

IAInstàncies - Classificador d'Instàncies amb IA per automatitzar el procés per a l'Ajuntament de Vilafranca del Penedès.

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

1.1. Objectiu de la PdC

S'ha creat aquesta PdC amb la intenció d'automatitzar el procés de classificació d'instàncies genèriques que entren a l'Ajuntament de Vilafranca del Penedès, per tal de reduir la càrrega de tasques del personal administratiu encarregat de derivar-les diàriament.

1.2. Context i justificació

En un context on el servei encarregat de revisar totes les instàncies que entren a l'ajuntament, es troba saturat amb altres tasques, es proposa d'automatitzar-la en la mesura del possible i així alliberar el personal administratiu d'aquesta feina diària.

1.3. Abast de la prova de concepte

2. Descripció del problema

2.1. Problema a resoldre

Diàriament es poden rebre unes 25 instàncies genèriques a través de la plataforma de tramitació de l'ajuntament, que s'han de derivar als departaments corresponents. Al disposar d'un ampli volum de dades històriques d'instàncies genèriques ja derivades, es proposa crear un model entrenat amb aquestes dades per tal de predir en noves instàncies el servei destinatari.

2.3. Serveis afectats

El principal servei afectat serà el de tramitació electrònica, en concret el gestor d'expedients que s'utilitza a l'Ajuntament, que és TAO de l'empresa TSYSTEMS

2.4. Beneficis esperats de la solució

- Pel ciutadà
Millorar els temps d'espera dels expedients, al derivar automàticament la petició al servei destinatari.
- Pel personal de l'administració
Alleugerar la tasca diària de revisió de les peticions que entren.
- Altres possibles beneficis i/o beneficiaris

3. Escenari de la prova de concepte

3.1. Requisits Funcionals / No Funcionals

3.2. Recursos necessaris (humans, tècnics, financers)

3.3. Ajuntaments i/o organismes col·laboradors a la PdC (funcions i contactes)

4. Metodologia

4.1. Enfocament General (e.g., supervisat, no supervisat)

Amb la intenció d'ajudar al personal administratiu de l'OAC en la tasca de classificació d'instàncies, el projecte es dividirà en tres parts:

- a. La primera determinarà si a la instància hi ha informat el codi d'expedient, anotació o notificació per poder classificar-la directament.
- b. En cas que no contingui aquesta informació es passarà pel model entrenat per obtenir 5 resultats principals de classificació per punts.
- c. Si el model dona un primer resultat amb una diferència de més de 20 punts respecte el segon resultat, la instància es derivarà directament al servei resultant. En cas que la diferència sigui menor de 20 punts, la instància quedarà pendent i s'afegirà una observació per tal que l'administratiu de l'OAC pugui veure quina predicció ha realitzat el procés i si és correcte o no.

4.2. Eines i tecnologies utilitzades

S'ha escollit l'algoritme de *Machine Learning Random Forest* per la classificació amb *scikit-learn* de *Python*.

Pandas per la depuració i tractament de les dades.

4.3. Procés de desenvolupament i proves

El primer punt és recopilar totes les dades històriques per després establir el procés de neteja, depuració i estructuració de les dades que formaran part de l'entrenament del model.

Un cop tenim aquestes dades preparades, buscar la parametrització més adient per crear el model i obtenir el millor resultat. Aquest punt és clau per obtenir el millor model possible.

Posteriorment programar tota la lògica de tractament de les instàncies noves que arriben. Primer comprovar que hi hagi informat el codi d'expedient, anotació o notificació dins la instància a través de patrons *REGEX*, en cas contrari, passar les dades pel model i obtenir els resultats de la classificació.

5. Implementació

5.1. Descripció de la solució proposta

El projecte està format per:

- Un model creat a partir de les dades recopilades.
- Un procés que s'executa cada 15 minuts per processar totes les instàncies que hagin entrat noves.
- Un *Dashboard* per fer proves d'instàncies, consultar mètriques del procés de classificació i llistat de totes les instàncies classificades amb opció d'afegir notes en cas que no hagi sigut correcte.

5.2. Arquitectura del sistema

6. Flux de dades i algorismes utilitzats

6.1. Recol·lecció de dades

S'obtenen aproximadament 28.000 Instàncies amb els camps "Exposo", "Demano" i el servei a qui es va redirigir manualment, de l'històric del gestor d'expedients.

6.2. Preprocessament de dades

Preparació de dades: Es passen les dades per una sèrie de processos per netejar els textos dels camps "Exposo" i "Demano" tals com treure accents, passar tot a

minúscules, eliminar paraules amb números i numeracions, eliminar articles i apòstrofs, espais extres, paraules d'una sola lletra i passar pel procés “*Stopwords*” configurat al català.

Posteriorment es passa per un procés de potenciador de paraules clau per tal de donar més importància a paraules que realment tinguin un significat dins el text.

6.3. Algorismes Utilitzats

6.4. Flux de processament

7. Resultats

7.1. Resultats Obtinguts

Després d'un període de dos mesos de proves on el personal administratiu del servei de l'OAC han anat comprovant els resultats de les prediccions que realitzava el model, i anar rectificant tant el model com els processos de tractament de les dades, hem posat en producció el projecte i podem determinar els següents resultats: en un any en funcionament el projecte ha processat aproximadament 5.000 instàncies, de les quals s'han derivat automàticament per identificació de codis 1.300, per classificació amb +20 punts de diferència 2.150 i s'ha informat de la predicció sense derivar unes 1.550. Això significa que s'han derivat automàticament unes 3.450. Amb un percentatge d'encert d'aproximadament 90%.

7.2. Mètriques d'avaluació

Hem obtingut els resultats dels *logs* del procés que junt amb el *dashboard* creat, podem veure els errors que el personal administratiu de l'OAC han detectat.

7.3. Comparació amb Objectius Inicials

Hem obtingut els resultat esperat.

8. Seguretat i privacitat de les dades. Avaluació de riscos

9. Anàlisi i Discussió

9.1. Lliçons Apreses

9.2. Limitacions de la PdC

10. Conclusions

10.1. Resum de troballes

10.2. Propers passos suggerits

Entrenar el model cada cert temps incloent les noves instàncies entrades per tal de mantenir un model actualitzat.

11. Recomanacions per a futures Implementacions ()**

11.1. Requisits funcionals

11.2. Cos ètic i de serveis

12. Referències (Si aplica)

12.1. Bibliografia i recursos consultats

12.2. Documentació addicional

13. Annexos (Si aplica)

13.1. Codis fonts

13.2. Dades de prova

13.3. Gràfics i visualitzacions