

# Système de recommandation de livres

# LIVRIA

|             |   |
|-------------|---|
| Référence   | Rapport_Final_LIVRIA                    |
| Projet      | Projet informatique individuel – LIVRIA |
| Date remise | 25/04/2019                              |

|                      |
|----------------------|
| Auteur               |
| Younès Ghennam - 2A  |
| Tuteur               |
| M. Simonazzi Nicolas |
| Tuteurs soutenance   |
| M. Simonazzi Nicolas |
| Mme. Clermont Edwige |

# TABLE DES MATIÈRES

|   |           |
|---|-----------|
| <b>I. Introduction</b>  | <b>3</b>  |
| I.1. Contexte du projet   | 3         |
| I.2. Pré-existant   | 3         |
| <b>II. Rappel des spécificités techniques du projet</b>           | <b>4</b>  |
| <b>III. Réalisations</b>  | <b>4</b>  |
| <b>IV. Annexe</b>   | <b>10</b> |
| <b>Hands-On Machine Learning with Scikit-Learn and TensorFlow</b> | <b>10</b> |
| Concepts, Tools, and Techniques to Build Intelligent Systems      | 10        |

# I. Introduction

La technologie émergente de l'Intelligence Artificielle, souvent appelée par son acronyme IA, repose sur des algorithmes relativement complexes, qui ne visent pas seulement à reproduire, à imiter les processus cognitifs humains, à accomplir une tâche proprement humaine et de manière plus efficace, mais aussi de manière plus générale, à imiter tout type d'intelligence réelle. Existant depuis les années 1960, la recherche s'est développée récemment au point de multiplier les applications : voitures autonomes, diagnostics médicaux, assistants personnels, construction de smart city, finance algorithmique, robots industriels, jeux vidéo, ou encore les systèmes de recommandations, pour ne citer qu'une partie des nombreux domaines d'application.

De plus en plus de plateformes reposant sur un système de recommandation ont vu le jour. Elles visent principalement à suggérer ou présenter un contenu susceptible d'intéresser l'utilisateur comme le font Youtube, Spotify, Netflix, Amazon et bien d'autres encore. Néanmoins, nous avons pensé pertinent de concevoir un outil similaire pour les férus de littérature en lui offrant la même visibilité grâce à un système de recommandation de livres nommé LIVRIA, destiné aux petits et grands lecteurs.

## ***I.1. Contexte du projet***

Dans le cadre de notre formation au métier d'ingénieur cognitif à l'ENSC (Ecole Nationale Supérieure de Cognitique), nous avons eu à réaliser un projet tutoré par groupes de 6 étudiants minimum, dont trois en seconde année et au moins le même nombre en première année du cycle d'ingénieur. Ce projet qui était à l'origine notre projet Transpromotion, j'ai décidé de le poursuivre, car je souhaitais bien intégrer les connaissances et les compétences que nous allions découvrir et approfondir en intelligence artificielle et en développement mobile lors de ce second semestre de deuxième année à l'ENSC.

L'idée de ce projet était à l'origine née du constat qu'il serait pertinent de mettre en place un site qui conseillerait des lectures en fonction du profil de l'utilisateur. Néanmoins, j'ai décidé de me tourner plus vers une implémentation d'application mobile pour l'interface utilisateur, car ces dernières sont de plus en plus répandues et je m'imaginais qu'il serait certainement plus pratique d'avoir accès à ce système de recommandation depuis son téléphone portable. Malheureusement, au moment où je rédige ce rapport, la partie interface développée en ReactNative n'est qu'à ses débuts, mais elle n'en restera pas là.

## ***I.2. Pré-existant***

Il existe actuellement un nombre considérable d'applications de recommandation de livres. Elles sont toutes pertinentes dans la mesure où elles fournissent un service assez proche de ce que l'on cherche à offrir. Toutefois, les méthodes utilisées diffèrent de celle que nous comptons mettre en place. En effet, une large majorité de ces applications opte pour des techniques de développement telles que les Tags, les algorithmes basés sur les précédentes lectures de l'utilisateur, ceux basés sur des classements selon le style, le genre, l'auteur, le mouvement littéraire ou encore le thème abordé par le livre. Les classements sont établis en tenant compte de plusieurs critères à savoir la popularité donnée par le nombre de lectures, les notes attribuées par les lecteurs, le nombre d'envie

de lire ce livre”, etc. Certaines applications utilisent un moteur de recommandation externe, tel que celui proposé par Amazon. D’autres approches sont à noter, notamment le “Trending Now”, technique qui consiste à donner plus de visibilité aux livres tendances et qui ont le plus de succès.

## II. Rappel des spécificités techniques du projet

Initialement, l’objectif du projet était de produire une application mobile codée en **React Native** et s’appuyant sur un modèle de Machine Learning implémenté en **Python** et utilisant différentes librairies, notamment celles de **Scikit-Learn** pour la partie IA.

Il s’avère que j’étais totalement étranger au langage Python, aux différents environnements de développement et aux outils mis à disposition pour développer des applications reposant sur ce qu’on appelle le “Machine Learning”. C’est ainsi que j’ai découvert, par la lecture de *Hands on Machine Learning with Scikit-Learn and TensorFlow* - ouvrage que je considère d’une grande qualité, des outils qui deviennent rapidement indispensables pour toutes les personnes travaillant dans ce domaine et qui codent en Python. J’ai découvert **Jupyter** et appris à utiliser ses notebooks comme environnement de travail durant tout ce second semestre pour ce projet. J’ai appris à configurer grâce à **Anaconda** mes propres environnements en affectant à chacun des packages particuliers pour ne pas avoir à en désinstaller et réinstaller systématiquement. Je me suis frotté à l’utilisation des librairies que sont matplotlib, numpy, pandas, pour ne citer que les plus récurrentes.

## III. Réalisations

Rappelons que ce projet et a été commencé dans le cadre du projet transposition au premier semestre de cette année. Tous les livrables concernant la conception orientée utilisateur (CCU) ont été réalisés à ce moment, et ces derniers comprennent :

- Réalisation d’un questionnaire, recueil et analyse des données,
- Personna,
- Storyboard,
- Wireframe et tests utilisateurs sur ce dernier,
- Analyse basée sur les critères de Nielsen.

A savoir qu’il faudra certainement reprendre certaines de ces réalisation et les adapter aux nouvelles contraintes, car le rendu ne pourra pas être le même sur un téléphone que sur un écran de PC. Néanmoins, j’ai laissé cette partie CCU pour me focaliser sur le côté plus technique de l’application, la partie IA étant la pierre angulaire du projet.

Vous pourrez retrouver l'ensemble de mon travail dans le dépôt Git. Ce dernier est composé d'une partie interface (cf. dossier **Interface\_ReactNative**) dont j'ai commencé ce qu'on pourrait considérer comme une petite ébauche, j'avais profité des enseignements en développement mobile pour commencer cette partie. D'autre part, vous retrouverez ce qui compose le plus gros de mon travail que j'ai prioritairement axé sur la partie intelligence artificielle (cf. dossier **Python Livria**). Il comprend l'ensemble des notebook jupyter qui permettent de prédire les thèmes susceptibles de plaire à un utilisateur en fonction de ses "critères". Ces notebook sont je pense vraiment bien commentés, je cherche à vous guider comme moi même j'aurais souhaité l'être lorsque je faisais mes premiers pas dans leur implémentation. C'est pourquoi je ne m'attarderai pas sur leur contenu que je vous invite à découvrir (cf. lien du dépôt git en **Annexe**). Je vous indique cependant le "sens de lecture" de ces notebook, qui rend compte de la méthodologie que j'ai mise en place pour réaliser ce travail :

| 0 / LIVRIA / Python Livria                    |  |       | Nom ↓ | Dernière Modification    | File size |
|---|--|-------|-------|--------------------------|-----------|
| ..  |  |       |       | il y a quelques secondes |           |
| data  |  |       |       | il y a 20 heures         |           |
| dataThemeCleaning.ipynb                       |  |       |       | il y a 21 heures         | 249 kB    |
| dataVisualization.ipynb                       |  |       |       | il y a 9 heures          | 107 kB    |
| dataVizualisation&Cleaning_GoodBooks10k.ipynb |  | Actif |       | il y a 7 heures          | 523 kB    |
| Goodbooks10k_Collaborative_filtering.ipynb    |  |       |       | il y a 10 heures         | 32.3 kB   |
| Livria_MultiLabel_classification.ipynb        |  | Actif |       | il y a 4 heures          | 51.7 kB   |
| Livria_recommender_system.ipynb               |  | Actif |       | il y a une heure         | 11.3 kB   |
| Themes_Collaborative_filtering.ipynb          |  | Actif |       | il y a 7 heures          | 66.5 kB   |

### Répertoire des fichiers .ipynb

Je vous conseille de commencer par lire le fichier **dataThemeCleaning**, j'y ai nettoyé et vectorisé les données pour pouvoir permettre leur utilisation avec les différentes fonctions de Scikit-Learn. Cela permettait aussi de mieux visualiser les données, deuxième étape dans la lecture des notebooks (Cf. dataThemeCleaning) pour en avoir une bien meilleure appréhension.

Durant ce projet, j'ai rencontré de nombreuses difficultés qu'il m'a fallu résoudre ou contourner d'une manière ou d'une autre. L'idée d'origine était de recommander des livres en fonctions des critères d'entrée que nous avons choisis de poser dans le questionnaire réalisé lors du transpromotion, mais comment recommander des livres avec un base de données ne regroupant que si peu de livres et où chaque individu ayant répondu au

questionnaire en recommande un différent par rapport au autres. Les avis ne se recourent pas sur les livres mentionnés dans les réponses au questionnaire, et nous manquons de donnée pour prédire une liste de livres en particulier. L'idée était alors de récupérer des données de livres via une API et j'ai donc réalisé un important travail de recherche sur la manière d'extraire les données relatives aux ouvrages que l'on souhaite recommander. Il fallait entraîner notre modèle de Machine Learning à prédire des thèmes plutôt que des livres, et à faire le lien entre ces thèmes et les livres caractérisés par ce(s) thème(s) récupérés par via l'API. Malheureusement, il n'existe aucune base de donnée aussi complète et je me suis retrouvé pendant un moment dans une impasse.

M. Simonazzi, qui a été mon tuteur pendant ce projet, m'a conseillé de m'inspirer sur ce que fait Amazon en terme de recommandation et m'a envoyé un notebook sur le filtrage collaboratif effectué sur une base de données de films (cf. MovieLens). Cela m'a débloqué car j'ai pu trouver par la suite un set de données similaire à celui sur les films, celui de Goodbooks-10, comprenant 10000 livres et environ de 6 millions de notes attribuées par un peu plus de 50000 utilisateurs. Ainsi, je vous invite à lire le notebook `dataVizualisation&Cleaning` puis `Gookbooks10k_Collaborativ_filtering` qui est la suite de mon travail. Les images qui suivent vous montre le choix des thèmes que j'ai fait sur le set de données de Goodbooks-10k.

Ces trois images montrent un classement des tags dans l'ordre décroissant en fonction du nombre de livres que chacun de ces tags caractérisent. Etant donné le fait qu'il y en a un grand nombre ne correspondant pas avec un genre littéraire ou encore un thème, j'ai du sélectionner ceux qui nous intéressent et je l'ai mis en valeur dans les petits encadrés rouges. Vous pouvez les compter comme moi, ils sont au nombre de 39, ce qui permet de recouvrir bien plus de thèmes que ceux que nous proposons dans le questionnaire qui étaient uniquement au nombre de 14.

| tag_name          | tag_name           | tag_name             | tag_name          | tag_name            | tag_name                |
|-------------------|--------------------|----------------------|-------------------|---------------------|-------------------------|
| to-read           | books              | classics             | cabre             | part-of-a-series    | science-fiction-fantasy |
| favorites         | adult-fiction      | read-2016            | crime             | books-i-have        | science                 |
| owned             | e-books            | read-2014            | did-i-finish      | ya-fiction          | friendship              |
| books-i-own       | read-in-2013       | contemporary-fiction | to-read-fiction   | mystery-suspense    | on-hold                 |
| currently-reading | book-club          | favorite-books       | fantasy-sci-fi    | kindle-books        | listened-to             |
| library           | audible            | drift                | nook              | own-to-read         | children's-books        |
| owned-books       | fantasy            | finished             | library-book      | read-2012           | 2014-reads              |
| fiction           | romance            | read-in-2011         | chick-lit         | mysteries           | general                 |
| to-buy            | audio-books        | read-in-2017         | 20th-century      | must-read           | ya-books                |
| kindle            | abandoned          | 5-stars              | paranormal        | need-to-buy         | my-favorites            |
| default           | novel              | historical-fiction   | school            | urban-fantasy       | sci-fi                  |
| ebook             | re-read            | paperback            | classic           | childhood           | action-adventure        |
| my-books          | have               | historical           | magic             | read-in-english     | humour                  |
| audiobook         | audio-book         | thriller             | mystery-thriller  | children            | biography               |
| ebooks            | mystery            | sci-fi-fantasy       | teen              | literary            | speculative-fiction     |
| wish-list         | borrowed           | american             | supernatural      | horror              | contemporary-romance    |
| my-library        | read-in-2012       | suspense             | recommended       | reread              | high-school             |
| audiobooks        | adventure          | reviewed             | nofiction         | childrens           | to-read-non-fiction     |
| i-own             | young-adult        | 4-stars              | favorite-authors  | young-adult-fiction | women                   |
| adult             | english            | unfinished           | realistic-fiction | thrillers           | philosophy              |
| audio             | did-not-finish     | read-2013            | literary-fiction  | read-2017           | 3-stars                 |
| favorites         | favorite           | home-library         | bookclub          | 2015-reads          | war                     |
| novels            | maybe              | library-books        | want-to-read      | on-my-shelf         | fantasy-sci-fi          |
| own-it            | sheffari-favorites | sci-fi               | read-in-2010      | british             | mystery-crime           |
| contemporary      | drama              | science-fiction      | thr               | 2016-reads          | usa                     |
| read-in-2015      | literature         | action               | unread            | kids                | sheffari-wishlist       |
| series            | general-fiction    | humor                | funny             | children-s          | comedy                  |
| e-book            | read-2015          | family               | bookshelf         | coming-of-age       | first-in-series         |
| read-in-2016      | read-2015          | history              | sci-fi-fantasy    | on-kindle           | juvenile                |
| read-in-2014      | ya                 | 2138                 | 1620              | 1262                | 1031                    |

| tag_name         | tag_name              | tag_name | tag_name               | tag_name | tag_name                          |             |                                     |     |
|------------------|-----------------------|----------|------------------------|----------|-----------------------------------|-------------|-------------------------------------|-----|
| stand-alone      | american-literature   | 1        | 612                    | vampire  | 557                               | want-to-buy | 493                                 |     |
| childrens-books  | children-s-literature | 882      | in-my-library          | 611      | 2017-reads                        | 557         | graphic-novels                      | 491 |
| england          | children-s-rit        | 869      | british-literature     | 610      | read-more-than-once               | 557         | 19th-century                        | 490 |
| ya-fantasy       | want                  | 865      | reference              | 608      | erotica                           | 557         | purchased                           | 490 |
| crime-mystery    | ya-rit                | 846      | epic                   | 603      | children-books                    | 556         | books-owned                         | 488 |
| kids-books       | overdrive             | 841      | 5-star                 | 602      | bought                            | 552         | 2015-read                           | 487 |
| detective        | paranormal-romance    | 840      | 2014-books             | 599      | to-read-own                       | 550         | movies                              | 485 |
| female-author    | to-re-read            | 839      | childhood-reads        | 599      | loved                             | 548         | kid-books                           | 485 |
| book-club-books  | murder                | 829      | new-adult              | 597      | american-rit                      | 548         | fantasy-fiction                     | 484 |
| personal-library | arc                   | 824      | collection             | 597      | 2006                              | 543         | graphic-novel                       | 484 |
| my-bookshelf     | dark                  | 824      | books-to-buy           | 595      | owned-to-read                     | 539         | read-aloud                          | 482 |
| 2013-reads       | psychology            | 810      | memoir-biography       | 592      | realistic                         | 536         | ya-romance                          | 479 |
| book-group       | to-read-nonfiction    | 806      | death                  | 591      | 1001-books                        | 535         | bio                                 | 478 |
| other            | politics              | 793      | alpha-male             | 588      | chic-rit                          | 528         | spiritual                           | 477 |
| crime-fiction    | europa                | 793      | classics-to-read       | 587      | classroom-library                 | 525         | scanned                             | 475 |
| youth            | sff                   | 777      | 2012-reads             | 586      | 1001                              | 524         | netgalley                           | 474 |
| faves            | middle-school         | 773      | school-books           | 582      | mystery-suspense-thriller         | 524         | classic-fiction                     | 474 |
| read-2011        | sci-fi-and-fantasy    | 771      | self-help              | 579      | 1001-books-to-read-before-you-die | 519         | modern-classics                     | 473 |
| modern-fiction   | 2016-books            | 769      | book-boyfriends        | 575      | adult-romance                     | 519         | comics                              | 469 |
| relationships    | murder-mystery        | 767      | animals                | 574      | pnr                               | 519         | my-ebooks                           | 469 |
| childhood-books  | high-fantasy          | 765      | mythology              | 572      | made-me-cry                       | 513         | political                           | 468 |
| couldn-t-finish  | hardcover             | 762      | children-s-fiction     | 568      | childrens-literature              | 509         | survival                            | 464 |
| read-in-2009     | read-for-school       | 762      | book-club-reads        | 567      | modern                            | 508         | biographical                        | 462 |
| middle-grade     | dystopian             | 755      | on-my-bookshelf        | 566      | 2015-reading-challenge            | 507         | me4                                 | 460 |
| memoir           | sff                   | 755      | chicklit               | 563      | spirituality                      | 503         | travel                              | 457 |
| gave-up-on       | biography-memoir      | 752      | to-get                 | 560      | fiction-to-read                   | 500         | military                            | 455 |
| religion         | classic-literature    | 752      | 2016-reading-challenge | 559      | translated                        | 497         | 1001-books-you-must-read-before-you | 453 |
| mine             | memoirs               | 744      | 2016-read              | 559      | mystery-thriller-suspense         | 496         | i-own-rit                           | 453 |
| crime-thriller   | teen-fiction          | 743      | uk                     | 558      | gave-up                           | 495         | read-2010                           | 451 |



| tag_name                |     | tag_name               |     | tag_name              |     | tag_name                    |     |
|-------------------------|-----|------------------------|-----|-----------------------|-----|-----------------------------|-----|
| biography-autobiography | 447 | life                   | 412 | women-writers         | 373 | romance-contemporary        | 344 |
| magical-realism         | 443 | graphic                | 411 | new-york              | 370 | summer-reading              | 343 |
| movie                   | 442 | to-read-fantasy        | 409 | psychological         | 370 | sub                         | 338 |
| standalone              | 441 | stories                | 408 | womens-fiction        | 368 | book                        | 337 |
| culture                 | 440 | religious              | 406 | 1001-import           | 368 | science-fiction-and-fantasy | 336 |
| steamy                  | 439 | finished-series        | 405 | fantasy-series        | 367 | brit-lit                    | 336 |
| cultural                | 439 | read-as-a-kid          | 404 | translation           | 364 | erotic-romance              | 335 |
| women-s-fiction         | 438 | adult-non-fiction      | 403 | children-young-adult  | 364 | horror-thriller             | 335 |
| on-my-kindle            | 438 | chapter-books          | 403 | completed-series      | 364 | children-s-book             | 334 |
| art                     | 436 | biographies-memoirs    | 402 | fiction-general       | 363 | ghosts                      | 333 |
| 2013-books              | 433 | crime-mystery-thriller | 400 | mystery-series        | 361 | 2013-read                   | 332 |
| home                    | 430 | abuse                  | 396 | demons                | 360 | comics-and-graphic-novels   | 331 |
| signed                  | 429 | 1990s                  | 396 | mysteries-thrillers   | 360 | 2017-reading-challenge      | 329 |
| 2014-read               | 428 | 2017-books             | 396 | personal-development  | 359 | my-collection               | 329 |
| book-boyfriend          | 425 | m-f                    | 394 | to-read-owned         | 357 | english-literature          | 329 |
| comic                   | 425 | vamps                  | 394 | christian             | 357 | fluff                       | 328 |
| sf-f                    | 423 | romantic               | 388 | ya-paranormal         | 357 | series-to-read              | 328 |
| love-triangle           | 422 | elementary             | 386 | 2011-reads            | 355 | could-not-finish            | 327 |
| my-childhood            | 422 | 2000s                  | 386 | not-read              | 355 | fantasy-science-fiction     | 327 |
| childrens-fiction       | 422 | picture-books          | 386 | mystery-detective     | 354 | dark-fantasy                | 327 |
| 1001-books-to-read      | 421 | adult-nonfiction       | 385 | comics-graphic-novels | 354 | roman                       | 327 |
| to-reread               | 419 | post-apocalyptic       | 385 | satire                | 354 | romantic-suspense           | 327 |
| witches                 | 418 | already-read           | 383 | short-story           | 353 | loved-it                    | 327 |
| classic-lit             | 418 | sociology              | 383 | read-in-2008          | 351 | books-read-in-2016          | 326 |
| bio-memoir              | 416 | young-adult-fantasy    | 381 | graphic-novels-comics | 349 | comics-manga                | 325 |
| erotic                  | 416 | favorite-author        | 380 | read-alouds           | 349 | lgbt                        | 325 |
| essays                  | 415 | business               | 377 | 2005                  | 349 | american-history            | 325 |
| self-improvement        | 413 | poetry                 | 377 | espionage             | 348 | comic-books                 | 325 |
| trilogy                 | 412 | fiction-fantasy        | 375 | family-relationships  | 345 | women-authors               | 324 |

Je me suis ensuite inspiré de cette méthode de prédiction pour les thèmes liés questionnaire, c'est pourquoi je vous invite ensuite à lire le notebook Themes\_Collaborativ\_filtering. Concernant les prédictions, il s'agissait en premier lieu de prédire les thème mais cette méthode m'a aussi permis de prédire des livres en se basant sur la similarité entre utilisateurs ou sur ce qu'il a déjà lu. Le lien fait entre les critères d'entrée et les thèmes en sortie repose sur mon travail réalisé dans le fichier Livria\_MultiLabel\_classification. Nous y testons différents types de modèle de classification multi-label et sélectionnons le modèle le plus performant. Enfin, nous utilisons ces modèles de prédiction dans le dernier notebook Livria\_Recommender\_system.

## IV. Annexe

**Code source :**

**<https://github.com/Younzer/LIVRIA>**

**Ressources :**

**Hands-On Machine Learning with Scikit-Learn and TensorFlow**  
Concepts, Tools, and Techniques to Build Intelligent Systems  
By Aurélien Géron