

Fonaments de Ciència de Dades

Pràctica 1 (25% nota final)

Criteris d'avaluació generals de la PAC.

És important tenir en compte les següents consideracions a l'hora de lliurar la pràctica:

- És obligatori i queda com a responsabilitat de l'estudiant revisar que l'arxiu lliurat al RAC és el correcte. Un arxiu buit o no pertinent es considerarà com no lliurat.
- Perquè el lliurament es consideri com realitzat, s'ha de completar almenys el 25% de tota l'activitat.
- No podrà modificar-se cap element de la pràctica passada la data de lliurament (repositori, arxius de Google Drive, etc.).
- Així mateix, també és responsabilitat de l'estudiant assegurar-se que, en el moment del lliurament de la pràctica, s'ha donat accés al professor als diferents elements privats que es lliurin (p. ex., repositori GitHub privat o arxius restringits de Google Drive). El professor indicarà al Tauler de l'aula el seu nom d'usuari en aquestes plataformes.

Autors: Antoni Espadas Navarro et Jordi Samaniego Vidal

1. Context

Explicar en quin context s'ha recollit la informació. Explicar per què el lloc web triat proporciona aquesta informació. Indicar l'adreça del lloc web.

En aquest projecte, s'ha decidit investigar sobre les interaccions de votacions entre els diferents països participants d'Eurovisió al llarg de la cronologia del festival. Des del 1956 s'han anat integrant països i de les set nacions inicials comptem, aquest 2022, amb 40 països participants, entre els quals se'n troben alguns de tan sorprenents com Austràlia. L'etiqueta "Eurovisió" no es refereix a països europeus, políticament ni geogràficament parlant, sinó als països membres de la Unió Europea de Radiodifusió. Es retransmet a tot el món, fins i tot a països no participants; i, premiant la seva fidelitat, de vegades es convida a participar certs països sorprenents com Austràlia.

Tot i això, s'acosta la propera edició del Benidorm Fest , el certamen que serveix com a avantsala d' Eurovisió i en què es decidirà qui ens representarà al Festival de la Cançó. Després de la polèmica decisió de l'any passat amb Chanel Terrero que, no obstant, es va alçar amb el millor lloc per al nostre país en molts anys, serà el moment de trobar un successor que aconsegueixi estar a l'alçada.

Per una banda, la gala d'obertura del Benidorm Fest tindrà lloc el diumenge 29 de gener de 2023. El seguirà la primera semifinal el dimarts 31 de gener i la segona semifinal el dijous 2 de febrer . La final serà el dissabte 4 de febrer , el guanyador del qual ens representarà a la gala eurovisiva del 13 de maig. S'emetrà a La 1, amb una durada de 60 minuts (semifinals) i 80 minuts (la final).

Per l'altra banda, Eurovisió 2023 ja té seu. Els 37 països participaran a Liverpool i serà la ciutat de la qual aculli la 67a edició del certamen davant la impossibilitat de celebrar-se a Ucraïna, guanyadora d'Eurovisió 2022, a causa de la invasió russa. La Unió Europea de Radiodifusió (UER) han confirmat també les dates en què se celebrarà les semifinals i la final del festival: els dies 9, 11 i 13 de maig , respectivament, al Liverpool Arena (també conegut com a M&S Bank Sorra).

2. Títol

Definir un títol que sigui descriptiu pel dataset:

Quin és el pronòstic del futur guanyador d'Eurovisió 2023 segons les seves aliances geopolítiques?

3. Descripció del dataset.

Desenvolupar una descripció breu del conjunt de dades que s'ha extret. És necessari que aquesta descripció tingui sentit amb el títol escollit.

Tal com expressa el títol, el dataset està basat sobre les dades més rellevants sobre històrics de votacions cronològics i 'win rates' de diferents cases d'apostes esportives d'enguany per veure el pronòstic com fluctua durant aquestes dates fins arribar el dia de celebració del festival. A més, per afegir valor al pronòstic, s'ha fet un rastreig fent crida a l'API rest de Twitter per detectar el número de Tweets recents amb l'etiqueta de '#Eurovision'. Concretament hi trobem una varietat de 5 arxius CSV, on trobem llistat de tots els països, els artistes, l'any en qüestió, les seves cançons, ranking, punts per votació (tant rebudes com efectuades) i el tipus vot (televisió, jurat i el total).

4. Representació gràfica.

Dibuixar un esquema o diagrama que identifiqui el dataset visualment i el projecte escollit.



5. Contingut.

Explicar els camps que inclou el dataset i el període de temps de les dades.

Countries.csv:

- country_code: nomenclatura del país
- country_name: nom complet del país

Countries_Editions_Songs.csv:

- country: idem country_name
- years: l'any de la edició del festival d'Eurovisió
- songs: cançó concursada
- artists: canta autor del festival
- places: posició en la ronda final
- points: vots rebuts
- qualification: qualificació en semi-finals

Odds_Rate_Table_2023.csv:

- country: ídem country_name
- winning chance: % probabilitats guanyar
- scrapped_day: dia quan s'emmagatzema la informació
- 'et als': floats de quota de pagament en casa d'apostes

Polled_Eurovision_2022.csv (enquesta de qui guanyaria el festival al 2022):

- country: idem country_name
- winrate: % de votants
- n_vote: num total de votants

Tweets_Count_With_Eurovision_Hashtag.csv (API Twitter):

- end_timestamp: enregistrament de la data final
- start_timestamp enregistrament de la data inicial
- tweet_count: número de Tweets amb l'etiqueta #Eurovision

Votes_From_Scoreboard.csv:

- edition_year: l'any de la edició del festival d'Eurovisió
- country_code_column: idem country_code
- value: total de punts
- country_code_row: idem country_code
- vote_type: tipus de votació

6. Propietari.

Presentar el propietari del conjunt de dades. És necessari incloure cites d'anàlisis anteriors o, en cas de no haver-n'hi, justificar aquesta cerca amb anàlisis similars. Justificar quins passos s'han seguit per actuar d'acord amb els principis ètics i legals en el context del projecte.

La pàgina web Eurovisionworld és una comunitat de fans per a tothom que tingui passió pel Festival de la Cançó d'Eurovisió. No estan afiliats a EBU ni a l'organització darrere del Festival de la Cançó d'Eurovisió. Totes les imatges d'eurovisionworld.com estan disponibles a Internet i es creu que són de domini públic. Es creu que les imatges publicades es publiquen d'acord amb la Llei d'ús just dels drets d'autor dels EUA (títol 17, codi dels EUA).

S'accepteu utilitzar eurovisionworld.com només per a finalitats lícites i d'una manera que no infringeixi els drets d'altres persones ni limiti o dificulti l'ús i el gaudi d'eurovisionworld.com per part d'altres persones. Tots els comentaris, suggeriments, aportacions, continguts, etc. són més que benvinguts, ens comenta la comunitat.

Tot i això, hi ha diverses anàlisis trobades a internet. Exemple d'aquests són:

- <https://www.oddschecker.com/tv/eurovision/winner>
- <https://www.nicerodds.es/>
- <https://eschome.net/>
- https://en.wikipedia.org/wiki/Eurovision_Song_Contest_2022

7. Inspiració.

Explicar per què és interessant aquest conjunt de dades i quines preguntes es pretenen respondre. És necessari comparar amb les anàlisis anteriors presentades a l'apartat 6.

A l'últim festival d'Eurovisió del dia 14 de maig de 2022, Ucraïna va aconseguir el micròfon de vidre després de destronar el Regne Unit i Espanya en un tomb inesperat durant el tele vot. Així, mentre que el jurat professional va jutjar que Ucraïna havia de quedar a la cinquena posició, el jurat popular va decidir donar-li més de 400 punts, la qual cosa el va col·locar a gairebé dos-cents punts de diferència amb la següent més votada.

Tot i això, no ha estat això el que ha generat tantíssima divisió en la victòria d'Ucraïna. Tot i que les xarxes socials han despertat amb el hashtag "#Tongazo", la gran majoria del públic ha acceptat la victòria del país del Do. El que realment ha causat una certa preocupació són el actes i reivindicacions polítiques no haurien d'estar permeses al Festival.

No obstant això, ja des de fa dècades experts avalen que Eurovisió, tal i com la coneixem, ha deixat de ser un concurs musical i que abans de conèixer música, és preferible tindre coneixements de geopolítica per encertar en les puntuacions finals. De fet, hi han estudis que s'endinsen en aquest sector, com El sociòleg israelià Gad Yair va analitzar les votacions a Eurovisió entre 1975 i 1992, i va identificar tres grans àrees basades en sentiments i interessos: el Bloc Occidental, el Bloc del Nord i el "difús" Bloc Mediterrani on hi hauria Espanya)¹.

En aquesta present és vol fer una aproximació d'aquelles interaccions cronològiques dels festivals d'Eurovisió i les seves votacions corresponent per mesurar fins a quin punt del mapa hi apareixen una forta relació entre diferents blocs intercomunitaris. Al mateix temps, analitzem les dades presents que puguin afectar de cara al pròxim festival a Liverpool que es celebrarà els propers dies 9, 11 i 13 de maig del 2023.

¹ Consultat a 10 de Novembre del 2022:

<https://www.lavanguardia.com/television/tv-or-not-tv/20160514/401790684306/geopolitica-bloques-paises-eurovision-espana.html>

8. Llicència.

Seleccionar una d'aquestes llicències pel dataset resultant i justificar el motiu de la seva selecció. Exemples de llicències que poden considerar-se:

- *Released Under CC0: Public Domain License.*
- *Released Under CC BY-NC-SA 4.0 License.*
- *Released Under CC BY-SA 4.0 License.*
- *Database released under Open Database License, individual contents under Database Contents License.*
- *Altres (especificar quina).*

Una possible llicència per a aquest conjunt de dades pot ser **CC BY-SA 4.0 License**. La elecció es basa en la idoneïtat de les clàusules que s'hi presenten en relació amb el treball realitzat on:

- Se'n permet l'ús comercial, cosa que incrementa les possibilitats que empreses puguin interessar-se en les dades generades, permetent així, la realització de nous projectes que reportin un reconeixement a l'autor original.
- Les noves contribucions han de ser publicades sota la mateixa llicència, ho que permet que se'l reconegui a l'autor original en tot moment i sota els mateix termes que van ser plantejats per ell.
- Es proveeix el nom del creador del conjunt de dades generat i s'indiquen els canvis realitzats sobre aquest. D'aquesta manera, es reconeix el treball de tercers i en quina mesura es van fer aportacions pel que fa al treball original.

9. Codi.

Codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

- *El codi haurà de situar-se a la carpeta /source del repositori.*
- *S'han d'indicar les llibreries i versions utilitzades. P. ex., en Python poden obtenir-se mitjançant la comanda `pip3 freeze > requirements.txt`*
- *Al document PDF s'han de comentar els aspectes més rellevants sobre com el codi realitza el procés de recol·lecció de dades, quines dificultats presenta el lloc web triat, i com les heu resolt.*

El codi va ser fet en llenguatge Python amb la implementació de la llibreria Selenium i BeautifulSoup:

<https://github.com/anthonyswords/ScrappingEurovision23>

10. Dataset.

Publicar el dataset obtingut en format CSV a Zenodo, incloent-hi una breu descripció. Obtenir i adjuntar l'enllaç del DOI del data set (<https://doi.org/...>). El dataset també haurà d'incloure's a la carpeta /dataset del repositori.

EL DOI del data set en format CSV: <https://doi.org/10.5281/zenodo.7323675>

11. Vídeo

Realitzar un breu vídeo explicatiu de la pràctica (màxim 10 minuts), que haurà de comptar amb la participació dels dos integrants del grup. Al vídeo s'haurà de realitzar una presentació del projecte, destacant els punts més rellevants, tant de les respostes als apartats com del codi utilitzat per a extreure les dades. Indicar l'enllaç del vídeo (<https://drive.google.com/...>), que haurà d'estar al Google Drive de la UOC.

L'enllaç del vídeo:

https://drive.google.com/drive/folders/1RYfT0NleyW--xX-hkMI_yjZX6t1yuC_X?usp=share_link

Contribucions	Signatura
Investigació prèvia	A. E. N. et J. S. V.
Redacció de les respostes	A. E. N. et J. S. V.
Desenvolupament del codi	A. E. N. et J. S. V.
Participació al vídeo	A. E. N. et J. S. V.