

Examen ' Conduite de projets en science des données et Predictive Analytics'

Instructions:

- Copier/coller ce texte dans un mail adressé à baudelth@fr.ibm.com à la fin de l'examen, et remplir les réponses en dessous de chaque question. DEPUIS VOTRE ADRESSE ENSC.FR (un filtre sur le mail sera appliqué pour trier les réponses).
- **Le nom de votre fichier réponse doit commencer par votre nom de famille.prénom.**
- 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.**

<<HANRIO, Solène>>

1: IA Décrire la différence entre approche réaliste et approche utilitariste dans la démarche scientifique.

Pour expliquer l'approche réaliste en science on peut commencer par définir le réalisme en épistémologie qui va être défini par l'idée que la connaissance porte sur des objets du monde réel indépendant de l'homme. C'est à partir de cette définition du réalisme que va être tirée le réalisme scientifique qui pose le principe que la science définit le monde réel. Les sciences permettraient ainsi d'expliquer et de prédire les phénomènes de notre monde.

On peut voir ici l'influence du dualisme de Platon et de la théorie des idées qui définit deux mondes, le monde réel et le monde des idées dans la dualité du corps et de l'esprit.

L'approche utilitariste quand à elle se définit comme l'approche cherchant à maximiser le bien-être du plus de personnes possibles. On va privilégier l'utilité commune.

La différence entre ces deux approches résiderait donc dans leur but fondamentale, pour l'une c'est l'explication du monde pour l'autre c'est l'utilité commune.

2: programmation logique/chainage avant

Dans un langage à base de règles simple en chaînage 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. Donner sa complexité.

L'algorithme permet de trier des variables entrées dans le tableau *input*. Il va ainsi parcourir *input*, chercher la plus petite valeur qui correspondra à la variable *tmp*. A la fin la variable correspondant à *tmp* est ajoutée dans le tableau *result* et est retirée du tableau *input*.

Le résultat est donc *result=[0,2,4,5,9]* et *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 (quelles fonctionnalités mettent-ils particulièrement en avant par rapport à la concurrence). Fournir les références utilisées.

Tout d'abord le logiciel *antsroute* qui propose l'optimisation des tournées de livreurs. Il prône la multiplication des services dont la planification de tournées, la gestion des clients et l'optimisation des parcours de livraison. Le mot « intelligent » revient beaucoup sur leur site internet.

Ensuite *Noticodeliv* qui propose l'optimisation des tournées, la démarche commerciale est sensiblement la même que le premier logiciel on retrouve encore en quantité le mot « intelligent » censée « attirer » le client certainement.

Enfin *Geoconcept*, propose la même chose que les logiciels précédents mais met en avant que la solution est sur le cloud et est donc facile d'accès.

<https://fr.geoconcept.com/>

<https://noticodeliv.com/fr/>

<https://antsroute.com/>

4: trafic routier

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

La densité de voitures présentes sur la route

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

Ce diagramme montre le nombre de véhicule par jour et par heure ainsi que l'occupation. Il permet d'avoir une base de prédiction du taux d'occupation des routes et du flux de voitures sur celle-ci.

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

- en zone urbaine

155.08

- sur voie rapide ou autoroute

118.46

- Pourquoi cette différence?

Les livreurs livrent à des adresses spécifiques situées en zone urbaine et non pas sur les voies rapides ou les autoroutes, ainsi on retrouvera bien plus de trafic dans ces zones.

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)

Les principales variables prédictives du temps de parcours d'un camion de livraison en ville sont définies dans ce tableau par ordre d'importance de gauche à droite :

Feature	Distance	RoundID	Static travel time forecast	Occupancy	Angle
Gini importance	77	11	8	2	1

b) Citer 2 facteurs potentiels affectant les temps de parcours et difficiles à mesurer avec les données fournies dans les fichiers fournis en TP.

Les deux facteurs pouvant affecter les temps de parcours et qui sont difficiles à mesurer avec les données fournies sont l'identifiant de livraison et les accidents ce sont des imprévus difficiles à prévoir.

6. Prescriptive Analytics

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.

Faux

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)

Faux

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

7. optimisation

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.

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

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

8. programmation par contraintes : 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 correspond à un algorithme de recherche heuristique qui définit les toutes les solutions possibles selon un arbre des possibilités. L'algorithme va ainsi stocker les solutions déjà explorées et leurs résultats et explorer les solutions à venir.

Le filtrage du domaine des variables permet de réduire le nombre de solutions possibles en prenant en compte les contraintes exercées sur le système.