RAPPORT PROJET 2

ETUDE DU GRAND DEBAT

Abdallah Essa, Mahamat Ismaïl Hour, Fayçal Rekibi

Kevin Sylla

Lien Colab : <https://colab.research.google.com/drive/1Z3GJMuKXLaRFXqgUJwwYsAHK-vZ8ZlOu>

1. Présentation et analyse du dataset

Le dataset est composé de réponses à des questions sur différents thèmes, dans le cadre du grand débat lancé par Emmanuel Macron, président de la république française. Les réponses sont des sortes de doléances et revendications sur des thèmes chers aux français. De ce fait, les réponses sont très diversifiées, à la fois dans leur fond, car les revendications des français sont très diverses, mais aussi dans leur forme, car certains français ont répondu de manière très succincte et d’autres avec de longues réponses. De plus, de nombreux individus n’ont répondu qu’à quelques questions et beaucoup de lignes sont composées en grande partie de vide. La tâche est non supervisée.

1. Notre approche

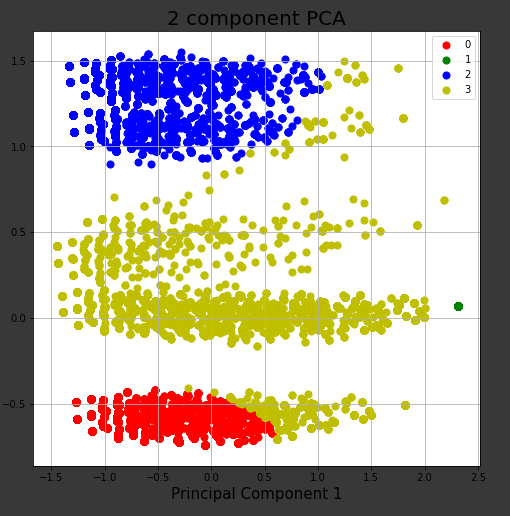
Notre principal objectif est de réussir à synthétiser les revendications des français et d’essayer de créer des clusters pour observer des différences de revendications.

1. Prétraitement

Nous avons sur les analyses de texte, éliminé les stop words de la langue française pour ne pas polluer l’identification des topics lors du LDA notamment. Nous avons également lemmatizé les mots français et mis les textes complètement en minuscule.

* 1. Création de clusters

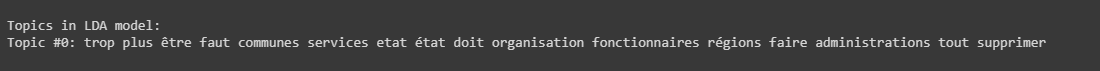
Pour séparer les données, nous avons isolé les questions fermées et avons utilisé des techniques de clustering après avoir binarisé les données de réponses à ces questions fermées.



On observe l’apparition de plusieurs clusters. Ensuite, les donnés sont séparées et nous allons traiter les données séparément en fonction des clusters. Ensuite sur chaque cluster et sur chaque question, nous appliquons un LDA pour identifier des topics et éventuellement observer des différences entre les clusters.

Voici quelques résultats :



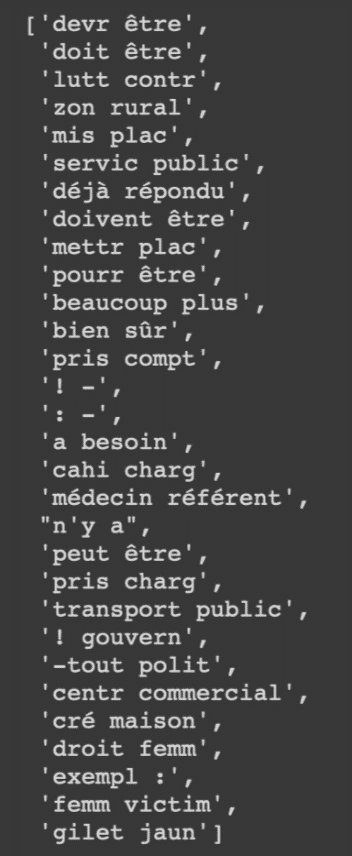
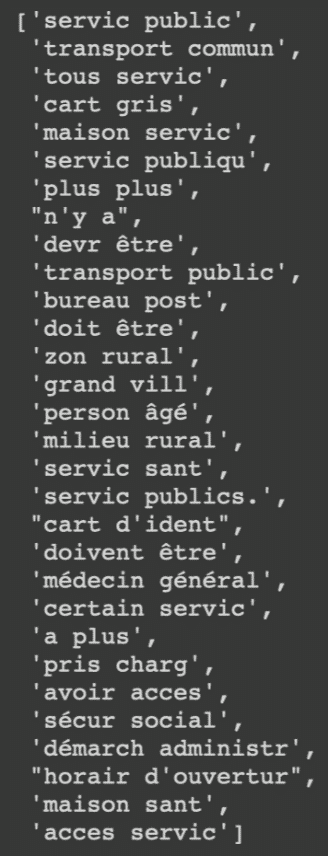
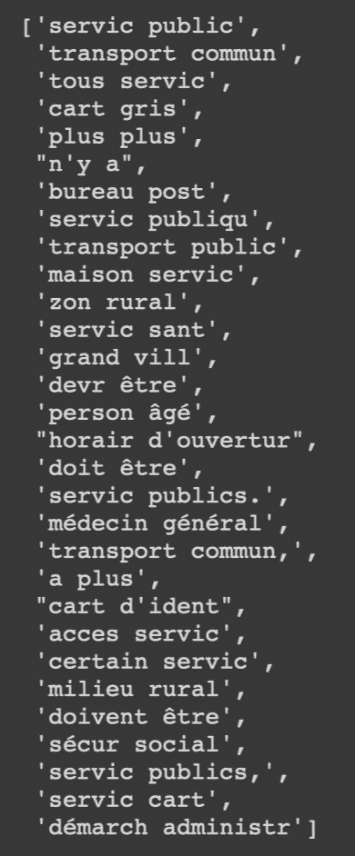






Nous avons également fait une analyse des bigram les plus frequents : voici les résultats :

**Si non, quels types de services publics vous manquent dans votre territoire et qu'il est nécessaire de renforcer ?**



Bi-Gram des 4 clusters de la question 7

On peut commencer à interpréter et comprendre les revendications principales à chaque question grâce à ces mots-clés. Néanmoins le format n’est pas clair et la suite va avoir pour but d’obtenir des phrases compréhensibles et bien interprétables.

L’objectif de cette deuxième partie est d’obtenir une synthèse de chaque réponse.

Ensuite une fois que nous avons fait cette analyse de topics, nous avons utilisé les mots isolés par les LDA pour chaque question pour identifier les commentaires les plus pertinents et les plus représentatifs de leur cluster respectif. Nous avons ensuite choisi les 10 réponses ayant le plus de mots en commun avec la liste de mots issue du LDA pour chaque question et chaque cluster. Nous avons divisé le nombre de mots fréquents par le log(nombre de mots de la phrase) pour ne pas être dépendant de la longueur de la phrase. Sinon, les phrases longues vont mécaniquement être plus susceptibles d’avoir des mots pertinents.

Ensuite, nous appliquons à ces 10 réponses une méthode de résumé extractif pour avoir une synthèse des 10 réponses les plus représentatives des topics les plus représentatifs du dataset.

C’est donc une tâche de résumé multi-document.

Ainsi, nous extrayons l’essence pour chaque cluster et chaque question, une forme synthétique des réponses, en se servant de toutes les réponses via le LDA pour « élire » les commentaires les plus représentatifs du dataset. Le fait de créer des clusters permet à minima d’obtenir plusieurs versions des commentaires les plus pertinents et à maxima de distinguer différentes revendications pour une question si le cluster offre de bons résultats. Etant donné que les résultats des modèles de clustering sont difficiles à interpréter, surtout sur des données textuelles, et probablement pertinent que sur quelques questions, seule une analyse des résultats finaux nous permettra de déterminer à quel point le clustering apporte de l’information.

Si ce n’est pas le cas, le clustering nous aura permis de diviser le dataset et d’avoir plusieurs versions des réponses.

Analysons les résultats pour une question :

**Si non, quels types de services publics vous manquent dans votre territoire et qu'il est nécessaire de renforcer ?**

CLUSTER 1 :

L’organisation de l’état \tSi on analyse le mouvement gilets jaunes on peut voir qu’un virage dans notre organisation de l’état et de notre république est nécessaire: une proportionnelle non intégrale doit être étudiée pour que chacun puisse y trouver son compte.\nVous allez me répondre il y a CAF pro, oui mais quand on travaille à l\'éducation nationale on n\'a pas CAF pro, pour aller sur la permanence de la caf il faut prendre rv surtout par internet avec son code confidentiel, il n\'y a plus de permanence où les usagers sont reçus sans rv....inutile de vous dire que pour des personnes isolées, voire pour certaines très proches de l\'illéttrisme, dans l\'incapacité de "ranger" les documents importants c\'est une démarche compliquée qu\'ils ne font pas (ils vont vous garder et vous apporter toutes les pubs de leurs boites aux lettres au milieu d\'autres courriers et ont perdu le plus souvent les plus importants - oui c\'est ça la réalité du travail social-) Du coup, faute d\'accès à leurs droits, leurs situations se dégradent et devient ensuite plus complexe à régler.'

CLUSTER 2 :

'Pour les villes proches de taille moyenne, une maison de santé regroupant un ou deux médecins , après les heures de fermeture des cabinets médicaux ; ces médecins devraient être les médecins exerçant dans ces villes , accomplissant leur service en alternance , et comme autrefois se déplaceraient à domicile en cas de besoin ; ce serait plus humain que le 15 pour des « urgences » lorsqu’ils s’agit de désagréments qui, sans être extrêmement graves, demandent des soins alors le malade n’est pas en état de se déplacer et de rester longtemps dans une salle d’attente Sauvegarder les maternités en milieu désertifié, et même si elles fonctionnent peu, on peut y ajouter d’autres services de proximité, d’utilité publique Les Urgences des hôpitaux qui sont surbookées, devraient vraiment concerner de vraies urgences, et retrouver le sens qui correspond à leur appellation.'

CLUSTER 3 :

2°/ Un service public du "déplacement" qui aurait pour objet de développer, d\'équilibrer et d\'harmoniser l\'offre de transport (notamment les solutions alternatives, avec des incitations) mais également de veiller à l\'égalité d\'accès et au moindre coût tout en respectant l\'exigence de confort, de sécurité et de qualité environnementale.les services publics en milieu rural sont de plus en plus compliqués , car de plus en plus les démarches se font par internet et dans notre pays nous avons encore des personnes qui n\'ont pas accès à internet et ou qui ne savent pas utiliser ses services ( pole emploi , caf , caisse de retraite, impôts, carte grise et maintenant les rdv médicaux ou paramédicaux..\nles services publics important en milieu rural : le bureau de poste , les mairies par contre le trésor public peut être mutualiser sur une communauté de commune ainsi que les services de gendarmerie .Que l\'ensemble des fonctionnaires (Etat, hospitalière, territoriaux) soient sur la même grille et passerelle entre-elles pour plus de flexibilité.\nEnfin, je crois beaucoup au développement des services publics itinérants indispensables pour aller à la rencontre des personnes fragiles et isolées notamment qui n\'ont pas toujours la pratique, les connaissances et les moyens de faire leurs démarches sur internet ou de se rendre en agglomérationHabitant en ville je n\'ai pas de problème d\'accès aux services public.'

CLUSTER 4 :

'Pour les « services aux français », nous avons besoin d’une supervision unique sous une haute autorité de l’ensemble de ces services (Pilotage, Organisation, Certification des Systèmes Qualité).\nJe pense que parmi les services publics que ce qui touche à l’emploi doit être profondément adapté à la situation : pôle emploi et toutes les associations contributives qui se sont crées pour en palier les insuffisances, les formations,….En tant que retraité cadre supérieur, j’ai aidé pendant quelques années des demandeurs d’emploi à retrouver un emploi ou à trouver des formations pour en retrouver et je suis effrayé par le système et par son coté une fois de plus beaucoup trop administratif d’état et pas assez concret de proximité.Ma réponse ne concerne pas ma situation personnelle mais l\'organisation des services publics sur le territoire.\nEn lien avec "La Poste", cette maison de santé doit permettre le suivi des personnes dépendantes à domicile.plus de services publics en Mairie , les citoyens sont plus proches de leur maire et de leur mairie donc donner plus de pouvoir aux mairies et d\'aide aux communes qui sont plus apte à connaitre les besoins publics et services pour les citoyensIl faut que les services Internet devenus obligatoires deviennent efficaces par obligation et surveillance impliquée de l\'état .'

Analyse :

En observant les résumés des clusters, nous observons que les résumés des clusters 1 et 4 concernent les problèmes de l’administration de l’état, par rapport à sa complexité et sa lourdeur.

A contrario, les clusters 2 et 3 concernent les services d’insertion de la ruralité et de la disponibilité des services tels que la santé etc.

De plus, les résumés des clusters 1, 3 et 4 se rejoignent dans les problèmes d’accès à internet.

Nous observons que sur cette question, les clusters arrivent à permettre de mettre en lumière des commentaires longs, exhaustifs et qui traitent de différentes problématiques. Les résumés générés peuvent servir de synthèse de réponse à la question posée, d’après notre analyse.