Bones Josep Maria!

Et comento algunes coses!

**PAPER:**

- L'abstract pot ser més complet, has de poder-ho llegir i entendre tots els conceptes. A més d'incloure els els resultats de paper!

- Hauries de justificar el text (no centrar) i parlat en 3r persona o impersonal

- Abans de parlar directament de la base de dades, l'has de presentar!!! No pots dir: La database conté això i això. Has de començar amb una presentació, com per exemple: L'objectiu del nostre estudi és els jumps, presentar qué son els jumps!! i després dir qui com i quan va ser recollida la informació

- Has de presentar millor les variables, no parlar d'elles abans de presentar-les i descriure bé cada component de forma adecuada:

Què és USL, LSl, ext... I perquè ho has utilitzat

- Aquesta frase no s'entén... "a histogram that collects the availability of offering the benefits required of it.".

- Hauries de posar quants vols has eliminat amb lo de 30 min, perquè aqui el que veig és la possiblitat d'eliminar precisament el teu objectiu. Si sempre tens vols "perfectes" que són puntuals, no seràn afectats per aquestes coses. I això haura d'anar a metodologia!

- No expliques bé el que és l'error de despegue, no ho has posat a la descripció de les dades però si que sembla significant... I pel que entenc tens els jumps i tots els errors dels junps. Per tal com expliques sembla que l'error de despegue del primer avió afecta els altres!

- Em costa una mica entendre el teu treball i les teves conclusions, hauries de repassar 2-3 vegades com dius les coses abans d'entregar

- Crec que els objectius, les metodologies i els resultats no estàn del tot lligats. Crec que la part de Jumps s'entén més, però referent a quines coses afecten més al retard, que entenc que és el segon objectiu, no s'entén com has arribat a aquesta conclusió! i no puc trobar-ho tampoc a la tasca. Hi ha maneres de trobar la importància en aquests casos de les variables.

- Gràcies per els acknowlegmens!

**TASCA:**

- Al readme s'haura d'explicar cada fitxer que conté

- Fer un ressample és molt perillós, estàs invenrtant 5 vegades més dades de les que existeixen! Hi ha mètodes d'undersapling i ressampling juntes que t'ajudaràn més!

- Aquí has fet un fallo **greu**! Has fet ressampling de TOTES les dades!! Això no és correcte. Només s'aplica ressampling a les dades d'entrenament!! Tu el que intentes és que el teu model trobi els patrons i per això augmentem les dades desbalancejades, però el test sempre ha de ser sobre dades reals!

- M'agradaria que provessis amb un test real les teves conclusions. La teva lògica és utilitzar loggistic regression per mirar els coeficients i veure si afecta o no el error del 3r jump. Podries haver utilitzat algo semblant a https://sefiks.com/2021/01/06/feature-importance-in-logistic-regression/ per trobar la importància.

De la manera que parles dels resultats em falta saber quina és per exemple l'accuracy o f1 en cada cas. Al ser classificació, és molt millor que R2.

Crec que el més important és que expliquis bé com has arribat a veure quines variables afecten el retard, facis bé el ressampling i donguis les mètriques adequades! Amb les matrius de confusió també!

Seguim,

Aina