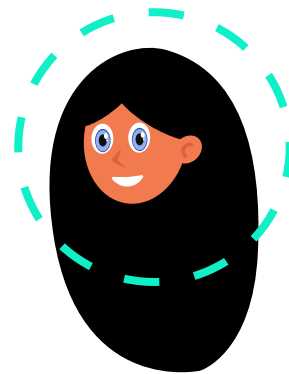


# ANIMATION : ALGORITHMES ET RECOMMANDATION



**Organisation générale :** en 4 équipes et en grand groupe

**Matériel :** pour chaque équipe, 2 fiches données “exemples” et 3 fiches données “à classer”. Pour l'enseignant-e : 1 indélébile foncé. Matériel de projection si possible.

**Espace/aménagement local :** Aménager l'espace en 4 îlots

**Durée :** 2x50 minutes

## DESCRIPTION ACTIVITÉ :

1

(10 minutes)

**S'intéresser aux usages des élèves: vous informez-vous en ligne ? Si oui, comment ? Via des applis liées à un média particulier (Rtl Info, Le Soir, autre) ? Via des réseaux sociaux (Facebook, Instagram, Twitter, autre) ? Via une application d'information (Google Actu, Apple News, autre) ?**



Il est important de ne pas être dans le jugement lors de cette phase. Si vos élèves ne s'informent pas parce qu'ils ne font pas confiance aux journalistes ou parce que cela les ennue, ou s'ils s'informent uniquement via les réseaux sociaux, ce sont des informations utiles pour vous afin de mener des débats intéressants à ces sujets par la suite.

2

(5 minutes)

**Mise en situation à expliquer aux élèves :**

**La radio de l'école décide de mettre en avant, chaque jour, 3 articles qui font l'actualité. Ces articles pourront ensuite être analysés par les profs dans le cadre de certains cours. Les enseignant-e-s ont décidé de laisser la responsabilité de cette sélection aux élèves. Cependant, pas question de perdre du temps quotidiennement à soumettre les articles du jour à un vote. Ils ont eu une idée : votre classe sera en charge d'imaginer un algorithme qui donnera un score aux articles pour établir un podium de façon automatisée.**

**Imaginer un algorithme :**

**Distribuer les données “exemples” à chaque équipe. Les élèves négocient pour choisir jusqu’à 5 types de données qu’ils jugent pertinents pour établir leur top 3. Une fois d’accord sur les données à utiliser, les élèves conçoivent un algorithme qui doit permettre de donner un score à chaque article pour établir un classement entre eux. Pour ce faire, ils doivent se mettre d’accord sur l’objectif qu’ils veulent atteindre avec leur algorithme.**

Avertir les élèves qu’une fois que l’algorithme est fixé, il ne pourra plus changer, il ne pourra qu’être exécuté.

Remarque 1 : préciser aux élèves que les articles donnés sont des exemples. Ils recevront ensuite d’autres articles sur lesquels appliquer leur algorithme.

Remarque 2 : préciser que les données se situent entre 1 et 20 parce qu’elles ont été pondérées pour éviter de manipuler de trop gros chiffres.

**Consignes pour la conception de l’algorithme :**

- **L’équipe peut tenir compte de maximum 5 critères, mais elle peut aussi décider d’en choisir moins. Si une équipe veut se baser sur 1 critère qui lui semblerait suffisant, c’est autorisé ;**
- **L’équipe doit prévoir une façon d’arbitrer les égalités ;**
- **L’équipe peut utiliser des additions, des soustractions, des multiplications, des divisions, mais il est aussi possible d’utiliser d’autres méthodes (par exemple : arbre de décision). La priorité est de parvenir à appliquer leur algorithme dans un délai relativement bref ;**
- **L’équipe peut introduire de l’aléatoire dans son algorithme. L’aléatoire peut être concrétisé par un lancer de dés, une pioche, etc ;**
- **L’équipe peut utiliser une calculatrice.**

**Exemple de processus de création d’un algorithme :**

Une équipe choisit de favoriser les médias qui semblent apporter une plus-value à l’information et traiter de sujets plutôt sérieux. Pour déterminer quels médias sont qualitatifs selon leurs critères, les élèves décident d’utiliser le nombre de visites mensuelles et de soustraire le nombre d’articles publiés par jour. Ainsi, si un média reçoit beaucoup de visites uniquement parce qu’il publie à la chaîne, il est pénalisé. L’équipe décide de ne pas tenir compte du nombre d’abonné-e-s car les élèves notent que d’expérience, ils connaissent des médias peu sérieux qui ont plus d’abonné-e-s que des médias qu’ils jugent de qualité.

L’équipe considère que la longueur d’un article (nombre de mots) est un indice qu’une plus-value journalistique a été apportée. Les médias qui copient-collent des informations légères et “putaclic” rédigent en général des articles super courts. Alors que quand il y a de l’analyse, les articles sont plus longs. Pour être sûrs d’éviter les articles putaclic sur lesquels les gens cliquent mais ne lisent que dans les grandes lignes pour avoir la réponse à une curiosité engendrée par le titre, les élèves décident de pénaliser les articles sur lesquels les gens passent un temps très court.

L’équipe a choisi d’utiliser 4 types de données : 1) Nombre de visites mensuelles ; 2) Articles publiés par jour ; 3) Nombre de mots ; 4) Temps resté sur la page.

**Et voici l'algorithme imaginé :**

$[(\text{visites/mois}) / (\text{articles/jour})] + [(\text{nombre de mots})] + [\text{pénalité de } -5 \text{ si le temps resté sur la page est inférieur ou égal à } 3]$  En cas d'égalité, c'est l'article qui a le plus de mots qui gagne. S'il y a encore égalité, on le joue au dé.

**Ce qui dans la phase d'application de l'algorithme donnerait :**

Ce qui donnerait en top 3: 1. "Le vélo électrique, vraiment propre ? Enquête sur les terres rares" (Score = 22) ; 2. "Les voitures de société délétères pour le climat ? Analyse des conséquences écologiques d'un système propre à la Belgique" (Score = 21) 3. "Washington-Pékin : une nouvelle étape franchie dans les relations entre les 2 superpuissances" (Score = 21).

L'article sur les voitures de société est passé devant car il avait un score de "20" sur le critère du nombre de mots (prévu en cas d'égalité).

Les articles qui ont obtenu le plus faible score sont : "McDonald offre son nouveau burger à tous les Léo et toutes les Léa" (Score = 1.25), "Selon une étude, les infusions de radis seraient très efficaces pour lutter contre le cancer de la prostate" (Score = 1.23) et "Le cadeau trop chou du prince Harry pour Meghan Markle (Score = 0.94)

Il faut s'assurer que les élèves n'avancent pas à l'aveugle et s'accordent, par la discussion et par essai-erreur à partir des données des exemples, sur une ou plusieurs finalités que leur algorithme doit pouvoir traduire. En effet, l'objectif principal de cette activité est de réfléchir aux enjeux posés par les algorithmes de recommandation.

**4****(15 minutes)****Appliquer son algorithme :**

**Distribuer les données "à classer" à chaque équipe et barrer à l'indélébile les colonnes que les élèves ont décidé de ne pas utiliser.**

**Les élèves appliquent leur algorithme pour attribuer un score à chacun des 20 articles de la fiche pour établir leur top 3. Les élèves doivent s'organiser pour se répartir les tâches et calculer les scores dans le délai imparti.**

**Quand l'équipe a fini de calculer les scores des articles, elle va consigner son top 3 au tableau.**

**5****(35 minutes)****Discussion en grand groupe :**

- Est-ce que vous pensez que votre algorithme reflète bien vos intentions ? Si non, comment expliquer l'écart entre vos intentions et le résultat ?
- Quelle équipe a intégré [tel type de donnée] dans son algorithme et pourquoi ? Que signifie cette donnée à vos yeux ?

- Si une autre équipe a utilisé la même donnée autrement, observer si elle lui a donné la même signification. Relancer au grand groupe : qui aurait apporté encore une autre signification et pourquoi ? Nourrir la réflexion en proposant des relances (cf. ci-dessous “interprétation des données”).

Quels autres types de données auriez-vous souhaité avoir à disposition ? Pensez-vous que ce type de données est disponible dans la réalité ? Si oui, comment pensez-vous qu’elles sont produites/récoltées ? Si non, que pourrait-on mettre en place pour les produire ? Est-ce facilement réalisable ?

- Quels sont les avantages et inconvénients d’une éditorialisation faite par des algorithmes par rapport à une éditorialisation faite par des humains ? Parmi les humains, qui est le mieux placé pour recommander ? Dans le cas de la mise en situation : plutôt des élèves ? des enseignant-e-s ? des personnes externes ? De quelle façon ces personnes pourraient-elles procéder pour faire les meilleures recommandations ?
- Comment pensez-vous que ça fonctionne sur des plateformes comme Facebook ou Twitter ? Quels problèmes peuvent découler de leur fonctionnement ? Comment pourrait-on les programmer pour qu’ils servent mieux nos intérêts en tant que société ? Quelles limites pourraient apparaître ?

L’interprétation de chaque type de données offre des pistes pour mener la discussion et questionner les choix faits par les élèves. Pour aller plus loin, lire au préalable le dossier enseignant-e du même parcours.

Interprétation des données :

#### **[Vues] :**

Cette mesure se calcule au nombre de clics. Un clic signifie avant tout que l’article a suscité la curiosité des internautes. Cela ne dit par contre pas si l’article a été lu avec attention, s’il a été jugé pertinent, s’il a été apprécié, etc. Souvent, le plus populaire est ce qui est le plus “sensationnel”, faisant appel aux émotions, le plus “putaclic”, etc.

#### **[Like, Love] :**

Le “like”, ou le “love” (emoji sur Facebook) peuvent avoir de multiples significations. Ils s’apparentent souvent à une forme de plébiscite, mais pas toujours. Dans tous les cas, il s’agit d’un moyen très rapide (voire pulsionnel) et à coût cognitif quasiment nul d’exprimer un soutien, un intérêt, une approbation, etc. Le like, comme le “love” sont visibles de ses amis, ou du grand public quand notre compte n’est pas privé. Il s’agit donc d’une expression personnelle qui implique un minimum d’engagement public. On imagine mal une personne “liker” un propos polémique, par exemple raciste, sexiste, ou homophobe si le réseau de cette personne est opposé à ces idées. Un article qui énonce une situation complexe, une triste réalité, ou une analyse distanciée (peu émotionnelle) d’un événement générera en général moins de like qu’une bonne nouvelle, un article rigolo, une curiosité, etc.

#### **[Dislike, Grrr, Triste] :**

Une réaction “négative” n’est pas forcément un signal de non appréciation d’un article. La réaction peut viser la situation relatée par l’article mais avoir grandement intéressé la personne. Facebook n’utilise pas de bouton “dislike” et lui a préféré l’ajout du “grrr” et du “triste”. Sur des plateformes comme Youtube ou Reddit, le “dislike” équivaut plutôt à une forme de signal négatif envoyé directement à l’algorithme par les internautes. Au plus il y aura un ratio de “dislike” élevé, au moins le contenu sera mis en avant par la plateforme. Il arrive régulièrement que ces plébiscites négatifs soient l’effet d’une campagne de groupe contre un sujet, ou contre un créateur de contenu (exemple : la communauté de fans de Bob l’éponge qui a mené une opération massive de dislike contre la vidéo de prestation de Maroon 5 au superbowl). Les annonceurs de publicité préfèrent être associés à du contenu positif, léger, non polémique (et donc pas trop marqué politiquement), les réactions “négatives” risquent d’être moins visibles par les algorithmes de recommandation des plateformes qui fonctionnent grâce à la publicité.



**[Commentaires] :**

Les articles d'opinion, les sujets clivants et polémiques vont générer plus de commentaires. La façon dont est rédigé l'article (surtout au niveau du titre et du chapeau) va également influencer. Un sujet polémique traité de manière plus "froide", analytique, risque de générer moins de commentaires qu'un contenu plus "chaud", qui prend moins de distance avec son sujet, qui fait appel aux émotions. Quand les sujets sont particulièrement clivants, les débats par commentaires interposés vont faire grossir rapidement le nombre total de commentaires. Les commentaires sont considérés par les plateformes comme une meilleure preuve d'"engagement" qu'un "like", car il demandent plus de temps et d'implication et participent à faire rester les internautes plus longtemps sur les plateformes. Dans un modèle d'économie de l'attention, le temps passé sur la plateforme permet de soumettre l'internaute à plus de publicité. Pour cette raison, les articles qui génèrent beaucoup de commentaires vont être beaucoup plus mis en avant par l'algorithme de recommandation.

**[Partages publics] :**

Les partages publics, ou semi-publics (à son réseau d'ami-e-s si on a un compte privé), sont un signal d'une volonté de donner plus de visibilité à un contenu. Il s'agit d'une forme d'engagement bien plus coûteuse que le simple "like". Même s'il suffit d'un clic pour partager, un partage implique de gérer ensuite les réactions de son propre réseau à l'article publié. Pour cette raison, les sujets qui seraient source de friction par rapport à un réseau d'ami-e-s vont être moins facilement partagés. Une fois qu'un article est partagé, c'est à charge de la personne qui partage de débattre par commentaires interposés avec tonton Robert qui trouve que "quand même c'était mieux avant", ou avec cette vague connaissance croisée lors des dernières vacances qui veut nous convaincre que "oui les infusions de radis sont vraiment très efficaces pour traiter tout type de maladie". Les motivations au partage public / semi-public d'un article peuvent être nombreuses : sensibiliser, faire rire, interpeller, questionner, etc. Dans tous les cas, il s'agira, d'une certaine façon, de se raconter sur les réseaux.

**[Partages privés] :**

À la différence des partages publics ou semi-publics, les partages privés (via Messenger, WhatsApp, groupes Facebook privés, etc.) n'impliquent pas une forme de "publicité" (au sens de rendre public). On prend moins de risque car on anticipe mieux la réception qui va en être faite : tel article pour telle personne ou tel groupe d'ami-e-s qu'on imagine que ça va intéresser particulièrement. Il y a aussi plus de liberté dans ce type de partage car, restant "privé", on ne verra pas la même auto-censure. Le contrôle social y est bien moindre. Ainsi, des journalistes ont pu révéler que des policiers français partageaient des propos haineux et racistes à la pelle sur un groupe privé. Cela ne pourrait jamais avoir lieu dans l'espace "public". Au-delà du fait que pour certaines personnes, l'image publique est encore plus importante (policier-e-s, enseignant-e-s, politicien-ne-s), certains propos ne se partagent pas publiquement, soit parce qu'ils sont interdits par la loi, soit parce qu'ils sont désapprouvés socialement. Finalement, ce qui semble circuler le plus dans les partages privés, c'est ce qu'on appelle des "bullshit news" (faits divers absurdes et trashes, informations choquantes, drôles ou bizarres). L'objectif de ces partages n'est alors pas d'informer mais plutôt de choquer ou d'amuser.

**[Nombre de mots] :**

Le nombre de mots utilisés dans un article peuvent donner une idée du travail journalistique nécessaire pour le produire. Des médias qui participent à une course au clic (financés par la publicité) vont devoir publier dans l'urgence pour être dans les premiers à sortir l'information. Dans ce cas, ils vont copier-coller la dépêche de presse et y ajouter quelques mots. Les dépêches étant les plus lacunaires possible, ce sera toujours très court. Des médias d'infodivertissement comme Melty ont un rythme de production effrené (entre 65 et 80 articles par jour) avec des articles très courts, compilant simplement des informations trouvables très facilement ailleurs sur le web. Un article long n'est pas nécessairement synonyme d'article de qualité, tout comme un article court pourra parfois être qualitatif et apporter un éclairage original ou un résumé bienvenu. On peut néanmoins observer que dans des rédactions qui ont un modèle économique mixte, avec des articles gratuits (financés par la publicité) et des articles payants (abonnement), ces derniers qui doivent justifier leur plus-value, seront généralement plus longs, apportant plus de contexte et d'analyse.



**[Temps sur la page] :**

Cette mesure a peu de valeur utilisée seule. Elle prend tout son sens lorsqu'elle est confrontée au nombre de mots. Un article avec beaucoup de mots qui présente une très faible mesure de temps passé sur la page est un article qui au mieux a été lu en diagonale, et au pire a été abandonné aux premières lignes. Cette mesure peut également être couplée à la mesure de nombre de vues. Un article qui fait énormément de vues sur lequel les gens passent très peu de temps peut être un signe de "piège à clic". "Le cadeau trop chou du prince Harry pour Meghan Markle" donnera envie de cliquer à de nombreuses personnes pour savoir de quoi il s'agit. La plupart des personnes quitteront la page après un "scroll", le temps de comprendre qu'il s'agit d'un petit chiot mignon et d'avoir la réponse à la curiosité et à l'insatisfaction laissée par la question de l'article.

**[Visites/mois] :**

Une telle mesure mériterait d'être affinée, et c'est évidemment le cas dans la réalité des médias qui disposent de statistiques très précises sur les internautes qui visitent leur site. Cette mesure permet néanmoins de dégager quelques tendances. Il est intéressant de relier cette donnée à celle du nombre d'articles publiés par mois. Il est logique qu'un site qui publie peu d'articles reçoive moins de visiteurs qu'un site qui publie des articles à la chaîne. On peut aussi faire l'hypothèse que des médias d'informations généralistes attireront une frange de la population plus large (et donc plus de visites mensuelles) que des médias d'information de niche. En effet, un média nommé Inside Rap Game, qui comme son titre le laisse entendre, traite de rap, ou encore Team Foot, auront un public cible plus restreint que des sites comme ZTL Info ou L'info qui se présentent plutôt comme des sites d'information généraliste.

**[Abonné-e-s réseaux sociaux] :**

Cette mesure nous renseigne sur la renommée en ligne d'un média. Elle est le plus souvent la marque d'une certaine adhésion au média et à ses contenus. Cette affiliation est "déclarative" car visible pour le réseau de la personne qui s'abonne. Un grand nombre d'abonné-e-s pour un média peut s'expliquer par plusieurs facteurs : il peut tout simplement s'agir d'un miroir de la popularité du média mais il peut aussi être la conséquence d'une stratégie rôdée pour les réseaux sociaux : contenus interactifs, quiz, jeux, etc.

**[ Articles/jour] :**

Des médias qui publient un très grand nombre d'articles par jour sont des médias qui produisent l'information, ou a minima une partie de ses articles, à la chaîne. On peut avoir des médias comme Melty qui produisent uniquement des articles à très faible plus-value. On peut cependant aussi avoir des médias au modèle mixte qui publient une série d'articles à la chaîne mais qui vont à côté de cela proposer des articles présentant une vraie plus-value journalistique. Un média qui engagera un long travail d'enquête pour chaque article publié aura nécessairement un rythme de production bien moindre, l'enquête journalistique demandant du temps et de l'argent. Le nombre d'articles publiés chaque jour peut aussi nous donner une idée de la taille d'une rédaction. Une rédaction qui fait travailler 40 personnes publiera inévitablement bien plus d'articles qu'une équipe de rédaction de 5 personnes.

6

(10 minutes)

En guise de conclusion, visionner Francetv slash / enquête, "Comment les réseaux sociaux se rendent indispensables - L'intox, c'est nous #5", Youtube, 2019, <https://bit.ly/3cGNpNa>

La synthèse élève reprend dans les grandes lignes les informations évoquées dans la vidéo.

CRITIQUE  
DE L'INFO  
L'OUTIL ULTIMEACTION  
MÉDIAS  
JEUNES