

Projet – Le Choixpeau magique (KNN)

Thématique : Algorithmique Niveau : 1^{ère} Type d'activité : projet Point(s) du programme traité(s) : Algorithme des K plus proches voisins (KNN) Résumé : Ce projet fait référence au monde fantastique du roman Harry Potter et de la scène du Choixpeau magique qui détermine dans laquelle des 4 maisons l'élève doit être envoyé. Cette mise en situation va être l'occasion de mettre en œuvre l'algorithme KNN. Prérequis : Traitement des données en tables (indexation, recherche, tri et fusion de tables à partir de fichiers CSV) et Interactions entre l'homme et la machine sur le Web (événements, formulaire,...)	Durée : 8h Matériel / logiciels : - Poste informatique équipé de l'IDE Python de votre choix et d'un navigateur web. Documents ressources : - Module de formation sur le traitement des données en tables (notebooks) - Module de formation sur l'algorithme KNN (notebooks) - L'ensemble des documents sont disponibles directement sur le gitlab : https://gitlab.com/david_landry/nsi
---	--

1) Le choixpeau magique dans le roman Harry Potter



Le Choixpeau magique (Sorting Hat) est un artefact doué d'une âme qui détermine de façon magique dans laquelle des quatre maisons de Poudlard chaque nouvel élève doit être envoyé. Ces quatre maisons ont pour nom : Gryffondor (Gryffindor), Poufsouffle (Hufflepuff), Serdaigle (Ravenclaw) et Serpentard (Slytherin).

Le Choixpeau est investi de l'intelligence des fondateurs et a le don de Legilimancie, lui permettant de lire dans les pensées de celui qui le met sur sa tête et de ressentir ses aptitudes ou ses humeurs et peut également répondre aux pensées de celui qui le porte.

Il est utilisé à chaque rentrée le 1er septembre lors de la cérémonie de répartition des élèves de première année, et ce depuis plusieurs siècles. Le Choixpeau choisit la maison adéquate pour chaque élève en lisant dans ses pensées, et déclare à haute voix sa décision finale prise en fonction des vertus, des projets et des centres d'intérêts de celui-ci.

https://harrypotter.fandom.com/fr/wiki/Choixpeau_magique

2) Le Choixpeau magique en NSI

L'objectif de ce mini-projet est de **modéliser un Choixpeau magique numérique**. Votre création numérique devra donc être en mesure de déterminer la maison la mieux adaptée à un profil d'« élève cible ».

Vous devrez donc :

- Construire un modèle, basé sur l'algorithme des k plus proches voisins, capable d'attribuer une maison lorsqu'on lui fournit un profil d'élève cible et des profils d'élèves en référence.
- Construire une procédure permettant de créer ce profil d'élève cible, en élaborant un QCM pertinent (2nde partie).
- Construire une interface graphique ergonomique, permettant la saisie du QCM (2nde partie), puis l'affichage du verdict du Choixpeau magique.

3) Le choixpeau magique : 1 ère partie

3.1 Objectif

La première partie consiste à concevoir le cœur du programme, celle qui utilisera l'algorithme des kPPV en s'appuyant sur les élèves références (Caracteristiques_des_persos.csv) et sur un profil d'élève cible prédéfini.

Le programme devra renvoyer les k plus proches voisins de plusieurs profils d'élèves donnés ci-dessous, puis leurs maisons attirées.

3.2 Cahier des charges

Votre programme doit être conçu pour faciliter son test et son évaluation. Il doit donc être complet, exécutable immédiatement avec un interpréteur Python.

L'utilisation du tri built-in de Python `sort()` est autorisée.

Il est à nouveau conseillé de construire une fonction par objectif, sans mélanger le traitement de données et l'affichage. Il faut également structurer au mieux votre programme (Encodage, commentaire introductif, variables globales, définitions des fonctions, appel des fonctions dans un corps de programme structuré).

Vos fonctions doivent être parfaitement documentées dans leur docstring.

Le choixpeau magique devra se prononcer sur les profils suivants :

- Courage : 9 Ambition : 2 Intelligence : 8 Good : 9
- Courage : 6 Ambition : 7 Intelligence : 9 Good : 7
- Courage : 3 Ambition : 8 Intelligence : 6 Good : 3
- Courage : 2 Ambition : 3 Intelligence : 7 Good : 8
- Courage : 3 Ambition : 4 Intelligence : 8 Good : 8

Les données doivent résulter d'une fusion entre les fichiers Characters.csv et Caracteristiques_des_persos.csv.

Il n'est pas autorisé de modifier les caractéristiques des personnages.

Les données étant en anglais, il est préférable d'utiliser des noms de fonctions et de variables en langue anglaise.

La distance par défaut sera la distance euclidienne.

La valeur de k sera fixée à 5 par défaut.

Votre IHM doit nécessairement automatiser les passages de chaque profil devant le Choixpeau magique, avec une visualisation claire des résultats obtenus (la maison choisie + les plus proches voisins et leurs maisons respectives).

L'IHM ne doit pas nécessairement passer par une interface graphique de type Tkinter, Pygame ou page web, même si cette option est envisageable (bonus).

Considérant que ce sera l'objet du cahier des charges obligatoire de la seconde partie, il paraît inutile d'y dépenser trop d'énergie dans un premier temps.

Un algorithme doit être rédigé en pseudo-code et ne doit décrire que le cœur de la problématique (l'algorithme kPPV qui permet de renvoyer la maison choisie), pas les créations de structures de données ni l'IHM.

Au besoin, votre algorithme pourra utiliser des appels de fonctions...

- ...dont il est inutile de détailler le pseudo-code.
- ...mais avec un nom suffisamment explicite et leur docstrings complet pour connaître leur(s) type(s) d'entrée(s) et de sortie(s).

Enregistrer le programme sous le format : kPPV_NOM1_NOM2_NOM3.py

Si vous avez un document PDF (ex : pseudo-code, lien vers vos outils collaboratifs, aide à l'utilisation de votre programme,...) : kPPV_NOM1_NOM2_NOM3.pdf

Si nécessaire, créer un fichier compressé (kPPV_NOM1_NOM2_NOM3.zip) pour organiser vos fichiers. Après décompression, la structure de dossiers / fichiers doit permettre une exécution facile du programme (tous les fichiers présents, pas de changement de nom de fichier ou de dossier à effectuer).

Votre projet doit être déposé sur l'espace de travail NSI d'e-lyco (« Vue d'ensemble », « tâches »). Vous devrez y créer votre groupe avant de déposer vos fichiers.

Remarques :

- les notebooks 4_11, 4_12, 4_13, P1, P2 et P4 pourront vous servir.
- La création d'index est uniquement intéressante pour des questions de performance mais inutile par ailleurs. On réservera donc les index en cas de bonus sur la validation croisée.

3.3 Bonus possibles

- La saisie manuelle d'un profil d'élève à tester par le Choixpeau.
- Une validation croisée est effectuée, avec affichage clair des résultats en pourcentage.
- Une optimisation des résultats est recherchée (à l'aide de la validation croisée, avec un travail sur les données, sur la valeur de k, ...).

3.4 Évaluation

Critères d'évaluation (note sur 10 points) :

- Respect du cahier des charges (6 point)
- Algorithme en pseudo-code (1,5 point)
- Bonnes pratiques PEP-8 (dont commentaires et docstrings) (1 point)
- Utilisation d'outils collaboratifs numériques (Gitlab, ...) (0,5 point)
- Point « subjectif » : implication, organisation du travail en équipe,... (1 point)
- Bonus (+ 2 points envisageables en fonction de l'ampleur du bonus)

4) Le choixpeau magique : 2 ème partie

4.1 Objectif

L'objectif est double :

- Construire un questionnaire permettant de créer le profil de l'utilisateur qui y répondra .
- Créer une IHM structurée sur une page web

4.2 Cahier des charges

4.2.1 Le programme du Choixpeau magique

Pour créer votre page web, vous devrez utiliser votre précédent programme (1 ère partie). Il convient donc de le corriger au besoin en fonction des remarques faites lors de son évaluation.

4.2.2 Le questionnaire

Le QCM doit comporter entre 10 et 20 questions.

Chaque question doit proposer un choix d'au moins 3 réponses.

Les questions doivent être, si possible, basées sur une mise en situation permettant d'établir une ébauche de profil psychologique. Il faut donc, dans la mesure du possible, éviter les questions qui suggèrent trop grossièrement un rapport avec les quatre maisons de Poudlard.

Chacune des réponses sera liée à un barème, par exemple sous la forme d'un tuple de valeurs numériques, correspondant aux 4 caractéristiques : (Courage, Ambition, Intelligence, Bonté).

Un exemple simpliste :

- Question : As-tu peur du noir ?
 - Oui, j'en suis terrifié (0, 5, 5, 5)
 - Non, ça ne me fait ni chaud ni froid (5, 5, 5, 5)
 - Non, au contraire, je m'y sens bien (8, 5, 5, 5)
- Cette question ne teste que le courage, mais rien n'empêche de construire des questions plus complexes, mettant en jeu plusieurs des caractéristiques à évaluer.

Les questions / réponses / barèmes peuvent être conservées dans un fichier CSV.

Les barèmes des réponses permettent de calculer un profil numérique de l'utilisateur, sous la forme d'un tuple de 4 entiers.

Un ajustement sera probablement utile pour obtenir des valeurs proches de la moyenne de celles des élèves de Poudlard, sinon les résultats risquent d'être déséquilibrés.

4.2.3 La page web

La page web doit être d'un design simple et efficace. Par exemple, le clic sur les réponses proposées paraît la solution la plus pertinente pour répondre à un QCM.

À l'issue de ce questionnaire, la page doit afficher :

- La maison attribuée à l'utilisateur.
- Les noms et les maisons des k plus proches voisins.

Elle doit se baser sur les langages HTML / CSS. Le CSS doit être réuni dans un fichier séparé.

Le programme, permettant de déterminer la meilleure maison, peut être conservé en langage Python, en utilisant le langage [Brython](#) pour intégrer Python à la page web.

Dans la mesure du possible...

- séparer le code Brython dans un fichier séparé.
- minimiser l'utilisation du HTML pour privilégier Brython.

4.2.4 Démarche de création du profil de l'utilisateur

Rendre un document texte au format PDF pour expliquer la démarche effectuée pour calculer le profil de l'utilisateur répondant au questionnaire (un profil est un tuple de 4 nombres flottants) :

- Comment avez-vous créé vos questions / réponses ?
- Comment avez-vous lié les réponses à une caractéristique ou à une autre (courage,...) ?
- Comment avez-vous calculé le profil de l'utilisateur à partir de ses réponses ?
- Comment avez-vous testé la fiabilité de votre Choixpeau magique ?
- Avez-vous eu besoin de faire des modifications ? Lesquelles ? Pour quel résultat ?

Bref, je vous demande de détailler votre réflexion sur ce point particulier du projet : préciser toute la démarche qui vous a permis d'obtenir un Choixpeau magique le plus efficace possible.

Remarque : comme dans tout document texte à rendre, il faut y mettre un titre et l'ensemble de vos noms (et, comme vous le savez, je préfère les polices sans serif).

4.2.5 Le rendu de projet

Enregistrer la page web sous le format : Choixpeau_NOM1_NOM2_NOM3.html

Enregistrer le fichier texte sous le format : Choixpeau_NOM1_NOM2_NOM3.pdf

Créer un fichier compressé (Choixpeau_NOM1_NOM2_NOM3.zip) pour organiser vos fichiers.

Après décompression, la structure de dossiers / fichiers doit permettre une exécution facile du programme (tous les fichiers présents, pas de changement de nom de fichier ou de dossier à effectuer).

Votre projet doit être déposé sur l'espace de travail NSI d'e-lyco (« Vue d'ensemble », « tâches »), en temps voulu. Vous devrez y créer votre groupe avant de déposer vos fichiers.

4.3 Bonus possibles

- Ajouter un bouton pour avoir la possibilité de rejouer.
- Ajouter un sélecteur pour changer la valeur de k.
- Ajouter 4 sélecteurs pour tester des profils types (ex : profil (7, 6, 3, 8)), sans passer par le questionnaire.

- Gérer les situations d'ex æquo.
- Mettre votre page web en ligne, à une adresse discrète, si possible non référencée.
- Pour les très courageux : recoder tout le cœur de votre programme Python en Javascript pour vous passer de Brython. Dans ce cas, il faudra faire preuve d'une grande autonomie...

4.4 Évaluation

Critères d'évaluation (note sur 10 points) :

- Respect du cahier des charges (6 points)
- Démarche de création du profil de l'utilisateur (1,5 points)
- Tests et fiabilisation des résultats (0,5 points)
- Bonnes pratiques PEP-8 (0,5 point)
- Utilisation d'outils collaboratifs numériques (Gitlab, ...) (0,5 point)
- Point « subjectif » : implication globale, organisation du travail en équipe, (1 point)
- Bonus (+ 2 points envisageables en fonction de l'ampleur du bonus)

Remarque : il ne vous est pas demandé d'algorithme en pseudo-code pour cette 2ème partie.