|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **titre du projet (dataset et tâche)**  **site web (source des data)** | **Étudiant.e** | (rapport,oral) Présentation du contexte, des données, des enjeux | **[Pre-processing, nettoyage des données, feature selection]** | **[visualisations des données]** | **(code) Exploitabilité du code** | [Technicité des approches envisagées] | **Exploration des hyper-paramètres** |
|  | Mr imparfait | 1 | 3 | 2 | 2 | 3 | 3 |
|  | Mme parfaite | 1 | 3 | 2 | 2 | 3 | 3 |
|  | Mr pas d’option | 1 | 0 | 0 | 2 | 0 | 3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| train/ validation / test ok? + overfitting absent/ maitrisé/ massif ? -1 à +1 (négatif permis) | autres bêtises (sur des sujets pourtant vus en cours) | **clarté des explications, des raisonnements** | **présentation, lisibilité du rapport (esthétique, présentation)** | **Qualité des figures : existence de figures, puis leur qualité (axe des x, des y, titre/légende, etc)** | [Autres mentions, bonus sur autres choses] | (code) Projet qui n’a globalement **pas** marché | Total brut |
| -1 | -2 | 2 | 2 | 3 | 1 | -2 | 17 |
| 1 | 0 | 2 | 2 | 3 | 1 | 0 | 23 |
| 0 | 0 | 2 | 2 | 3 | 0 | 0 | 13 |

bareme du projet

Ce barème est approximatif: c’est plus un guide pour homogénéiser la notation entre les profs, qu’une règle absolue.

Par ailleurs, avoir fait certaines choses, a moitié, ne donne pas les points. Exemple typique: utiliser des métohdes très complexes, mais sans avoir compris comment elles fonctionnent, ne rapporte pas plus de point qu’avoir utlisé seulement des méthodes simples (voire, c’est pire.. il vaut mieux bien expliquer qq méthodes simples)

Certaines colonnes sont en négatif: ce sont seulement des pénalités possibles, la “note max” de ces colonne est 0.

Certaines colonnes sont entourées de [crochets] : cela signifie que l’élément évalué est optionnel. Typiquement, 1 projet aura toujours certains de ces éléments optionnels présents, et sera donc noté sur 20 (et non sur 13). La somme de ces points optionnels est de 10 points

J'ai eu des questions intéressantes, je met mes réponses ici pour que tous en profitent:

la quantité de travail demandé est proportionnelle au nombre d'étudiants dans le groupe ?

- oui, on attend plus de choses d'un groupe de 4 que de 3, et plus d'un groupe de 3 que de 2, en termes de quantité de travail en tout cas (et éventuellement de qualité, mais même à 2 on attend la qualité !)

- non, le projet à 1 n'est pas une bonne chose. Mais à 2, oui, c'est autorisé.

- pour un projet à 2 personnes, on peut faire quelque chose de qualitatif mais plus réduit en quantité, par exemple une tache ou il y a bcp de pre-proc mais pas d'autres difficultés, alors que pour un groupe de 4 il faut qu'il y ait plus de choses, par exemple un pre-proc riche ET une optimisation des hyper-paramètres un peu plus poussée que juste en faire varier 1 (ce qui est un minimum).

Et également j'aimerais savoir si avoir un dataset et un objectif plus sophistiqué ou original est plus valorisé ?

Ou on peut tranquillement rester sur un dataset et un projet basique comme par exemple prédire le taux de crise cardiaque de patients ou quelque chose dans le style sans être dévalorisé.

Oui, en pratique on va mettre des notes qui vont monter plus haut sur des projets plus intéressants, même si ce n'est pas indispensable. En fait c'est souvent les bons étudiants qui prennent des projets plus intéressants ET font un travail de qualité dessus. Alors que les plus paresseux prennent un dataset+tache très simple ET font un travail de qualité moyenne dessus.

Cependant, mon conseil c'est de préférer un objectif raisonnable, mais faire un travail de qualité dessus, que chercher un objectif très sophistiqué, et s'y casser les dents (ça ne s'est pas trop vu en pratique, car on est là pour vous éviter les sujets piégeux)

Remarque: le heart failure dataset c'est quand même vraiment un poncif, on le tolère en dernier recours mais alors il faut faire qqchose de vraiment propre et maîtriser tout votre code pour valider (ce dataset a vraiment été bcp utilisé, et il est assez trivial d'approche).