



# Série d'activités d'Actua sur l'IA

## Activité 2

# Les arbres décisionnels : classer des objets spatiaux

# Les arbres décisionnels : classer des objets spatiaux

**Vous avez accédé directement à cette activité? Saviez-vous que notre série en compte huit autres? Consultez notre site web pour en connaître les détails et l'ordre recommandé pour les réaliser. Elles présentent toutes des notions d'intelligence artificielle ainsi qu'un vocabulaire propre au domaine. [Un glossaire](#) permettant de vérifier le sens des mots nouveaux ou inconnus les accompagne. Amusez-vous tout en réparant une station spatiale!**

*Toi et ton groupe faites partie d'une équipe d'astronautes et de scientifiques en mission à bord de la station spatiale Actua. Comble de malchance, la station vient d'être bombardée par des rayons magnétiques! Le seul qui peut vous sauver, c'est DANN\*, le logiciel d'IA de la station; toutefois, il a un peu perdu la carte depuis l'incident. Révisez vos connaissances techniques et apprenez quelques notions d'intelligence artificielle pour sauver l'équipage!*

*Après avoir appris quelques notions de base dans l'activité [«Introduction : Qu'est-ce que l'intelligence artificielle?»](#), vous voici prêts à entreprendre la réparation de DANN! La station orbitale Actua est équipée d'un appareil de détection qui surveille et repère les objets spatiaux proches et lointains. Malheureusement, les dommages causés par les rayons magnétiques semblent avoir remis son classifieur à zéro. Afin de pouvoir le réinitialiser, le Centre de contrôle de mission vous a demandé entre autres de proposer un arbre décisionnel et de le tester. Une fois que ce sera fait, vous pourrez procéder à l'analyse des résultats d'une expérience scientifique dans l'activité suivante intitulée [«Analyse de régression : faire des prédictions à l'aide de données»](#).*

*\*Pour «Dedicated Actua Neural System»*

## Présentation de l'activité

Dans cette activité, les élèves sont amenés à concevoir et à évaluer un arbre décisionnel. Les arbres décisionnels offrent un moyen de trier des objets ou des données dans différentes catégories en posant des questions à leur sujet.



Les élèves se serviront d'un jeu de données échantillon afin de formuler une série de questions. Ils évalueront ensuite la validité de leur arbre en déterminant si ce dernier permet d'étiqueter correctement un jeu de données de test.

Activité conçue par Actua, en 2022.

Contexte de l'activité	Durée	Public cible
En classe	75 minutes	Niveaux 9-12 (13-18 ans)

## Résultats d'apprentissage

À la suite de cette activité, les participants pourront :

- créer et évaluer des arbres décisionnels;
- classer et tester des ensembles de données;
- déterminer la meilleure façon d'utiliser et de trier les données à différentes fins;
- comprendre le rôle des arbres de décision dans les algorithmes informatiques.

OUTILS	COMPÉTENCES	ATTITUDES
<p><b>Connaissances, ressources et expériences</b></p> <ul style="list-style-type: none"> <li>• Créer et évaluer des arbres décisionnels.</li> <li>• Classer et tester des jeux de données.</li> </ul>	<p><b>Compétences numériques, compétences en STIM, employabilité et aptitudes essentielles à la vie</b></p> <ul style="list-style-type: none"> <li>• Analyse computationnelle</li> <li>• Pensée critique</li> <li>• Reconnaissance de formes</li> </ul>	<p><b>Intelligence numérique, action communautaire et pensée computationnelle</b></p> <ul style="list-style-type: none"> <li>• Déterminer le meilleur moyen de trier des données en fonction de différents usages.</li> <li>• Comprendre le rôle des arbres décisionnels dans les algorithmes</li> </ul>



OUTILS	COMPÉTENCES	ATTITUDES
		informatiques.

## Logistique (durée, taille du groupe, matériel)

Titre de la section	Durée	Taille du groupe	Matériel
<b>Introduction : Jeu des vingt questions</b>	10 min	Petits groupes	Pour chaque groupe : <ul style="list-style-type: none"> <li>• Stylo/crayon</li> <li>• Papier</li> </ul> Toute la classe : <ul style="list-style-type: none"> <li>• Liste d'éléments pour le jeu des vingt questions (voir l'étape 2 de la marche à suivre)</li> </ul>
<b>Activité n° 1 : Un arbre pour prendre des décisions</b>	15 min	Toute la classe	
<b>Activité n° 2 : Classer des objets spatiaux</b>	30 min	Chacun pour soi ou en petits groupes	Pour chaque élève ou groupe : <ul style="list-style-type: none"> <li>• Accès à un ordinateur</li> <li>• Copie des données d'apprentissage (<i>annexe C</i>)</li> <li>• Stylo/crayon</li> <li>• Papier</li> </ul>
	10 min	Chacun pour soi ou en petits groupes	<ul style="list-style-type: none"> <li>• Copie des données de test (<i>annexe B.2</i>)</li> <li>• Stylo/crayon</li> <li>• Calculatrice</li> </ul>
<b>Réflexion et récapitulation</b>	10 min	Toute la classe	

## Consignes de sécurité

Les consignes de sécurité ci-dessous ne sont pas exhaustives. Veillez à passer en revue l'activité et à inspecter l'environnement où elle sera réalisée afin de déterminer si des mesures additionnelles sont requises pour assurer la sécurité des élèves.

### Sécurité en ligne

Certains volets de cette activité nécessitent l'usage d'appareils connectés à Internet.

- Examinez au préalable les vidéos, les sites web et le matériel prévus afin de vous assurer qu'ils conviennent à vos élèves.
- Au besoin, rappelez aux jeunes de se concentrer sur la tâche à faire et d'utiliser uniquement les liens fournis pour l'activité.

## Marche à suivre

### Introduction : Jeu des vingt questions

Le jeu des vingt questions consiste à tenter de deviner un objet ou un animal en posant une série de questions. Chaque question vise à réduire le nombre de réponses possibles. Deux par deux ou en petits groupes :

1. Désignez une personne qui jouera le rôle de « répondant ». Elle ne peut répondre aux questions que par « oui » ou par « non ».
2. Le répondant choisit un élément dans la liste ci-dessous sans le dévoiler aux autres.

- |                 |                     |
|-----------------|---------------------|
| a. Aigle        | g. Navette spatiale |
| b. Avion        | h. Ours             |
| c. Bus scolaire | i. Pigeon           |
| d. Camion       | j. Pupitre          |
| e. Cellulaire   | k. Roche            |
| f. Lune         | l. Tortue           |



3. À tour de rôle, les autres élèves tentent de deviner son choix en posant seulement des questions auxquelles on peut répondre par « oui » ou par « non ».
4. Demandez à quelqu'un de prendre toutes les questions en note.
5. On peut poser jusqu'à 20 questions.
6. Si vous devinez la réponse, désignez un nouveau répondant et répétez le jeu à partir de l'étape 2. Donnez à tous les élèves la chance de jouer ce rôle avant de poursuivre.
7. Une fois le jeu terminé, passez les questions en revue.
  - a. Certaines questions pourraient-elles s'appliquer à plus d'un élément?
  - b. Comment pourriez-vous utiliser ces questions pour répartir les éléments de la liste en deux groupes? En trois groupes? En quatre groupes?

## Activité n° 1 : Un arbre pour prendre des décisions

Imaginez une pile d'objets disparates disposés sans ordre particulier. Les arbres décisionnels permettent de distinguer des éléments les uns des autres et de les répartir dans des groupes parfois appelés « classes » ou « étiquettes ». Pour les différencier, on se pose des questions sur leurs caractéristiques, c'est-à-dire sur les traits particuliers qui permettent de les identifier. Le but d'un arbre décisionnel est de répartir les objets de la pile dans des groupes aussi distincts que possible.

1. Pour en savoir plus sur les arbres de décision, examinons ensemble un exemple. Nous allons essayer de répartir les objets célestes ci-dessous.

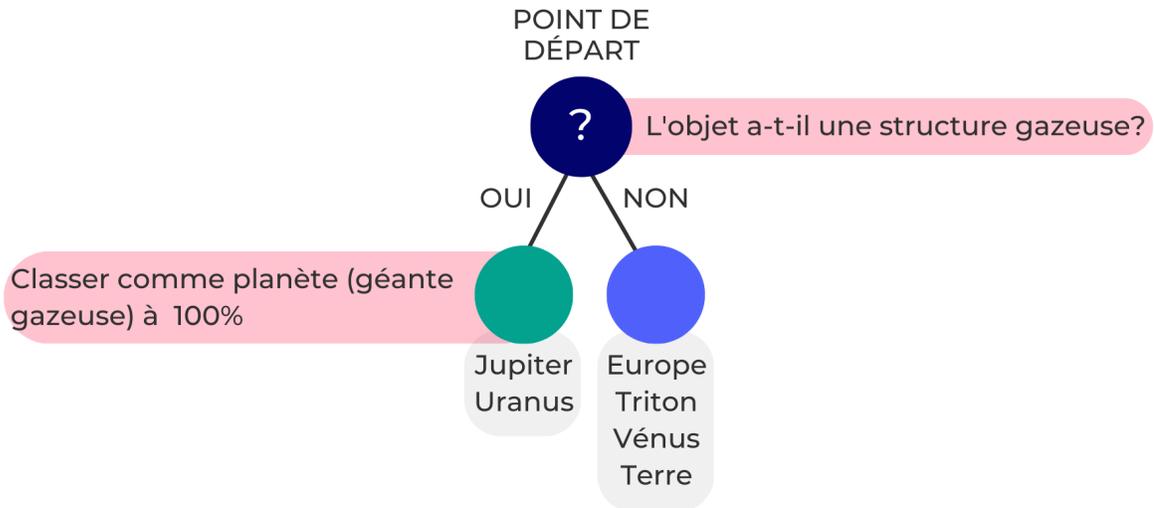
Caractéristiques		Étiquette	Nom (extra)
Structure	Orbite		
solide	planète	lune	Europe



solide	planète	lune	Triton
gazeux	Soleil	planète (géante gazeuse)	Jupiter
gazeux	Soleil	planète (géante gazeuse)	Uranus
solide	Soleil	planète (terrestre)	Vénus
solide	Soleil	planète (terrestre)	Terre

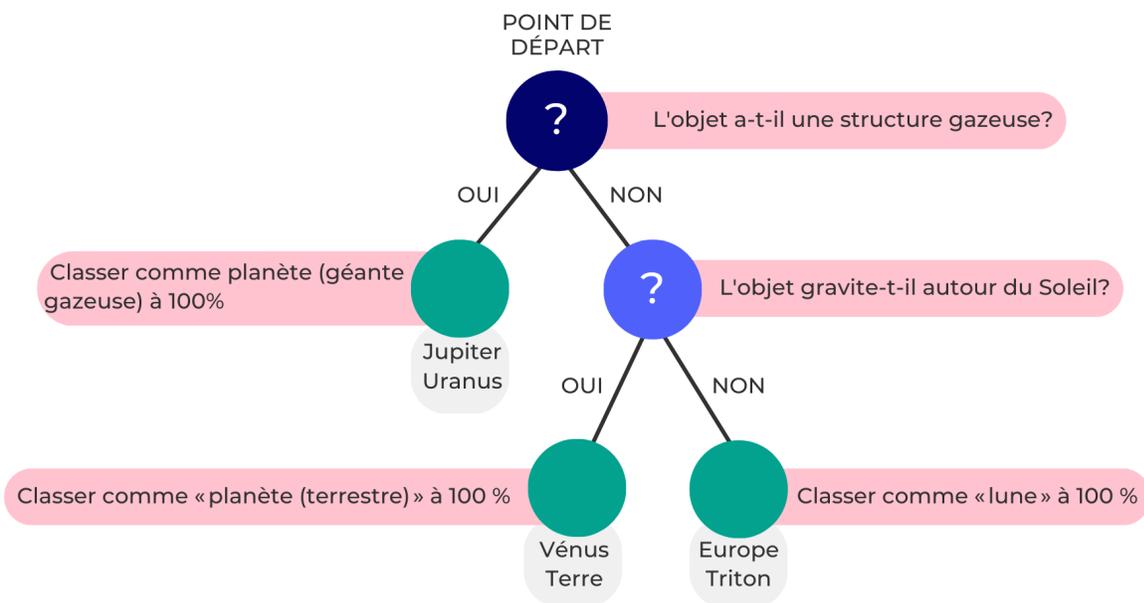
2. Étudiez le tableau ci-dessus. Combien y a-t-il de sortes d'objets spatiaux? (*Indice : regardez dans la colonne «Étiquette»*)
  - a. Le tableau contient trois sortes d'objets différents : des lunes et deux types de planètes (géantes gazeuses et terrestres).
  
3. L'idée consiste à utiliser l'information dans le tableau pour formuler des questions qui permettront d'identifier chaque type d'objet. Étudiez la colonne «structure» et réfléchissez à la première question qu'il faudrait poser pour commencer à trier les données.
  - a. La question devrait ressembler à ceci : «Est-ce que l'objet a une structure gazeuse?»
  
4. Vous pouvez maintenant commencer à construire votre arbre décisionnel :
  - a. Tracez le point de départ de l'arbre. On l'appelle le «nœud racine». Écrivez la 1<sup>re</sup> question à droite de celui-ci.
  - b. Tracez ensuite deux lignes à partir du nœud racine. Ce sont les «branches». Il doit y avoir une branche pour chaque réponse possible à la 1<sup>re</sup> question («oui» ou «non»). Étiquetez les branches.
  - c. À l'aide de la 1<sup>re</sup> question, triez tous les objets dans la liste (on pose la question et on répartit les objets dans le groupe «oui» ou dans le groupe «non» en fonction de la réponse).

- d. Une fois le tri terminé, examinez les groupes qui se sont formés. Y a-t-il des groupes qui ne comptent *qu'un seul type* d'objet?

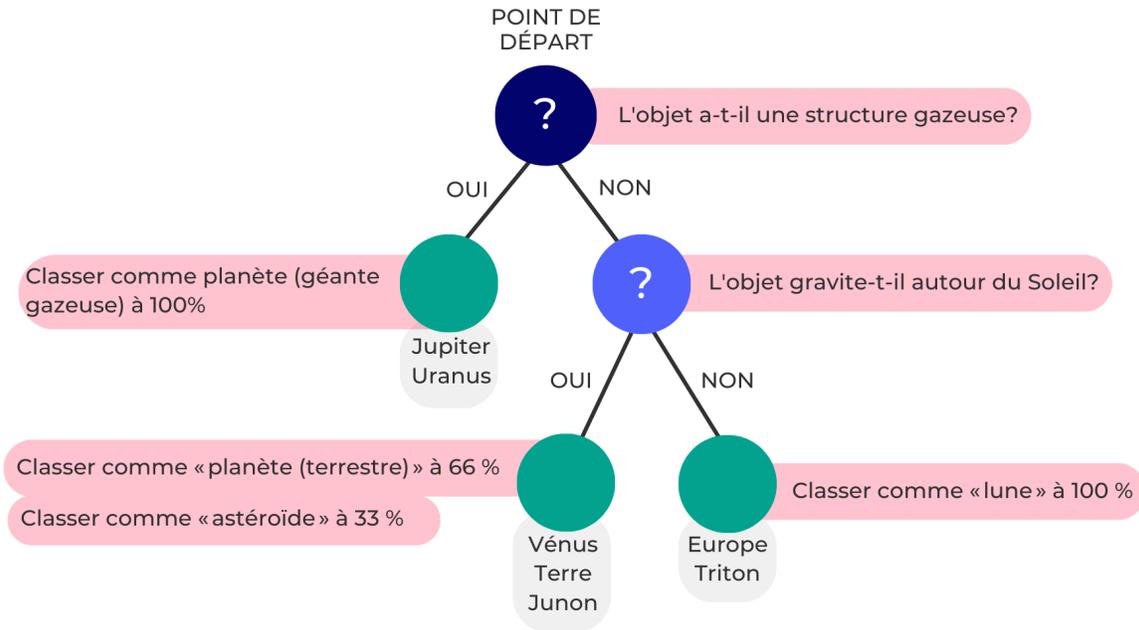


5. Vous devriez obtenir un groupe comptant *uniquement* des géantes gazeuses. C'est une bonne chose, puisque nous avons réglé le cas d'un type d'objet parmi ceux que nous avons au départ.
6. Vous aurez peut-être remarqué la phrase «Classifier comme planète (géante gazeuse) à 100 %». Qu'est-ce que cela veut dire à votre avis? À quoi correspond ce pourcentage?
- Il nous indique qu'on peut cesser de poser des questions, puisqu'on juge qu'on ne peut pas différencier les objets davantage. On peut donc les «classer», c'est-à-dire leur attribuer une étiquette.
  - Le pourcentage de 100 % indique notre degré de certitude envers notre décision. Le fait que cette catégorie ne contient que des géantes gazeuses confirme la justesse de notre raisonnement.
7. Le 2<sup>e</sup> groupe, par contre, contient encore plusieurs sortes d'objets. Essayons de trouver d'autres questions pour les différencier encore davantage. À votre avis, quelle devrait être la question suivante? (Indice : étudiez le tableau et pensez aux caractéristiques dont nous avons déjà tenu compte et à celles qui restent.)

- a. La question devrait ressembler à celle-ci : « L'objet gravite-t-il autour du Soleil? »
8. Continuez à tracer votre arbre en suivant les mêmes étapes que précédemment.
- a. À l'aide de la 2<sup>e</sup> question, classez les objets du second groupe.
  - b. Une fois le tri terminé, examinez les groupes qui se sont formés. Y en a-t-il qui ne comptent *qu'un seul type* d'objet? Si oui, ajoutez d'autres décisions, comme nous l'avons fait avec les géantes gazeuses.



9. L'exemple ci-dessus est intéressant parce qu'il a suffi de deux questions pour répartir tous les objets entre nos trois classes de sorte que chaque objet ne figure que dans une seule classe (c'est la raison pour laquelle nous sommes certains à 100 % que chacune de nos décisions était la bonne). Ceci n'est pas toujours le cas. Il arrive en effet qu'on soit incapable de différencier entièrement des données.
10. Qu'arrivera-t-il si on ajoute un nouvel objet spatial? Prenons l'astéroïde Junon (solide, orbite : Soleil) et tentons de le classer. Cela modifiera-t-il la structure de notre arbre? Voici à quoi pourrait ressembler notre dernier schéma :



11. On constate ici que notre degré de confiance en notre décision a changé :

- Puisque deux objets sur trois dans ce groupe sont des planètes, il n'y a que 66 % de chances de classer correctement un objet comme planète terrestre dans cette catégorie.
- Puisqu'un objet sur trois dans ce groupe est un astéroïde, il n'y a que 33 % de chances de classer un objet comme astéroïde dans cette catégorie.

## Activité n° 2 : Classer des objets spatiaux

Examiner le jeu de données

- Jetons maintenant un coup d'œil au jeu de données sur les objets spatiaux avec lequel nous allons travailler. Ce jeu de données a été créé pour ressembler aux données que la station spatiale utilisera. Il regroupe cinq types d'objets célestes étiquetés (astéroïde, comète, débris, météorite, satellite) et classés en fonction de trois caractéristiques :
  - leur composition matérielle : **organique** (contient de l'eau ou d'autres composés organiques) ou **inorganique** (composé de métal et de roche)



en fonction des caractéristiques du jeu de données sur les objets spatiaux.

2. À partir de vos questions, commencez à construire votre arbre décisionnel. Choisissez des questions et décidez de l'ordre dans lequel vous allez les poser.
3. Appliquez les **données d'apprentissage** à votre arbre en prenant note de la composition des groupes à chaque étape et des probabilités relatives aux décisions de classification finales (sous forme de fraction ou de pourcentage pour chaque type d'objet).

### Tâche n° 2 : Caractéristiques catégorielles et continues

1. Formulez des questions basées sur les caractéristiques **continues**. *Si vous hésitez sur la façon de procéder, voici un truc pour vous aider : examinez l'étendue des valeurs (minimum et maximum) correspondant à la taille de chaque objet. Pourriez-vous vous en servir pour les différencier? Certains objets sont-ils visiblement plus grands ou plus petits que d'autres?*
2. Créez un arbre décisionnel en associant les deux types de caractéristiques.
3. Appliquez les données d'apprentissage à votre arbre en prenant note de la composition des groupes à chaque étape et des probabilités relatives aux décisions de classification finales (sous forme de fraction ou de pourcentage pour chaque type d'objet).

### Activité n° 3 : Valider un arbre décisionnel

Ensuite, servez-vous du jeu de **données de test** pour déterminer si votre arbre décisionnel est valide.

1. Appliquez les données de test pour chaque objet afin de déterminer où il se classerait. Prenez les résultats en note, car vous en aurez besoin à l'étape suivante.
2. Après avoir classé, en procédant rangée par rangée, les données de test dans les nœuds terminaux de votre arbre, calculez le résultat pour



chaque feuille en multipliant le nombre d'objets par le degré de confiance (la probabilité d'une bonne réponse). Voici un exemple :

- a. Si vous avez 5 astéroïdes dans un nœud terminal avec un degré de confiance de 100 % pour la décision « Classifier comme un astéroïde » :  
 $5 \times 100 \% = 5 \times 1 = 5.$
  - b. Si vous avez 3 comètes et 2 météorites dans un nœud terminal avec un degré de confiance de 60 % pour la décision « Classifier comme une comète » et de 40 % pour la décision « Classifier comme une météorite » :  
 $(3 \times 60 \%) + (2 \times 40 \%) = (3 \times 0,6) + (2 \times 0,4) = 2,6.$
3. Additionnez les notes obtenues dans tous les nœuds terminaux. Le résultat ne devrait **pas** excéder le nombre d'objets contenus dans le jeu de données de test. Le chiffre le plus élevé devrait être 10 dans le cas présent, mais il suppose que les données ont été classées parfaitement, ce qui pourrait ne pas être le cas.
  4. Si vous avez le temps, essayez d'appliquer cette procédure de validation à vos deux arbres décisionnels, puis comparez les résultats.

## Réflexion et récapitulation

Après avoir testé votre arbre décisionnel, comparez vos résultats et vos arbres en petits groupes ou tous ensemble. Une note élevée indique que l'arbre permettra de classer les données plus efficacement. Réfléchissez aux questions suivantes :

1. Quelle méthode avez-vous utilisée pour décider des questions à poser?
2. Comment avez-vous déterminé l'ordre des questions?
3. Combien de strates/niveaux contenait votre arbre? Pouvez-vous observer un rapport entre le nombre de strates/niveaux et le résultat du test?
4. Vous est-il arrivé de répéter une question?



5. Répondez aux questions suivantes en tenant compte à la fois des données de l'échantillon et des données de l'exercice :
- Quel effet les données d'apprentissage ont-elles sur la structure et la conception d'un arbre décisionnel?
  - Quel effet les caractéristiques des données ont-elles sur la structure et la conception d'un arbre décisionnel?
  - Comment pouvez-vous *démontrer* que votre arbre est aussi efficace que possible?

Les arbres décisionnels ne sont qu'un exemple de stratégie servant à classer ou à trier des données. Toutefois, les données ne sont pas toujours présentées dans un format aisément lisible comme un tableau ou un tableur. Dans les activités subséquentes, nous allons étudier d'autres méthodes de classification possibles, y compris des modèles utilisés en vision et en écoute artificielle dans le but de reconnaître des images et des sons.

## Possibilités d'adaptation

Il est possible d'adapter différents aspects de cette activité (durée, environnement, matériel, taille du groupe ou instructions) pour la rendre plus accessible ou plus complexe. Les **modifications** ci-dessous vous permettront de diminuer le niveau de difficulté de l'activité et les **ajouts**, d'augmenter sa durée ou son niveau de difficulté.

### Ajouts

#### Étendre le jeu de données fourni à titre d'exemple (planètes et lunes du système solaire)

- Notre exemple ne comporte qu'un petit sous-ensemble de données sur les objets et les catégories du système solaire. Au lieu d'utiliser le jeu de données fourni, demandez aux élèves de créer un nouvel ensemble d'objets.

#### Optimiser l'arbre décisionnel

- La méthode de notation décrite dans l'activité n° 3 vise à donner une idée générale de la validité d'un arbre décisionnel et à permettre les



comparaisons. Si vous comparez les arbres des différents groupes, lesquels ont donné les meilleurs résultats?

- Y a-t-il un rapport entre le nombre de strates (la hauteur) d'un arbre et son degré d'exactitude?

## **Modifications**

### **Simplifier le jeu de données**

- Le jeu de données fourni dans cette activité a été conçu de façon à ce qu'il ne soit pas possible de les différencier parfaitement (les valeurs de « débris » et « satellite » pour la taille se recoupent et ils ont les mêmes attributs « proche » et « inorg »). Si vous pensez que cela peut créer de la confusion, vous pourriez supprimer l'une des catégories ou demander aux élèves d'en créer une seule (par ex. « objets de fabrication humaine »).

## Conditions d'utilisation

Avant de réaliser cette activité en tout ou en partie, vous reconnaissez et acceptez ce qui suit :

- Il vous appartient de passer en revue toutes les sections du présent document et la documentation connexe ainsi que d'appliquer les consignes de sécurité nécessaires à la protection de toutes les personnes concernées;
- Les mesures précisées à la rubrique « Consignes de sécurité » du présent document ne sont pas exhaustives ni ne remplacent votre propre cadre d'examen de la sécurité;
- Actua n'est pas responsable des dommages attribuables à l'usage du présent contenu;
- Vous pouvez adapter ce document à vos besoins (le remanier, le transformer ou créer du matériel à partir de celui-ci), à condition d'indiquer qu'Actua en est l'auteur original et que vous y avez apporté des changements. Ce contenu ne peut être transmis à de tierces parties sans la permission écrite d'Actua.

## À propos d'Actua

Représentant plus de 40 universités et collèges à travers le pays, Actua est le principal réseau de sensibilisation des jeunes aux sciences, à la technologie, à l'ingénierie et aux mathématiques (STIM) au Canada. Chaque année, 350 000 jeunes prennent part à des ateliers pratiques, à des camps et à des projets communautaires inspirants dans plus de 500 localités d'un océan à l'autre. Actua met l'accent sur la participation de jeunes sous-représentés dans le cadre de programmes destinés aux Autochtones, aux filles et aux jeunes femmes, aux jeunes à risque ainsi qu'à ceux vivant dans des communautés nordiques ou éloignées. Pour de plus amples renseignements, consultez notre site web à [actua.ca](http://actua.ca) et suivez-nous sur [Twitter](#), [Facebook](#), [Instagram](#) et [YouTube](#)!



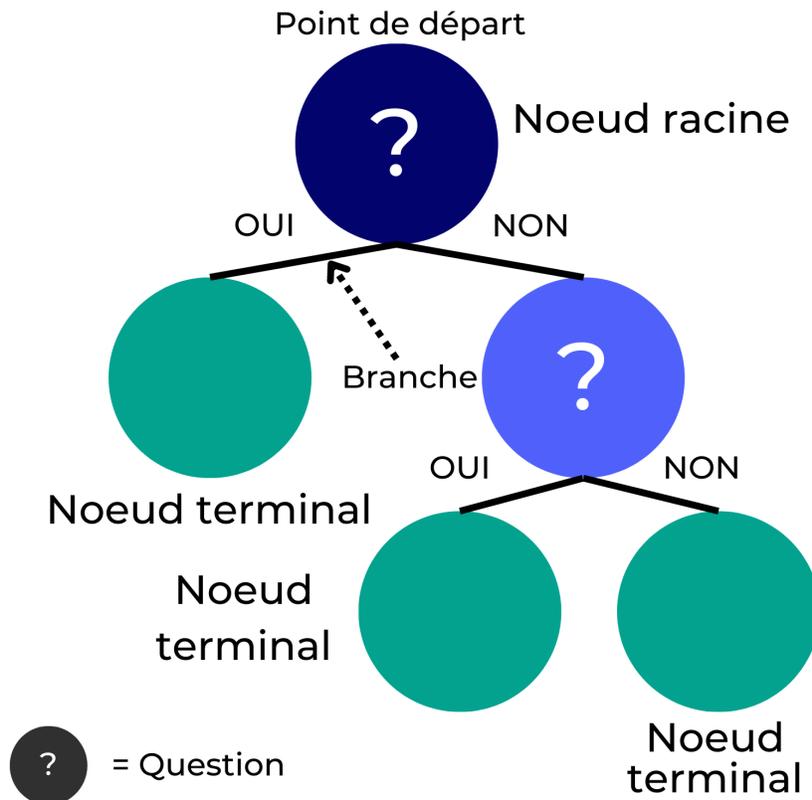
## Annexes

### Annexe A – Liens carrières/mentorat

- Statisticien/statisticienne
- Programmeur/programmeuse
- Analyste de données
- Programmeur/programmeuse en IA

### Annexe B – Information documentaire

#### Arbres décisionnels



Un arbre décisionnel a la structure d'un arbre inversé. Le point de départ s'appelle « racine » ou « nœud racine ». Chaque nœud correspond à une question. La 1<sup>re</sup> question est posée au niveau du nœud racine. Les deux branches tracées à partir de celui-ci représentent chacune une réponse et se terminent par un nouveau nœud. Lorsqu'il ne reste plus de questions à poser,

on cesse d'ajouter des branches. Le dernier nœud d'une série s'appelle « nœud terminal » ou « feuille ». Même s'il peut parfois y avoir plusieurs réponses à une question, la structure des arbres décisionnels est le plus souvent binaire, ce qui signifie qu'il n'y a que deux réponses possibles permettant de trier les données.

## Annexe C : Documents d'accompagnement

### Échantillon de jeu de données

Jeu de données d'apprentissage sur les objets spatiaux

Caractéristiques			Étiquette
Composition	Taille (d, mètres, approx.)	Distance (proche ou lointain)	
inorg	27,32	lointain	astéroïde
inorg	164,06	lointain	astéroïde
inorg	169,13	lointain	astéroïde
org	657,43	lointain	comète
org	879,37	lointain	comète
org	905,20	lointain	comète
org	957,84	lointain	comète
inorg	0,21	proche	débris
inorg	2,51	proche	débris
inorg	3,68	proche	débris
inorg	4,40	proche	débris
inorg	9,33	proche	débris
inorg	9,44	proche	débris
inorg	4,03	proche	météorite
inorg	4,38	proche	météorite
inorg	4,92	lointain	météorite
inorg	6,55	lointain	météorite
inorg	24,40	proche	satellite
inorg	25,76	proche	satellite

inorg	26,87	proche	satellite
-------	-------	--------	-----------

### Jeu de données de test sur les objets spatiaux

Utilisez les données ci-dessous pour évaluer le degré de précision de votre arbre décisionnel.

Caractéristiques			Étiquette
Composition	Taille (d, mètres, approx.)	Distance (proche ou lointain)	
inorg	157,02	lointain	astéroïde
inorg	181,30	lointain	astéroïde
org	656,60	lointain	comète
org	700,02	lointain	comète
org	939,23	lointain	comète
inorg	1,48	proche	débris
inorg	4,02	proche	météorite
inorg	12,28	proche	satellite
inorg	23,04	proche	satellite
inorg	26,87	proche	satellite

