



IAB - CE Intel·ligència Artificial i Big Data
M03 - Programació en Intel·ligència Artificial
U5 - Gestió de projectes d'intel·ligència artificial

EAC6
(Curs 2024-25 / 1r semestre)

Presentació i resultats d'aprenentatge

Aquest exercici d'avaluació continuada (EAC) es correspon amb els continguts treballats a la unitat 5 Gestió de projectes d'intel·ligència artificial.

Forma i data de lliurament

Un cop finalitzat l'exercici d'avaluació continuada heu d'enviar el document des de l'apartat **M6B1 Lliurament EAC1** de l'aula, dins del termini establert. Tingueu en compte que el sistema no permetrà fer lliuraments després de la data i hora indicades.

El nom del fitxer serà el següent: **DAW_M06B1_EAC1_Cognom1_Inicial del cognom2.zip** Els cognoms s'escriuran sense accents. Per exemple, l'estudiant *Joan García Santos* posaria el següent nom al seu fitxer de l'EAC1: **DAW_M06B1_EAC1_Garcia_S.zip**.

Substituïu **Nom i cognoms** de la capçalera per les vostres dades personals.

El termini de lliurament finalitza a les **23:55 h** del dia **25/10/2024**. La proposta de solució de l'EAC es publicarà el dia **28/10/2024** i les qualificacions el dia **04/11/2024**.

En aquest document trobareu un conjunt d'exercicis que es corresponen al darrer exercici d'avaluació contínua (EAC6) del mòdul.

Entregareu el fitxer **IAB_M03_EAC6_cognom_nom.zip**, on *cognom* i *nom* els haureu substituït pel vostre cognom i nom respectivament, sense utilitzar accents i caràcters especials. Per exemple, l'alumne *Jordi Caño Garcia* entregaria el fitxer **IAB_M03_EAC6_cano_jordi.zip**.

El fitxer zip inclourà dues carpetes: la carpeta *portcanto/* i la carpeta *solucio/*. La carpeta *portcanto/* és pròpiament la solució del projecte Python, codi que també estarà pujat a Github tal com es demana el darrer exercici. La carpeta *solucio/* contindrà una carpeta per cada exercici (P1...P11), i dins de cada carpeta incloureu allò que es detalla en cada exercici (fitxers plans de Python, captures de pantalla, fitxers plans de text amb explicacions,...). Aquest EAC6 no es pot fer en quadern de Jupyter (ja sigui local o Colab).

- **És necessari incloure comentaris** del vostre codi, que expliquin com s'ha implementat la solució del problema plantejat.
- S'ha d'utilitzar rutes de fitxers relatives. Típicament el dataset estarà dins la carpeta *data/* de l'arrel principal del projecte.
- En totes les captures de pantalla que es demanen, l'alumne ha de demostrar que s'han generat en la màquina de l'alumne. Ha de sortir el nom de l'alumne, per demostrar l'autoria.

	Codi: I71	Exercici d'avaluació contínua 6	Pàgina 1 de 5
	Versió: 02	CEI_AB_M03_EAC6_Enunciat_2425S1	Lliurament: 20/12/2024

- Cadascun dels apartats inclou la seva puntuació. La puntuació no es correspon només si el codi funciona correctament i el resultat és el correcte, sinó que també es puntuaran aspectes com els comentats anteriorment (ús de comentaris, etc). La puntuació indica el pes que té cada apartat sobre la nota final de l'EAC.
- **És imprescindible** citar les referències consultades per a realitzar l'activitat. Es valorarà que el codi proporcionat solucioni el problema proposat i també la qualitat del codi (comentaris, llegibilitat, claredat, ús de les estructures de dades adequades, bona nomenclatura de les variables i funcions).
- Recordeu que en el mòdul heu acceptat els compromisos d'honestedat acadèmica. No es pot compartir informació entre els companys; s'ha de citar les fonts utilitzades; el fòrum de l'aula és el canal correcte per posar dubtes.
- A l'aula s'especifica el dia i hora límits per a l'entrega. No s'admetran entregues fora de termini ni per correu electrònic.

Enunciat

El **Port del Cantó** és un port de muntanya que uneix les comarques de l'Alt Urgell (Adrall) i el Pallars Sobirà (Sort). Són 18Km de pujada i 18Km de baixada, que típicament es puja entre 54 i 77min; i es baixa entre 24 i 36min. Es generaran dades sintètiques que simularan una cursa ciclista entre Adrall i Sort. Es treballarà sobre aquestes dades.

- Per fer els següents exercicis es proporciona el fitxer **portcanto_esquelet.zip**, que descomprimireu en la carpeta *portcanto/*, que serà l'arrel del projecte. Aquesta carpeta contindrà l'estructura bàsica del projecte, amb scripts python que haureu de completar.

P1 (1p). script generardataset.py

Dins del script generardataset.py es proporciona el diccionari que s'utilitzarà per generar el dataset, i s'adjunten els comentaris suficients per entendre el que es pretén.

Dins de la carpeta *P1/* s'adjuntarà:

- Document de text (fitxer llegir.txt), on explicareu el vostre entorn on heu executat el projecte (local, venv, docker, i aquella informació que creieu que pugui ser necessària per al professor i la correcció de l'exercici).
- script *generardataset.py*
- captura de pantalla de l'execució del script
- captura de pantalla del dataset generat: ciclistes.csv
- fitxer ciclistes.csv

P2 (1p). script clustersciclistes.py

El script clusterciclistes.py realitza diferents tasques:

	Codi: I71	Exercici d'avaluació contínua 6	Pàgina 2 de 5
	Versió: 02	DAW_M06B1_EAC1_Enunciat_2223S1	Lliurament: 20/12/2024

- càrrega del dataset
- Anàlisi exploratori de les dades
- Neteja de les dades (elimina les columnes id i tt).
- Extracció dels labels: guardem les etiquetes dels ciclistes (BEBB, ...), i seguidament eliminar el tipus del dataframe
- definir la funció `visualitzar_pairplot()`, de manera que ja es pot apreciar de forma qualitativa si existeixen clústers.
- model de clustering amb KMeans de scikit-learn.
- guardar el model amb el nom `model/clustering_model.pkl`
- Calcular l'homogeneïtat, completa i v-measure. Guardar aquests models a `model/scores.pkl` en forma de diccionari.
- visualitzar els clusters
- afegim la columna label al dataframe
- com que obtindreu 4 labels (0, 1, 2, 3), s'ha de saber cadascun dels labels a quin cluster es correspon (BEBB, BEMB, MEBB, MEMB). Aquesta informació la guardareu en una estructura de dades adient, i la guardareu al fitxer `model/tipus_dict.pkl`.
- generació d'informes: fitxers `informes/BEBB.txt`, etc.
- classificació de nous valors. Es proporciona 4 valors de ciclistes (id, temps de pujada, temps de baixada, temps total), i s'han d'associar al cluster corresponent.

Dins de la carpeta P2/ s'adjuntarà:

- script `clustersciclistes.py`
- captura de pantalla de l'execució del script,
- contingut de la carpeta `img/`, amb les imatges que s'han generat

P3 (1p). carpeta model/

Amb la llibreria *pickle* hauràs hagut de generar els següents fitxers:

- `clustering_model.pkl`
- `scores.pkl`
- `tipus_dict.pkl`

Expliqueu amb les vostres paraules per què serveix guardar els models; per què és necessària una llibreria com *pickle*.

Expliqueu amb les vostres paraules com heu fet l'associació entre els 4 tipus de ciclistes i les classes que ha trobat el model. Adjunteu aquesta informació dins de la carpeta P3/ en un fitxer de text.

P4 (1p). Generació dels informes (carpeta informes/)

Expliqueu amb les vostres paraules el contingut dels fitxers que hi ha en aquesta carpeta. Adjunteu aquesta informació dins de la carpeta P4/, en un fitxer de text.

	Codi: I71	Exercici d'avaluació contínua 6	Pàgina 3 de 5
	Versió: 02	DAW_M06B1_EAC1_Enunciat_2223S1	Lliurament: 20/12/2024

P5 (0.5p). pylint: anàlisi de codi estàtic

- Fer l'anàlisi estàtic dels scripts *generardataset.py* i *clustersciclistes.py*
- sortida de pantalla de l'anàlisi estàtic que heu realitzat, amb un score ≥ 9 (adjunteu captura de pantalla)
- Comenteu 5 correccions que heu hagut de fer en el vostre codi (adjunteu fitxer de text)

P6 (0.5p). Documentació

Heu documentat correctament les funcions amb docstrings

- Executeu les comandes per generar la documentació
- Adjunteu la captura de pantalla amb la sortida de les comandes que acabeu d'executar
- Adjunteu les captures de pantalla amb la documentació generada

P7 (1p). Tests

Dins la carpeta *tests/* creeu el script *testportcanto.py* (se us proporciona l'esquelet), on codificareu dues classes:

classe *TestGenerarDataset*, on definireu tres funcions:

- `test_longituddataset`
- `test_valorsmitjatp`
- `test_valorsmitjatb`

classe *TestClustersCiclistes*, on definireu tres funcions:

- `test_check_column`
- `test_data_labels`
- `test_model_saved`
- Executeu les comandes per passar els tests
- Adjunteu la captura de pantalla amb la sortida de les comandes que acabeu d'executar, on es demostra que es passen els 6 tests.
- Expliqueu amb les vostres paraules la importància de fer tests unitaris (document de text en la carpeta P7/).

P8 (1p). Predicció de nous valors

	Codi: I71	Exercici d'avaluació contínua 6	Pàgina 4 de 5
	Versió: 02	DAW_M06B1_EAC1_Enunciat_2223S1	Lliurament: 20/12/2024

```
# Classificació de nous valors
nous_ciclistes = [
    [500, 3230, 1430, 4660], # BEBB
    [501, 3300, 2120, 5420], # BEMB
    [502, 4010, 1510, 5520], # MEBB
    [503, 4350, 2200, 6550] # MEMB
]
```

- Adjunteu la sortida per pantalla on demostreu que heu classificat correctament aquests nous valors

P9 (1p). MLflow (script)

- Codifiqueu el script `mlflowtracking-K.py` (se us proporciona l'esquelet), on es crea a MLflow l'experiment "K sklearn ciclistes", i on es fa una variació del paràmetre K des de K=2 fins a K=8, i es guarden els resultats a MLflow. Adjunteu aquest script.

P10 (1p). MLflow (interfície gràfica)

- Feu una captura de pantalla amb els resultats que heu obtingut en l'experiment que es defineix en l'script `mlflowtracking-K.py`.

Comenteu els resultats (document de text en la carpeta *P10/*). Concretament:

- Quin valor de K té millor v-measure? És K=4 el millor per fer el clustering?
- Quin valor de K té millor completesa? A partir de la imatge que has guardat a la carpeta *img/*, és una bona opció agafar K=2?

P11 (1p). Github

- En un document de text dins la carpeta *P11/*, escriu la url del vostre projecte a Github.
- Expliqueu el flux de treball que heu de fer per exportar el vostre projecte a Github.
- Escriu les comandes per fer el *push* de tot el projecte a Github. Expliqueu breument què fa cada comanda.
- Escriu la url del teu projecte penjat a Github.
- Adjunteu una captura de pantalla del vostre projecte a Github
- No us oblideu dels fitxers *README.md* (ben documentat), *requirements.txt* i *LICENSE.txt*.

	Codi: I71	Exercici d'avaluació contínua 6	Pàgina 5 de 5
	Versió: 02	DAW_M06B1_EAC1_Enunciat_2223S1	Lliurament: 20/12/2024