

DANS LA PEAU D'UN ALGORITHME

QU'ILS NOUS GUIDENT DANS NOTRE CONSOMMATION COMMERCIALE, CULTURELLE OU INFORMATIONNELLE, LES ALGORITHMES DE RECOMMANDATION SONT DEVENUS INCONTOURNABLES POUR NOTRE NAVIGATION SUR LE WEB ET RESTENT LARGEMENT OPAQUES POUR CEUX ET CELLES QUI LES UTILISENT.

Ce jeu propose aux participant·e·s de se mettre, tour à tour, dans la peau des ingénieur·e·s de YouTube devant créer leurs propres algorithmes de recommandation devant classer les clips vidéo musicaux à partir de leurs données statistiques.



PRÉSENTATION DU JEU

Dans la peau d'un algorithme est un dispositif pédagogique visant à stimuler, chez les participants, une réflexion critique sur les plateformes numériques, les données qu'elles collectent, les algorithmes qui les composent et les usages que nous en avons.

Les participant-e-s, en étant confronté-e-s aux problèmes pratiques se posant lors de l'automatisation de la recommandation, sont amené-e-s à développer un questionnement sur les enjeux médiatiques, techniques et socio-culturels liés à son fonctionnement.

PUBLIC

Dans la peau d'un algorithme est parfaitement adapté aux groupes de **6 à 24 personnes**, **âgés de 14 ans ou plus** – les grands groupes ou classes requièrent évidemment une organisation plus laborieuse.

TEMPS

Le jeu nécessite **2 heures** et la discussion requiert **1 heure**, l'ensemble pourrait éventuellement être ramené à 2 heures, par des animateurs expérimentés.

MATÉRIEL

Il s'agit d'une activité déconnectée. Le seul matériel nécessaire comprend : **de quoi écrire** et les **fiches de jeu** imprimées pour les participant-e-s. Une calculatrice peut également être utilisée par les groupes qui le souhaitent.

MISE EN PLACE ET POSTURE

Lors de la première phase du jeu, le groupe est divisé en **3 équipes** de tailles équivalentes. Chaque équipe doit pouvoir travailler autour d'une table sans être dérangée par les autres.

Les équipes ne doivent pas communiquer entre elles durant le jeu. Lors de la dernière phase du jeu, le groupe entier se réunit en un large cercle pour pouvoir discuter tous ensemble.

Le rôle de l'animateur-trice consiste à :

- Donner les consignes et distribuer les documents aux différentes équipes à chaque étape.
- Être disponible pour les 3 équipes lors de l'étape de création de l'algorithme.
 - Pouvoir répondre à leurs questions
 - Donner des pistes aux équipes bloquées
 - Questionner les choix des équipes qui auraient terminé leur algorithme bien avant l'échéance
- Encourager les 3 équipes durant les étapes de calculs et veiller à ce qu'elles terminent leurs opérations dans les temps en s'assurant de la participation de tout le monde.

ASTUCE

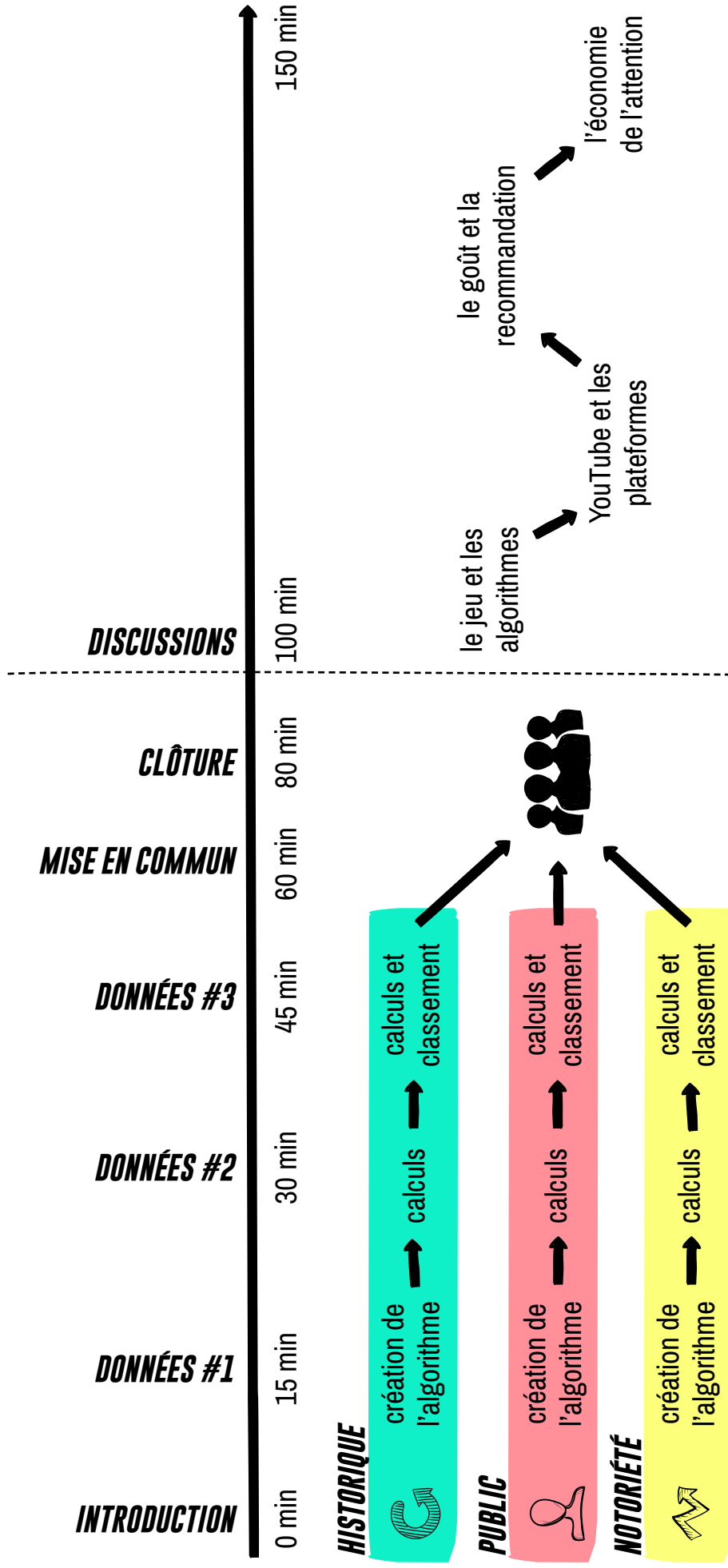
Il est fortement recommandé d'avoir pris connaissance de l'ensemble des fiches distribuées aux équipes avant de démarrer le jeu.

ASTUCE

Attention de ne pas orienter les participant-e-s dans une direction particulière. Il faut pouvoir les aider à formuler leurs propres idées sans pour autant indiquer implicitement une "bonne direction." L'idéal est de privilégier l'attitude la plus neutre possible.



SÉQUENCE DE JEU



Peut faire l'objet d'une deuxième session

DÉROULEMENT

INTRODUCTION

Commencez par lire l'introduction du jeu au groupe, divisez le ensuite en 3 équipes pour expliquer les règles.

Bienvenue chez YouTube ! Vous avez été recruté-e-s pour former trois équipes de conception des nouveaux algorithmes de recommandation de vidéos.

Dans une première phase, les équipes seront chacune tour à tour ingénieur-e-s et algorithmes. En tant qu'ingénieur-e-s, d'abord, vous devrez créer l'algorithme en question. En tant qu'algorithmes, ensuite, vous devrez l'appliquer pour classer une série de vidéos. Attention ! Vous devrez être efficaces : votre temps est limité.

Dans une deuxième phase, les trois équipes seront réunies afin que tout le monde puisse présenter, discuter et comparer les différents algorithmes et les classements qui en sont sortis. Nous, responsables techniques et commerciaux, désignerons alors le meilleur algorithme.

ASTUCE

Il est préférable de NE PAS préciser au groupe que les données reçues sont différentes pour chaque équipe.

DONNÉES #1

Chaque équipe va recevoir des données liées à des clips musicaux publiés sur YouTube, ainsi qu'une fiche résumant les instructions propres à leurs données.

Chaque équipe dispose ensuite de 15 minutes pour établir son algorithme en sélectionnant les données qu'elle estime pertinentes et en imaginant une règle pour combiner ces données de façon à produire un classement des clips musicaux. Une description plus détaillée et des exemples de formalisation de règles algorithmiques se trouvent aux pages **8 à 16**.

A la fin du temps imparti, la règle est fixée et NE pourra PLUS être modifiée pendant le reste du jeu.

DONNÉES #2

Chaque équipe va alors recevoir de nouvelles données. Elles devront s'organiser pour arriver à appliquer leur algorithme aux nouvelles vidéos en 15 minutes.

ASTUCE

Sugérez aux équipes de se répartir le travail de calcul pour être plus efficace.

DONNÉES #3

Chaque équipe reçoit les dernières données ainsi que 1 ou 2 fiches-classement pour ordonner les vidéos en fonction de leur score. L'équipe dispose encore une fois de 15 minutes.

Demander à chaque équipe de désigner en leur sein un-e porte-parole qui présentera les données de l'équipe ainsi que son algorithme lors de la mise-en-commun



MISE EN COMMUN

Inviter les équipes à s'installer en cercle pour la mise en commun.

- Chaque équipe devra à tour de rôle présenter son algorithme et le classement qui en est ressorti. Quelles sont les données que l'équipe a reçues ? Quelles sont les données que l'équipe a jugé pertinent d'utiliser ? Quelle règle de calcul a été construite à partir de ces données ? Quel est – ou quels sont (pour les équipes ayant dû recommander des vidéos pour deux profils différents) le(s) classement(s) produit(s) par cet algorithme ?

- Les membres des 2 autres équipes sont ensuite invité·e·s à critiquer de façon constructive l'algorithme présenté et la logique qui a mené à sa conception. Le raisonnement est-il pertinent ? Les objectifs de recommandation poursuivis par l'équipe sont-ils clairs ? La traduction de ceux-ci en une règle de calculs est-elle logique ? Auraient-ils fait autrement ?

- Une fois que les 3 algorithmes ont été présentés et critiqués, le groupe doit les hiérarchiser. Faudrait-il les prendre tous les trois en compte ou y en a-t-il qui ne sont absolument pas pertinents ? Certains sont-ils plus importants que d'autres à prendre en compte ? Chaque personne est invitée à voter pour l'algorithme qui selon elle prédira au mieux les vidéos sur lesquelles les profils-cibles vont cliquer en premier.

CLÔTURE DU JEU

L'animateur·rice est amené·e à clôturer le jeu en évaluant les algorithmes des différentes équipes. L'enjeu de cette évaluation est de désigner un algorithme gagnant, permettant de clôturer le jeu, et de donner une explication rudimentaire sur la façon dont YouTube évalue ses propres algorithmes. Vous pouvez lire le texte suivant ou vous le réapproprier avec vos mots.



« Nous voilà arrivés au moment fatidique. Nous, en tant que responsables techniques et commerciaux de YouTube, devons évaluer les mérites de vos algorithmes de recommandation. L'enjeu est d'identifier les algorithmes les plus prometteurs, afin de déterminer ceux sur lesquels nous devons concentrer nos efforts. Nous avons entendu vos présentations, quelques détails nous ont probablement échappé. Mais ce n'est pas bien grave, nous n'avons pas besoin de savoir comment vos algorithmes fonctionnent, pour pouvoir mesurer leurs performances – pour attribuer à chacun un score.

La seule véritable façon d'évaluer la performance d'un algorithme de recommandation est de faire des expériences en ligne. Leur but est de voir comment différentes personnes réagissent à différents algorithmes. Les algorithmes seront jugés plus performants si l'engagement des utilisateurs est plus important. Et on a plein de façon de mesurer cet "engagement" : Est-ce que les gens restent plus longtemps sur la plateforme ? Est-ce qu'ils voient ou cliquent sur plus de publicité ? Etc.

Le déploiement en ligne de nouveaux algorithmes sur nos infrastructures digitales installées aux quatre coins du monde est bien trop coûteux. On parle de centaines d'heures de travail et d'importantes transformations d'infrastructure. On ne peut pas tester sur des millions d'utilisateurs·trices n'importe quelle idée sortie de la tête d'un petit groupe d'ingénieur·e·s. Nous n'en sommes pas encore là, pour l'instant. On va d'abord évaluer la performance de vos algorithmes de recommandation via des expériences hors ligne, beaucoup plus accessibles et moins coûteuses. Leur but est de voir si vos algorithmes de recommandation parviennent à prédire les prochaines vidéos visionnées par vos profils cibles.

Il y a là quelque chose que nous vous avons caché. Nous avons un plus grand historique des personnes auxquelles vous deviez recommander des vidéos. Et nous savons si elles ont interagi ou non avec ces vidéos que vous étiez censées classer. On va donc pouvoir vérifier si vos prédictions visaient juste en comparant vos prédictions.



Comment faire ? La mesure hors ligne la plus utilisée de performance de l'algorithme, et bien, elle est très simple. On appelle cette métrique, le "Recall" :

- D'abord, on ne considère que les 3 premières vidéos des classements de chaque équipe (« classements prédictifs ».) Ce sont les seules qui s'affichent dans la barre de droite YouTube, quand on regarde une vidéo- et les seules sur lesquelles les gens cliquent effectivement.
- Ensuite, on cherche à identifier, pour chaque équipe, les prédictions qui s'avèrent correctes. Pour ce faire, on compte le nombre de fois que chaque vidéo présente dans les 3 premières positions de chaque classement prédictif se retrouve aussi dans les 3 premières vidéos vues par les profils cibles (« classements réels ».)
- Une vidéo présente à la fois dans le top 3 du classement prédictif et du classement réel rapporte un point, et un point supplémentaire si elle se trouve à la même place dans le classement.

ASTUCE

Pour l'équipe « notoriété », il faut donc comparer le même classement prédictif avec les deux classements réels.

PALMARES HISTORIQUE #3	VIDÉOS
	1. ARCADE FIRE
	2. BEATLES
	3. BIG FLO ET OLI
	4.
	5.
	6.
	7.
	8.
	9.
	10.

PALMARES HISTORIQUE #3	VIDÉOS
	1. JUL
	2. ANGÈLE
	3. ARIANA GRANDE
	4.
	5.
	6.
	7.
	8.
	9.
	10.

PALMARES HISTORIQUE #3	VIDÉOS
	1.
	2.
	3.
	4.
	5.
	6.
	7.
	8.
	9.
	10.

Le Recall, cette méthode rudimentaire, décrit plutôt fidèlement la façon dont les ingénieur-e-s procèdent. Il est évident que YouTube combine souvent les différents types de données et d'algorithmes, que YouTube travaille sur un nombre de vidéos et d'utilisateurs beaucoup plus grands, et que YouTube prend en compte d'autres types de métriques. Une fois qu'un algorithme de recommandation semble satisfaisant hors ligne, il peut être raffiné et déployé. Il sera ainsi testé en ligne et on pourra mesurer, de différentes façons, la qualité de l'engagement des utilisateurs. Voilà comment YouTube assure son objectif principal : augmenter ses revenus, maximiser l'engagement des utilisateurs, améliorer ses prédictions.



COMMENT TRANSFORMER LES DONNÉES EN ALGORITHMES ?

Voici pour chaque équipe des exemples d'algorithmes auxquels les équipes pourraient aboutir sur base des données reçues. Ces méthodes et ces exemples sont là pour que vous puissiez aider les équipes en panne d'inspiration et ne visent pas l'exhaustivité : il existe une infinité de façons plus originales et créatives de combiner les données pour créer des règles. La seule chose qui importe est d'aboutir à un classement des vidéos. Ces méthodes sont suivies par une description des données de chaque équipe et des manières de les interpréter, mais également des biais que celles-ci peuvent entraîner.

ASTUCE

Pour gérer les égalités probables, il est important de déterminer à l'avance comment régler ces cas

ÉQUIPE PUBLIC

Méthodes

Les données "Public" concernent les informations démographiques de l'audience des vidéos YouTube. L'équipe doit d'abord choisir quels sont les critères jugés pertinents et comment les interpréter. Elle doit ensuite formaliser une comparaison de ces données avec celles du profil-cible.

ASTUCE

L'équipe peut combiner plusieurs méthodes en un algorithme.

L'équipe doit prévoir une façon d'arbitrer les égalités.

L'équipe peut délibérément choisir de ne pas tenir en compte de tout d'un critère.

L'équipe peut également décider d'inverser le pourcentage d'un critère.

L'équipe peut introduire de l'aléatoire dans le classement.

Méthode n°1 : Score en fonction du pourcentage du profil-cible

1. On prend le pourcentage qui correspond au profil-cible pour chaque critère
2. On pondère ce pourcentage pour chaque type de données selon l'importance du critère
3. On peut différencier cette pondération au niveau de la localisation en fonction du pourcentage de population dans les "autres pays" que le top 3

LOCALISATION → Si pays d'origine du profil-cible dans le top 3, son % $\times 5$

1^{er} CRITÈRE → Si "autres" $> 40\%$, alors ce % $\times 3$
↳ car le groupe veut favoriser le contenu qui aurait un succès "international", qui dépasse les régions

ÂGE → % de la tranche d'âge correspondant au profil-cible $\times 2$

2^e CRITÈRE

GÉNÈRE → on n'en tient pas compte
↳ car le groupe considère que ce n'est pas pertinent

APPLICATION SUR « POOKIE » @ Juliette

→ Pays d'origine (BEL) pas dans TOP 3 = 0

→ Autres $> 40\%$ $48 \times 5 = 240$

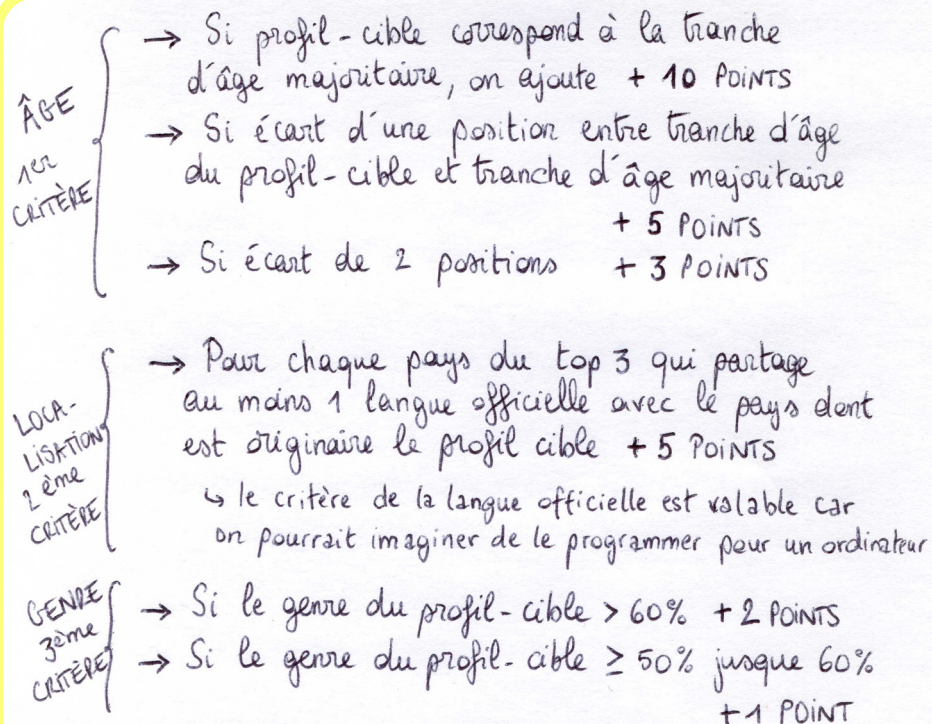
→ 13-17 ans $39 \times 2 = 78$

$\left. \begin{array}{l} 0 \\ 240 \\ 78 \end{array} \right\} = 318$



Méthode n°2 : Score en fonction de la position du profil-cible

1. On donne un score pour chaque critère en fonction de la position du profil-cible
2. On pondère ce score en fonction de l'importance donnée au critère



APPLICATION SUR «POOKIE» 🎯 Juliette

- Profil-cible = tranche d'âge majoritaire + 10
 - Profil-cible → BELGIQUE Langues off: FR, NL, ALL
 - France → FR ✓ + 5
 - RDC → FR ✓ + 5
 - Mali → FR ✓ + 5
 - Genre profil-cible = 57% + 1
- = 26

Comment interpréter les informations démographiques ?

- Le pourcentage de vues en fonction du **genre** présente de manière binaire la proportion d'hommes et de femmes qui ont regardé la vidéo.

- Le pourcentage de vues en fonction de la **localisation** indique le top 3 des pays depuis lesquels les utilisateurs et utilisatrices ont visionné la vidéo ainsi que la proportion des personnes l'ayant visionnée ailleurs. Plus la proportion "Autres" est élevée, plus la vidéo touche une audience internationale et peut donc être considérée comme mainstream. À l'inverse, une vidéo présentant un pourcentage élevé de son audience dans un seul pays pourra être considérée comme une vidéo touchant une audience locale. L'équipe peut également inférer une ou plusieurs langues parlées dans le pays concerné en considérant cette donnée comme plus pertinente que la localisation géographique en soi.

- Le pourcentage de vues en fonction de la **tranche d'âge** est un graphique en bâtons représentant la proportion de chaque catégorie d'âges des personnes ayant visionné la vidéo. L'équipe peut utiliser le pourcentage de la catégorie d'âge du profil-cible mais également l'écart entre la catégorie d'âge majoritaire et celle du profil-cible.



Les biais d'interprétation

Le biais principal qui peut apparaître en utilisant les données "Public" consiste en la boucle de rétroaction qui se met en place lorsque l'on recommande une vidéo à une catégorie de personnes qui est déjà majoritaire dans l'audience de cette vidéo. L'exemple le plus évident est celui des données de genre : l'écart entre le nombre d'hommes et de femmes qui visionnent une vidéo aura tendance à augmenter mécaniquement si on recommande cette vidéo principalement au genre déjà majoritaire des personnes qui l'ont regardée.

Les pays les plus peuplés et les plus connectés dans le monde sont naturellement plus souvent présents dans le top 3 du pourcentage de vues en fonction de la localisation. Il y a également des difficultés à inférer la langue d'une audience uniquement en fonction de sa localisation : par exemple, un visionnage depuis la Belgique pourrait concerner une vidéo pour une audience francophone, néerlandophone ou germanophone si on se réfère aux langues nationales.

Le pourcentage de visionnages en fonction de l'âge ne prend pas en compte la proportion de personnes connectées pour chaque catégorie.

ÉQUIPE NOTORIÉTÉ

Méthodes

Les données "Notoriété" reprennent les informations non-personnalisées liées à chaque vidéo. L'équipe chargée de les traiter doit donc fournir un classement de vidéo absolu indifférencié en fonction de l'audience des vidéos. Ce sont les données les plus fournies, qui nécessitent donc un travail important de sélection de données pertinentes et de hiérarchisation de celles-ci. Elles sont divisées en deux catégories : à gauche, les données propres à la vidéo et à droite celles qui concernent la chaîne qui a publié la vidéo.

ASTUCE

L'équipe peut simplifier (arrondir) les grands chiffres pour les manipuler plus facilement. L'équipe peut introduire de l'aléatoire dans le classement.

Méthode n°1 : mise en équation

1. L'équipe choisit les données qui lui semblent pertinentes à utiliser et écarte les autres
2. L'équipe établit des opérations entre chaque donnée
3. Ces opérations peuvent être pondérées

$$\frac{\text{(Vues vidéo / mois mise en ligne)}}{\text{Abonné·e·s à la chaîne}}$$

APPLICATION SUR «POOKIE»

$$\frac{275\,729\,549 / 22}{5\,390\,000} = 2,32$$

→ le groupe décide de favoriser les vidéos récentes ayant eu un certain succès

→ le groupe décide de pénaliser les grosses chaînes dont le nombre de vues à une vidéo peut s'expliquer par un nombre de fans conséquent



→ le groupe décide que les like et les dislike ont la même valeur en tant que signe des "réactions" suscitées par une vidéo

le groupe décide de pondérer la donnée "vues vidéo" afin de limiter son impact

$$\frac{\text{Like} + \text{Dislike} + \text{Vues vidéo}}{100\ 000}$$

Mois mise en ligne

APPLICATION SUR «POOKIE»

$$\frac{1700 + 140 + 2757}{22} = 208,9$$

Comment interpréter les données de notoriété ?

- Le **nombre de vues d'une vidéo** est un marqueur de popularité. Le nombre de vues est corrélé au nombre de mois depuis lequel la vidéo est en ligne et sera influencé par le nombre d'abonnements à la chaîne, ceux-ci étant alertés lorsqu'une nouvelle vidéo est publiée sur la chaîne.
- Le **nombre de mois depuis la mise en ligne de la vidéo** peut être utilisé pour rapporter les autres données à sa longévité : le nombre de vues ou de "likes" par mois, par exemple. Prendre en compte cette donnée permet de valoriser une vidéo en fonction de son ancienneté ou de sa nouveauté.
- Les **nombre de "likes" et de "dislikes"** correspondent à la popularité déclarée d'une vidéo. S'ils peuvent être considérés comme des signes de plébiscite ou de rejet, ils sont également un marqueur de l'engagement provoqué par la vidéo, c'est-à-dire de réactions indifféremment positives ou négatives. On peut donc les utiliser comme des données s'opposant l'une à l'autre ou, à l'inverse, les considérer de manière indifférenciée.
- Le **nombre de commentaires** est un marqueur d'engagement. On ne peut en effet pas les analyser de manière qualitative (le contenu du commentaire) mais uniquement de manière quantitative (le nombre total de commentaires.)
- Le **nombre d'abonnements à la chaîne** peut être considéré comme un marqueur de popularité de la chaîne. Une personne s'abonne à une chaîne pour être informée des nouvelles vidéos publiées par cette chaîne. Ce nombre est corrélé au nombre de mois depuis la mise en ligne de la chaîne.
- Le **nombre de vidéos publiées** est, corrélé au nombre de mois depuis la mise en ligne de la chaîne, un marqueur de productivité de la chaîne.
- Le **nombre de mois depuis la mise en ligne de la chaîne** peut être utilisé pour rapporter les autres données à sa longévité : le nombre de vues ou d'abonnements par mois, par exemple. Prendre en compte cette donnée permet de valoriser une chaîne en fonction de son ancienneté ou de sa nouveauté.
- Le **nombre de vues de la chaîne** est un marqueur de popularité. Le nombre de vues est corrélé au nombre de mois depuis lequel la chaîne est en ligne et également au nombre de vidéo publiées et du nombre d'abonnements, ceux-ci étant prévenus lors de la publication d'une vidéo.



Les biais d'interprétation

Le nombre de vues une donnée très sensible aux boucles de rétroaction, puisque plus on recommande une vidéo vue de nombreuses fois, plus le nombre de vues augmentera.

Il n'y a pas d'informations à propos de la répartition des interactions (vues, "likes", "dislikes", commentaires et abonnements) sur la période de mise en ligne. Cela biaise donc les combinaisons calculées entre le nombre de vues et le nombre de mois depuis la mise en ligne (d'une vidéo ou d'une chaîne.)

Les interactions avec une vidéo peuvent être manipulées de différentes façons (création d'un bot qui "like" ou "dislike" des vidéos automatiquement, concours impliquant un abonnement ou un commentaire, création de multiples comptes,...)

ÉQUIPE HISTORIQUE

Méthodes

Pour interpréter les données "Historiques", il faut imaginer que le profil-cible et les quatre profils sélectionnés ont visionné de nombreuses vidéos en commun. Les vidéos à gauche du tableau ont été vues au delà de la moitié par le profil-cible et ont été choisies parce qu'elles présentent des variations de visionnage parmi les 4 profils affinitaires.

Les 5 vidéos à droite du tableau sont celles qui doivent être notées et classées et n'ont donc pas encore été recommandées au profil-cible mais ont été, pour certaines, déjà vues (en deçà ou au delà de la moitié) par les profils affinitaires.

Pour classer ces 5 vidéos, il faut leur donner un score en fonction de qui les a vues, et donc attribuer un indice de pondération à chaque profil affinitaire et à chaque type de visionnage (pas du tout, moins de la moitié, plus de la moitié) On peut donc diviser le calcul en 3 étapes.

Méthode n°1 : Calcul matricielle

1. Donner un indice aux types de visionnage
 - L'équipe décide comment elle interprète les différents types de visionnage et leur attribue un indice.
 - Cet indice peut être absolu ou relatif aux différents profils
2. Donner un indice aux profils affinitaires
 - L'équipe donne un indice à chaque profil affinitaire en additionnant ses indices de visionnage pour toutes les vidéos vues par le profil-cible (à gauche du tableau)
3. Attribuer un score aux vidéos pour le profil-cible
 - L'équipe calcule le score de chaque vidéo à recommander en fonction des indices des profils, affinitaires et des types de visionnage de ceux-ci pour les vidéos à recommander



↳ D'abord, on attribue un indice aux \neq types de visionnage

$\text{diagonal} = -3$ $\square = 3$ $\blacksquare = 3 \rightarrow$ le groupe décide de pénaliser les vidéos qui n'ont pas retenu l'adhésion jusqu'au bout

↳ Sur cette base, on calcule un score pour chaque profil

$$\text{Maité} = -3 - 3 - 3 + 3 - 3 \Rightarrow -9$$

$$\text{Thomas} \Rightarrow 9 \quad \text{Bilal} \Rightarrow 15 \quad \text{Julie} \Rightarrow 15$$

↳ Enfin, on croise les scores des profils à la partie droite du tableau (c-à-d le visionnage qu'ils ont fait de chaque vidéo) en les ADDITIONNANT

↳ Δ Pour gérer les égalités probables, on décide de faire passer devant la vidéo qui a le moins de diagonal . Et s'il y a encore égalité, on le joue au dé.

APPLICATION SUR «POOKIE» \odot Juliette

$$\left[\begin{array}{c} \text{diagonal} \\ -3 \end{array} - 9 \right] + \left[\begin{array}{c} \text{diagonal} \\ -3 \end{array} + 9 \right] + \left[\begin{array}{c} \blacksquare \\ 3 \end{array} + 15 \right] + \left[\begin{array}{c} \blacksquare \\ 3 \end{array} + 15 \right] = 30$$

Maité Thomas Bilal Julie

↳ D'abord, on attribue un indice aux \neq types de visionnage

$\text{diagonal} = 1$ $\square = 5$ $\blacksquare = 20$

↳ Sur cette base, on calcule un score pour chaque profil

$$\text{Maité} = 1 + 1 + 1 + 5 + 1 \Rightarrow 9$$

$$\text{Thomas} \Rightarrow 36 \quad \text{Bilal} \Rightarrow 40 \quad \text{Julie} \Rightarrow 70$$

↳ Enfin, on croise les scores des profils à la partie droite du tableau (c-à-d le visionnage qu'ils ont fait de chaque vidéo) en les MULTIPLIANT

APPLICATION SUR «POOKIE» \odot Juliette

$$\left[\begin{array}{c} \text{diagonal} \\ 1 \end{array} \times 9 \right] + \left[\begin{array}{c} \text{diagonal} \\ 1 \end{array} \times 36 \right] + \left[\begin{array}{c} \blacksquare \\ 20 \end{array} \times 40 \right] + \left[\begin{array}{c} \blacksquare \\ 20 \end{array} \times 70 \right] = 2245$$

Maité Thomas Bilal Julie

Comment interpréter le type de visionnage ?

- Une vidéo **vue au delà de la moitié** par un profil affinitaire peut être interprétée comme ayant été appréciée par celui-ci. Plus deux profils ont de nombreuses vidéos vues au delà de la moitié en commun, Plus ils peuvent être considérés comme proches.
- Une vidéo qui **n'a pas été vue du tout** par un profil affinitaire peut être interprétée comme n'ayant pas éveillé l'intérêt de celui-ci.
- Une vidéo **vue en deçà de la moitié** par un profil affinitaire laisse davantage d'incertitude. On peut l'interpréter comme ayant suscité l'intérêt du profil pour qu'il clique dessus mais n'ayant pas gardé son attention au delà de la moitié.

Les biais d'interprétation

- On ne peut que spéculer sur la signification exacte du type de visionnage. Il peut très bien y avoir plusieurs personnes qui utilisent le compte du profil affinitaire pour visionner des vidéos, que la vidéo a été lancée mais que le profil n'est plus devant son écran, que la vidéo a été coupée car la personne a été interrompue dans son visionnage, que la personne a cliqué involontairement sur un lien, etc.

DISCUSSIONS

Déroulement

Une fois les phases de **mise en commun** et **clôture du jeu** terminées, la partie ludique s'achève et la discussion peut être entamée.

Celle-ci est structurée en **4 thématiques** différentes. Nous avons établi une liste non exhaustive des questions qui peuvent être posées pour chacune d'entre elles. Cette proposition consiste en un cadre non contraignant : chaque groupe est différent et identifie des enjeux variés en fonction de son expérience du jeu. Certains sujets suscitent de nombreuses réactions et d'autres laissent les participant-e-s plutôt indifférents. Il est donc conseillé de s'attarder sur les points qui génèrent spontanément des débats et de survoler les autres, en particulier lorsque le temps d'animation est limité. Vous pouvez également orienter cette discussion en fonction de vos objectifs éducatifs spécifiques et de votre cadre d'apprentissage. C'est pourquoi il n'est pas forcément nécessaire d'explorer en profondeur toutes les questions de chaque thématique, ni de suivre l'ordre établi par les fiches.

ASTUCE

Vous pouvez également utiliser les informations des fiches "Comment passer des données aux algorithmes" et plus particulièrement des biais d'interprétation pour relancer les discussions.

Il est possible que vous n'ayez pas toutes les réponses aux questions des participant-e-s. Pas de panique : ce sont des questionnements complexes qui évoluent rapidement. Il peut être alors constructif de séparer les questions techniques et politiques. Les premières peuvent faire l'objet d'une recherche approfondie par la suite tandis que les secondes ne peuvent se résoudre que par le débat, en évacuant les jugements moraux et autres considérations normatives. C'est toute l'idée des enjeux proprement démocratiques des technologies algorithmiques.



Thématique 1

“DANS LA PEAU D'UN ALGORITHME” [retour sur le jeu]

Cette partie se focalise sur l'expérience du jeu en soi. L'objectif est d'inviter les participant-e-s à formaliser leur ressenti et leur compréhension de ce qu'est un algorithme.

D'après vous, pourquoi avons-nous joué à ce jeu ? Quels sont ses objectifs éducatifs ?

Qu'avez-vous compris ou appris en jouant à ce jeu ?

Après avoir joué à ce jeu, comment définiriez-vous un algorithme avec vos mots ? A quoi cela sert-il ? Comment cela fonctionne-t-il ?

Comment avez-vous vécu la partie ?

Quelles ont été vos principales difficultés ?

Avez-vous eu envie de modifier l'algorithme pendant les calculs, et pourquoi ?

Est-ce que vous avez toujours été d'accord entre vous sur les choix à faire ?

Est-ce que vous pensez que votre algorithme reflète bien vos intentions ?

Quelles données auriez-vous voulu avoir à disposition ? Ces données vous semblent-elles récoltables ?



Thématique 2

“YOUTUBE” [plateforme]

Cette partie s'intéresse à YouTube, à l'utilisation qu'en font les participant-e-s et à leurs représentations de son fonctionnement.

Comment arrivez-vous sur une vidéo ? Est-ce que vous y êtes conduits via d'autres plateformes ? Est-ce que vous cherchez ? Est-ce que vous cliquez sur les recommandations ?

Qui utilise YouTube ? A quelle fréquence ? Qu'est-ce que vous y regardez ?

Est-ce que la recommandation est nécessaire ?

Qui est content des recommandations YouTube ? Qui les suit ? Qui ne les suit pas ?

Avant les réseaux sociaux, quels étaient les moyens de recommander ?

Existe-t-il d'autres types de recommandation ? Comment fonctionnent-elles ?

Est-ce que ce que vous voulez regarder sur YouTube est ce que vous regardez effectivement ?

Est-ce que vous avez une chaîne YouTube ? Pourquoi avez-vous choisi YouTube ?

Quelles sont les stratégies que l'on peut mettre en place pour mettre son contenu YouTube en avant ?

Comment YouTube gagne-t-il de l'argent ?

D'après vous, comment pensent les ingénieurs YouTube ? Quels sont leurs objectifs ?

Pensez-vous que votre algorithme est similaire à celui de YouTube ? Poursuit-il les mêmes intentions ?

Imaginez votre plateforme de partage de vidéos idéale. Comment fonctionnerait son algorithme ? Pourrait-on se passer d'un algorithme ? Comment ?



Thématique 3

“LES GOÛTS” [recommandation]

Cette partie développe la réflexion introduite à la fin de la précédente et met en évidence les tensions entre une intention et un calcul : comment recommander des vidéos à un public selon des critères objectifs et des données quantitatives ? Dans ce sens, on essaye de caractériser ce qu’est une “bonne” recommandation et les différents objectifs qui la sous-tendent.

Pour vous, que signifie “recommander” quelque chose à quelqu’un ?

Qu’est-ce qu’une “bonne” recommandation ?

Vous arrive-t-il de recommander du contenu culturel ? (film, série, musique, livre, sketch,...)

Qu’est-ce qui est difficile dans le fait de recommander ?

A qui faites vous des recommandations ?

Sur quoi vous basez-vous pour recommander un contenu ?

Comment se construisent les goûts ? Quels sont les facteurs qui les influencent ? Ceux-ci se retrouvent-ils dans les données du jeu ?

Qu’est-ce que le goût (culturel) ?

Quel est le rôle de la recommandation dans la construction des goûts ?

Quel est le rôle des algorithmes dans la construction des goûts ?



Thématique 4

“L'ÉCONOMIE DES PLATEFORMES” [médias]

Cette partie propose, en guise de conclusion, une mise en perspective des algorithmes de recommandation dans une vision plus large des plateformes numériques et de leurs modèles économiques. Elle vise à élargir la problématique des algorithmes et de leur incidence sur nos pratiques quotidiennes – nos interactions sociales, nos manières de nous informer, nos espaces de débat – à des questions de citoyenneté et de démocratie.

Comment ces différentes plateformes gagnent-elles de l'argent ?

Pouvez-vous nommer d'autres plateformes numériques qui utilisent des algorithmes ? (réseaux sociaux, streaming de musique et de vidéo, moteurs de recherches, sites de vente,...)

A quoi servent et comment sont utilisés les algorithmes sur ces différentes plateformes ?

Quel impact cela a-t-il sur notre façon de nous informer et donc notre vision du monde ?

**Dans quel cas faudrait-il modifier radicalement leur fonctionnement ?
Voire éviter d'en utiliser ? Par quoi pourrait-on les remplacer ?**