

### \*\*\* Comment est structuré ce répertoire ? \*\*\*

Ce répertoire contient trois sous dossiers :

- **Synthèse** : Contient le rapport et la présentation de synthèse de l'étude
- **Ressources** : Contient une présentation du langage Bengali pat Bengali.AI qui est disponible sur Kaggle et également le modèle final retenu au format HDF5.
- **Notebooks** : Contient à la racine les Notebooks principaux (EDA, modélisation et analyse de l'erreur), d'autres pistes de modélisation dans le dossier **pistes de modélisation** et le notebook de preprocessing des images dans le dossier **utilitaires**

### \*\*\* Comment utiliser ce répertoire ? \*\*\*

**Pour consulter les résultats :**

Se rendre dans le dossier synthèse et lire le rapport

**Pour faire des prédictions sur un nouveau set de données:**

- Charger le modèle qui se trouve dans le dossier **ressources**
- Se baser sur le code se trouvant à la fin du notebook **modélisation** (partie « inference ») et l'adapter à vos données

**Pour créer un nouveau modèle\* :**

- Ouvrir le notebook **modélisation**
- Modifier la fonction **create\_multi\_outputs\_model()** dans la partie « **model definition** » en faisant attention à bien conserver les trois couches de sortie
- Exécuter le notebook

**\* Attention à l'emplacement de vos données si vous effectuez l'entraînement en local sur votre machine**

### \*\*\* Les prérequis \*\*\*

**Les données :**

Disponible sur le site de kaggle pour la compétition « Bengali.AI Handwritten Grapheme Classification »

adresse : <https://www.kaggle.com/c/bengali-ai-cv19/data>

## **Hardware :**

Pour utiliser les notebooks tels quels, la configuration suivante est conseillée :

- Processeur + Minimum 16 Gb de RAM
- GPU avec 16Gb de RAM (optionnelle mais accélère fortement l'entraînement, possible d'utiliser une GPU avec moins de RAM en diminuant la taille des Batch)

Le plus simple est d'utiliser les Kernel Kaggle CPU / GPU qui ont les caractéristiques nécessaires.

## **Les librairies :**

### **Utilitaire :**

- tqdm
- pyarrow

### **Manipulation d'images :**

- OpenCV

### **Manipulation de donnée et algèbre :**

- Pandas
- Numpy

### **ML :**

- Scikit Learn
- Tensorflow
- Keras

### **Visualisation :**

- **seaborn**
- **matplotlib**