

# Examen 'IA et Predictive Analytics'

Instructions:

- Copier/coller ce texte dans un mail adressé à [thomas.baudel@esiee.fr](mailto:thomas.baudel@esiee.fr) à la fin de l'examen, et remplir les réponses en dessous de chaque question.
- 5 à 6 lignes de texte par question sont généralement suffisantes pour obtenir une bonne note. Des points supplémentaires sont attribués pour des réponses plus détaillées. Chaque question apporte 2 points. Il n'est pas nécessaire de répondre aux questions marquées [BONUS] pour obtenir la note maximale mais des réponses justes à ces questions rapportent des points supplémentaires. Certaines questions sont beaucoup plus faciles que d'autres, et elles ne sont pas (toutes) par ordre de difficulté croissante...
- **Lorsque vous utilisez une réponse trouvée sur internet, donner l'hyperlien des sources utilisées.**

<<Dussaussois, Marine>>

Conduite de projets en science des données

## 1: IA

a) Décrire la différence entre approche réaliste et approche utilitariste dans la démarche scientifique.

Dans l'IA, l'approche réaliste correspond à une approche symbolique. On exprime le raisonnement de manière logique par de la programmation logique (exemple Prolog), avec des règles logiques et des contraintes. L'approche utilitariste tend vers une approche statistiques. Autrement dit, on entraîne les données avec par exemple du Machine learning, des réseaux de neurones.

b) Watson Studio: décrire une caractéristique importante de Watson Studio comme environnement de développement en science des données, qui le rend utile en situation de développement en entreprise.

Watson Studio est une plateforme d'apprentissage automatisée qui aide entre autres à la prise de décision pour les entreprises. Elle est intéressante pour les entreprises puisqu'elle permet le développement en environnement Cloud.

## 2: programmation logique/chainage avant

Dans un langage à base de règles simple en chainage avant (on appelle cela un [système de production](#)) on a le programme:

```
var input=[] , result=[] , i=1 , tmp=0;
```

```
when input.length>0 and i >= input.length then result.append(input[tmp]), input.removeAt(tmp),  
i=1, tmp=0;
```

```
when input[i] < input[tmp] then tmp=i, i=i+1;
```

```
when input[i] >= input[tmp] then i=i+1;
```

Lorsque l'instruction `input=[2,0,5,4,9]`; est exécutée, que contiendront les variables *result* et *input* en retour? Expliquer l'algorithme.

A la fin, *result* prend pour valeur [0,2,4,5,9] et *input* est une liste vide.

L'algorithme tri de manière croissante les valeurs contenues dans *input*, les affecte dans une nouvelle liste *result* et vide la liste *input*.

On cherche d'abord à localiser la plus petite valeur dans *input* grâce à *tmp*. Puis on sort cette valeur en l'affectant à la liste *result*. Et ainsi de suite avec le reste des valeurs de la liste *input*.

**3: Smart City:** Trouver sur internet 3 logiciels commerciaux **professionnels** destinés à remplir un rôle similaire à celui du projet SmartDeliveries. En vous basant sur la présentation commerciale, identifiez leurs principales caractéristiques démarquantes (quels fonctionnalités mettent-ils particulièrement en avant par rapport à la concurrence). Fournir les références utilisées.

- Mapotempo : <https://www.mapotempo.com/>

Logiciel et application mobile clé en main, solution 100 % en ligne

- 1) Planification & Optimisation
- 2) Suivi & Alertes en temps réel
- 3) Rapports & Analyses
- 4) Intégration d'API

- Antsroute : <https://antsroute.com/optimisation-des-tournees-de-livraison/>

Logiciel d'optimisation des tournées de livraison et d'intervention

Mêmes fonctions que Mapotempo

- DANEM : <https://danem.fr/ref/logiciel-d-optimisation-de-tournee.html>

#### **4: trafic routier**

a- Quelles sont les principales variables mesurées par un détecteur de trafic?

Taux d'occupation

b- Qu'est ce que le diagramme fondamental d'un détecteur de trafic, pourquoi est-il utile pour mesurer et prévoir la congestion?

Il s'agit du graphique de « occupancy » en fonction du « flow ». Autrement dit, le débit en fonction de la vitesse. Il est important pour mesurer et prévoir la congestion car lors d'un bouchon ou lorsque la route est quasiment vide, on a un débit faible (voiture à l'arrêt) mais pourtant pas la même vitesse.

c- quel est le débit typique maximal d'un tronçon à une voie

- en zone urbaine
- sur voie rapide ou autoroute
- Pourquoi cette différence?

## 5: temps de parcours

a) Quelles sont les principales variables prédictives du temps de parcours d'un camion de livraison en ville, par ordre d'importance décroissante? (déterminées en cours)

- la distance
- le type de véhicule
- l'expérience du conducteur
- la motivation du conducteur
- circonstances locales non prises en compte par le modèle

b) Citer 2 facteurs potentiels affectant les temps de parcours et difficiles à mesurer avec les données fournies.

- la congestion
- 

## 6: géoréférencement

On a un fichier d'adresses tel que vu en cours:

ID, BASE, KIND, CITY, FROM, TO

1, Docteur Bouchut, Rue du, Lyon, 1, 100

etc...

a) Décrire la logique d'une fonction python qui permet de retrouver l'identifiant de tronçon (ID) correspondant à chaque adresse, en supposant que la base d'adresses est stockée dans un tableau, et que l'on a une fonction distance(a, b) qui retourne la distance de Levenshtein entre 2 chaînes.

17, rue Bouchut, Lyon

17 rue du Docteur Bouchot, Lyon

b) proposer cette fonction en python (améliorant celle vue en cours)

linkid(s, addresses):

```
""" addresses contient une liste de tronçons [id, base, kind, city, from, to], s l'adresse à trouver
"""

return id
```

Prescriptive Analytics

### Question 1.

Les affirmations suivantes sont-elles vraies ou fausses:

a- Un problème de décision est dans la classe de complexité NP si et seulement si il n'existe pas d'algorithme polynomial pour le résoudre.

Vrai

b- Dans l'industrie, la majorité des problèmes d'ordonnancement sont résolus grâce à des heuristiques.

Vrai

c- Le problème suivant possède exactement trois solutions:

```
u in {1,3}
v in {1,2}
w in {3,4}
x in {1,5}
y in {4,5}
allDifferent(u,v,w,x,y)
```

Vrai

d- L'algorithme de résolution de CP-Optimizer est un algorithme exact: si un problème d'optimisation est faisable, il garantit de trouver une solution optimale.

Vrai

## Question 2.

a- En cherchant sur internet, décrivez un problème d'optimisation combinatoire non vu dans le cours dont la version de décision est un problème NP-Complet.

Plus longue sous séquence commune

b- Décrivez une petite instance particulière de ce problème d'optimisation (avec des valeurs pour chacune des données).

Exemple :

« abcde » et « ceij »

Plus longue sous séquence commune : « ce »

c- Donnez une solution faisable non-optimale et une solution optimale de cette petite instance.

Non-optimal : force brut

optimal : programmation dynamique

## Question 3.

Deux principes fondamentaux de la Programmation par Contraintes sont (1) la recherche arborescente et (2) le filtrage du domaine des variables. Décrivez brièvement ces principes, leurs rôles et la façon dont ils sont mis en oeuvre durant la résolution.

La recherche arborescente consiste à énumérer toutes les possibilités et vérifier si elles violent ou non les contraintes. Le filtrage consiste à déduire à partir des contraintes les valeurs impossible. Lorsqu'une variable ne possède plus qu'un candidat celle-ci lui est affectée. Cela permet d'éviter une énumération exhaustive.

## Question 4.

Un problème classique en ordonnancement est le problème d'open-shop pour lequel un ensemble de  $n$  jobs doivent être exécuté sur  $m$  machines. Chaque job consiste en un ensemble de  $m$  opérations de durée connue, devant être exécutées dans un ordre arbitraire sur les machines. Plus précisément, si l'opération  $o_{ij}$  désigne la  $j$ ème opération du job  $i$ , cette opération utilise la machine  $j$  et sa durée est  $D_{ij}$ . L'ordre des opérations du job  $i$  n'est pas connu d'avance: les  $m$  opérations  $o_{ij}$  d'un job  $i$  donné doivent être ordonnées mais cet ordre est libre. D'autre part, une machine ne peut pas effectuer plus d'une opération à la fois. L'objectif est de déterminer les dates de début et de fin de chaque opération de manière à minimiser la date de fin du plan. Ce problème d'optimisation est NP-difficile.

Les données d'entrée du problème sont donc:

- $n$ , le nombre de jobs
- $m$ , le nombre de machines
- $D_{ij}$  ( $i \in [1, n]$ ,  $j \in [1, m]$ ), la durée de la  $j$ ème opération du job  $i$

a- Décrivez un modèle CP Optimizer à base de variables d'intervalles et de contraintes noOverlap pour résoudre le problème d'open-shop.

b- [BONUS] Décrivez un modèle CP Optimizer pour une variante du problème d'open-shop pour laquelle les machines peuvent effectuer au plus deux opérations en parallèle.

Un problème de décision est dans la classe de complexité NP si et seulement si il n'existe pas d'algorithme polynomial pour le résoudre