

En présence de :

- Monsieur Mohand-Saïd HACID
- Monsieur Emmanuel COQUERY

Déroulement de la séance :

Durant cette séance, j'ai d'abord commencé par présenter l'actuel avancement de mon projet de fin d'étude, intitulé "Étude comparative des algorithmes d'arbres de décision dans un environnement *Big Data*". Une fois la présentation terminée, j'ai eu les remarques suivantes :


— Revoir les slides suivants :

1. N°10 : les caractéristiques d'un RDD.

Revoir le troisième point

RDD ?

- ☐ Liste de partitions stockée en **RAM**
- ☒ Caractéristiques
- ☐ Tolérants au pannes
- ☐ Distribués
- ☒ Inchangeable

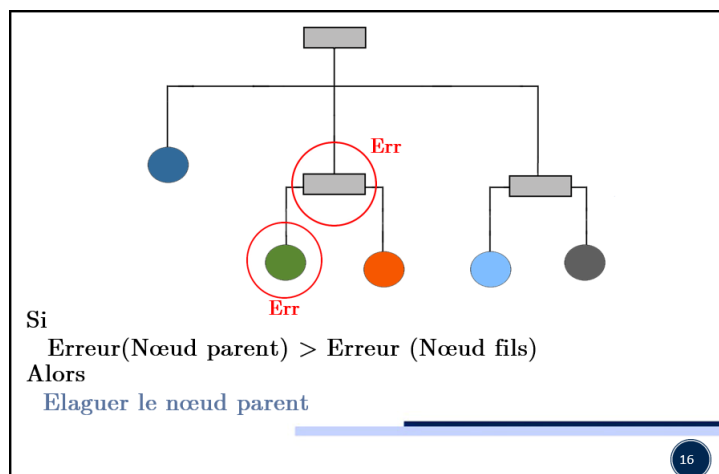


10

2. N°16 : Élagage de l'arbre de décision.

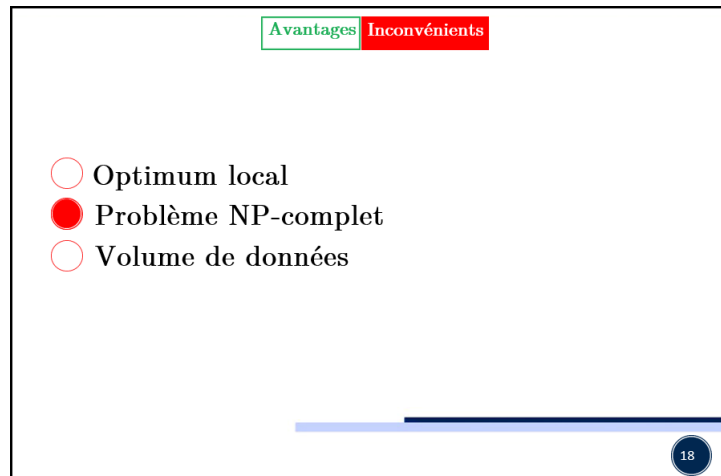
En se basant sur ce que montre le slide, un sous arbre est élagué puis remplacé par un nœud fils, d'autres nœuds (appartenant au sous arbre) seront éventuellement perdus, ce qui affectera la précision de l'arbre de décision, qui a la base, est supposée être optimisée avec l'élagage, ce qui est contradictoire.

Cette idée est à revoir avec une recherche encore plus approfondie.



3. N°18 : avantages et inconvénients des arbres de décision.

- Sur quel critère se baser pour estimer qu'un arbre de décision est optimal par rapport à un autre? (temps?, vitesse?, taille?, autres?).



- Quelle importance doit-on attribuer à chacun de ces critères? et par rapport à quel contexte?
- Ne serait-il pas intéressant de calculer/se rapprocher de l'optimum global, même au détriment de son coût de calcul, dans des cas où le besoin d'analyse n'est pas régulier.
- Les caractéristiques de l'ensemble de données d'apprentissage peuvent-ils avoir un effet sur la qualité de l'arbre de décision produit? Dans ce cas, quel serait le critère d'un bon jeu de données? cette idée a-t-elle été abordée sur un des articles des approches proposées?

4. N° 24, 25 : Les deux approches MReC4.5 et MRID4.

Les figures qui ne présentent pas de comparaison entre l'approche proposée et celles existantes ne devraient pas être mises.

- A préparer pour le mercredi 21/03/2018 :

1. Pour chacun des critères de division cités :
 - Une définition claire, nette et précise.
 - Déroulement d'un exemple.
 - Expliquer dans quel contexte tel critère est meilleur qu'un autre.
2. Définition formelle d'un arbre de décision, ainsi que les critères qui permettent d'évaluer, de mesurer et de comparer la qualité des arbres de décision.
3. Explication détaillée de l'idée de la nouvelle approche, et décrire le contexte pour lequel elle pourrait être meilleure que celles déjà existantes.