

CIDOBOT: Classificació temàtica automatitzada d'anuncis oficials mitjançant IA

L'objectiu d'aquest document és explicar els punts clau que hauran de recollir-se perquè una Prova de Concepte (PdC) quedi correctament documentada per a la seva possible utilització posterior. Es detallen a continuació cadascun dels apartats clau.

1 Introducció

- **Objectiu de la PdC**

Disposar d'una solució basada en IA per automatitzar la classificació temàtica d'anuncis oficials, millorant l'eficiència dels processos interns del servei .

- **Context i justificació**

La classificació manual d'anuncis oficials suposava una càrrega de treball elevada amb poc valor afegit. La PdC respon a la necessitat d'optimitzar aquest procés mitjançant tecnologies emergents, contribuint a la modernització de l'administració pública.

- **Abast de la prova de concepte**

Classificació temàtica d'anuncis provinents de BOPB, BOPG, BOPL, BOPT, DOGC i BOE en àmbits com processos selectius, subvencions, contractació, normativa local i convenis col·lectius.

2 Descripció del problema

- **Problema a resoldre**

La gestió manual de grans volums d'anuncis oficials dificultava l'eficiència del servei. Calia una solució automatitzada capaç de classificar correctament els anuncis i reduir la càrrega de treball.

- **Serveis afectats**

Servei del Butlletí Oficial de la Província i Documentació i Publicacions Oficials (SBOPiDiPO).

- **Beneficis esperats de la solució**

- Pel ciutadà
Accés més ràpid i fiable a la informació pública.
- Pel personal de l'administració
Reducció de tasques manuals i dedicació a funcions estratègiques.
- Altres possibles beneficis i/o beneficiaris
Millora de la transparència i reducció de costos operatius.

3 Escenari de la prova de concepte

- **Requisits Funcionals / No Funcionals**

Requisits funcionals:

- Classificació temàtica automatitzada
- Processament multilingüe (català i castellà)
- Informació del grau de confiança
- Recollida d'errors per reentrenament

Requisits no funcionals: consideracions sobre rendiment, seguretat, etc

- Alta precisió
- Escalabilitat
- Integritat de la informació
- Adaptabilitat a diferents contextos

- **Recursos necessaris (humans, tècnics, financers)**

- Humans: Personal SBOPiDiPO per entrenament i validació
- Tècnics: Xarxes neuronals, interfície web, servidor corporatiu
- Financers: Cost de desenvolupament assumit pel proveïdor

- **Ajuntaments i/o organismes col·laboradors a la PdC (funcions i contactes)**

No aplica

4 Metodologia

- **Enfocament General (e.g., supervisat, no supervisat)**

Aprenentatge automàtic supervisat

- **Eines i tecnologies utilitzades**

Deep Learning amb xarxes neuronals

- **Procés de desenvolupament i proves**

Entrenament amb resums d'anuncis classificats manualment durant 18 mesos.
Implementació de xarxes neuronals per cada butlletí.

5 Implementació

- **Descripció de la solució proposta**

CIDOBOT analitza automàticament els resums d'anuncis i retorna la classificació temàtica amb grau de confiança.

- **Arquitectura del sistema**

- Xarxes neuronals per cada butlletí
- Diccionari de paraules més repetides
- Sistema escalable i replicable

6 Flux de dades i algorismes utilitzats

- **Recol·lecció de dades**

Anuncis oficials diaris

- **Pre-processament de dades**

Vectorització de textos, eliminació de puntuació, diccionari de paraules

- **Algorismes Utilitzats**

Xarxes neuronals amb aprenentatge supervisat

- **Flux de processament**

Entrenament, predicció i retroalimentació per millora contínua

7 Resultats

- **Resultats obtinguts**

- 700 anuncis diaris processats
- Estalvi anual equivalent a 1 mes de treball
- Alta precisió en la classificació

- **Mètriques d'avaluació**

- Grau de confiança
- Capacitat de reentrenament

- **Comparació amb Objectius Inicials**

Objectius assolits: eficiència, precisió, transparència

8 Seguretat i privacitat de les dades. Avaluació de riscos

No s'apliquen protocols específics, ja que es tracta d'informació pública.

9 Anàlisi i Discussió

- **Lliçons Apreses**

- Disseny centrat en l'usuari
- Adaptació lingüística per butlletí
- Importància del cicle de millora contínua

- **Limitacions de la PdC**

No aplicables. Resultats satisfactoris.

10 Conclusions

- **Resum de troballes**

CIDOBOT ha demostrat ser una solució efectiva per automatitzar la classificació temàtica d'anuncis oficials, amb impacte positiu en eficiència, precisió i transparència.

- **Propers passos suggerits**

- Classificació per ODS
- Generació automàtica de resums
- Automatització de la gestió documental

11 Recomanacions per a futures Implementacions (**)

- **Requisits funcionals**

Disposar de dades prèviament classificades per entrenament

- **Cos ètic i de serveis**

Orientació a l'usuari i valor afegit per al personal públic

12 Referències (Si aplica)

- **Bibliografia i recursos consultats**

No aplica

- **Documentació addicional**

No aplica

13 Annexos (Si aplica)

- **Codis fonts**
No aplica
- **Dades de prova**
No aplica
- **Gràfics i visualitzacions**
No aplica