VIII edició Premis La Unió a la Innovació en Gestió

Qüestionari 2017

ENTITAT: Hospital Plató Fundació Privada

Adreça: Carrer Plató, 21
Població: Barcelona
Persona de contacte: Marta Barahona Orpinell
Telèfon: 932023922 / 645809089
e-mail: marta.barahona@hospitalplato.com

Persones responsables del projecte (Nom, càrrec i e-mail)

Jordi Altés, Director Mèdic Hospital Plató. jordi.altes@hospitalplato.com
Marta Barahona, Directora de Gestió de Clients Hospital Plató. marta.barahona@hospitalplato.com
Gemma Gallardo, Directora de Infermeria Hospital Plató. gemma.gallardo@hospitalplato.com
Anna Jover, Directora de Gestió Social i Corporativa Hospital Plató. anna.jover@hospitalplato.com
Montserrat Monge. Cap Unitat Admissions i Gestió de Clients Hospital Plató. montserrat.monge@hospitalplato.com
Alejandro Nieto. Director de Sistemes de Informació Hospital Plató. alejandro.nieto@hospitalplato.com
Rafel Xavier Vidal. Director Econòmic i de Gestió Hospital Plató. rafel.vidal@hospitalplato.com
Damià Pujol, Soci Consultor Athenea Solutions. damia.pujol@atheneasolutions.com
Christian Fons-Rosen, Co-Director Master's Degree on Data Science, Universitat Pompeu Fabra, Barcelona Graduate School of Economics. christian.fons-rosen@upf.edu

Modalitat a la qual es presenta:

- ☐ **Innovació en polítiques d'eficiència en recursos**
- ✓ **Innovació en gestió dels professionals o gestió assistencial**
- ☐ **Innovació en gestió d'aliances estratègiques**

Guió per a la redacció del Cas d'Innovació

TITOL:

Orientats cap a l'eficiència: *Data Science* per predir el risc de readmissió i el temps per l'alta hospitalària

ABSTRACT (Resum de l'experiència en un màxim de 10 línies):

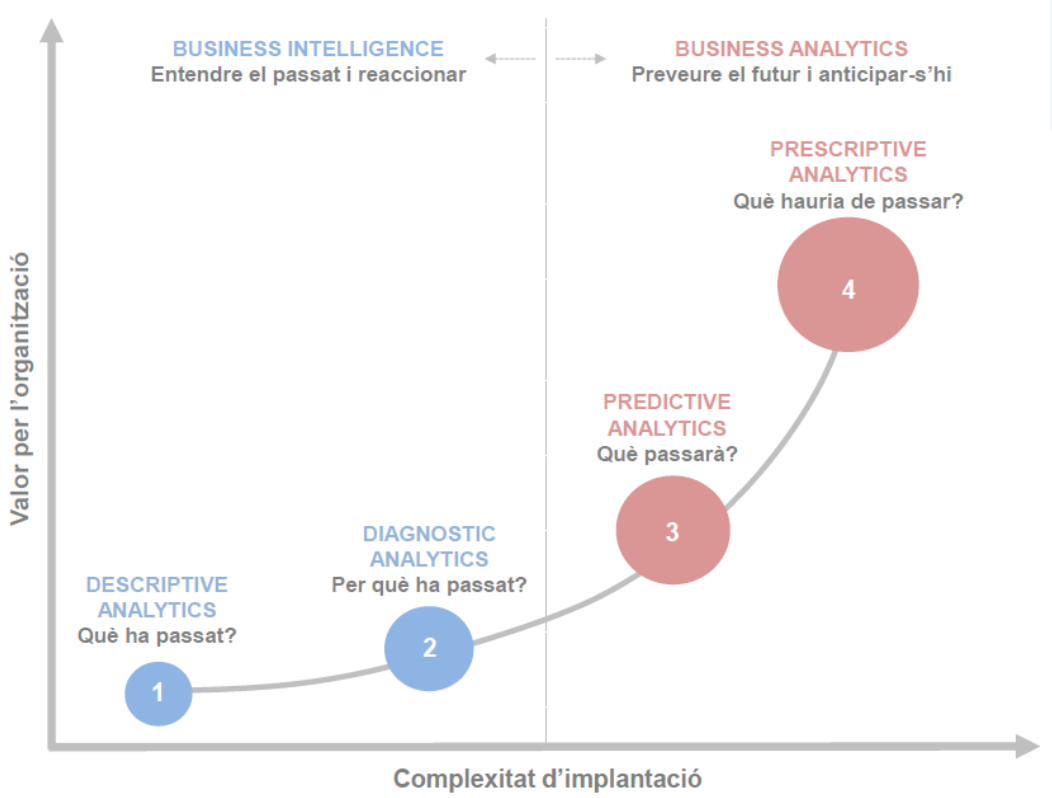
L'Hospital Plató ha promogut un partenariat amb la consultoria Athenea Solutions i la Universitat Pompeu Fabra amb l'objectiu d'aplicar solucions de *Data Science* per millorar la qualitat assistencial, optimitzar i millorar l'eficiència. El present document explica l'experiència viscuda en el primer projecte dut a terme per aquesta aliança enfocat a la predicció del risc de readmissió i l'estimació dels dies pendents per donar l'alta als pacients ingressats a l'hospital.

Els resultats obtinguts en la prova pilot han estat exitosos, mostrant com les tècniques de *Data Science* poden ajudar a prendre accions preventives per evitar readmissions, traduint-se en una millora de la qualitat assistencial i de l'eficiència en el procés d'hospitalització. Per altra banda, la predicció dels dies per l'alta a temps real contribueix a una planificació òptima de la distribució de pacients a les plantes de l'hospital, per tal d'optimitzar l'ús del recurs del llit.

ANTECEDENTS I OBJECTIUS

Avui en dia, el sector sanitari genera unes quantitats enormes de dades. Molts hospitals i altres centres clínics catalans disposen d'eines que els permeten visualitzar la informació històrica de la seva activitat per tal de prendre decisions. Aquests sistemes donen resposta bàsicament a les preguntes "què ha passat?" o "per què ha passat?", i es basen en l'explotació de les bases de dades mitjançant consultes ad-hoc, reports i *dashboards*, entre d'altres. Tradicionalment, aquests sistemes s'engloben dins el concepte de *Business Intelligence*.

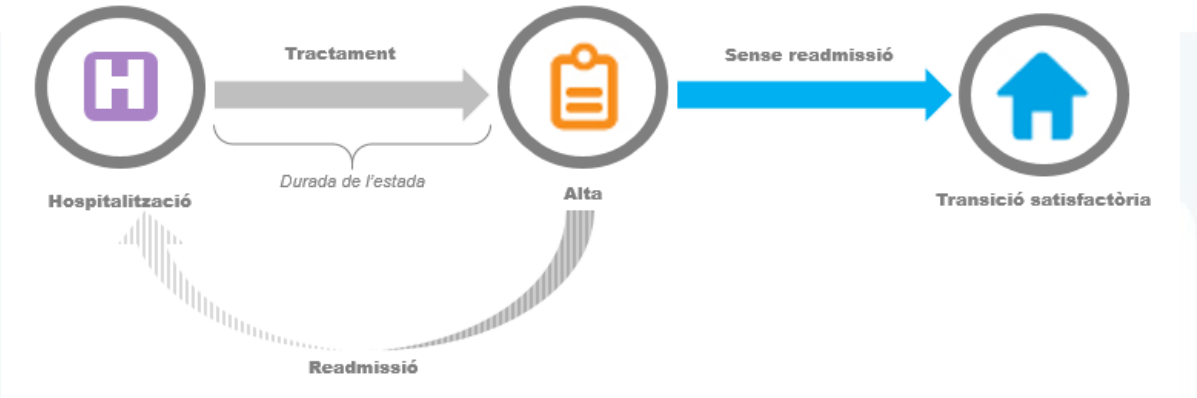
En els darrers anys ha començat a guanyar força el terme de *Business Analytics*, per descriure totes aquelles tècniques que busquen preveure els resultats futurs per tal de prendre decisions que permetin anticipar-s'hi, contestant a la pregunta "què passarà?", o fins i tot "què hauria de passar?". Aquests sistemes fan un ús més extensiu de les dades i es basen entre d'altres en l'anàlisi estadístic, l'anàlisi de tendències o els models predictius, emmarcats dins de l'anomenada ***Data Science***. Es tracta doncs d'una evolució del *Business Intelligence* que genera un coneixement no explícit molt rellevant a l'hora de millorar la qualitat assistencial, optimitzar els recursos i reduir costos.



El *Business Analytics* en el sector sanitari encara no té una presència important a Catalunya, tot i que a molts països el seu ús està cada cop més consolidat i les experiències viscudes demostren el seu indubtable èxit. El 2016, en la línia de la seva forta aposta per la innovació, la Direcció General de l'**Hospital Plató** va promoure un partenariat amb la consultoria **Athenea Solutions** i la *Barcelona Graduate School of Economics* de la **Universitat Pompeu Fabra** a través del seu *Master's Degree in Data Science*, amb l'objectiu d'introduir solucions de *Data Science* en el seu centre. La Comissió de Projectes Estratègics i d'Innovació de l'hospital va prioritzar el projecte i per tal d'implicar al personal del centre en les tasques necessàries per dur a terme la innovació va crear un grup de treball format per la Direcció de Gestió Assistencial, la Direcció Mèdica, la Direcció d'Infermeria, la Cap d'Admissions i el Director de Sistemes d'Informació Hospitalària.

El primer projecte realitzat dins d'aquesta aliança aborda dos dels problemes més recurrents en l'àmbit hospitalari: el risc de readmissions i l'estimació dels dies pendents per donar l'alta a un pacient ingressat. La solució d'aquests dos problemes té una aplicabilitat directa a la pràctica diària clínica i és d'un gran interès per qualsevol hospital, ja que afecta la qualitat assistencial proporcionada al pacient, la gestió dels recursos de l'hospital (reserva d'habitacions, tancament de plantes, etc.) i la gestió de costos econòmics (costos d'hospitalització, penalitzacions econòmiques, etc.).

Com hem dit, un dels problemes adreçats és el del **risc de readmissió** d'un pacient ingressat. Les readmissions o reingressos evitables (*preventable readmissions*) suposen un dels principals causants de l'augment de costos d'hospitalització i, a més, impliquen penalitzacions econòmiques. Segons el CatSalut, es considera readmissió hospitalària el reingrés no programat d'un pacient per la mateixa malaltia (o grup de patologies) de base que va provocar el primer ingrés, durant els següents 30 dies. En aquest projecte, però, considerem també les readmissions a 60 dies.

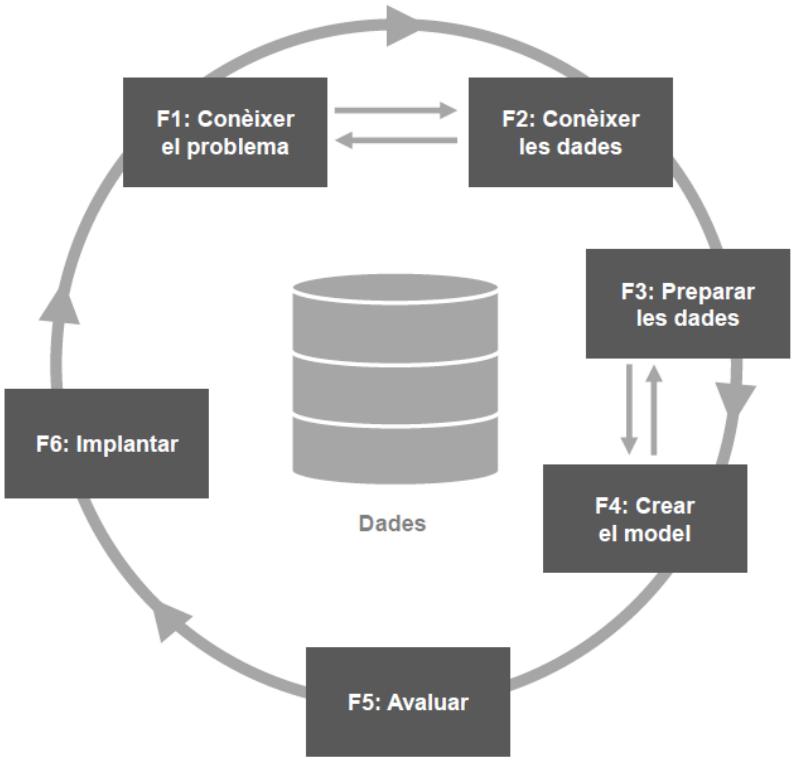


A més de les implicacions econòmiques i l'ocupació de recursos que suposa el reingrés inesperat d'un pacient, existeixen implicacions directes sobre la salut i qualitat de vida del pacient. Hi ha readmissions que simplement no es poden preveure, però en molts casos sí que ens podem anticipar i evitar que al donar una alta el pacient torni a reingressar. Per això és necessari detectar quan un pacient susceptible a ser donat d'alta presenta un risc de readmissió i actuar en conseqüència (allargar l'estada, fer un seguiment a domicili, etc.).

El segon problema adreçat en aquest projecte és, com hem dit, l'estimació dels dies pendents per donar l'alta a un pacient ingressat. La durada de l'estada (*length of stay*) d'un pacient afecta directament el cost de l'hospitalització (ocupació de recursos, facultatius, equipament, etc.). No sempre és possible estimar el nombre de dies que un pacient estarà ingressat doncs cadascun d'ells té les seves particularitats i el curs del seu ingrés pot ser imprevisible. Poder realitzar prediccions temps real i el màxim acurades possible sobre els **dies pendents per donar l'alta** a un pacient és vital per tal de realitzar una bona planificació de recursos de personal, disponibilitat de llits, tancament de plantes, etc.

DESENVOLUPAMENT DEL PROJECTE

La metodologia que s'ha utilitzat per realitzar aquest projecte s'anomena CRISP-DM i consta de les següents sis fases:



En una exhaustiva iteració entre les fases F1 i F2, es van decidir els dos problemes a tractar i es van recopilar les diferents fonts de dades necessàries que formen l'històric d'ingressos a l'hospital des del 2013 fins a l'actualitat contemplant els següents tipus de dades:

Dades generals del pacient	Edat, sexe, cobertura, etc.
Dades mèdiques del pacient	Diagnòstics, PCC, MACA, etc.
Dades generals de l'ingrés	Data i hora d'ingrés i d'alta, etc.
Dades clíniques de l'ingrés	Servei, planta, origen d'admissió, etc.
Dades mèdiques en el moment de l'alta	Medicacions administrades, nombre de proves de laboratori i l'evolució dels seus resultats, Prealt, etc.
Dades d'utilització prèvies a l'ingrés	Visites a urgències en l'últim any, ingressos en l'últim any, readmissions anteriors, etc.

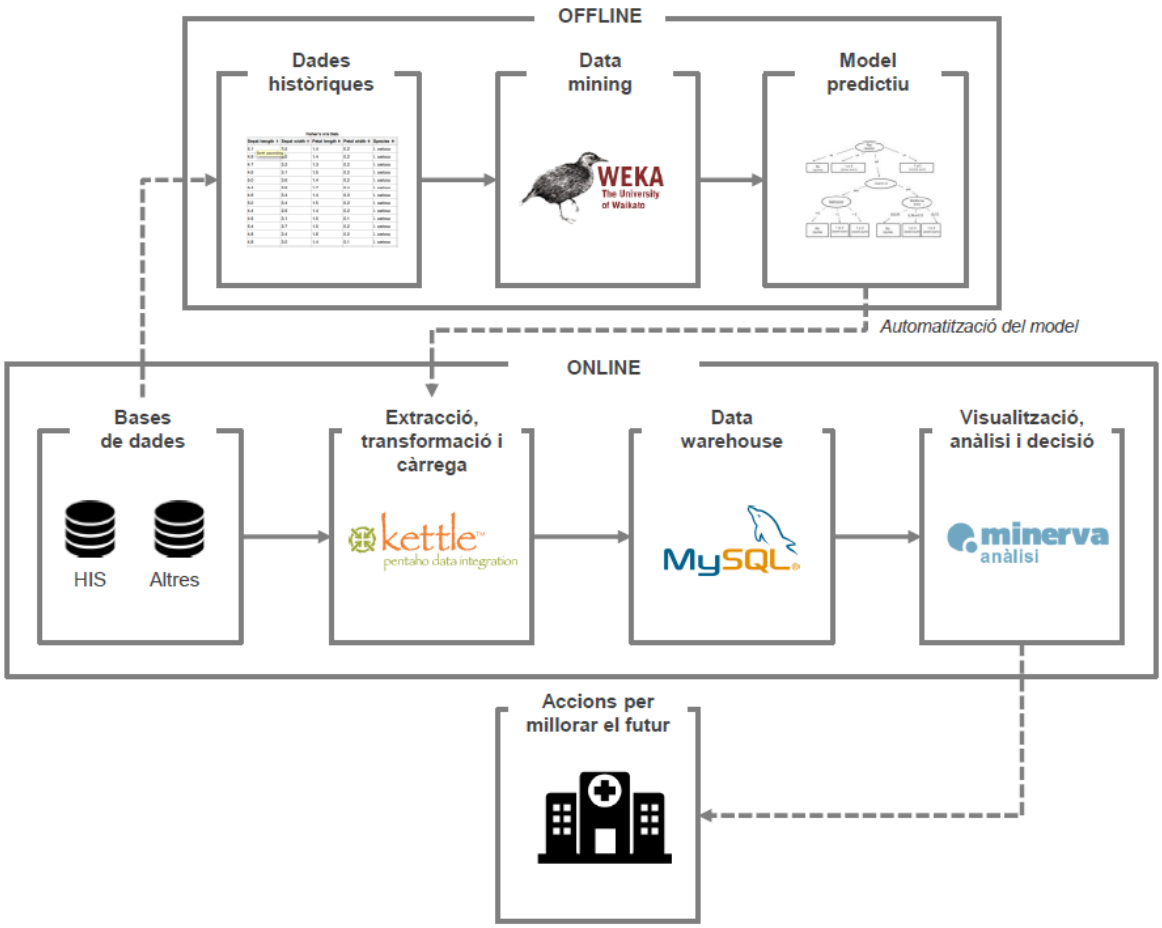
En la fase F3, totes aquestes dades provinents de desenes de fonts de naturalesa molt diferent han estat preprocessades i homogeneïtzades en dos conjunts de dades finals corresponents als dos problemes tractats en aquest projecte. Els diferents elements d'aquests conjunts de dades s'han etiquetat d'acord amb el problema a tractar (futura readmissió (sí/no) i dies per l'alta). Aquests dos conjunts de dades es coneixen com a conjunts d'entrenament ja que serveixen per "entrenar" un model predictiu per tal que "aprengui" a fer prediccions.

Així doncs, en la fase F4, s'han construït dos **models predictius**, també anomenats classificadors, mitjançant l'aprenentatge supervisat dels dos conjunts d'entrenament. La idea general és que els

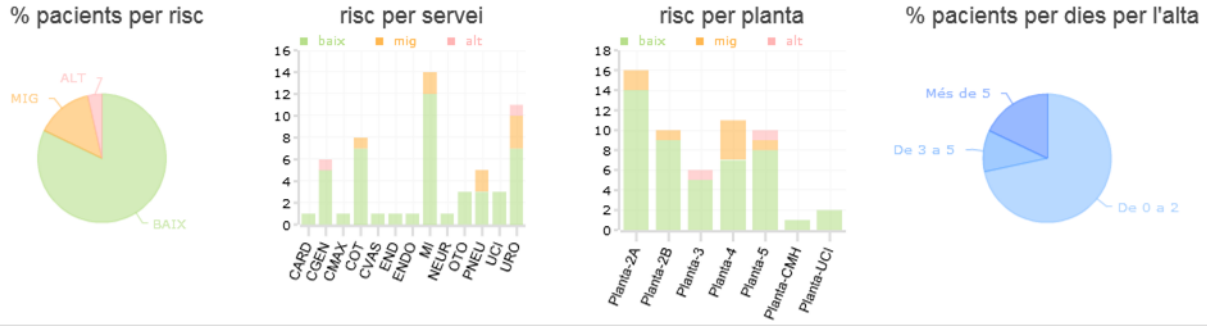
conjunts de dades que contenen tots els ingressos històrics etiquetats es van presentant a un algorisme que utilitza tècniques estadístiques i de *data mining* (regressió logística, arbres de decisió, *random forests, voting*, etc.) per anar construint un model que finalment serà capaç de fer prediccions sobre nous ingressos.

En una fase F5 s'ha avaluat la capacitat predictiva dels dos models amb tècniques de validació creuada (*crossvalidation*) que consisteixen en realitzar successives divisions dels conjunts de dades entre un conjunt d'entrenament (per construir el model) i un conjunt de validació (per avaluar-ne la capacitat predictiva). Les mesures utilitzades per mesurar la capacitat de predicció dels models han estat mesures típiques estadístiques: matriu de confusió, sensibilitat, especificitat, MCC, etc.

Totes les fases descrites fins el moment corresponen a la part *offline* del projecte, o sigui, la part que es desenvolupa al principi de tot, es manté estable un cop realitzada i no es veu alterada per les dades a temps real. La part *online* seria la que utilitza el model generat en la part *offline* per tal de fer prediccions a temps real amb les dades del dia a dia, com podem veure en el següent esquema:



La part *online*, doncs, es correspon a la fase F6 d'implantació. En aquesta fase s'ha creat una solució d'extracció, transformació i càrrega (ETL) de les dades que es generen diàriament, per tal de ser passades pels dos models predictius. Les prediccions d'aquests models són visualitzades en un quadre de comandament (*dashboard*) de l'eina Minerva d'Athenea Solutions per ser utilitzades pel personal mèdic a temps real. Tot seguit podeu veure una captura de pantalla del *dashboard*:



Llit	Planta	Servei	Edat	Sexe	Cobertura	Origen admissió	Quirúrgic	PCC/MACA	Data ingrés	Durada ingrés	Dies per l'alta	Risc readmissió	
219:B	Planta-2B	CGEN	64	H	SCS	Urgent	N	N/N	13/10/2017	4	De 0 a 2	BAIX	=
405:A	Planta-4	MI	99	D	SCS	Urgent	N	S/N	16/10/2017	1	De 0 a 2	BAIX	▼
405:B	Planta-4	MI	87	D	SCS	Urgent	N	S/N	16/10/2017	1	De 0 a 2	BAIX	▼
202:B	Planta-2A	MI	81	D	SCS	Urgent	N	N/N	09/10/2017	8	De 3 a 5	BAIX	=
407:00	Planta-4	PNEU	71	H	SCS	Urgent	N	S/N	26/09/2017	21	Més de 5	MIG	=
517:B	Planta-5	COT	68	D	SCS	Urgent	N	N/N	14/10/2017	3	De 0 a 2	BAIX	=
227:B	Planta-2B	MI	99	D	SCS	Urgent	N	N/S	16/10/2017	1	De 0 a 2	BAIX	▼
205:A	Planta-2A	MI	93	D	SCS	Urgent	N	N/N	16/10/2017	1	De 0 a 2	BAIX	=
227:A	Planta-2B	PNEU	75	D	SCS	Urgent	N	S/N	13/10/2017	4	De 0 a 2	BAIX	▼

Com es pot veure, la predicció del risc de readmissió s'indica en uns rangs de baix, mig, alt i pels dies per l'alta en els rangs de 0 a 2 dies, de 2 a 5 dies i més de 5 dies. A la part superior es dóna una visió global en forma de gràfiques sobre la distribució de riscs de readmissió i el percentatge de pacients per cada rang de dies per l'alta. A la part inferior del *dashboard* es mostra una llista amb alguns detalls de tots els pacients que formen l'actual cens de l'hospital. Clicant a qualsevol pacient es pot accedir al seu detall on es mostren diverses dades de caràcter general, mèdic i històric:

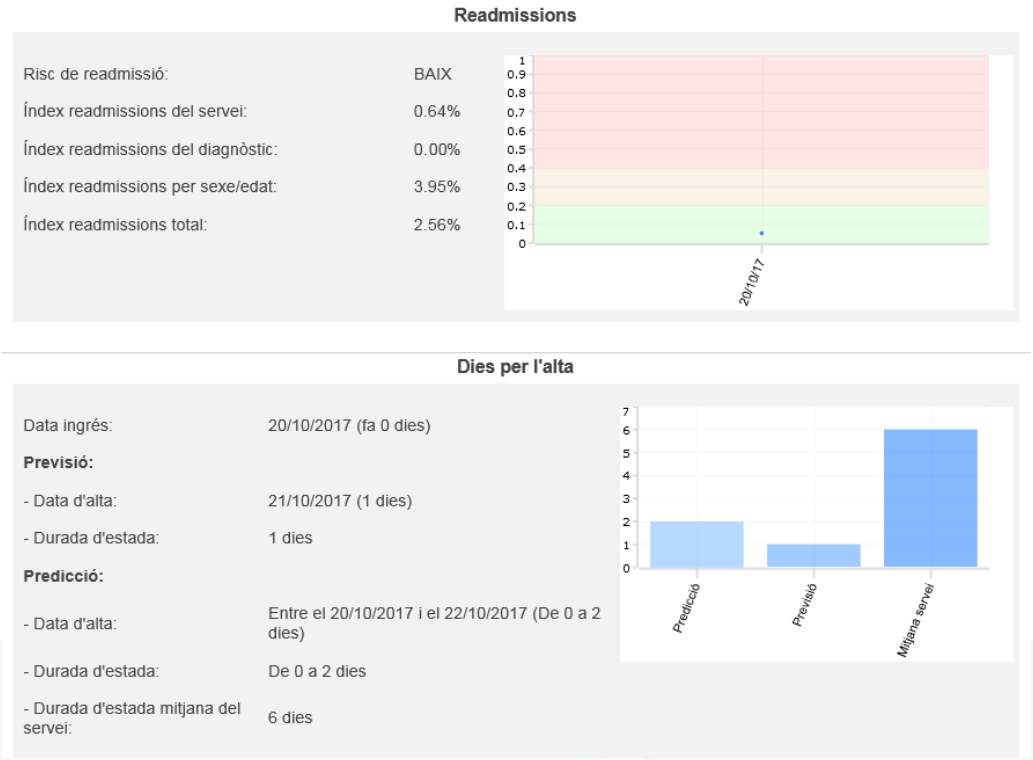
Pacient: (92 anys) Sexe: ♂

Perfil

Llit:	213:A (Planta-2A)
Servei ingrés:	END
Diagnòstic principal:	Altres malalties dels intestins i del peritoneu
Procediment principal:	N/A
Quirúrgic:	N
Origen admissió:	Urgent
PCC/MACA:	N/N
Cobertura:	SCS

Número readmissions passades:	0
Durant l'ingrés:	
- Número medicaments diferents:	5
- Número proves laboratori:	2
- Número altres proves:	1
- Última prova hemoglobina:	N/A (=)
- Última prova creatinina:	No patològica (▼)
- Última prova sodi:	No patològica (▼)
- Última prova potassi:	No patològica (▼)
Últim any:	
- Número admissions:	0
- Número visites:	11
- Número urgències:	3

I també dades corresponents al risc de readmissió com l'evolució del seu risc al llarg de l'ingrés o l'índex de readmissions per pacients similars, i la predicció de dies per l'alta juntament amb la data prevista d'alta (en cas que consti) o la mitjana d'estada per el servei relacionat:



En aquest projecte, dividim la fase F6 en una fase de prova pilot i una fase d'utilització definitiva. Actualment el projecte es troba en la fase de prova pilot. En aquesta fase els resultats del *dashboard* s'han analitzat per part d'un grup de professionals mèdics donant la seva valoració sobre aquests. Per altra banda, s'estan realitzant tasques de validació estadística de les prediccions realitzades.

En la futura fase d'utilització definitiva es proporcionarà accés al *dashboard* a tot el personal involucrat i s'utilitzaran les prediccions per realitzar accions concretes (per exemple, fer una trucada telefònica a un pacient donat d'alta per tal d'evitar una readmissió o planificar la distribució dels pacients en les plantes en funció de les prediccions sobre els dies per l'alta de cada pacient ingressat).

RESULTATS

L'avaluació de la fase pilot projecte, s'ha realitzat a través de 5 indicadors que mesuren l'impacte de la innovació a través de la capacitat predictiva dels models construïts. Pel cas de les readmissions partim de la següent taula de resultats:

Real/Predicció	Risc baix	Risc mig	Risc alt
No readmissió	2061	97	1
Readmissió	43	61	37

Cada cel·la conté el número de pacients dels que s'ha fet la predicció que tenen un risc baix, mig o alt (columna) i si han causat readmissió o no (files). Per exemple, veiem que 2061 pacients han tingut risc baix i al final no han readmès.

Partint de l'anterior taula, hem tret dos indicadors:

- **Indicador 1.1: Fiabilitat de la predicció d'absència de risc de readmissió**: 98%.
  En la gran majoria de casos (98%) que el model ha fet la predicció que un ingrés no causaria readmissió, ha estat finalment així. Aquest fet permetrà limitar a un petit grup de pacients (etiquetats amb risc mig-alt) les accions per evitar readmissions, d'una manera molt segura.
- **Indicador 1.2: Fiabilitat de la predicció de presència de risc alt de readmissió**: 97%.
  Aproximadament el mateix percentatge d'encert s'obté pels pacients que són etiquetats amb risc alt. Sabent que gairebé sempre que un pacient és etiquetat amb risc alt causarà readmissió, podem monitoritzar aquest grup de pacients d'una manera molt personalitzada i fer les accions addicionals necessàries per evitar la readmissió.

Pel cas dels dies per l'alta, mostrem alguns resultats en la següent taula:

Real/Predicció	Més de 5 dies	3-5 dies	0-2 dies
Més de 5 dies	23.885	3.486	473
3-5 dies	3.653	9.243	4.671
0-2 dies	474	3.806	25.810

Igual que abans, cada cel·la d'aquesta taula conté el número de pacients dels que s'ha fet la predicció que trigarien 0-2 dies, 3-5 dies o més de 5 dies a rebre l'alta (columna) i els dies que realment han trigat a rebre-la (files).

- **Indicador 2.1: Fiabilitat de la predicció d'alta en més de 5 dies**: 85%.
  Si el model prediu que un pacient estarà més de 5 dies, serà efectivament així en un 85% dels casos.
- **Indicador 2.2: Fiabilitat de la predicció d'alta en menys de 2 dies**: 83%.
  Un encert aproximat obtenim en les prediccions d'alta en 0-2 dies.
- **Indicador 2.3: Fiabilitat de la predicció d'alta en 3-5 dies**: 56%.
  Més complicat serà prendre decisions en els pacients que el model indica que estaran 3-5 dies ja que la predicció es complirà en un 56% dels casos. Com a part positiva, només un 22% del pacients rebran aquesta predicció.

CONCLUSIONS

Vist en perspectiva, veiem que aquest projecte ha estat ideal com a primera experiència en el món del *Data Science*. Tal com hem vist en la secció anterior, els resultats de la prova pilot han estat exitosos i ens animen a seguir endavant. Els indicadors del risc de readmissió ens permeten, entre altres coses, plantejar un enfocament més dirigit cap els pacients als quals el professional hauria d'aplicar accions preventives, amb el guany de temps i reducció de cost que això suposa. A més, el fet de poder-se centrar exclusivament en un grup de pacients més petit, millorarà l'eficàcia de les accions empreses i això s'hauria de traduir en una **reducció de les readmissions** i, per tant, en una **millora de la qualitat assistencial i de l'acompliment dels objectius econòmics pactats amb el CatSalut**.

Pel què fa a la predicció de la durada d'estada, aconseguim un gran encert en les dues categories més interessants que són les altes imminents i els ingressos de llarga estada, ja que permeten una bona planificació de les plantes de l'hospital a nivell setmanal. Millores en la decisió d'accions importants com per exemple el tancament de plantes el cap de setmana, implica **reducció de costos a nivell de recursos humans i materials**.

A part d'haver obtingut resultats exitosos, l'experiència ens ha permès afrontar i superar algunes de les principals dificultats amb les que ens podem trobar en el futur en un projecte d'aquestes característiques. Concretament, el problema de les readmissions utilitza un conjunt de dades extremadament no balancejat. Això significa que el volum d'ingressos de pacients que no han causat readmissió és molt més gran que el d'ingressos que sí que n'han causat. Els conjunts de dades no balancejats suposen una dificultat addicional en les tècniques de *data mining*, que implica la utilització de matrius de costos o algorismes de balancejat per facilitar la construcció de models de predicció vàlids. A més, l'índex de readmissions a l'Hospital Plató és baix en comparació amb altres centres hospitalaris, així que detectar possibles readmissions i prevenir-les en un centre on ja s'estan obtenint uns bons resultats, és encara més complicat.

Una altra dificultat ha estat l'acceptació per part del personal mèdic de models de caixa negra. Algunes de les tecnologies més exitoses en el món del *data mining* com el *Random Forest* construeix models que segueixen un raonament de predicció de difícil (sinó impossible) explicació, sobretot pels usuaris no familiaritzats amb *Data Science*. La implantació escalonada amb una prova pilot, la pedagogia mitjançant simplificació dels algorisme utilitzats, així com la pròpia capacitat d'encert dels models proporcionats, han permès millorar molt en aquest sentit realitzant una **gestió del canvi satisfactòria**.

Per altra banda, una de les lliçons que hem après d'aquest projecte és que en la fase inicial és extremadament necessari fer un anàlisi exhaustiu per determinar en quin punt del temps disposem de cada una de les dades. És evident que no és el mateix treballar amb les dades històriques del conjunt d'entrenament (que disposem de totes elles) que treballar amb dades a temps real que no sempre són disponibles quan es necessiten. En la majoria de casos, aquest anàlisi és molt intuïtiu. Per exemple, podem saber fàcilment que no disposarem del "servei en l'alta" en el moment de l'ingrés. En altres casos, en canvi, no és tan evident. Un exemple és el cas dels diagnòstics del pacients. Intuïtivament podem pensar que el diagnòstic d'un pacient ingressat estarà disponible en el moment de voler realitzar una predicció, però la realitat és que en molts casos aquest diagnòstic encara no ha estat registrat a la base de dades i cal recórrer a altres fonts (per exemple, agafar el diagnòstic de la urgència del pacient en cas que sigui un pacient provinent d'urgències).

En una fase futura es planteja portar aquest projecte un pas més enllà i ja no parlar només de prediccions sinó de prescripcions (*Prescriptive Analytics*). Per exemple, en el cas de les readmissions, no ens limitaríem a etiquetar un pacient amb un risc baix, mig o alt, sinó que l'eina desenvolupada podria mostrar accions recomanades per aquest pacient. Mitjançant l'anàlisi de les probabilitats així

com dels costos associats a l'hospitalització i a les diferents accions per prevenir una readmissió, el *dashboard* podria proposar una sèrie d'accions al professional juntament amb una justificació. Per exemple, una proposta podria ser allargar l'estada un dia més o programar alguna acció addicional per l'endemà a l'alta (per exemple, contacte telemàtic). Cal tenir clar que el *dashboard* és només un ajut a la decisió i que aquestes accions són propostes que han de ser consensuades amb els professionals.

Finalment, a banda del tema de les readmissions i els dies per l'alta, els àmbits en els que es pot aplicar el *Data Science* en el sector de la salut són molt variats. Des de la previsió de nombre d'urgències fins a l'optimització de les reserves de quiròfans, se'ns obre un ventall de possibilitats per explotar en el futur pròxim que esperem que tinguin una repercussió molt positiva en la qualitat assistencial i l'eficiència del nostre hospital.