

Lutter contre la masculinité algorithmique et les biais toxiques

À trop se farder, l'on finit par se cacher. Depuis des décennies le *male gaze* caricature et occulte les femmes au cinéma. À l'instar de la vague *Me Too*, la voix des femmes se fait de plus en plus entendre. L'utopique projet de ce mémoire est d'essayer de les montrer. L'avènement des plateformes de streaming uniformise nos goûts et nos choix. Au moyen d'un algorithme de recommandation biaisé, je propose dans ce projet de donner plus de visibilité aux contenus féminins. Pour cela, j'ai extrait la base de données du site TMDb sur laquelle j'ai effectué une *Content-Based filtering*. L'effet de ce biais, certes ténu mais présent, permet d'apporter un faible éclairage sur ces visages féminins.

Lorsqu'Adèle Haenel quitte, écoeurée, la cérémonie des César, Virginie Despentes scande dans un pamphlet : « Désormais on se lève et on se barre ! ». Aujourd'hui j'aimerais dire : « Désormais on se lève et on se montre ! ». Façonnées par le regard des hommes depuis tant d'années, la place des femmes et leurs représentations au cinéma surpassent une dimension artistique. Tantôt speakerine, parfois miss météo, à la télévision comme au cinéma la femme a souvent le choix entre demoiselle en détresse ou femme fatale. Des James Bond Girls, à Charlie dans *Top Gun*, en passant par la potiche Marilyn, l'imaginaire masculin a dessiné un archétype galvaudé des femmes. Si l'on retrouve parfois des personnages féminins cultivés, telle que Romy Schneider dans *Les choses de la vie*, rapidement le regard du réalisateur contemple son corps dénudé et l'enferme dans son rôle de « femme de ». La femme pourrait-elle être montrée véritablement autrement que comme le faire-valoir d'un homme, menacée par le « *male gaze* » ? La lutte féministe engagée englobe des choix tant sociétaux qu'économiques et nous pousse à reconsidérer notre répartition des rôles.

Au-delà des rôles, l'égalité de représentations des genres est aussi à la peine au cinéma. Une étude publiée en 2021 par l'équipe d'Antoine Mazières a montré, *via* une analyse en Intelligence Artificielle de 3500 films, que seulement 35% des visages apparaissant à l'écran étaient féminins. Cette inégalité de répartition est d'autant plus frappante selon les genres cinématographiques avec seulement 30% de visages féminins au sein de films policiers. « Couvrez ce sein que je ne saurais voir ! » Dans l'industrie cinématographique les

femmes sont pourtant bien représentées et dominant dans les métiers de scripte, costumière, ou maquilleuse. En revanche, les réalisatrices ne représentent qu'un tiers des films produits. Mais alors pourquoi cachons-nous à ce point les femmes derrière la caméra ? Si le cadre de ce mémoire n'a pas pour objet de répondre à cette interrogation, pourrions-nous plutôt chercher des pistes sur comment permettre aux femmes d'être vues ?

Afin de répondre à cette question, deux limites sont à considérer. Tout d'abord, le manque de femmes à l'écran. Ne pouvant montrer ce que nous ne possédons pas, nous devons garder en tête cette triste réalité. De plus, le biais algorithmique imposé par l'utilisateur peut invisibiliser les femmes. Si celui-ci s'avère friand de films aux mâles prédominants, les algorithmes de recommandations, tels qu'ils sont construits, recommanderont à l'utilisateur un contenu similaire. Mal paramétrés, les algorithmes de recommandations peuvent mener les utilisateurs à tourner en rond. En effet, il faut savoir que 80% du contenu Netflix visionné n'a pas été cherché par l'utilisateur mais directement recommandé par la plateforme, et seul 10% du catalogue Spotify est réellement utilisé. Cette « sur-recommandation » mène alors à moins d'exploration pour les consommateurs, et plus de standardisation pour les auteurs du contenu et du public visé.

En l'état, notre seule façon de révéler la présence des femmes serait de jouer avec ces algorithmes. L'objectif de ce projet est de contourner ce biais algorithmique pour proposer aux utilisateurs du contenu jusqu'alors inexploré. En 2014, l'idée d'imposer au CSA des algorithmes mettant en avant des contenus français a été envisagée. Laissons alors germer cette idée et tentons de déformer les algorithmes pour pouvoir mettre en lumière comédiennes et réalisatrices !

Analyse exploratoire

Le site The Movie Database (TMDb) a été choisi comme base de travail permettant d'extraire, *via* un API, les données de plus de 25 000 films. Pour chaque film, différentes informations sont récoltées comme l'année de production, le genre, le nom des acteurs, etc. Seuls les 70% de films les mieux notés du dataset ont été conservé dans ce projet. En parallèle, un outil d'analyse de représentation des femmes a été utilisé : le test de Bechdel. Développé par Alison Bechdel, ce test se décompose selon 3 critères permettant d'allouer à chaque œuvre un score allant de 0 à 3 : « *Le film doit mettre en scène au moins deux femmes nommées dans l'œuvre, qui parlent ensemble et dont le sujet de conversation ne se centre pas sur un homme* ». Lorsqu'il existe, le score de Bechdel de chaque film a été récupéré *via* une API.

Dans une première analyse exploratoire, j'ai cherché à récolter un maximum d'informations nécessaires à la recommandation d'un film. L'extraction et la catégorisation de ces informations m'a permis tout d'abord de classer les films selon la décennie ou la région de production (**Fig. 1**). La majorité des films à ma disposition couvre essentiellement la période 1980 à nos jours et représente principalement des films Nord-Américain ou Européens. La présence prépondérante de films Américains pourrait être, par la suite, un facteur limitant pour une recommandation adaptée pour chaque continent.

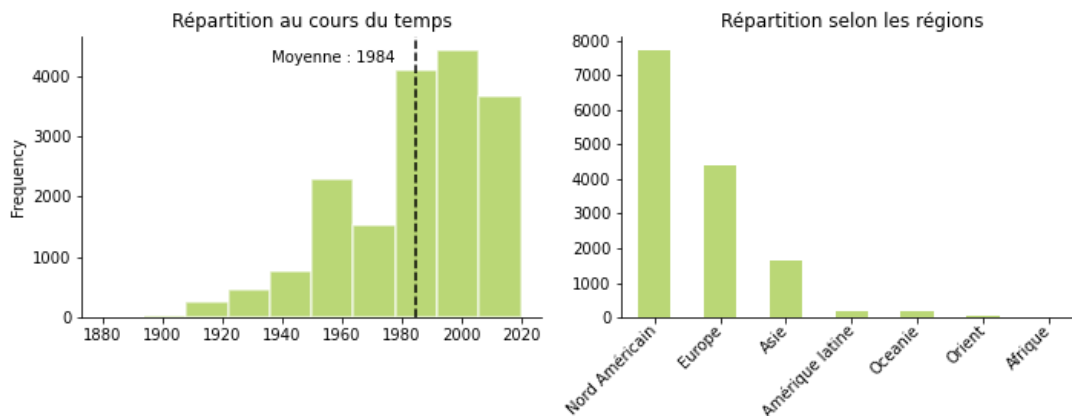


Fig. 1 – Visualisation de la répartition spatio-temporelle du dataset

Sachant les disparités de représentation des genres selon les catégories de films, j'ai choisi d'extraire les différentes catégories de films (**Fig. 2**). Dans cette base de données, les films largement représentés appartiennent aux genres Drame, Comédies, Action, ou Horreur.

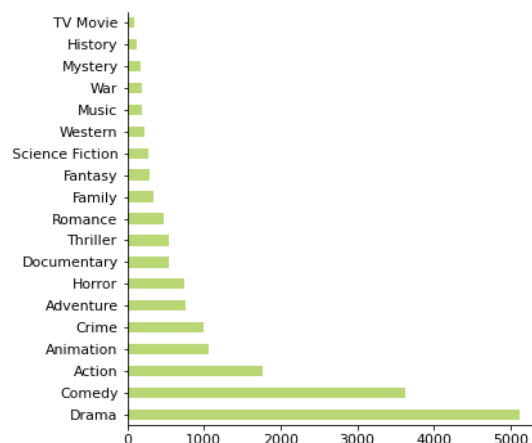


Fig. 2 – Répartition des différentes catégories de films

Je récupère ensuite les informations concernant le casting du film. J'extrais le genre des comédiens, ainsi que leur nom et leur côte de popularité (définir selon le site TMDb). De manière peu surprenante, on retrouve une écrasante majorité d'hommes pour les premiers rôles. Au fur et à mesure que l'on s'éloigne dans le casting, cette inégalité tend à diminuer (Fig. 3).

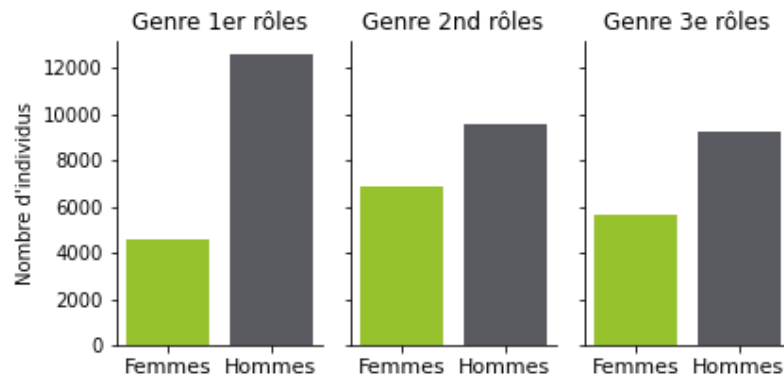


Fig. 3 – Répartition des genres selon l'ordre d'apparition au casting

Cet écart de répartition est d'autant plus marqué concernant la réalisation des films où les femmes sont quasi absentes (Fig. 4). Ce biais de répartition s'avère être un facteur limitant dans la mise en évidence de contenu féminins.

En outre, l'absence des femmes influe aussi sur la vision que les utilisateurs se font d'elles. Le site TMDb collecte pour chaque artiste le nombre de vue des utilisateurs afin de déterminer un niveau de popularité. Lorsque l'on s'arrête sur ce paramètre au regard du genre des acteurs.trices ou des réalisateurs.trices, on constate une popularité moindre concernant le genre féminin (Fig. 5). Cette moindre popularité pourrait jouer un rôle déterminant dans la réception du public face à un contenu féminin.

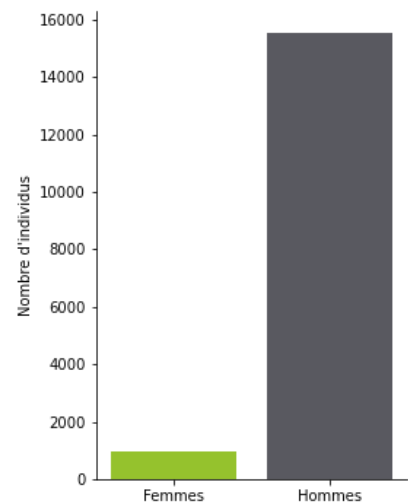
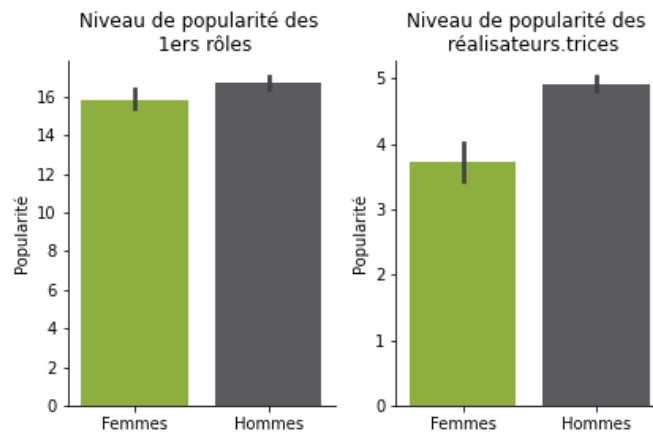


Fig. 4 – Répartition des genres chez les réalisateurs.trices



Enfin, de sorte à mieux catégoriser chaque film, j'extrais une liste de trois mots-clés pour chacun. J'ai ainsi construit un nuage de mots me permettant de visualiser les mot-clés les plus fréquents dans l'ensemble de mon dataset (**Fig. 6**).

Algorithme de recommandation

L'objet de ce travail est de filtrer l'ensemble des informations collectées pour bâtir une matrice permettant de prédire un contenu similaire susceptible de plaire à l'utilisateur. En matière d'algorithme, il est possible d'exploiter les intérêts des utilisateurs selon leurs profils ou leurs notations. Cependant, notre base de données ne contient que la note moyenne par film. J'ai donc opté pour une recommandation Objet ou *Content-Based filtering*. En me fondant

sur les propriétés intrinsèques des films, je vais déterminer des patterns redondants et corrélérer les caractéristiques d'un film avec d'autres similaires. Cette méthode permet notamment de palier les problèmes dits de « cold start » lorsque l'utilisateur n'a pas encore fourni suffisamment d'informations pour nourrir l'algorithme.

Pour construire mon algorithme, je me suis inspirée de la notion de similarité basée sur des mots-clés. Je peux ainsi décrire chaque film selon un vecteur de valeurs binaires (où 1 = le mot apparaît et 0 = le mot n'apparaît pas). L'idée sera donc de mesurer la ressemblance des vecteurs entre deux films. La plupart des algorithmes utilisant un index de similarité sont basés sur le résumé d'une œuvre. Pour ma part, j'ai choisi non pas d'utiliser le résumé de chaque œuvre mais de recréer mon propre corpus. Ainsi, pour chaque film je récupère la décennie de production, la région, les catégories, le nom des trois premiers acteurs, le nom du/de la réalisateur·trice, et les trois mots-clés associés. J'obtiens ainsi une phrase contenant tous les éléments essentiels à ma recommandation.

J'ai choisi d'employer une technique de *Natural Language Processing* (NLP) pour construire mon algorithme. Sur la phrase précédemment créée, j'utilise un *CountVectorizer* qui me permet de vectoriser chaque mot de cette phrase et de créer une matrice où chaque mot est représenté dans une colonne. À la fin, chaque cellule contiendra une valeur correspondant au nombre de fois où le mot apparaît dans cette phrase. Sur la base de cette matrice, j'emploie une technique classique de fouille de texte : la *Cosine Similarity*. L'objectif est de mesurer la similarité entre deux séquences de nombres et ainsi de regrouper les contenus similaires. Enfin, je définie une fonction qui va récupérer 30 films présentant le meilleur score de similarité au film recherché.

Sur la base de ces 30 films recommandés, je vais extraire 10 films avec le meilleur score de Bechdel, 10 films en triant les actrices de genre féminin en premier, et 10 films en triant en priorité les réalisatrices. J'obtiens ainsi 3 *DataFrame* réunissant un maximum de critères pour mettre en avant des femmes. Pour ma recommandation finale, je ne sélectionne que 10 films de cette recommandation biaisée, trié de sorte à montrer en priorité des actrices féminines (**Fig. 7**).

Dans un premier temps, j'ai voulu voir l'effet du biais sur un de mes films préférés : Kill Bill (**Fig. 8**). Je constate après biais une légère augmentation du score de Bechdel (1,8 à 2,8), ainsi qu'une légère diminution concernant les genres des acteurs.trices (1,7 à 1,4). Au regard des résultats encourageants, j'ai observé l'effet sur l'ensemble de mon dataset (**Fig. 9**).

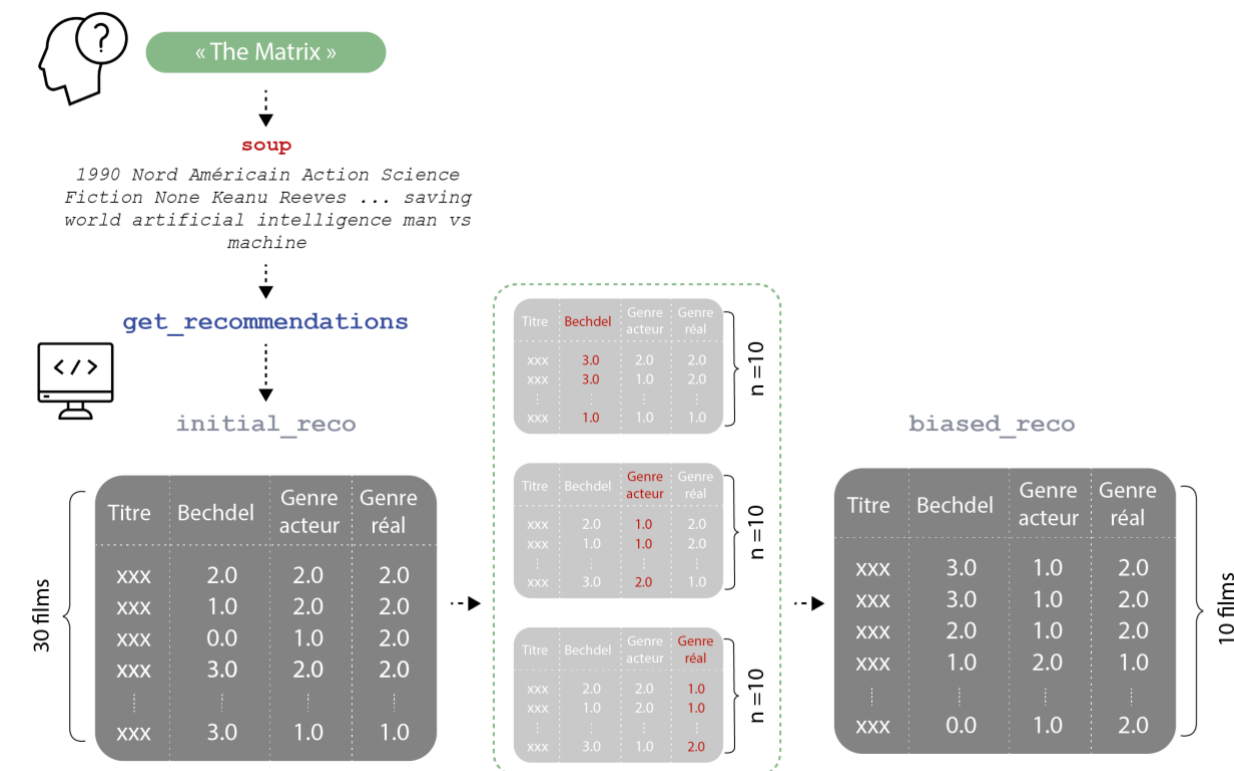


Fig. 7 – Chaîne de traitement algorithme de recommandation

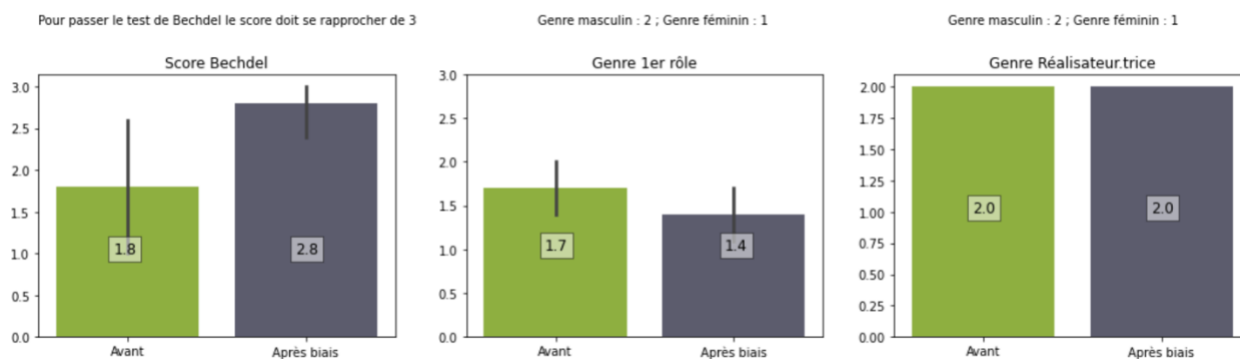
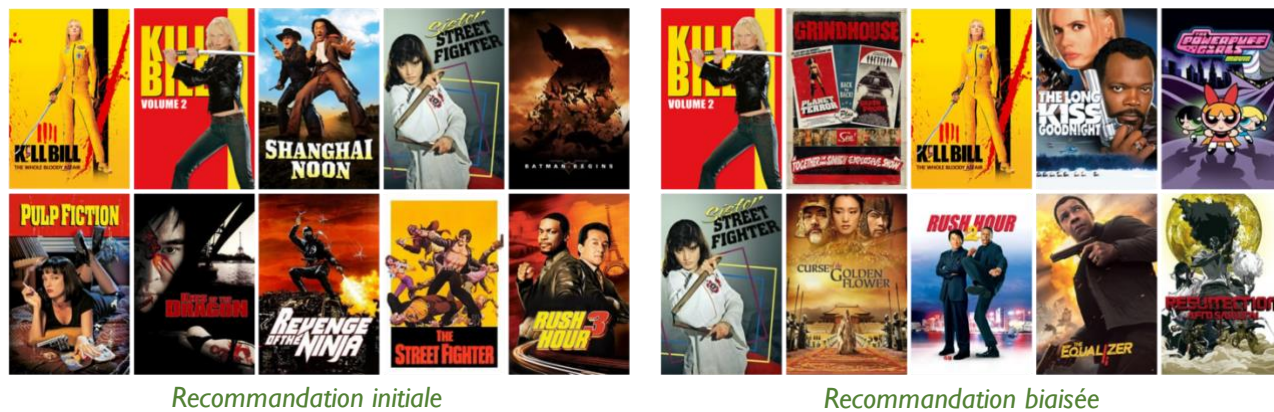


Fig. 8 – Effet du biais algorithmique sur la mise en avant de représentations féminines pour le film Kill Bill Vol. I

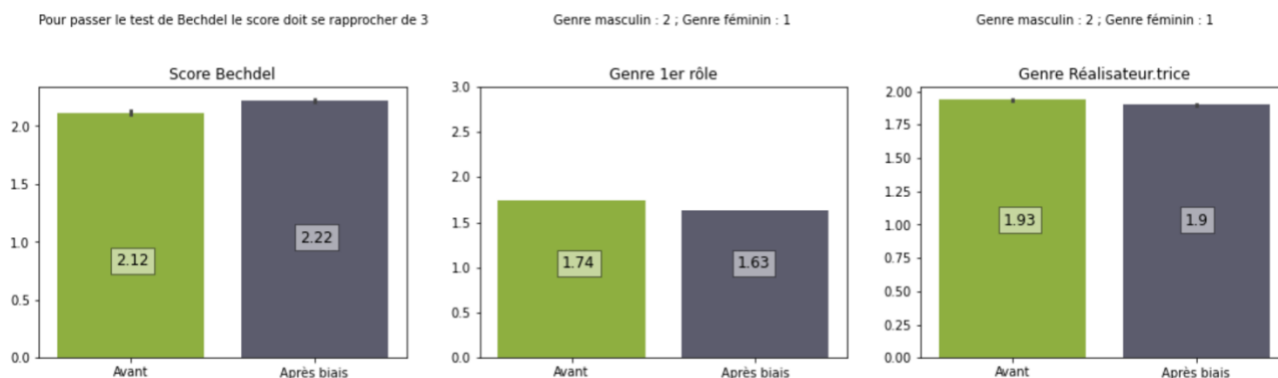


Fig. 10 – Effet du biais algorithmique sur la mise en avant de représentations féminines sur l'ensemble du dataset

Un petit biais pour l'algorithme, un petit pas pour les femmes

Sans pouvoir changer le monde et la répartition des genres, notre biais algorithmique montre qu'il est possible de défier la masculinité toxique.

Dans le cadre de cet exercice, je me suis imaginée travailler pour une plateforme de streaming. L'objectif étant toujours de proposer aux utilisateurs un contenu de qualité et adapté à leurs demandes. L'effet de mon biais peut paraître léger, mais l'objectif est d'influencer subtilement le choix des utilisateurs sans les amener à quitter la plateforme. J'imagine ainsi pouvoir ainsi contribuer aux changements des mentalités progressivement. Au-delà de cet aspect socio-économique, le manque cruel de femmes que ce soit en tant qu'actrices ou en tant que réalisatrices freine considérablement notre marge de manœuvre.

Enfin, s'il a le mérite d'exister et permet de proposer un outil pour analyser la représentation des femmes, le score de Bechdel est critiquable. À mon grand regret, Matrix 1, par exemple, passe ce test simplement parce que deux femmes échangent deux répliques. Un film peut passer haut la main ce test avec simplement 10 secondes de dialogue féminins. L'avènement du féminisme ne passe pas inaperçu, et peut mener à une sorte de « *pinkwashing* » contre-productif. Les sociétés de productions savent qu'elles n'auront que des cases à cocher pour occulter leur misogynie.

Si l'effet semble moins marqué pour l'ensemble du dataset, il ne faut pas oublier la présence de films de guerre ou de western où l'absence de femme est considérable. Mon algorithme se noie dans la pluralité des genres et l'effet se dissipe. Au-delà de ces aspects, mon algorithme reste extrêmement simple. L'idéal serait d'employer un système de recommandation hybride et de mettre des utilisateurs dans la boucle pour fournir une recommandation plus adaptée.

Références

- Github du notebook et des données : <https://github.com/JustineFacchini/Projet-DESU>
- Mazières, A., Menezes, T. & Roth, C. Computational appraisal of gender representativeness in popular movies. *Humanit Soc Sci Commun* **8**, 137 (2021).
<https://doi.org/10.1057/s41599-021-00815-9>
- CNC - La place des femmes dans l'industrie cinématographique
- Bechelttest.com