CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Presentació del Projecte d'Impacte

"BusCA – Anàlisi de les alteracions del servei de bus metropolità "

Socis promotors

Generalitat de Catalunya · Barcelona · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación · conversia compliance is our business · CVC Centre de Visió per Computador · eurecat Centre Tecnològic de Catalunya · forvis mazars & infini Digital Efficiency · i2cat · softeng Maximize the cloud. Securely. · UPC Universitat Politècnica de Catalunya BarcelonaTech · IDEAI Intelligent Data Science and Artificial Intelligence Research Center

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Presenta l'acte



Mireya Saludes

Responsable de negoci i d'ecosistemes
CIDAI

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Benvinguda institucional



Ramon Pruneda

CTO

AMB Informació

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Benvinguda institucional



Sergi Masip

Coordinador de l'Estratègia Catalunya IA 2030 de la Secretaria de Polítiques Digitals

Generalitat de Catalunya

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

AGENDA

10:10 h | **El repte de la gestió de les alteracions en el servei d'autobús metropolità**

- **Ramon Campoy,** especialista de Gestió de la Informació del Transport d'**AMB Informació**

10:25 h | **Desenvolupament del projecte BusCA: anàlisi de dades i modelització predictiva**

- **Francisca Rosell,** investigadora de la Unitat Applied Artificial Intelligence (AAI) d'**Eurecat**

11:00 h | **Resultats, recomanacions i viabilitat d'un sistema d'alertes**

- **Javier Fernández,** especialista de Dades d'**AMB Informació**

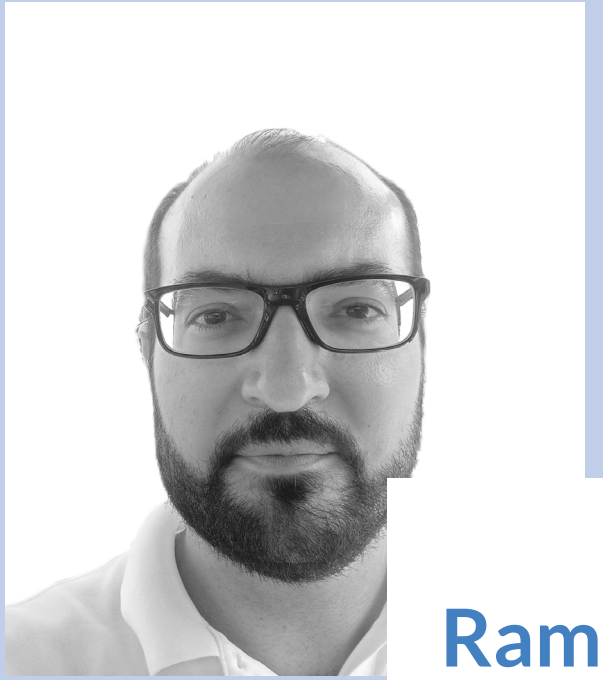
11:10 h | **Torn obert de preguntes**

11:30 h | **Cafè & Networking**

11:45 h | **Fi de l'acte**

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

El repte de la gestió de les alteracions en el servei d'autobús metropolità



Ramon Campoy
Especialista Gestió de la Informació del Transport
AMB Informació

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

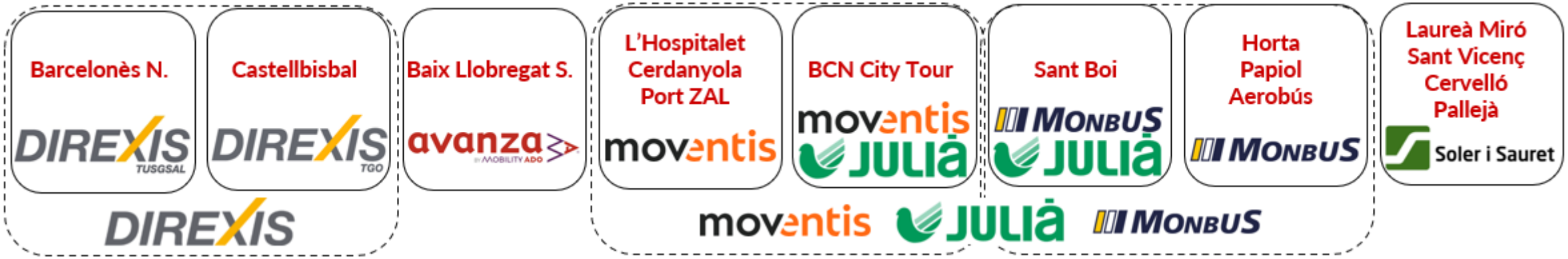
Gestió de les alteracions al Bus Metropolità

1. L'ecosistema del Bus Metropolità
2. La gestió de les alteracions
3. Un calendari d'alteracions anuals?
4. El projecte BusCA

L'ecosistema del Bus Metropolità

Contractes, operadores

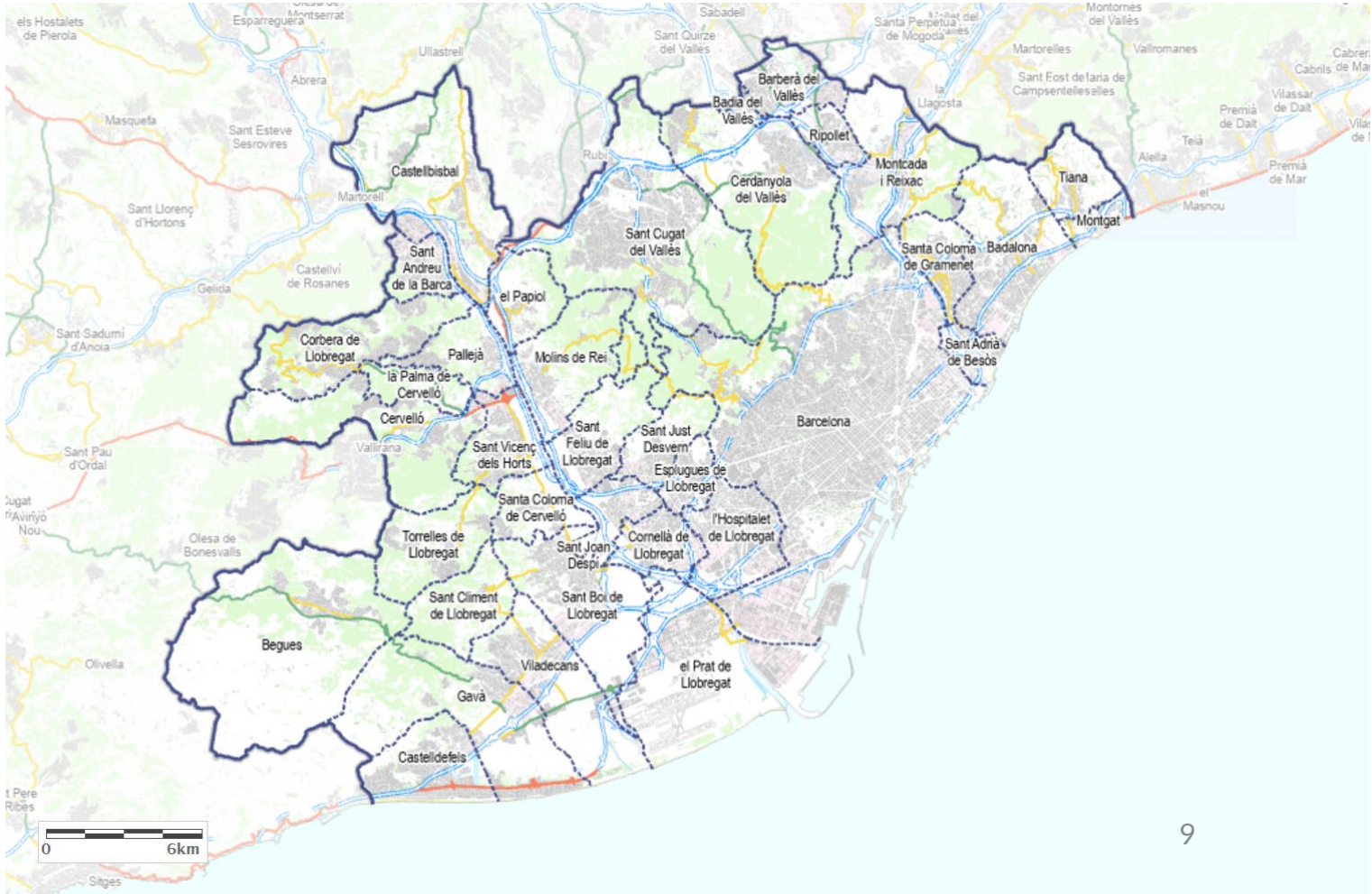
- ✓ Serveis de gestió indirecta de Bus Metropolità
- ✓ 15 contractes
- ✓ ~8 operadores
- ✓ Servei diürn i nocturn
- ✓ 125.000.000+ viatgers anuals
- ✓ 147 línies → 119 diürnes, 24 nocturnes, 4 especials
- ✓ 3.000+ parades
- ✓ **2.000+ incidències anuals**



L'ecosistema del Bus Metropolità

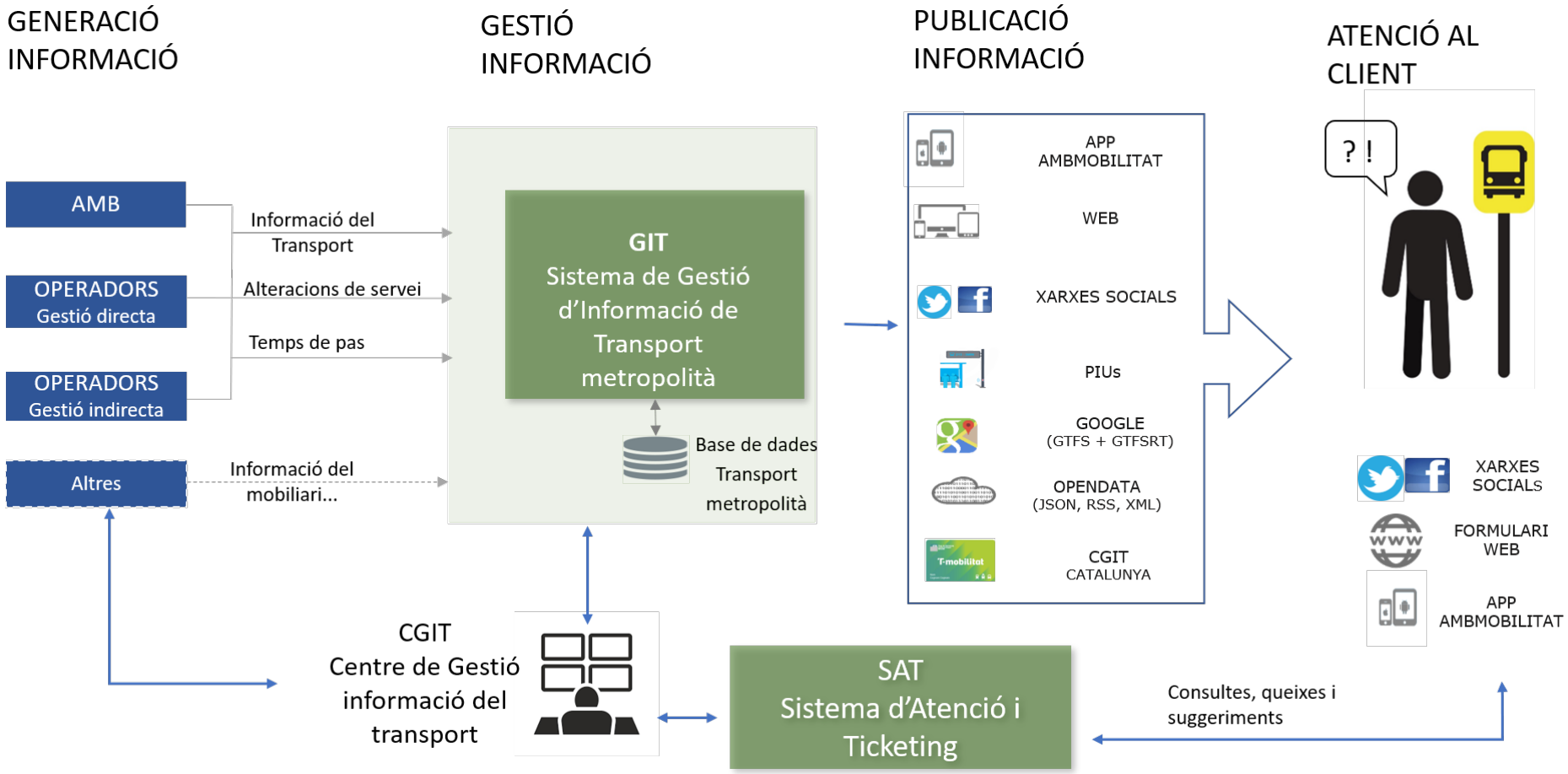
Municipis

- ✓ AMB: 36 municipis
- ✓ Bus Metropolità: en 28 municipis l'AMB és la titular del servei de bus + 3 no titular + 2 fora de l'AMB
- ✓ Agents:
  - ✓ Els ajuntaments (policia local, tècnics...)
  - ✓ Les operadores: coneixedores de la realitat local, del dia a dia sobre el terreny
  - ✓ AMB/AMB Informació (CGIT): gestió de la informació del transport



La gestió de les alteracions

AMB Informació i el CGIT



CIDAI Centre of Innovation for Data tech and Artificial Intelligence

i2cat

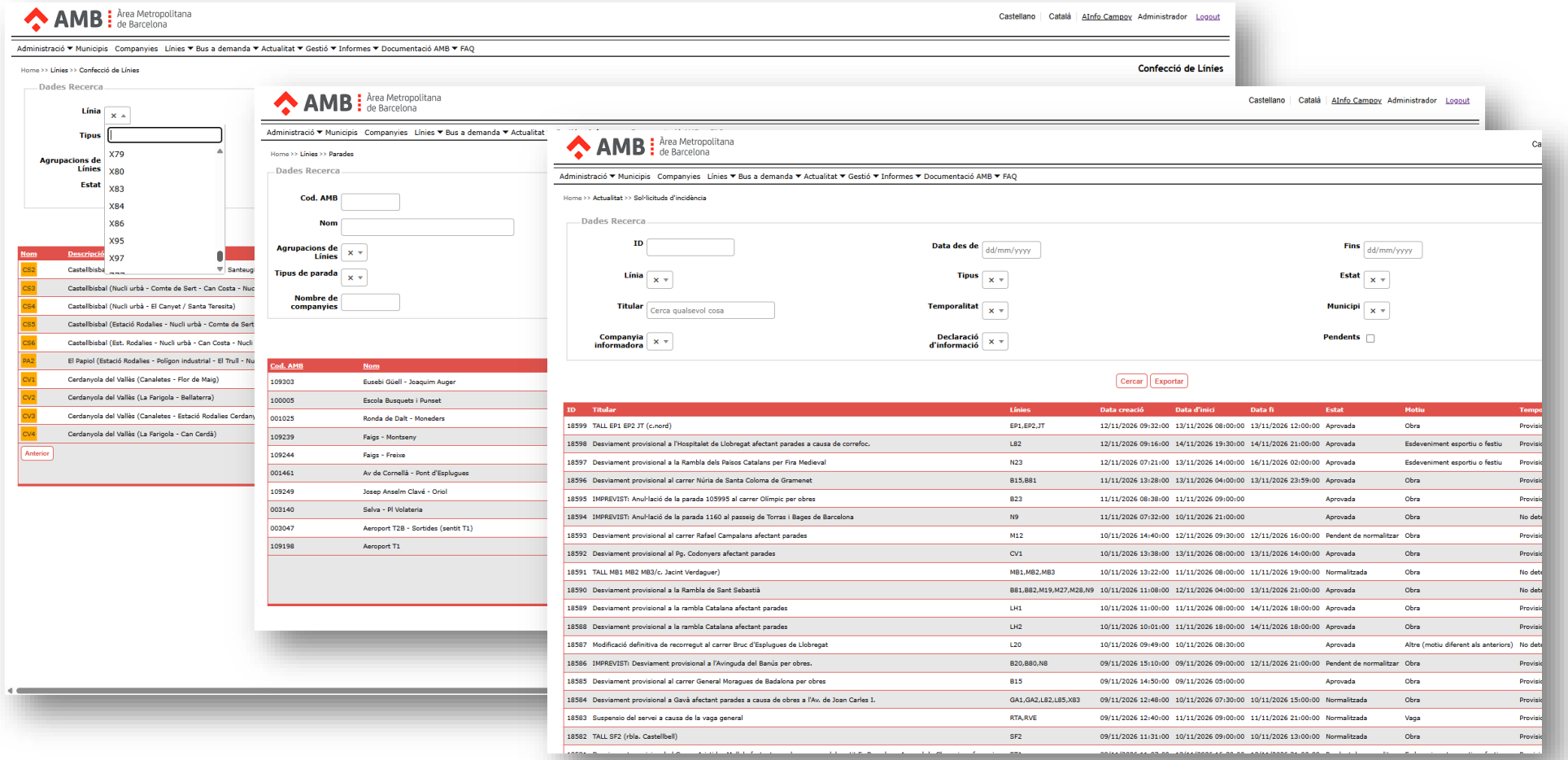
AMB INFORMACIÓ

La gestió de les alteracions

Gestió de la informació. Operativa

Manteniment de:

- ✓ BDD parades
- ✓ BDD mobiliari
- ✓ BDD línies
- ✓ BDD incidències
- ✓ Etc.



eurecat

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

i2cat

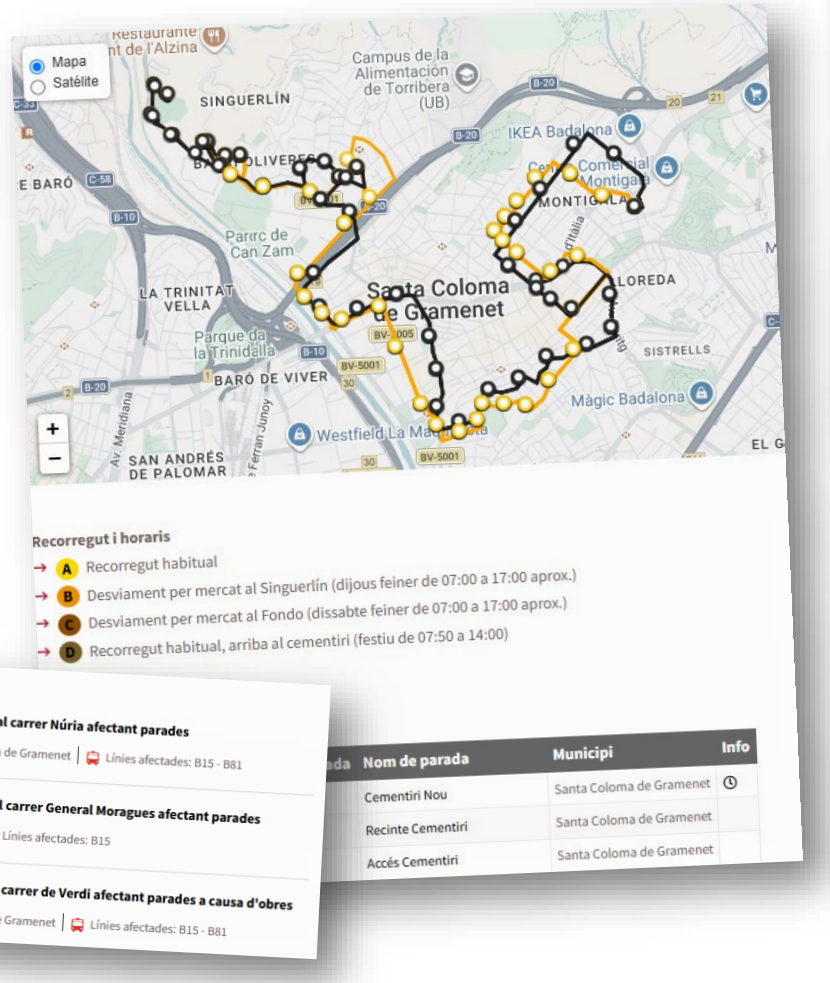
AMB INFORMACIÓ

La gestió de les alteracions

Gestió de la informació. Operativa

Publicació a:

- ✓ Web
- ✓ App
- ✓ XXSS
- ✓ PIUs
- ✓ GMaps
- ✓ Altres









eurecat

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

i2cat

AMB INFORMACIÓ

La gestió de les alteracions

Publicació de la informació. Què esperen les usuàries?

- ✓ Informació completa
- ✓ Informació exacta
- ✓ Informació actualitzada

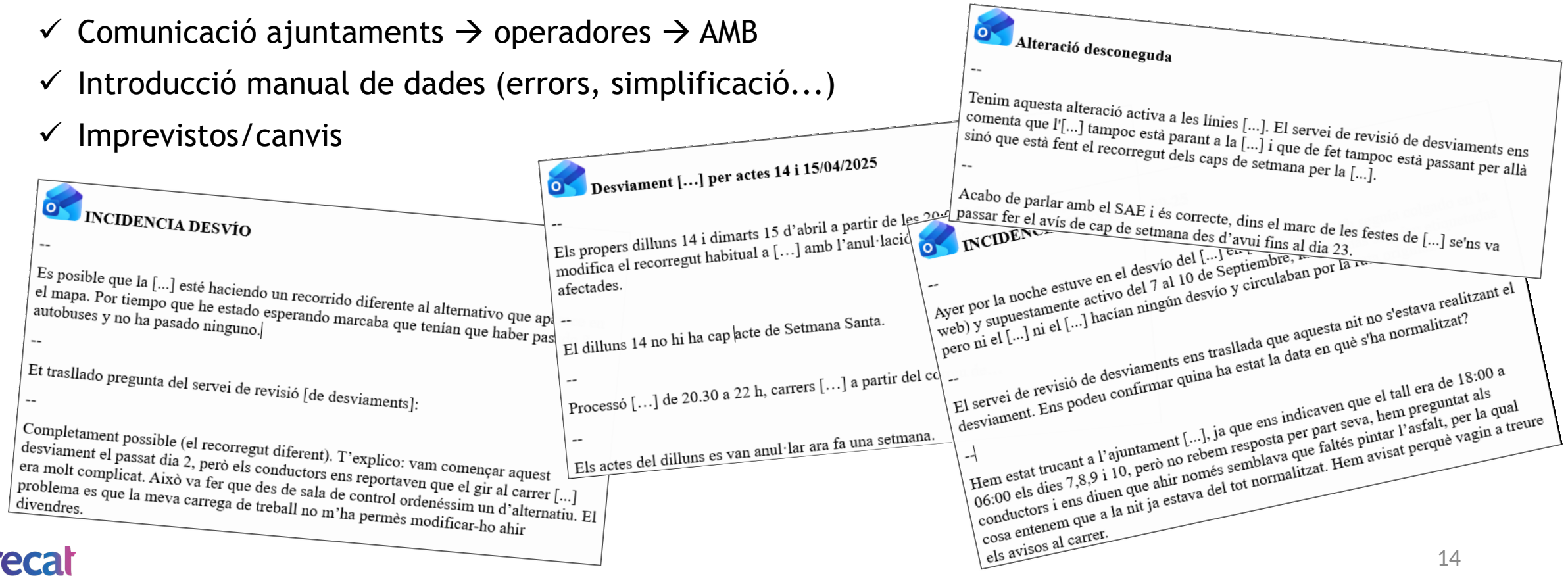


eurecat

La gestió de les alteracions

Generació de la informació. Quins reptes ens trobem?

- ✓ Recursos ajuntaments/operadores/AMB (recursos, disponibilitat...)
- ✓ Comunicació ajuntaments → operadores → AMB
- ✓ Introducció manual de dades (errors, simplificació...)
- ✓ Imprevistos/canvis



INCIDENCIA DESVÍO

--

Es posible que la [...] esté haciendo un recorrido diferente al alternativo que ap
el mapa. Por tiempo que he estado esperando marcaba que tenían que haber pas
autobuses y no ha pasado ninguno.

--

Et trasllado pregunta del servei de revisió [de desviaments]:

--

Completament possible (el recorregut diferent). T'explico: vam començar aquest
desviament el passat dia 2, però els conductors ens reportaven que el gir al carrer [...]
era molt complicat. Això va fer que des de sala de control ordenéssim un d'alternatiu. El
problema es que la meva carrega de treball no m'ha permès modificar-ho ahir
divendres.

Desviament [...] per actes 14 i 15/04/2025

--

Els propers dilluns 14 i dimarts 15 d'abril a partir de les 20
modifica el recorregut habitual a [...] amb l'anul·lació
afectades.

--

El dilluns 14 no hi ha cap acte de Setmana Santa.

--

Processó [...] de 20.30 a 22 h, carrers [...] a partir del cc

--

Els actes del dilluns es van anul·lar ara fa una setmana.

Alteració desconeguda

--

Tenim aquesta alteració activa a les línies [...]. El servei de revisió de desviaments ens
comenta que l'[...] tampoc està parant a la [...] i que de fet tampoc està passant per allà
sinó que està fent el recorregut dels caps de setmana per la [...].

--

Acabo de parlar amb el SAE i és correcte, dins el marc de les festes de [...] se'ns va
passar fer el avís de cap de setmana des d'avui fins al dia 23.

INCIDENC

--

Ayer por la noche estuve en el desvío del [...]
web) y supuestamente activo del 7 al 10 de Septiembre,
pero ni el [...] ni el [...] hacían ningún desvío y circulaban por la

--

El servei de revisió de desviaments ens trasllada que aquesta nit no s'estava realitzant el
desviament. Ens podeu confirmar quina ha estat la data en què s'ha normalitzat?

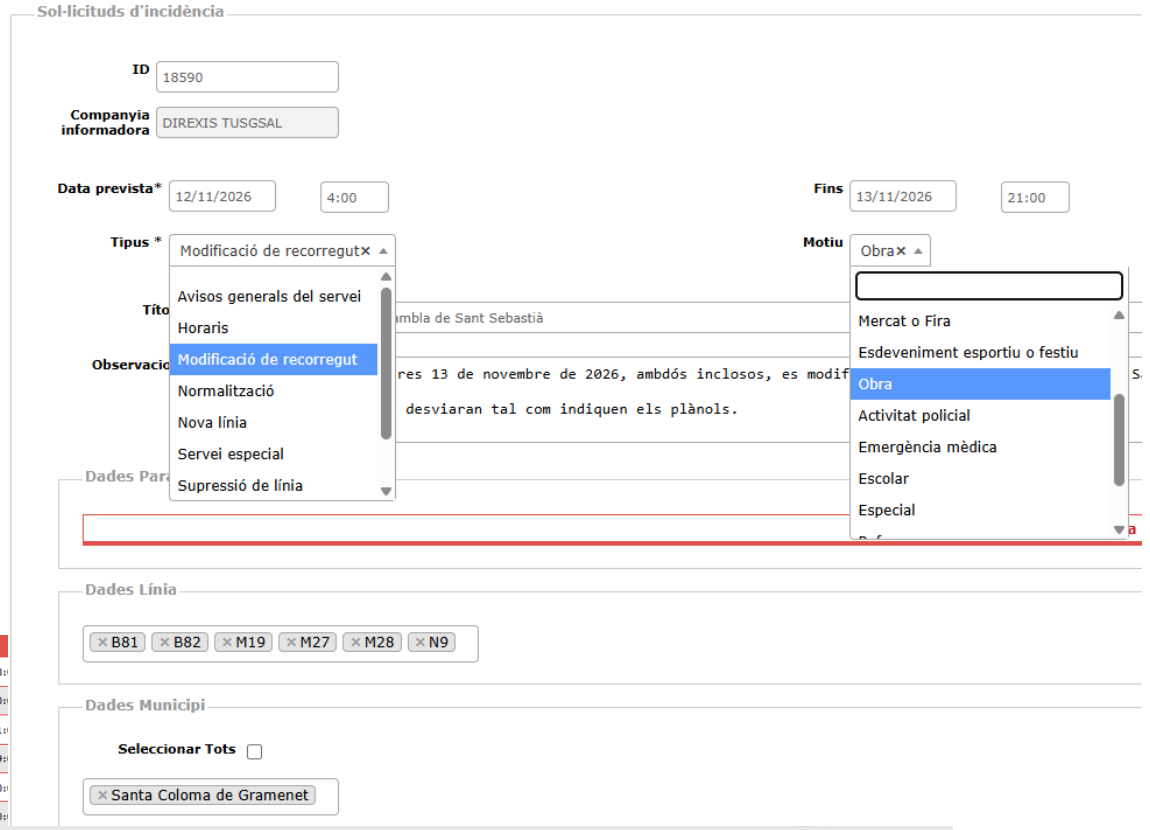
--

Hem estat trucant a l'ajuntament [...], ja que ens indicaven que el tall era de 18:00 a
06:00 els dies 7,8,9 i 10, però no rebem resposta per part seva, hem preguntat als
conductors i ens diuen que ahir només semblava que faltés pintar l'asfalt, per la qual
cosa entenem que a la nit ja estava del tot normalitzat. Hem avisat perquè vagin a treure
els avisos al carrer.

La gestió de les alteracions

Gestió de la informació. Quines oportunitats tenim?

- ✓ BDD d'alteracions/incidències amb camps d'informació estandarditzats
- ✓ Dades històriques des de 2006, més representatives quant més actuals
- ✓ **Possibilitat de fer servir les dades històriques per predir el futur**
- ✓ Però: les línies canvien, les alteracions van evolucionant, manca part de les dades o els mateixos criteris...

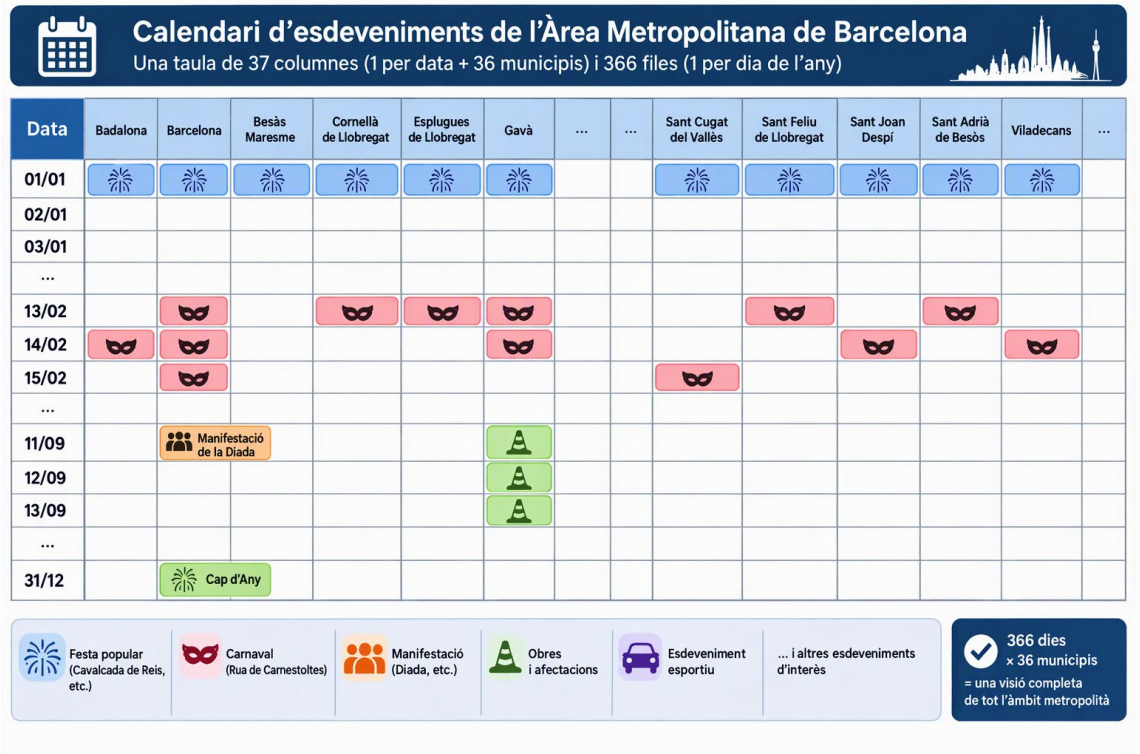


ID	Titular	Línies	Data creació
18590	Desviament provisional a la Rambla de Sant Sebastià	B81,B82,M19,M27,M28,N9	10/11/2026 11:08
18589	Desviament provisional a la rambla Catalana afectant parades	LH1	10/11/2026 11:00
18588	Desviament provisional a la rambla Catalana afectant parades	LH2	10/11/2026 10:01
18587	Modificació definitiva de recorregut al carrer Bruc d'Esplugues de Llobregat	L20	10/11/2026 09:49
18586	IMPREVIST: Desviament provisional a l'Avinguda del Banús per obres.	B20,B80,N8	09/11/2026 15:10
18585	Desviament provisional al carrer General Moragues de Badalona per obres	B15	09/11/2026 14:50
18584	Desviament provisional a Gavà afectant parades a causa de obres a l'Av. de Joan Carles I.	GA1,GA2,L82,L85,X83	09/11/2026 12:48:00	10/11/2026 07:30:00	10/11/2026 15:00:00	Pendent de normalitzar	Obra	Provisional	Avanza
18583	Suspensio del servei a causa de la vaga general	RTA,RVE	09/11/2026 12:40:00	11/11/2026 09:00:00	11/11/2026 21:00:00	Aprovada	Vaga	Provisional	BARCELONA CITY TOUR
18582	TALL SF2 (rbla. Castellbell)	SF2	09/11/2026 11:31:00	10/11/2026 09:00:00	10/11/2026 13:00:00	Pendent de normalitzar	Obra	Provisional	Soler i Sauret

Un calendari d'alteracions anuals?

Per què podria ser útil, i com?

- ✓ Suport a operadores + CGIT
- ✓ Classificar per municipi (alguns a una o més operadores)
- ✓ A Barcelona subdividir per zones (corredor Diagonal, corredor Gran Via-centre, línies del Port, línies d'Horta/Gràcia i línies que entren per tota la meitat est)
- ✓ Separar dia/nit (servei nocturn concentrat en 2 contractes)



Un calendari d'alteracions anuals?

Esdeveniments recurrents + esdeveniments puntuals

- **Exemples d'esdeveniments puntuals:**
  - Esdeveniments esportius o festius: Tour de France, visita del Papa...
  - Manifestacions: vaga general, per conflictes sectorials (transport públic, taxi, agricultura...), després d'esdeveniments de gran importància (inici guerra, temes polítics...)

- **Fonts de dades:**
  - Webs dels ajuntaments, de l'AMB, de mitjans de comunicació locals, dels organitzadors esportius, de festes a Catalunya, de sindicats, associacions... per manifestacions, de notícies per actualitat global (lligat amb manifestacions), del temps/meteorologia (lligat amb curses/obres que es cancel·len)...
  - Calendaris oficials de dies festius anuals
  - Xarxa social X
  - Cercadors web en general

- **Model d'IA generativa**

Un calendari d'alteracions anuals?

Esdeveniments recurrents

- **Exemples d'esdeveniments anuals:**
  - Esdeveniments esportius o festius: Cavalcades de Reis, Desmuntatge llums de Nadal, Rues de Carnestoltes, Volta a Catalunya, Cursa Corte Inglés, curses/cercaviles/altres de festes majors, curses de Sant Silvestre, etc.
  - Obres: túnels que es tallen per manteniment setmanal/mensual, carrers "habituals" per manteniment recurrent...
  - Mercats o fires: fires a la zona de Pl. Espanya, fires municipals...
  - Manifestacions: 8 de març, 1 de maig, 11 de setembre, 12 d'octubre (afectació reduïda)

- **Fonts de dades :**
  - BDD alteracions

- **Model d'IA tradicional (Machine Learning)**

El projecte BusCA

BusCA (Bus Control d'Alteracions): avaluació i predicció de les interrupcions del transport públic

- Explotació de les BDD de sol·licituds d'incidència + incidències publicades
- Incorporació de dades externes? BDD de festius municipals
- Model d'IA tradicional. Cerca de "patrons" històrics
- Predicció: quines alteracions es produiran, quan, a on (a quins municipis), etc.

El projecte BusCA sorgeix per donar resposta a la necessitat de predir alteracions futures al servei de Bus Metropolità per tal de millorar la gestió operativa i la informació a les usuàries

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Desenvolupament del projecte BusCA: anàlisi de dades i modelització predictiva



Francisca Rosell
Investigadora de la Unitat Tecnològica d'Intel·ligència Artificial Aplicada
EURECAT

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

BusCA
Anàlisi de les Alteracions en el Transport Públic

1 Context i objectius del projecte
2 Fonts de dades utilitzades
3 Anàlisi exploratòria / EDA
4 Modelització 1: predicció incidència si/no
5 Modelització 2: predicció del tipus d'incidència
6 Conclusions tècniques

Context i objectius del projecte

Context

- AMB disposa d'una plataforma on es recullen dades relacionades sobre possibles alteracions en els serveis dels autobusos de l'àrea metropolitana (no inclou els serveis de TMB) a mig-curt termini
- Operadors i gestors AMB introdueixen i validen la informació
- Un cop validades les dades introduïdes, les afectacions als serveis es publiquen als sistemes d'informació per als usuaris

Objectius

- **Anàlisi de les afectacions dels serveis** i el seu impacte a la xarxa de transport metropolità
  - ✓ Classificar i tipificar les diferents alteracions, detectant periodicitats i patrons
  - ✓ Detectar i analitzar altres fonts de dades que puguin ajudar a preveure alteracions futures
- **Prova de concepte**, segons els resultats de l'anàlisi previ:
  - ✓ Estudi d'un possible model de predicció per les alteracions que són conseqüència d'activitats planificades
  - ✓ Estudi d'un sistema d'alertes per als gestors /operadors per informar de les possibles afectacions en el servei

Fonts de dades utilitzades

Contingut

- ✓ Data set d'incidències operatives (2022-2025)
  - Informació **heterogènia**
    - **Incidències a nivell de línia**
    - **Incidències a nivell de parada**
  - Contingut per 38 municipis

- ✓ Open Data
  - Informació d'esdeveniments d'alguns municipis, principalment del municipi de Barcelona
  - Calendaris municipals

Anàlisi exploratòria / EDA

Dataset d'incidències

- Cobertura temporal: **2022-2025, 4266 registres data -->** cal incloure les 'no incidències'
- Abast territorial**: 38 municipis**
- Alta concentració: **46%** d'**incidències** a **5 municipis**
- Infraestructura: **115 línies** amb més de **2.700 parades**
- Neteja necessària per **elevat percentatge de nuls**

Anàlisi exploratòria / EDA

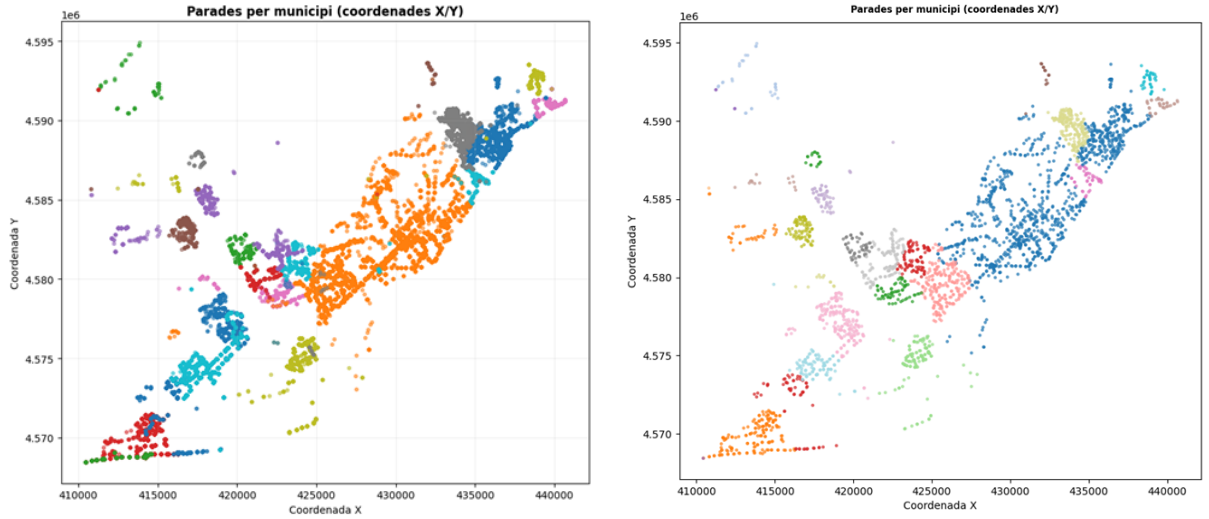
Neteja i consistència temporal

Correcció de dates e inconsistències:

- DataInici -> data registre
- DataFi -> data validació
- Durades negatives -> DataFi<DataInici

Assegurar coherència temporal i evitar errades en variables de duració

Normalització territorial



- Estandardització nom municipis
- Correcció d'errors de codificació
- Revisió de coordenades (dades geoespacials)

Anàlisi exploratòria / EDA

Reclassificació d'incidències

Reducció de categories poc informatives:

- 10,86% de nuls
- Categories genèriques (Desconeguda, especial, policial)...

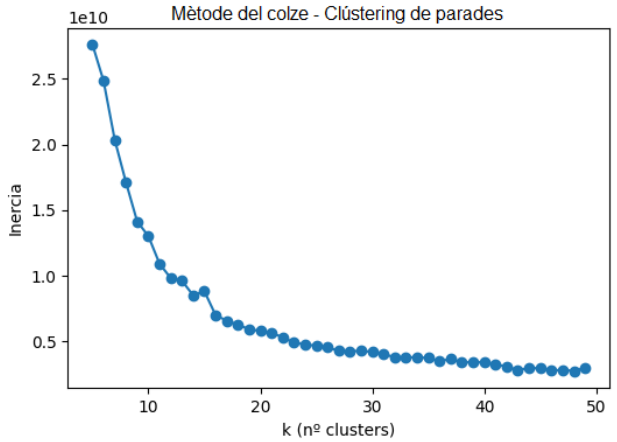
Solució:

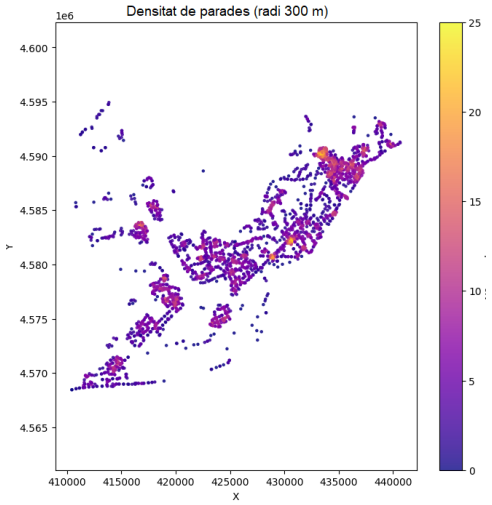
- Reclassificació amb el suport del camp de text 'Observacions'

Enginyeria de variables

Creació de variables per donar context als models:

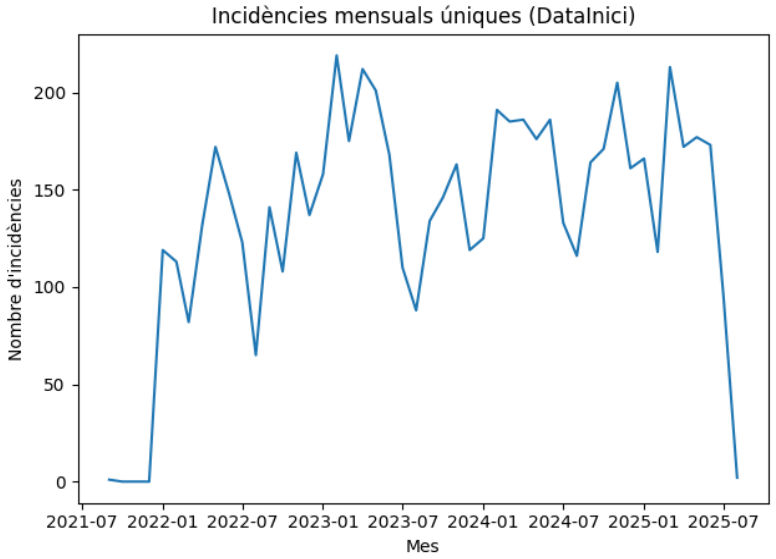
- Temporals (any, mes, dia, setmana...)
- Calendaris i festius (autonòmics o locals, dia laborable, pont...)
- Espacials (clustering de parades, densitat de xarxa...)
- Històriques (freqüència d'incidències per parada, mitjanes mòbils, EMAs..)
- Tipologia d'incidència

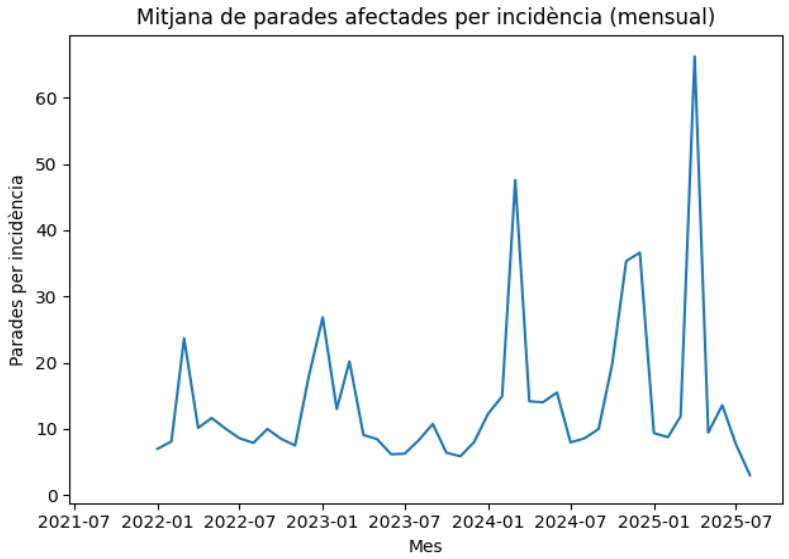




Anàlisi exploratòria / EDA

Visualització general de la sèrie temporal

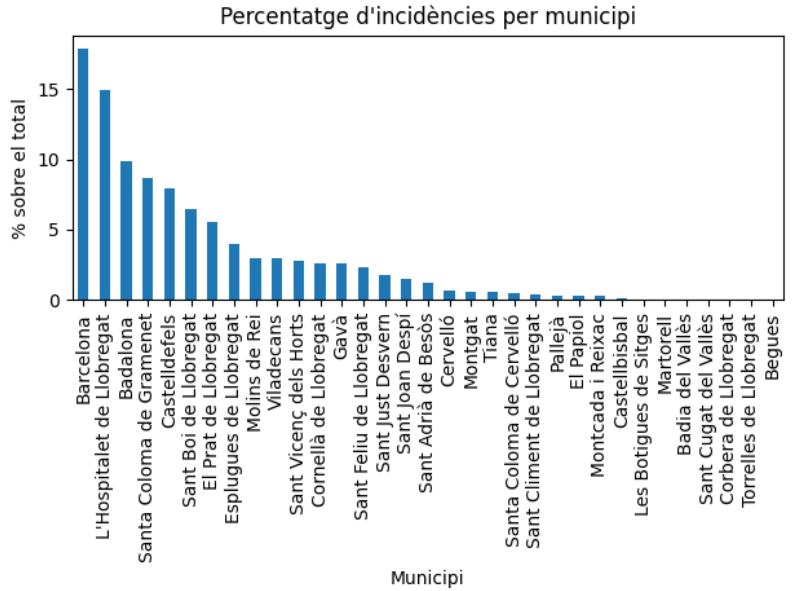


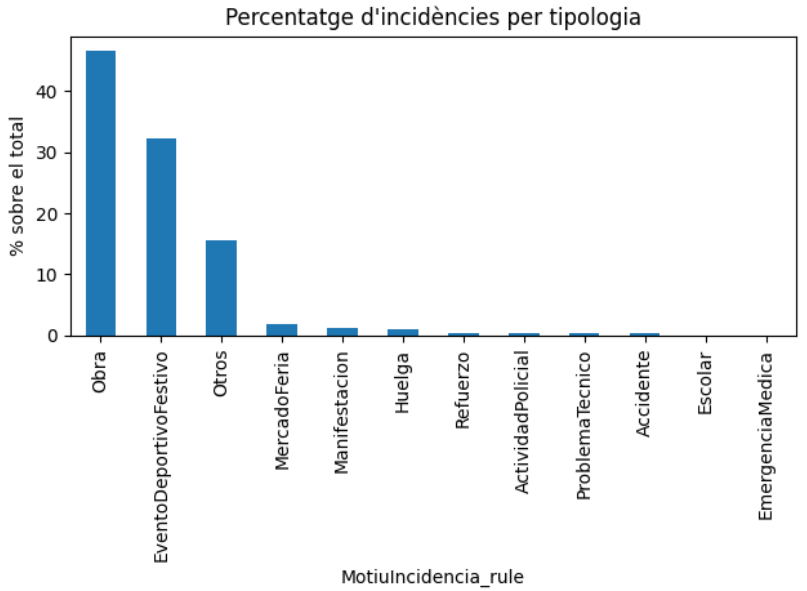


- Increment progressiu del volum d'alteracions
- Pics puntuals en determinats mesos (inici d'any)
- Variable mensual rellevant, sense estacionalitat regular

Anàlisi exploratòria / EDA

Distribució de les incidències



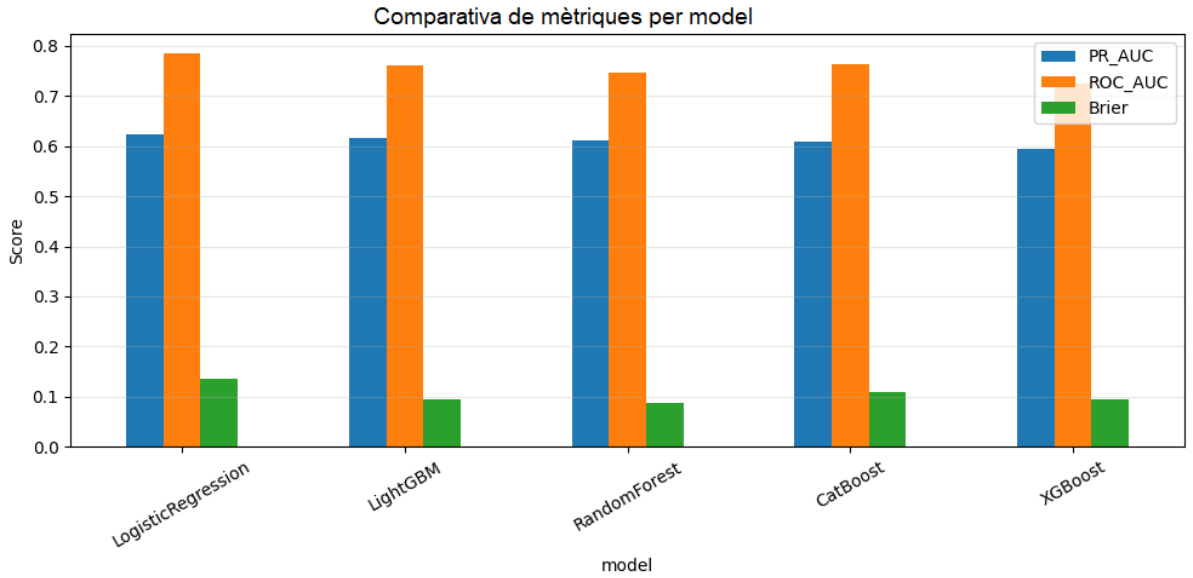


- Pocs municipis concentren gran quantitat de les incidències
- Obra concentra el 45% del total d'incidències, esdeveniment el 30% i altres el 15%

Modelització 1: predicció incidència si/no

Selecció i validació del model

- ✓ Es comparen els models amb i sense Open Data:
  - Logistic Regression
  - Random Forest
  - LightGBM
  - CatBoost
  - XGBoost



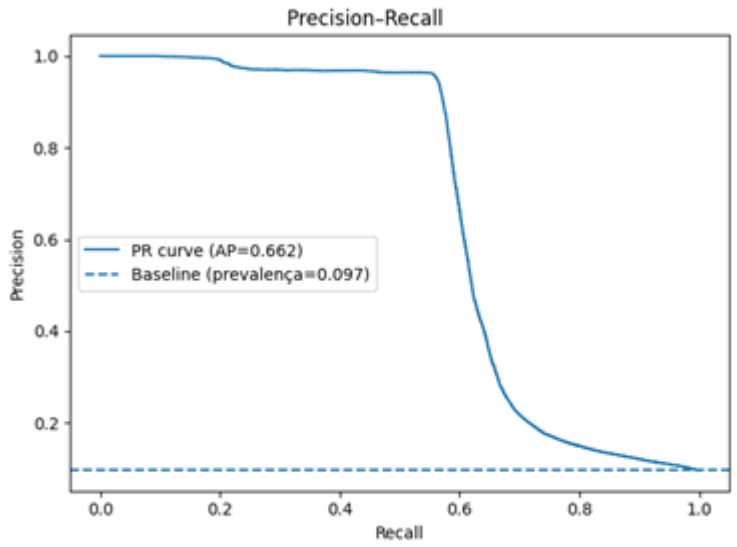
- ✓ Open Data empitjora el rendiment del model, principalment perquè una parts dels esdeveniments no disposen d'assignació espacial, introduint soroll al data set.
- ✓ LightGBM model seleccionat per les raons:
  - Rendiment equilibrat
  - Alta escalabilitat
  - Millor comportament en gran volum de dades

Modelització 1: predicció incidència si/no

Model final

- ✓ Rendiment final:
  - PR-AUC en test 0.66 (enfront a prevalença 0.10, és a dir 6.6 cops millor que aleatori)

- ✓ Interpretació:
  - Identifica incidències d'alta probabilitat amb gran precisió.
  - Funciona molt bé en municipis amb molta senyal
  - Menys capacitat en incidències esporàdiques o amb baixa freqüència

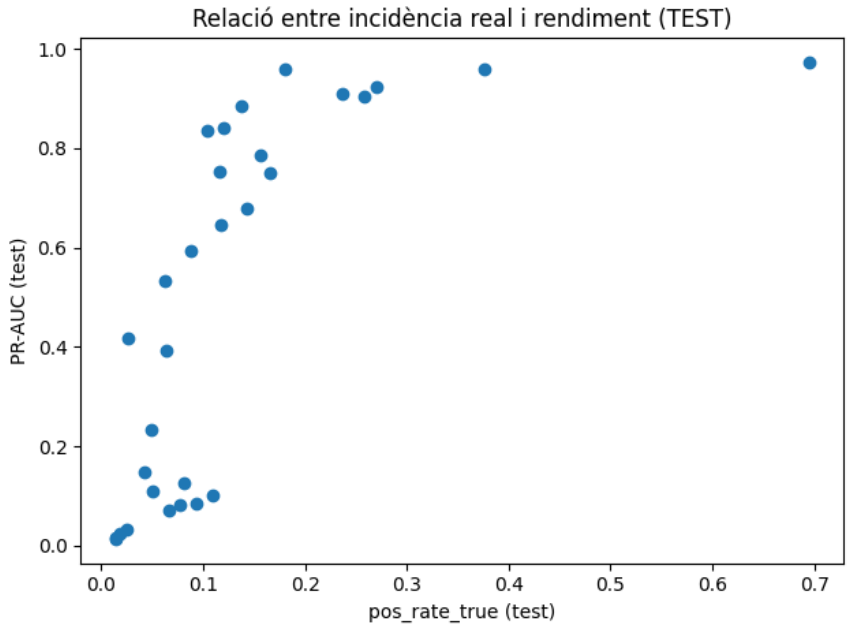


Modelització 1: predicció incidència si/no

Rendiment territorial i estratègia de desplegament

- ✓ El rendiment varia segons la incidència i el volum de dades disponible:
  - Més positius -> millor capacitat predictiva
  - Baixa prevalença –> PR-AUC més variable

- ✓ Desplegament selectiu en municipis amb millor senyal predictiva, amb llindar ad-hoc

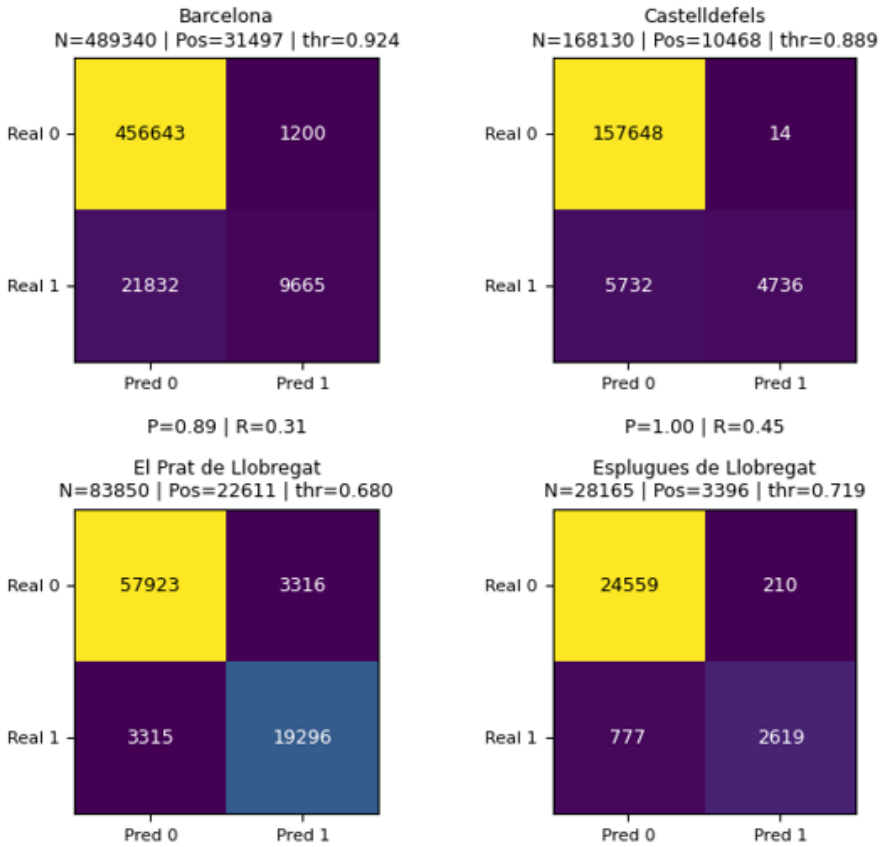


Modelització 1: predicció incidència si/no

Implicacions operatives

- ✓ Model com a eina de suport a la decisió
  - Permet anticipar incidències en municipis amb suficient històric
  - Orientat a minimitzar falses alarmes
  - Ús de llindars específics per municipi

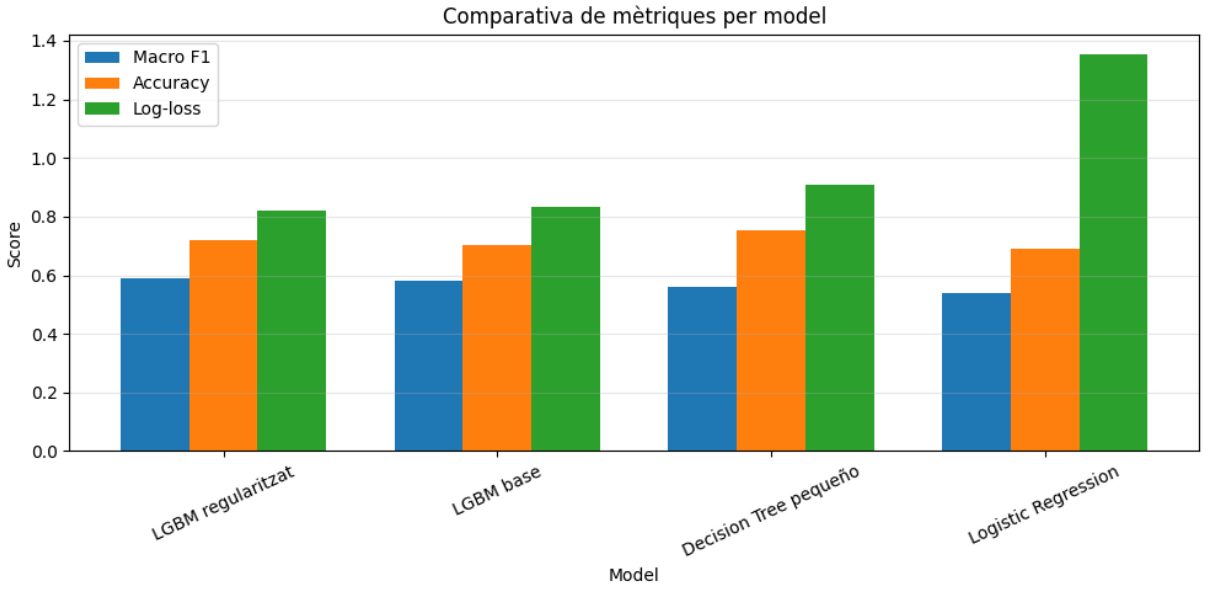
- ✓ Limitacions
  - El rendiment depèn del volum de dades i freqüència de les incidències
  - Dificultat per predir incidències esporàdiques o causades per factors externs
  - Recomanació de desplegar només en 20 municipis



Modelització 2: predicció del tipus d'incidència

Selecció i validació del model

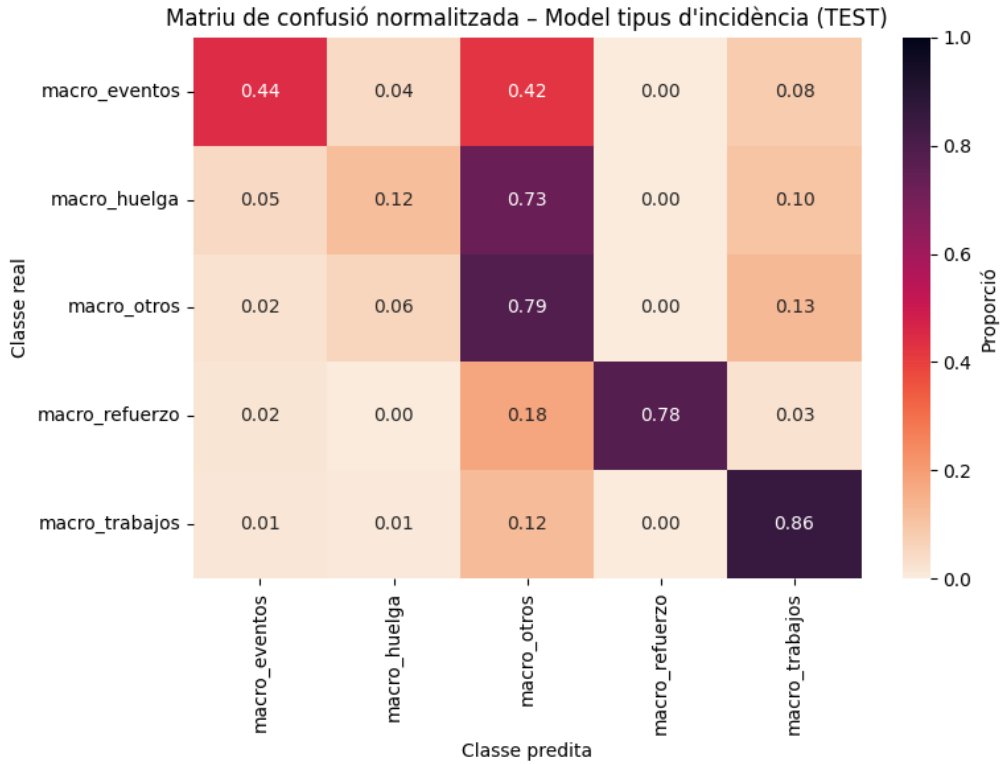
- ✓ Classificació multiclase del tipus d'incidència
  - Treball, esdeveniments, vaga, reforç i altres
- ✓ Es comparen els models:
  - Decision Tree
  - Logistic Regressión
  - LightGBM
  - LightGBM regularitzat
- ✓ LightGBM regularitzat model seleccionat:
  - Macro-F1 més alt, millor LogLoss (més petit)



Modelització 2: predicció del tipus d'incidència

Model final

- ✓ Rendiment final:
  - Accuracy 0.66 en test
  - Macro-F1 0.61
- ✓ Millor rendiment per tipologia:
  - Macro_Trabajos 0.81
  - Macro_refuerzos 0.86



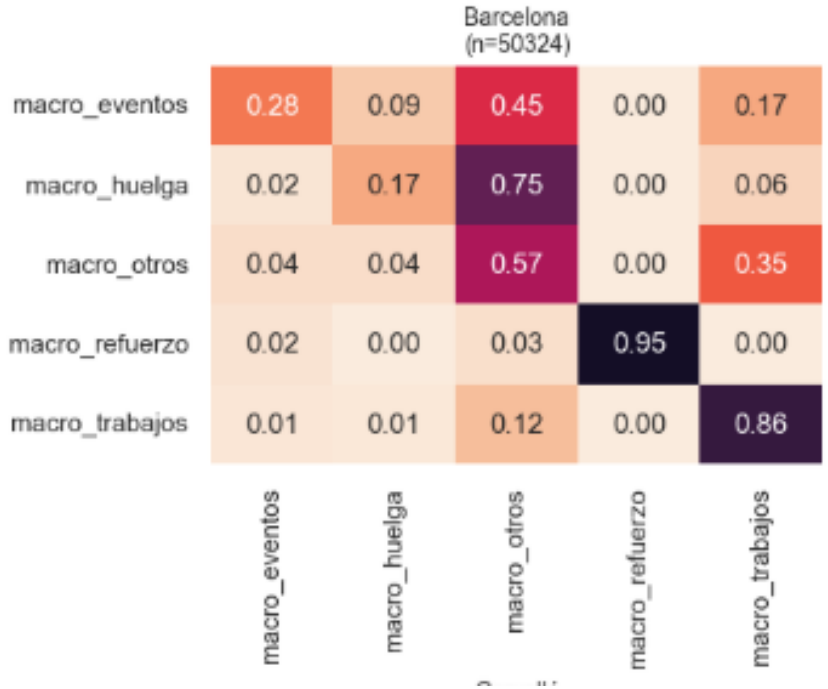
- ✓ El model mostra millor comportament en incidències amb patrons recurrents i estructurats
- ✓ Model conservador: ús de la tipologia 'Altres' per defecte

Modelització 2: predicció del tipus d'incidència

Rendiment territorial i estratègia de desplegament

- ✓ Desplegament condicionat per municipi degut al rendiment condicionat per:
  - Volum històric disponible
  - Diversitat de les tipologies
  - Qualitat de les etiquetes

- ✓ Patrons observats:
  - Confusions entre esdeveniments, vagues i altres
  - Pràcticament no hi ha confusions entre obres i reforços



Conclusions tècniques

Conclusions

- ✓ S'ha desenvolupat un **sistema predictiu** capaç d'anticipar **incidències** amb un horitzó de 15 dies combinant informació temporal, espacial i històrica.
- ✓ El model de detecció presenta **bon rendiment** en municipis amb **volum significatiu** de dades històric i incidències.
- ✓ Els **llindars específics** per municipi milloren l'equilibri entre **detecció d'incidències** i **control de falses alarmes.**
- ✓ El **rendiment no es homogeni** territorialment, es recomana un **desplegament selectiu per municipis**.
- ✓ La incorporació de dades Open Data no ha millorat el rendiment del model, **manca geolocalització dels esdeveniments**.
- ✓ L'especialització del model en municipis amb alta incidència no aporta millores estructurals, el **model global presenta millor capacitat de generalització**.

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Resultats, recomanacions i viabilitat d'un sistema d'alertes



Javier Fernández
Especialista de Dades
AMB Informació

AMB INFORMACIÓ i2cat eurecat Centre Tecnològic de Catalunya

Hi haurà una alteració?

Es recomana desplegar el model en 20 dels 30 municipis avaluats

El model avisa només quan n'està segur: prioritza que els avisos siguin fiables, encara que se n'escapin alguns

ELS 30 MUNICIPIS

50% **ENCERTA** dels avisos que dona

61% **DETECTA** de les alteracions reals

- **Sense model**, avisant a cegues, s'encertarien **10 de cada 100** avisos
- Només **1 de cada 10** parades-dia té alteració
- La mitjana **amaga la diferència** entre els municipis on funciona i on no

ON ES DESPLEGA

86% **ENCERTA** dels avisos que dona

65% **DETECTA** de les alteracions reals

- **Alteracions freqüents:** 1 de cada 9 dies a cada parada
- **Recurrents:** es repeteixen a les mateixes parades i èpoques, i el model **les aprèn**
- Hi passen el **87% de les alteracions** de la xarxa
- **S'escapen les esporàdiques:** no tenen precedents que les anunciïn

ON NO ES DESPLEGA

7% **ENCERTA** dels avisos que dona

32% **DETECTA** de les alteracions reals

- **Alteracions escasses:** 1 de cada 22 dies a cada parada
- **Esporàdiques** o amb massa pocs casos: **sense patró** per aprendre
- **93 de cada 100** avisos serien **falses alarmes**

Sobre les alteracions esporàdiques: **no indica un mal funcionament del model, sinó una restricció inherent a la predictibilitat del fenomen**

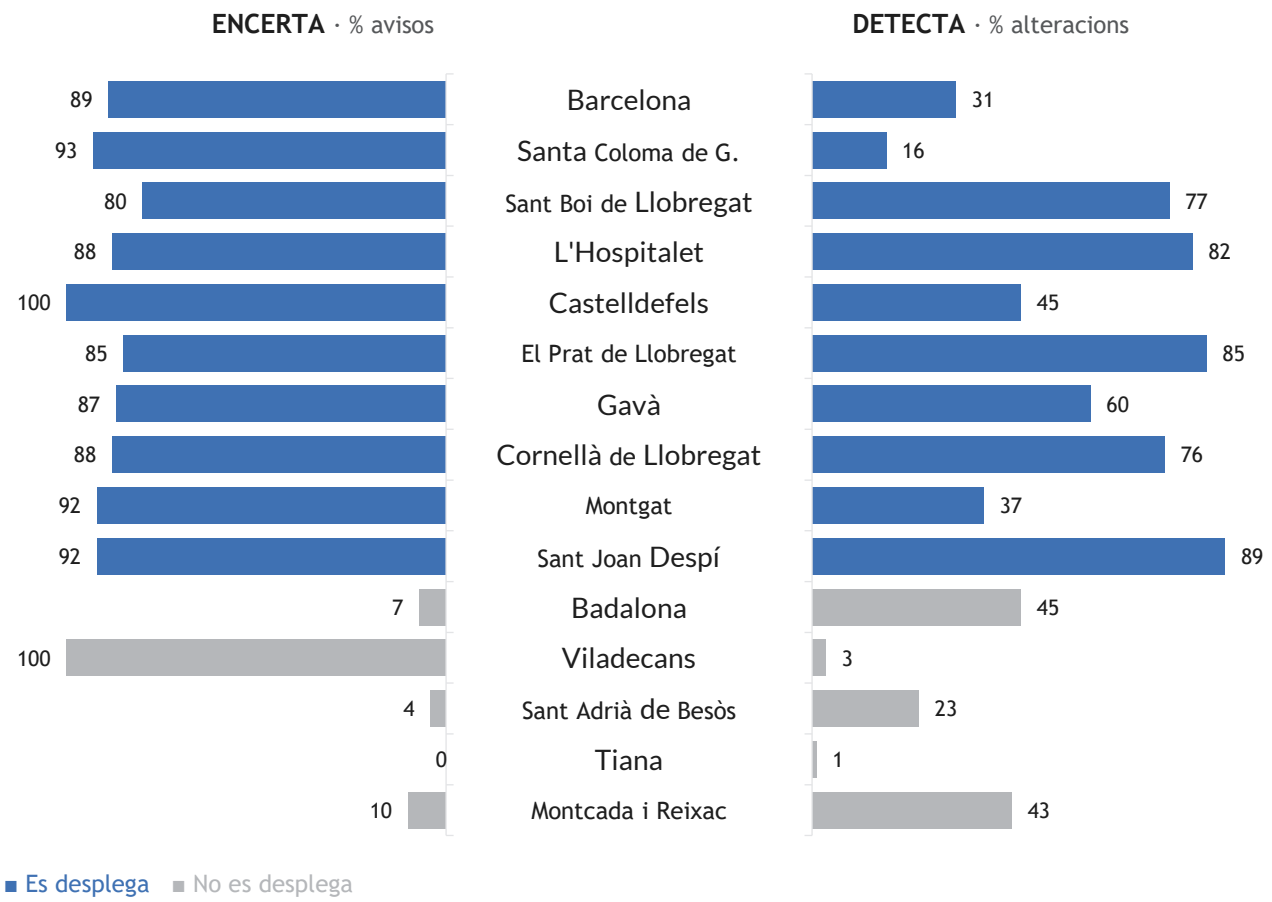
COM LLEGIR-HO

Per a cada parada i dia, amb 15 dies d'antelació.
Encerta: de cada 100 avisos, quants eren alteracions reals.
Detecta: de cada 100 alteracions reals, quantes es van avisar.

Municipi a municipi

Es desplega on el model pot garantir 85 encerts de cada 100 avisos

Funciona on hi ha senyal: alteracions freqüents i recurrents. Tenir molt volum de dades no n'hi ha prou.



QUÈ ÉS LA SENYAL? CALEN TRES COSES

1 **Que hi hagi alteracions sovint.** Si una parada en té poques, no hi ha prou exemples.

2 **Que es repeteixin,** a les mateixes parades i èpoques, com les obres llargues o les festes de cada any.

3 **Que n'hi hagi prou en total** per aprendre i comprovar-ho.

L'Hospitalet · funciona bé
Encerta 88 i detecta 82 de cada 100.
Alteracions freqüents (cada parada en té 1 de cada 4 dies) i recurrents.

Barcelona · encerta, però detecta poc
Encerta 89, però només detecta 31 de cada 100.
Molt volum repartit entre moltes parades: cadascuna té alteració 1 de cada 16 dies. El model només avisa quan n'està molt segur.

Badalona · no es desplega
Tercer municipi en volum, però de cada 100 avisos 93 serien falses alarmes. Cada parada té alteració 1 de cada 20 dies i sense un patró que es repeteixi.

De quin tipus serà?

Als municipis on es desplega, el model també intenta dir la causa de l'alteració

FUNCIONA BÉ

Obres i reforços

86 % **IDENTIFICA** de les obres reals

12 % **ACABA A 'ALTRES'** de les obres reals

- Les obres són **gairebé la meitat** de les alteracions
- **Reforços**: n'identifica 78 de 100, però són **menys del 2%** dels casos

AMB LIMITACIONS

Esdeveniments

44 % **IDENTIFICA** dels esdeveniments reals

42 % **ACABA A 'ALTRES'** dels esdeveniments reals

- Són **1 de cada 3** alteracions
- **Canvien de lloc i data**, i les dades externes **no diuen a quina parada afecten**

FORA D'ABAST

Vagues

12 % **IDENTIFICA** de les vagues reals

73 % **ACABA A 'ALTRES'** de les vagues reals

- **Menys del 2%** dels casos
- Es decideixen i s'anuncien **amb pocs dies d'antelació**: cap dada que tinguem les anticipa

MANCANÇA

Quan dubta, el model diu 'Altres': és prudent, perquè evita donar una causa equivocada, però al tècnic no li indica la causa.
Aprèn d'una etiqueta imperfecta: **1 de cada 4 registres de GIT no té un motiu útil** i s'ha hagut de reconstruir a partir del text.

Les mancances són a les dades

Cada limitació que hem vist té el seu origen en com registrem la informació

1 Motiu poc informat

1 de cada 4 registres no tenen un motiu útil (nuls, 'Altres', 'Desconegut').
S'ha hagut de reconstruir amb regles sobre el text.

→ Contribueix al fet que esdeveniments i vagues acabin a 'Altres'

2 Incidències sense parada

El 42% dels registres no indiquen parada: es donen d'alta a nivell de línia.
Hi ha parades que **no corresponen al municipi**.

→ Part de les alteracions no es poden situar a la parada, que és la unitat que prediu el model

3 Dades externes dispars i sense històric

Només alguns municipis en tenen, registrades de manera incompleta i **cadascun amb el seu format** (adreça, coordenades o sense ubicació).
Són **massa recents**: poc històric d'esdeveniments i molt escàs d'obres.

→ No es poden vincular a la parada: afegir-les va empitjorar el model

4 Registre incomplet i canviant

Si una alteració no es registra, per al model no ha passat.
Els criteris de classificació **han anat canviant** amb els anys.

→ El model aprèn del passat: si el registre canvia, el que ha après deixa de servir

Del model a la pràctica

Contrastat amb alteracions reals, el resultat no es va traslladar directament

SOBRE EL PAPER

Avisos fiables
Identifica incidències d'alta probabilitat amb precisió

Bona identificació de les obres
El tipus d'alteració més habitual

Un sol model, ajustat a cada municipi
El grau de seguretat per avisar s'adapta a cada territori per evitar falses alarmes

POSSIBLES MOTIUS

Per què no es trasllada directament

Esdeveniments que l'històric no recull
Curses o festes majors que es repeteixen cada any, però **no constaven en anys anteriors**

Prediccions poc concretes
El risc es reparteix entre moltes parades i dies; el tècnic necessita **una alerta concreta**

Mètriques de laboratori
Compten encerts parada a parada, **no si l'avís és útil per decidir**

LLIÇÓ

Validar amb els usuaris finals des del principi:
Acordar què és una alerta útil, amb quin detall (línia, parada o dia) i com s'ha de visualitzar.

Recomanacions sobre les dades

Què recomana el projecte per a cada mancança

1 · MOTIU POC INFORMAT

Revisar l'esquema de motius

Categories que **descriguin la causa real** de l'alteració i **protocols de registre més clars**.

2 · INCIDÈNCIES SENSE PARADA

Validar en el moment del registre

Menys text lliure: relació parada, línia i municipi, coordenades i camps mínims obligatoris.
En incidències de línia, generar totes les parades afectades.

3 · DADES EXTERNES

Començar ja l'històric

Esdeveniments, festius i obres **amb coordenades**, recollits **de manera automàtica** amb l'escolta d'agendes locals i xarxes socials. **Si s'incorporen tard, no permeten aprendre patrons.**

4 · REGISTRE INCOMPLET I CANVIANT

Millorar el registre i seguir el model

Protocols de registre més clars i seguiment periòdic del model, amb reentrenament i actualització dels llindars.

CONCLUSIÓ Una millor captura de les dades inicials **redueix el cost, millora els models i facilita l'escalabilitat** del sistema.

Línies de treball a AMB InformaciÓ

Accions per millorar les dades, començant per les nostres

1 GIT: motiu i coordenades

QUÈ

Explorar una **categorització més detallada del motiu sense perdre l'actual**, que segueix l'estàndard GTFS, i registrar **coordenades precises**.

PER QUÈ

Reduir **nuls i categories genèriques**, i poder **situar cada incidència al mapa**.

2 Base de dades d'esdeveniments

QUÈ

Consolidar les dades obertes dels ajuntaments en una **base de dades metropolitana** d'esdeveniments, festius i obres.

PER QUÈ

Les dades obertes actuals estan **disperses i sense coordenades**.

3 Escolta d'agendes i xarxes socials

QUÈ

Recollir **de manera automàtica** els esdeveniments i avisos publicats a les agendes locals i a xarxes socials, i **guardar-ne l'històric**.

PER QUÈ

Captar alteracions **que no arriben a GIT** i començar a acumular-ne **històric**.

4 Incorporar TMB

QUÈ

Incorporar les dades de TMB per tenir **Barcelona completa**. S'està explorant compartir-les a través de l'**Espai de Dades M3**.

PER QUÈ

Sense TMB, Barcelona **només inclou les línies de l'AMB**.

PRIMER PAS **Començar per GIT**: és la dada que controlem i la base de qualsevol model futur.

On som i què falta

Malgrat les limitacions, el sistema pot ser una eina efectiva de suport a la decisió

ON SOM

Una eina útil, però limitada

Identifica **amb 15 dies d'antelació** on és probable que calgui intervenir, als municipis i en els tipus d'alteració **on hi ha senyal**.

No preveu tot el que passa.

QUÈ FALTA PER A UNA SOLUCIÓ COMPLETA

1. **Millors dades de registre:** un motiu que expliqui la causa, la parada afectada i coordenades precises a GIT.
2. **Dades externes homogènies i amb històric:** esdeveniments, obres i festius de tots els municipis, amb el mateix format i amb ubicació.
3. **Alertes complementàries per a l'imprevist:** les alteracions sense precedents o esporàdiques no es poden preveure a partir del passat. Cal detectar-les a curt termini amb altres fonts, com avisos, xarxes socials o comunicacions oficials.
4. **Una alerta pensada per a l'usuari final:** acordar què avisar, amb quin detall i com visualitzar-ho.

Viabilitat d'un sistema d'alertes

És viable per fases, amb les dades com a condició

PRIMER PAS

Un pilot acotat

- Avisos **a 15 dies** als municipis amb prou històric
- Alteracions **planificades**, com les obres
- **Definit amb els tècnics GIT**: què avisar, amb quin detall i com veure-ho

SI MILLOREN LES DADES

Més tipus i més territori

- Avisos **també d'esdeveniments**
- **Més municipis**, a mesura que millori i creixi l'històric
- **Barcelona completa**, amb TMB

SOLUCIÓ COMPLETA

Preveure i alertar

- El model a 15 dies, **per al que es repeteix**
- Alertes a curt termini, **per a l'imprevist**
- **Seguiment i reentrenament** periòdic del model
- Una eina que **no depengui només de la IA**

PROPERS PASSOS

- **Millores al registre de GIT**
- **Base de dades metropolitana** d'esdeveniments, festius i obres

Tres idees
per emportar-se

1. El model és una **eina útil de suport a la decisió** on les alteracions són **freqüents i recurrents**.
2. El que el limita són **les dades d'origen**, no l'algoritme.
3. Un sistema d'alertes és **viable per fases**, començant **pel registre**.

Gràcies.

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

AMB INFORMACIÓ

La IA comença al registre

Abans d'un bon algoritme, una bona dada

Javier Fernández
Especialista de Dades
AMB Informació

Projecte BusCA

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Torn obert de preguntes

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Gràcies per la vostra assistència

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Cafè & Networking

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Presentació del Projecte d'Impacte

"BusCA – Anàlisi de les alteracions del servei de bus metropolità "

Socis promotors

Generalitat de Catalunya · Barcelona · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación · conversia compliance is our business · CVC Centre de Visió per Computador · eurecat Centre Tecnològic de Catalunya · forvis mazars & infini Digital Efficiency · i2cat · softeng Maximize the cloud. Securely. · UPC Universitat Politècnica de Catalunya BarcelonaTech · IDEAI Intelligent Data Science and Artificial Intelligence Research Center