

# Sorbonne Université

## Examen M2 DAC - REDS

Jeudi 9 février 2023

### Consignes générales

- Documents autorisés (papier uniquement).
- Durée : 1h30

## 1 Méthodologie données (6 points)

### Question 1

Expliquez l'intérêt de la validation croisée par rapport à un découpage train/test fixé une fois pour toute.

### Question 2

Quel est l'intérêt du leave-one-out par rapport à la validation croisée classique ? Discuter de l'influence du nombre de folds.

### Question 3

Un modèle a été déployé en production, après avoir montré des performances satisfaisantes lors de sa mise au point. Proposez une méthode pour s'assurer qu'il continue à fonctionner correctement pendant toute sa durée de vie ?

## 2 Méthodologie recherche (10 points)

On dispose d'un modèle baptisé *Frometon* : il s'agit d'un modèle de langue spécialement dédié à la langue française. Il a été conçu pour obtenir de meilleures performances que des modèles génériques comme Bert ou GPT lorsqu'il est utilisé sur du texte rédigé en français. L'objectif est de pouvoir l'entraîner sur un corpus de taille beaucoup plus réduite que d'autres modèles. De plus on souhaite qu'il soit particulièrement performant lorsqu'il est utilisé sur des recettes de cuisine. Du point de vue de l'architecture, il repose sur deux nouvelles couches baptisées *Reblochon* et *Munster* (évidemment fictives, et leur éventuel fonctionnement sera sans importance dans l'exercice).

Remarque : on ne s'intéressera pas à une tâche précise, si besoin vous pouvez fixer une tâche de NLP classique et l'utiliser dans vos réponses.

On s'intéresse à la rédaction d'un article en vue d'une soumission à une conférence internationale. Les textes à rédiger peuvent l'être en français ou en anglais.

**Question 1**

Proposez une liste de contribution.

**Question 2**

Proposez une liste d'expériences pour valider les différentes contributions. (On suppose que les datasets nécessaires pour vos expériences sont disponibles.)

**Question 3**

Proposez un titre pour l'article.

**Question 4**

Rédigez un abstract pour cette soumission (taille classique, 150 à 200 mots environ).

### 3 Docker (4 points)

**Question 1**

Etablir un fichier Dockerfile afin de construire une image pour un serveur http avec le fichier app.py qui se trouve dans votre home sous le répertoire app :

À partir d'une image de base python en version 3.8

Vous indiquerez un nom quelconque de développeur. Il sera nécessaire :

- d'installer les paquets python-pip et python-dev
- d'insérer le fichier app.py contenant le code
- d'établir un répertoire de travail app avec un fichier requirements.txt contenant les modules Python à installer (flask, Flask-Bootstrap, ipython)

Le serveur sera accessible dans le conteneur sur le port 8888.

Utilisez le template suivant :

```
FROM .....
MAINTAINER .....
COPY .....
WORKDIR ....
RUN .....
EXPOSE ....
ENTRYPOINT
CMD .....
```

**Question 2**

Donner la commande qui permet de construire l'image exam avec pour tag la version v0 à partir du dockerfile précédent

**Question 3**

Donner la commande qui permet de lancer l'image exam version v0 en mode daemon avec comme nom exam en exportant le port 8080 de la machine hôte :

**Question 4**

Comment obtenir l'adresse IP associée au conteneur ?

**Question 5**

À quoi correspond cette adresse IP ?