**Collège de Bois de Boulogne**

**420-A57-BB Écosystème IA**

**Examen final**

Partie théorique : les documents ne sont pas autorisés

**Durée théorie : 30 min**

**Durée totale de l’examen : 3 heures**

**Exercice 1 (20 pts)**

1. On vous demande de concevoir une chaine de développement permettant de développer un modèle IA ainsi que l’application (java, python ou autre) qui va le servir. Fournir un schéma conceptuel de votre solution CI/CD pour les deux parties en utilisant une approche Docker. Pour chaque composante utilisée, fournir une description.

**Exercice 2 (20 pts)**

Répondre aux questions suivantes sachant qu’il peut y avoir une ou plusieurs réponses possibles.

1. Les bénéfices de la mise en place d’une chaine d’intégration continue sont les suivants :
2. Résolution de problèmes qui peuvent surgir pendant le projet
3. Donne de la visibilité au projet
4. Limite le nombre de participants au projet
5. Toutes les réponses ci-dessus

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Un système de mise en production continue permet :
2. De résoudre rapidement les bugs qui se produisent dans le projet
3. D’intégrer de nouvelles données dans le modèle
4. De limiter le nombre de mise en production
5. Toutes les réponses ci-dessus

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Dans la phase de monitoring d’un modèle en production, il est important que l’outil de monitoring nous permette de savoir les prédicteurs qui sont réellement utilisés ?
2. Vrai
3. Faux

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Avant de mettre un modèle en production, il est important de savoir les informations suivantes
2. Le nombre de prédicteurs présents dans le modèle
3. Le nombre réel de prédicteurs utilisés
4. Les seuils de suivi de ces prédicteurs
5. Toutes les réponses ci-dessus

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Dans le développement d’un modèle de données, il est préférable d’avoir un gestionnaire de code source et un gestionnaire de données séparé
2. Vrai
3. Faux

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Le format python pickle peut être utilisé dans le déploiement de modèle mais en s’assurant que l’application de service est capable de lire ce format
2. Vrai
3. Faux

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. L’outil MLFlow peut être utilisé pour le suivi des métriques en développement et en production :
2. Vrai
3. Faux

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Le framework Spark est intéressant en ML car :
2. Il permet de prendre en charge de gros volume de données
3. Il dispose d’implémentation d’algorithmes ML
4. Il est efficace pour n’importe quel type de données
5. Toutes les réponses ci-dessus

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Docker est une approche de containerisation :
2. Pour remplacer les machines virtuelles
3. Pour simplifier les problèmes de déploiement
4. Pour éviter des problèmes de dépendances de packages
5. Toutes les réponses ci-dessus

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Dans Docker, un container représente une instance d’une image et sera par exemple l’exécutable pour lancer un outil d’une chaine de développement ML :
2. Vrai
3. Faux

Réponse \_\_\_\_\_\_\_\_\_\_\_\_\_\_