

ANÀLISI DE DADES

Març de 2023

Informe realitzat per:

Pere Girona Campi

 peregirona.com

 github.com/perot84

 pere@peregirona.com

 linkedin.com/in/peregirona

ACAPPS

Informe

Aquest informe analitza i exposa els resultats del tractament de dades extrets del CRM de l'entitat a 10 de març de 2023.

S'ha treballat a partir de tècniques de ciència de dades i de "Business Intelligence" per a analitzar les dades en tot el seu conjunt.

Detalls de l'informe

Entitat	ACAPPS
Data d'extracció de les dades	15/03/2023
Tècniques utilitzades	Data Science i Business Intelligence

Objectius

- Agilitzar l'entrada de dades
- Millorar l'entrada de dades
- Millorar el processament de les dades
- Facilitar la creació de justificacions i projectes
- Millorar la presa de decisions a partir de l'anàlisi i predicció de dades

Procediments

- Unió de totes les dades vinculades als usuaris, centrant-nos en les persones que es troben en cerca de feina.
- Predicció de quant de temps triga una persona a trobar feina, des que entra al servei.
- Detecció de millores en la recollida i gestió de les dades
- Creació d'un panell en Power BI on es visualitzen diferents dades

Resultats

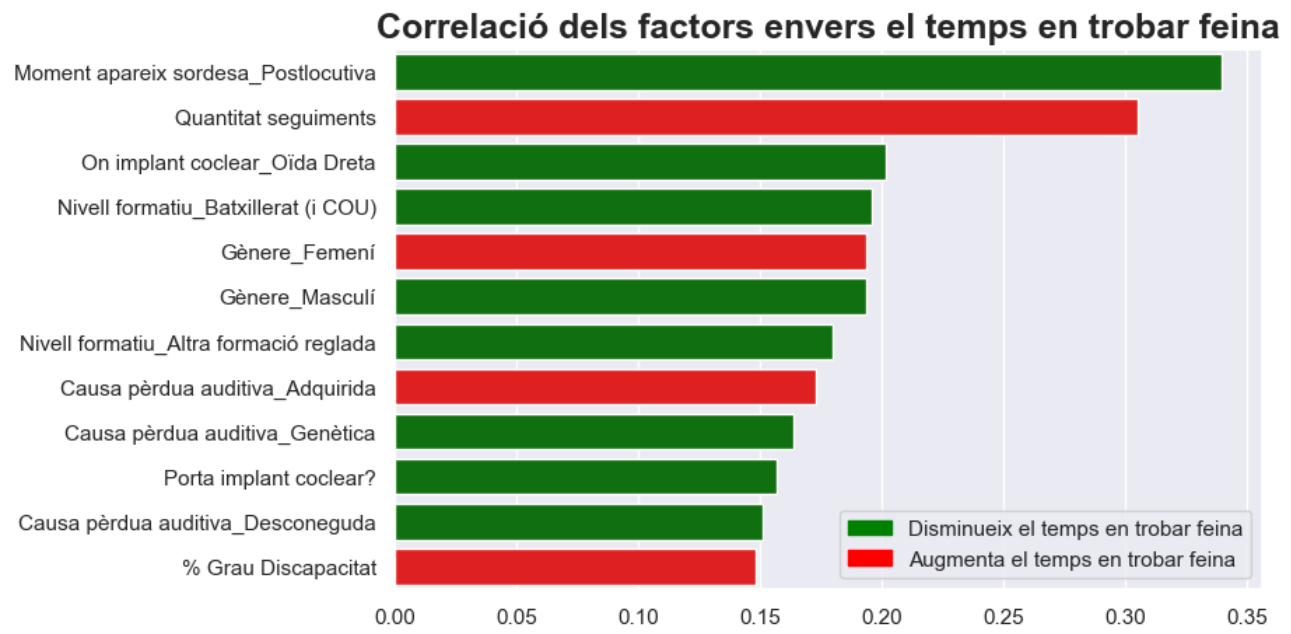
Base de dades amb totes les dades pròpies de cada persona incloses les noves columnes creades

Predicció del temps que trigaria una persona trobarà feina des que entra al servei.

Després de la neteja de dades s'ha treballat amb una mostra de 77 persones amb un marge d'error de 102 dies.

Aquest resultat implica que el model té un error de 102 dies en la seva predicció. Cal destacar que hem treballat amb poques entrades (77) i eliminant algunes variables per no tenir suficients dades o ser de poca qualitat. Creiem que si es millora la qualitat i quantitat de dades es podrien millorar de manera significativa els resultats.

Anàlisi dels factors més determinants en el temps que triga una persona a trobar feina



Com podem observar en el gràfic el fet que una persona aparegui la sordesa després d'aprendre a parlar és el factor que més impacta en disminuir el temps en què triga a trobar feina. Per altra banda, la quantitat de seguiments és el factor amb una correlació més gran envers l'augment en trobar feina d'una persona. En aquest cas cal tenir en compte la característica de correlació i no de causalitat, ja que possiblement el fet que una persona trigui més a trobar feina provocarà que rebi un major nombre de seguiments.

Anàlisi en clau de gènere de diferents aspectes de l'entitat (enllaç Power BI)

Possibilitats de millora per part del servei

Crear més variables (columnes)

Fer altres prediccions. Per exemple:

A quantes ofertes s'ha d'inscriure una persona per a trobar feina

Evolució de la quantitat d'usuaris, d'empreses, de socis...

Predicció de subvencions

Factors que determinen l'aprovació o no de les subvencions

Automatitzar la creació de la base de dades i de les prediccions a partir de la creació d'una API (protocol que connecta la base de dades del CRM amb el codi que crea les base de dades i les prediccions).

Informe de propuestas de millora en la recollida i gestió de les dades

Aquest informe recull les diferents propostes de millora en el registre i recollida de dades

Millores detectades generals

Situació inicial	Accions a realitzar	Situació final
Diferents criteris en l'entrada dels camps	Establir criteris comuns Utilitzar camps tancats	Possibilitat d'explotar les dades
Camps buits	Establir obligatorietat en certs camps	Disminuir camps buits
Camps amb masses valors	Determinar quines categories són importants i seleccionar	Augmentar la possibilitat d'explotar les dades i disminuir la dificultat per escollir a quin camp pertany la categoria.

En la mesura del possible caldria limitar al mínim els camps oberts i establir que si certs camps no estant omplerts el sistema els registri com a 0 per tal de poder treballar millor amb les dades.

Millores detectades per base de dades

Base de dades: Persones

Variable	Situació inicial	Accions a realitzar	Situació final
ID	Hi ha varies persones assignades a diferents ID (número únic), la qual cosa provoca duplicats en el propi CRM i errors en l'explotació de dades (possible error en la migració)	Eliminar duplicats al CRM	Facilitar l'explotació de dades i evitar errors per duplicitat
Codi postal, província i població	Dades recollides de codi postal, província i població per separat.	Mantenir només el codi postal i deduir les altres a partir d'un diccionari dins el CRM.	Evitar temps i errors.

Base de dades: Assistències

Variable	Situació inicial	Accions a realitzar	Situació final
	Les assistències no es vinculen a un usuari, si no que se'n crea un de nou	Incloure les assistències dins la mateixa base de dades que persones	Poder crear les dades i fer més fàcil la inscripció als projectes o activitats.

Base de dades: Organitzacions

Variable	Situació inicial	Accions a realitzar	Situació final
Codi postal, província i població	Dades recollides de codi postal, província i població per separat.	Mantenir només el codi postal i deduir les altres a partir d'un diccionari dins el CRM.	Evitar temps i errors.

Base de dades: Dades formatives

Variable	Situació inicial	Accions a realitzar	Situació final
Nivell formatiu	<p>Categorització dels nivells formatius no comparables a dades oficials.</p> <p>Camp omplert manualment</p>	<p>Utilitzar un diccionari oficial (IDESCAT)</p> <p>Crear variable a base de dades persones on reculli la formació de més alt nivell.</p>	<p>Dades fàcilment comparables a dades oficials</p> <p>Possibilitat d'establir el nivell formatiu de la persona, en el cas que es reculli més d'una formació</p> <p>Creació automàtica de columnes de nivell de formació.</p>
Codi postal, província i població	<p>Dades recollides de codi postal, província i població per separat.</p>	<p>Mantenir només el codi postal i deduir les altres a partir d'un diccionari dins el CRM.</p>	<p>Evitar temps i errors.</p>

Base de dades: Experiència laboral i Candidatures

Variable	Situació inicial	Accions a realitzar	Situació final
Llocs de feina	<p>Dades recollides en camp obert</p>	<p>Utilitzar un diccionari oficial (IDESCAT)</p>	<p>Dades fàcilment comparables a dades oficials</p> <p>Possibilitat d'explotar les dades</p> <p>Creació automàtica de columnes de tipus de feina.</p>

Base de dades: Dades discapacitat

Variable	Situació inicial	Accions a realitzar	Situació final
	Actualització de dades incloses com a registre nou (possible error en la migració)	Actualitzar enlloc de crear noves dades	Disminuir errors i facilitar per explotar les dades

Base de dades: Reserves

Variable	Situació inicial	Accions a realitzar	Situació final
	La base de dades no permet accedir al recurs reservat	Vincular el recurs reservat a la base de dades	Analitzar l'ús dels recursos

Annexes

[Classificació catalana d'ocupacions 2011 \(ca\).](#)

[Classificació catalana d'educació 2020 \(classificació del nivell de formació assolit\)_\(ca\).](#)