

Étape S1-1: Question(s) de recherche

Consignes :

- Discuter entre vous (remue-méninges)
- Poser beaucoup de questions au commanditaire qui parfois ne sait pas trop ce qu'il veut, ou ce qu'il est possible de faire.
 - Qui est-elle ?
Elodie Cayuela est docteure en histoire de l'art et a rédigé la thèse : Théorie et pratique du portrait en France et en Grande-Bretagne au XVIIIe siècle : regards croisés.
 - Où travaille-t-elle ?
Elle est attachée temporaire d'enseignement et de recherche (A.T.E.R.) à l'université Paul Valéry Montpellier 3.
 - Quel est son rôle dans sa «compagnie» ?
Elle est professeure, doctorante.
 - Pourquoi le problème est important pour elle, pour son entité, pour la société ?
Elle aurait envie de pouvoir constater les liens de manière objective entre les différents tableaux et artistes du XVIIIème siècle visuellement parlant.
 - Qu'est-ce qui vous est demandé exactement de résoudre ?
Le but est de développer un site web permettant de récupérer les données (=télécharger un pdf du rendu par exemple), de visualiser les différents nuages de points créés, de charger les différents jeux de données et de changer les différentes variables de calculs.
Pour ce faire, il faudra entraîner un modèle (de manière supervisée ou non (non supervisé est plus compliqué)). Pour chaque image, il la transformera en vecteur et une fois tous les vecteurs positionnés, on aura une réduction de dimension utilisée de différentes manières. Il faudra permettre le changement de variables.
 - Pourquoi pensez-vous que l'on vous demande de résoudre ce problème en particulier ?
Elodie Cayuela est très attachée à l'histoire de l'art et l'étudie depuis un moment notamment lors de sa thèse. Elle s'intéresse particulièrement à voir les influences de certains artistes sur d'autres, et la data science permet de déceler des choses qu'un historien ne peut pas faire seul.
 - Seriez-vous intéressé à résoudre autre chose plutôt ?
Le réel problème que nous avons peut se détacher de l'idée globale de l'histoire de l'art du XVIIIème siècle et peut se généraliser à tous les problèmes bien que cela soit notre porte d'entrée, nous avons pour objectif de le rendre général.
 - Quel est le contexte actuel dans lequel on vous demande de résoudre ce problème ?

Le problème se place dans un contexte de recherche et de poursuite de la thèse de Elodie Cayuela où elle s'est rendue compte, à l'aide de travaux complémentaires d'une autre thèse, que lorsqu'on donnait en entrée d'un système de machine learning des tableaux, il créait naturellement (sans même avoir des informations autres que la peinture) des groupes tels que "anglais vs français".

- Pourquoi est-ce important de résoudre ce problème maintenant ?
A priori, il n'y a pas encore d'études sur ce sujet précis et aucun outil (site web) à disposition permettant de créer soit même son tableau de bord et le rendu de ses données.
- Reformuler le problème de façon non technique et valider cette formulation auprès du commanditaire
 - Traduire leur requête ambiguë en un problème concret et bien défini
Créer un site web permettant d'extraire un graphique résumant de manière objective les similitudes entre (tous ?) les tableaux du XIIIème siècle (anglais + français)
 - Pouvez-vous expliquer simplement le problème à quelqu'un d'autre ?
Pour mieux comprendre les relations entre les œuvres, le plus simple est de réussir à leur donner une appartenance à un groupe suivant certaines similitudes comme le type de peinture (peinture à huile, aquarelle...), le style utilisé... Pour ce faire, nous voulons visualiser les œuvres au sein d'une même page.
 - Prioriser les questions par ordre d'importance
Développer une plateforme de visualisation de l'espace de représentation dans un premier temps pour notre sujet d'étude en ayant en tête un développement uniforme pour les futurs sujets
Avoir une interactivité entre les différentes variables
Permettre la flexibilité des données et de pouvoir les uploader et entraîner avec un même schéma
 - Formuler des questions qui définissent le «business problem» et qui peuvent être attaquées par une technique de data science (prédiction, classification, clustering, recommandation, modélisation statistique, etc.)
Comment mettre en évidence les similitudes et les différences entre les tableaux ?
Comment faire en sorte de voir les inspirations entre les artistes ?
Comment reconnaître les impacts qu'ils ont eu sur leurs confrères ?
Les situations géographiques des artistes ont-elles des impacts sur leurs productions ?
Comment avoir un modèle assez performant qui pourra reconnaître les différents vecteurs de nos tableaux afin de pouvoir les classer au mieux ?
- Identifier les variables clefs et des métriques permettant de juger du succès de votre future analyse

Lors de l'entraînement de nos modèles nous devons faire en sorte d'avoir le modèle qui réussit au mieux le traitement de nos jeux de données de test. Il sera donc important de regarder la "précision" de notre modèle ainsi que la logique et la pertinence des visualisations (ex : traverser un nuage de points (1 point = 1 tableau) sur une carte géographique et visualiser les tableaux (au passage de la souris sur le point), leurs similitudes et une "lente transformation" à l'oeil nu, à l'instar d'un folioscope)

- Formuler une liste de questions plus spécifiques qui peuvent être résolues en utilisant des nombres et du texte:
 - Combien (régression)?
A quel point deux éléments proches sont-ils similaires ?
Combien de paramètres sont relativement similaires suivant la distance entre leurs points ? Combien de paramètres sont relativement différents ?
Quel est le nombre minimal de données à avoir afin que nos modèles fonctionnent ?
Comment éviter un surapprentissage ?
 - Quelle catégorie (classification)?
Quel critère choisir pour séparer nos données ?
Combien de critères choisir ?
Quel but se cache derrière les critères sélectionnés ?
 - Quel groupe (clustering)?
Que signifient les tailles de nos clusters ? Ces clusters peuvent-ils être traduits par des mouvements artistiques typiques de la période ?
Il y a t'il un lien entre les clusters ?
 - Est-ce anormal (détection d'anomalies)?
Est ce normal d'avoir des erreurs dans nos modèles ?
D'où peuvent provenir ces erreurs ?
 - Quelle option choisir (recommandation)?
Quel type de modèle semble être le mieux ? Et pourquoi fonctionne-t-il mieux que les autres ?
Quelles sont les visualisations les mieux adaptées au jeu de données et aux questions que l'on se pose ? (Idée de solution pour les similitudes entre les tableaux : map avec nuage de points et pop-up au survol des points avec les images et les caractéristiques des oeuvres)