SOSTENIBILITAT	"sense descripció"	Si CRITERI DE SOSTENIBILITAT=missing
PETITA CAPACITAT	"sense descripció"	Si PETITA CAPACITAT=missing
DATA ACTUALITZACIÓ CENS	"01/01/1900"	Si DATA ACTUALITZACIÓ CENS =missing
DATA ACTUALITZACIÓ CAPACITAT	"01/01/1900"	Si DATA ACTUALITZACIÓ CAPACITAT=missing
DATA CLASSIFICACIÓ ZOOTÈCNICA	"01/01/1900"	Si DATA CLASSIFICACIÓ ZOOTÈCNICA =missing
DATA CANVI ESTAT SUBEXPLORACIÓ	"01/01/1900"	Si DATA CANVI ESTAT SUBEXPLORACIÓ =missing

Taula 7. Regles lògiques del processament del fitxer Enquestes de Bioseguretat

Variable	Nou valor	Regla lògica
Classificació zootècnica	"sense descripció"	Si Classificació zootècnica=missing
Qualificació sanitària	"sense descripció"	Si Qualificació sanitària=missing
ADS	"sense descripció"	Si ADS=missing
Integradora	"sense descripció"	Si Integradora =missing

Tractament dels valors extrems

Es van identificar valors extrems en la quantitat de dosis de diverses explotacions, de manera que es va consultar amb la persona experta al camp i va indicar que eren valors extrems de la població. Per tant, les anàlisis es van fer sense excloure els outliers. Calia esperar que hi hagués valors extrems del consum d'antibiòtics si succeeixen males praxis en l'ús d'aquests, de manera que no poden ser inclosos de l'estudi, ja que es vol identificar les característiques d'aquestes explotacions, entre d'altres.

Enginyeria de variables

S'ha fet un codi automatitzat no segons l'usuari final, sinó mitjançant la selecció de les variables més importants de tota la base de dades, incloent-hi fitxers principals i auxiliars. Així, es van crear 3 fitxers nous amb les dades agregades pel tipus de medicació, graus de localització geogràfica, registre de mortalitat i enquesta de bioseguretat, per tal d'obtenir una base de dades més enriquida que servirà d'input al panell de visualització en Power BI.

Taula 8. Resum dels fitxers resultants de l'enginyeria de variables

Nom	Descripció
TablaMaestraF1	Taula amb les variables d'F1 amb 11 variables més afegides provinents d'F3
TablaMaestraF1-Extendida	Taula amb les variables de TablaMaestraF1 i 10 variables més provinents de taules auxiliars
F1Maestra-clust-Extendida	TaulaMestraF1-Estesa amb la variable c4 dels grups afegida

Com que cada taula està continguda o imbricada a la següent, es mostren a continuació les cinc primeres observacions d'aquelles variables afegides per a cada fitxer.

Figura 11. Imatge dels 5 primers registres de TablaMaestraF1

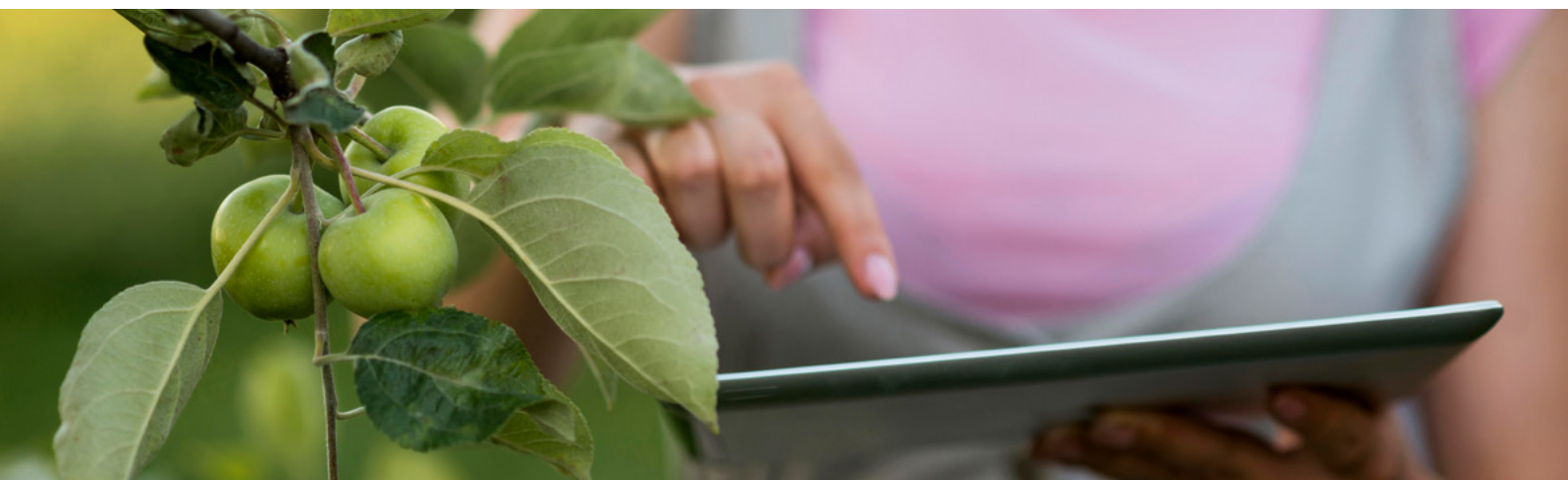
Anio_Trimes	Peso	peso2	mgConsumo_B	mgConsumo_C	mgConsumo_D	pct_mgConsumo_B	pct_mgConsumo_C	pct_mgConsumo_D	CantidadSA_B	CantidadSA_C	CantidadSA_D
2021_2	0	0	0	0	0	0	0	0	0	0	0
2021_2	65	65	250000	6300100	22311000	0,87	21,83	77,3	200	1366,7	2194
2021_2	73,8235294	142,5	44250	7020000	10112564	0,26	40,87	58,87	385	2000500	3289,6
2021_2	43,8048799	38,3333333	1505000	19895150	38330000	2,52	33,31	64,17	830	504963,5	6270
2021_2	65	65	0	0	5566000	0	0	100	0	0	400
2021_2	65	65	50000	150000	750000	5,26	15,79	78,95	100	300	300

Figura 12. Imatge dels 5 primers registres TablaMaestraF1-Extendida

Codi_postal	Servei_territorial_explotació	Provincia_explotació	Comarca_explotació	Municipio_explotació	Cadavers_Total_animals	Promedio_bloc1	Promedio_bloc2	Promedio_bloc3	Prom_punt_total
8630	Barcelona	Barcelona	Baix Llobregat	Abrera	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA
25651	Lleida	Lleida	Pallars Jussà	Abella de la Conca	NA	NA	NA	NA	NA

Figura 13. Imatge dels 5 primers registres de TablaMaestraF1-Extendida

Promedio_bloc1	Promedio_bloc2	Promedio_bloc3	Prom_punt_total	c4
NA	NA	NA	NA	4
NA	NA	NA	NA	3
NA	NA	NA	NA	3
NA	NA	NA	NA	3
NA	NA	NA	NA	4
NA	NA	NA	NA	4



Resum de les noves variables calculades

Aquí es descriuen les variables transformades que corresponen als tres fitxers agregats pel tipus d'antibiòtic que es va crear a l'enginyeria de variables.

Taula 9. Resum de les noves variables transformades en la fase de l'enginyeria de variables

Variables	Descripció	Fitxer del què deriven
Año_Trimestre	Concatenació de les variables Any i Trimestre	F1
Peso2	Càlcul del pes amb UR/Cens	F1/F2
mgConsumo_B	Consum de l'antibiòtic tipus B en mg	F3
mgConsumo_C	consum de l'antibiòtic tipus B en mg	F3
mgConsumo_D	consum de l'antibiòtic tipus B en mg	F3
pct_mgConsumo_B	percentatge de consum de l'antibiòtic tipus B	F3
pct_mgConsumo_C	percentatge de consum de l'antibiòtic tipus C	F3
pct_mgConsumo_D	percentatge de consum de l'antibiòtic tipus D	F3
CantidadSA_B	Quantitat d'antibiòtic de tipus B segons dosis	F3
CantidadSA_C	Quantitat d'antibiòtic tipus C segons la dosi	F3
CantidadSA_D	Quantitat d'antibiòtic de tipus D segons la dosi	F3
Cadavers_Total_animals	Nombre de cadàvers de porcins total per any i trimestre	Recollida de cadàvers
Promedio_Bloc1	Puntuació mitjana del bloc I per any i trimestre	Enquestes de bioseguretat
Promedio_Bloc2	Puntuació mitjana del bloc II per any i trimestre	Enquestes de bioseguretat
Promedio_Bloc3	Puntuació mitjana del bloc III per any i trimestre	Enquestes de bioseguretat
Prom_punt_Total	Puntuació mitjana total per any i trimestre	Enquestes de bioseguretat
C4	Classes les dades identificades i assignades al clustering	



Data mining

S'ha treballat amb tècniques factorials i biplots per extreure les relacions presents de les variables mitjançant un anàlisi multivariant de les dades i per comprendre en un mapa de reducció de dimensionalitat les relacions de les variables incidents en cas d'estudi.

ACP. Anàlisi de les relacions globals entre les variables

El primer experiment de l'ACP (Lebart, 1995) amb les variables numèriques de la taula resum F1 va mostrar una independència total de les variables UR i mgUR respecte a la resta. Per veure si els resultats estaven afectats per la mida de l'explotació (efecte grandària), es va decidir experimentar escalant les variables d'alguns dels experiments per la variable Cens per eliminar aquest efecte.

$$UR = Censo \times Peso$$

$$mgUR = \frac{mgTotal}{UR}$$

D'aquesta manera, l'anàlisi de components principals es va realitzar amb diferents variables numèriques dels fitxers d'F1, TablaMaestraF1 i TablaMaestraF1-Estesa segons si les dades estaven escalades pel Cens.

Taula 10. Resum dels 4 experiments de l'ACP, les dades i variables numèriques

Experiment	Dades	Variables numèriques	Justificació
1	F1 inicial	IdCz, Censo, UR, mgTotals, N_Prescripcions, mgUR, IR	Experiment inicial amb F1 subministrada per l'usuari final
2	F1 escalat	IdCz, UR, mgTotals, N_Prescripcions, mgUR, IR	El mateix experiment anterior, però fent un estudi per càpita dividint pel cens per eliminar el possible efecte grandària de les explotacions
3	TaulaMestraF1 amb consums absoluts sense escalar	IdCz, Censo, UR, mgTotals, N_Prescripcions, mgUR, IR, Año_Trimestre, Peso, mgConsumo_B, mgConsumo_B, mgConsumo_B.	Enriquiment dels experiments 1 i 2 en fer l'enginyeria de variables
4	TaulaMestraF1 amb les proporcions dels consums	IdCz, UR, mgTotals, N_Prescripcions, mgUR, IR, Año_Trimestre, Peso, pct_mgConsumo_B, pct_mgConsumo_B, pct_mgConsumo_B.	Per suggeriment de l'expert, l'experiment número 4 es va repetir canviant el consum d'antibiòtics absoluts per les proporcions

A3. Metodologia del model de governança i sostenibilitat

La metodologia que s'ha emprat per a l'elaboració del Pla de Governança i Sostenibilitat s'ha basat en 4 pilars fonamentalment. El primer ha estat l'anàlisi de la protecció de la propietat intel·lectual (existent i esperada), on s'ha treballat amb totes les institucions involucrades en el projecte per tal de registrar tota la propietat intel·lectual que s'ha aportat al projecte i tota la propietat intel·lectual que sorgirà del projecte. El segon, tercer i quart, han estat obtinguts mitjançant entrevistes amb els actors claus del sector (granges ramaderes, cooperatives i associacions del sector). Aquests tres últims pilars són la valoració general de l'algorisme, la valoració del Dataspase, i els suggeriments aportats pels entrevistats.

Anàlisi de la protecció de la propietat intel·lectual

L'objectiu principal de l'anàlisi és identificar, documentar i protegir la propietat intel·lectual associada al PAI. Aquest procés assegurarà la preservació del valor del projecte i la gestió adequada dels actius intangibles:

1. **Preservació del Valor del Projecte.** Mitjançant la identificació, documentació i protecció de la propietat intel·lectual, es preserva la inversió i els recursos dedicats al desenvolupament del projecte.
2. **Minimització de Riscos.** L'avaluació de la protecció de la propietat intel·lectual ajuda a minimitzar els riscos associats a la divulgació no autoritzada o l'ús indegut de la tecnologia desenvolupada. Aquest anàlisi permet prendre mesures preventives per protegir els actius intangibles del projecte.
3. **Negociació i Col·laboració amb Entitats Externes.** L'anàlisi de la propietat intel·lectual és crucial per a les negociacions i per establir acords clars sobre els drets de propietat. Aquesta transparència facilita la col·laboració i evita conflictes legals futurs.
4. **Compliment Legal.** L'atenció a la protecció de la propietat intel·lectual assegura que el projecte compleixi amb les lleis i regulacions pertinents.
5. **Facilitació de Futurs Desenvolupaments.** L'anàlisi de la propietat intel·lectual proporciona una base per decidir sobre estratègies de llicència, comercialització o altres formes de valorització de la tecnologia.

6. **Alineació amb Objectius Estratègics.** L'anàlisi està alineat amb els objectius estratègics del projecte, assegurant que les decisions relacionades amb la propietat intel·lectual estiguin en consonància amb les fites a llarg termini.

Els passos metodològics són els següents:

- **Identificació de Tecnologies i Elements de Propietat Intel·lectual:**
 - Creació d'una llista exhaustiva de totes les tecnologies i elements de propietat intel·lectual relacionats amb el projecte.
 - Especificació del nom de cada tecnologia o element.
- **Autoria:**
 - Identificació dels autors de cada tecnologia o element.
 - Detall de la contribució específica de cada autor al desenvolupament de la propietat intel·lectual.
- **Titularitat:**
 - Determinació de la titularitat de cada tecnologia o element.
 - Especificació de les entitats o persones que són els titulars de la propietat intel·lectual.
- **Condicions d'Ús/Explotació:**
 - Establiment de les condicions d'ús i explotació de cada tecnologia o element.
 - Inclusió de les restriccions o llicències associades amb la propietat intel·lectual.
- **Previsió del Foreground:**
 - Creació d'una taula similar per a les tecnologies o elements de propietat intel·lectual que es preveu que es generin en el futur (foreground).
 - Inclusió de les mateixes categories: Nom de la Tecnologia, Autors, Titularitat, Condicions d'Ús/Explotació.

Valoració general de la solució (algoritme)

Presentació de la solució

A partir de la instal·lació d'uns sensors/càmeres en un corral de porcs, que faciliten la posició de cada porc en cada moment del temps, el software fa un seguiment individualitzat de l'estat dels animals (benestar, malalties...) i del seu comportament (temps dedicat a beure, a menjar, a caminar, etc.). Això permet treure una sèrie d'estadístiques o KPIs que poden ser d'utilitat pels propietaris/gestors de les granges ramaderes a diferents nivells (optimització d'espais, productivitat, control i gestió de malalties, compliment de normatives, obtenció de certificacions o segells de qualitat, etc.).



Valoració

- **Valoració i nivell d'interès general per les explotacions ramaderes:** Avantatges i inconvenients, potencial i utilitats.
- **Valoració de la informació:** Valoració de les estadístiques i KPIs que es proporcionen. També altres informacions que poden ser interessants (a curt, mig i llarg termini).
- **Valoració de la interface:** Importància de la personalització vs. que sigui estàndard.
- **Valoració del pipeline (procés des que una explotació vol participar fins l'obtenció de resultats):** Instal·lació de càmeres (valoració i implicacions), processament de les dades i obtenció dels KPIs, timings.
- **Valoració global de la solució:** Elements més rellevants, perfil d'explotacions que poden estar interessades, permeables a la introducció de la tecnologia. Requisits i condicions. Preu que s'estaria disposat a pagar.
- **Valoració de l'espai de dades compartides (espai de dades /dataspace):**
 - **Governança de les dades.** Es planteja el cas en que un usuari vulgui "contractar" el servei, per tant hauria de signar un contracte que li permetria tindre accés al servidor/núvol on es gestionen les dades. Per tal de poder fer ús de l'algoritme, l'usuari ha de proporcionar les imatges necessàries o si no té els medis, instal·lar els sensors/càmeres al recinte i les imatges es pujaran automàticament al servidor. A partir d'aquí, l'empresa obtindrà els KPIs o estadístiques que s'hagin definit. Amb la contractació del dataspace, les explotacions (contractin o no el servei de l'algoritme), poden accedir al banc d'imatges/vídeos i dades d'altres explotacions per a poder treure informació rellevant. Alhora, totes les dades aportades serveixen per anar entrenant l'algoritme. D'aquesta manera el model té més prestacions i és més estable.
 - **Valoració de l'espai de dades.** Avantatges i inconvenients, suggeriments de millora.
 - **Utilitats de l'accés de les dades compartides de l'espai de dades.** Per a explotacions ramaderes, per a cooperatives i associacions
 - **Accés per pagament vs. accés públic i gratuït.** La possibilitat de que el servei sigui mitjançant un únic pagament o pagaments recurrents, versus la possibilitat de que sigui un servei gratuït d'accés públic.

Suggeriments i conclusions

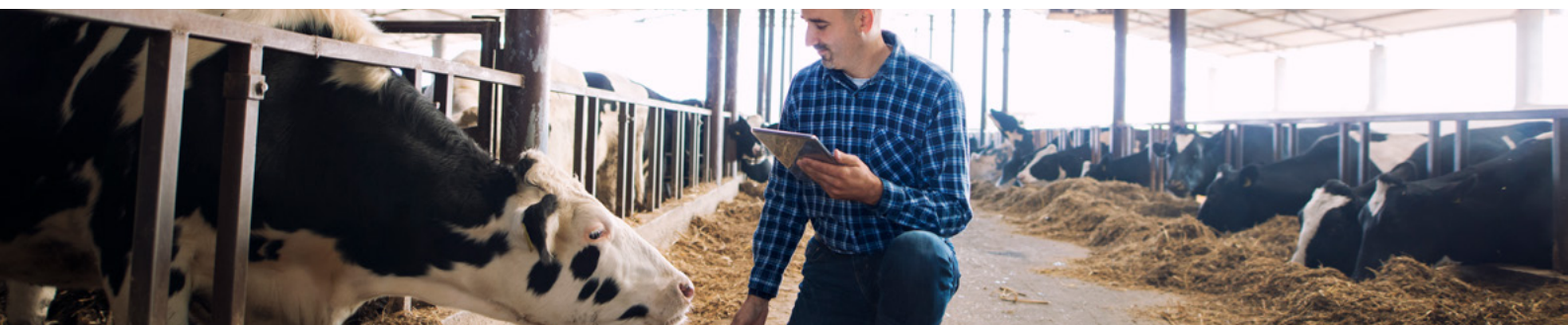
A través d'una metodologia basada en quatre pilars essencials, hem desenvolupat el Pla de Governança i Sostenibilitat per al PAI sobre el sector agropecuari. Des de l'anàlisi de la propietat intel·lectual fins a les interaccions amb els actors clau, cada fase ha estat dissenyada per establir unes bases sòlides per a futurs desenvolupaments. Amb aquest enfocament, aspirem a protegir la nostra propietat intel·lectual i crear una solució algorítmica amb un impacte positiu i sostenible.

Suggeriments

1. **Optimització del Procediment d'Anàlisi de Propietat Intel·lectual.** Explorar eines o tecnologies que puguin agilitzar la recopilació de dades de propietat intel·lectual, com a plataformes de gestió de propietat intel·lectual o sistemes automatitzats de registre.
2. **Ampliació de la Participació.** Incentivar la participació d'un nombre més gran d'institucions i entitats en l'anàlisi de la propietat intel·lectual, afavoriria una visió més completa i diversificada.
3. **Desenvolupament Continu del Foreground.** L'establiment d'un marc per al seguiment continu del foreground, actualitzant les taules a mesura que es generin noves tecnologies o elements de propietat intel·lectual.
4. **Fomentar la Col·laboració en l'Espai de Dades Compartides.** La implementació d'iniciatives per fomentar la col·laboració i la contribució activa d'empreses i explotacions ramaderes a l'espai de Dataspace, milloraria la qualitat i quantitat de dades disponibles.
5. **Garantir Transparència en la Governança de Dades.** El reforç dels mecanismes de governança de dades, asseguraria que els usuaris comprenquin i estiguin d'acord amb les polítiques de l'ús de les seves dades en el context del Dataspace.

Conclusions

1. **Valor Afegit de l'Anàlisi de Propietat Intel·lectual.** La metodologia d'anàlisi de propietat intel·lectual és fonamental per a la preservació del valor del projecte i la mitigació de riscos, proporcionant una base sòlida per a la gestió eficient dels actius intangibles.
2. **Importància de les Entrevistes Sectorials.** Les entrevistes amb actors claus del sector enriqueixen la comprensió global del projecte, afegint noves perspectives valuoses que contribueixen a la valoració de l'algoritme, el Dataspace i altres aspectes crítics.
3. **Solució Algorítmica Amb Repercussions Positives.** La solució algorítmica presentada ofereix beneficis significatius per a les explotacions ramaderes, millorant la gestió i el benestar dels animals, i generant estadístiques rellevants que poden impulsar la productivitat i la qualitat dels productes.
4. **Dataspace com a Recurs Estratègic.** L'espai de Dataspace és un recurs estratègic, que permet l'accés a dades compartides, contribuint al desenvolupament del model algorítmic i oferint utilitats tant per a explotacions ramaderes com per a altres actors del sector.
5. **Perspectives Futures.** Les perspectives futures inclouen l'exploració de noves tecnologies, la millora contínua dels processos i la promoció d'un entorn col·laboratiu que afavoreixi la innovació i el desenvolupament sostenible en el sector de la ramaderia.



Referències



- Catalunya, G. d. (sense data). *Les agendas compartides de la RIS3CAT 2030*. Recollit de <https://fonseuropeus.gencat.cat/ca/ris3cat/2030/agendes-compartides/index.html>
- Computador, C. d. (2022). Recollit de <https://cdn.eurecat.org//PDF/CIDAI/PoC/POC-CVC-110523.pdf>
- Gaia-X. (sense data). *Pontus-X*. Recollit de Streamlined interoperability accross data ecosystems: <https://www.pontus-x.eu>
- Fayyad, U. P.-S. (1996). *From Data Mining to Knowledge Discovery in Databases*. 17.
- Gibert K, S.-M. M. (2016). A survey on pre-processing techniques: relevant issues in the context of environmental data mining. 29(6).
- Gibert, K. I.-M.-R. (2018). Which method to use? An assessment of data mining methods in Environmental Data Science. *Environmental modelling & software*, 110, 3-27.
- Lebart, L. M. (1995). *Statistique exploratoire multidimensionnelle (Vol. 3)*. Paris: Dunod.
- RStudio_Team. (2020). *RStudio: Integrated Development for R*. Recollit de <http://www.rstudio.com/>
- Gibert, K. G.-R.-S. (2008). The role of KDD Support-Interpretation tools in the conceptualization of medical profiles: An application to neurorehabilitation. *Acta Informatica Medica*, 16(4), 178.

Promotors



Membres

