

Présentation du problème

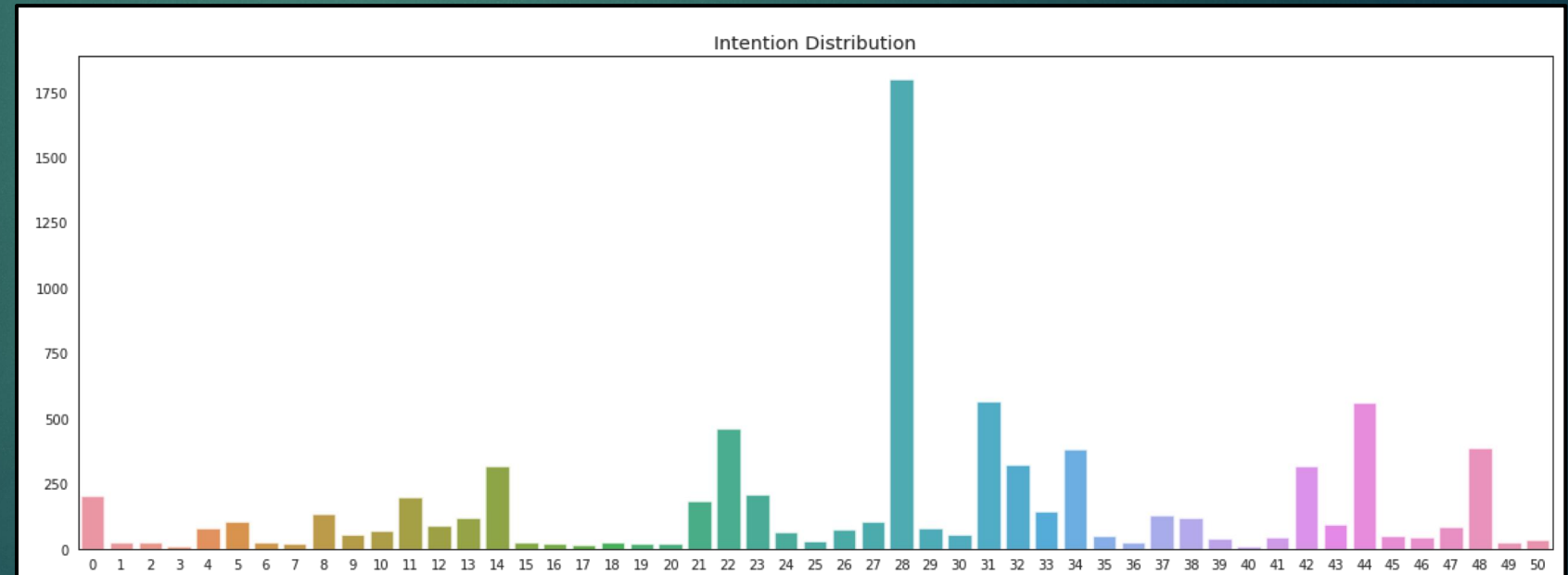
POSOS



8028 questions (train)
2035 questions (train)

51 catégories
(anonymisées)

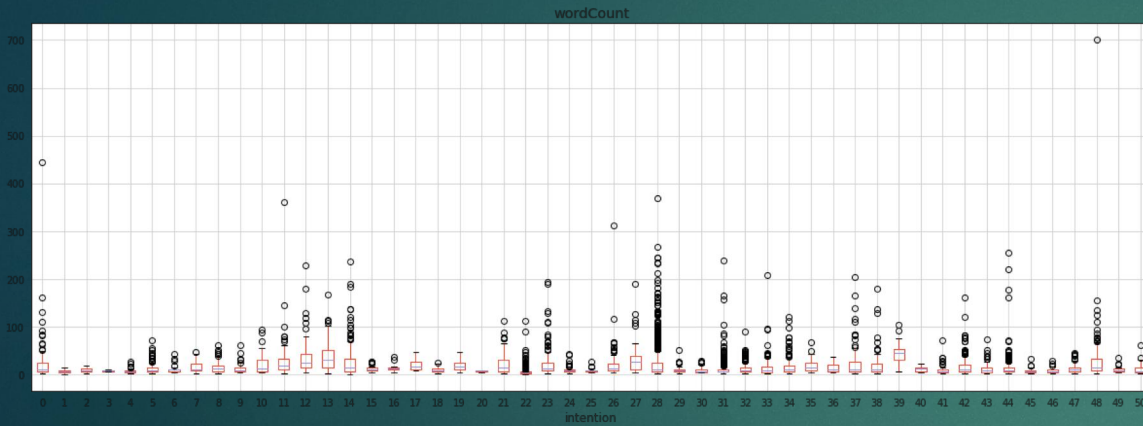
question	intention
bonjour, je m suis trompé de forum pour ma question alors je la repose ici. je pris pour la première fois hier du paroxétine et ce matin c'est une catastrophe. picotement dans tous le corps annonciateur de sueur froide très très massive et de vomissement. j'en suis à deux crises depuis 5 heure du mat. la cela semble passer mes mes mains reste moites et chaude estce normal pour la première fois merci a tous	28
est ce que le motilium me soulagera contre les nausées?	31
mon médecin m'a prescrit adenyli. au 2ème cachet des maux de tête terribles et au 3ème palpitations, sueurs froides, chaleur intense dans la tête, tremblements, fourmillements dans la lèvre supérieure, difficultés à respirer.. dès l'arrêt du médicament tous les symptômes ont disparu. cela est-il déjà arrivé à quelqu'un??	28
Est-ce qu'il existe une forme adaptée aux enfant de 5ans du Micropakine ?	44
mon medecin me soigne pour une rhino pharyngite et m'a prescrit du amoxicilline comme anti biotique. Est-ce vraiment pour cette indication?	31



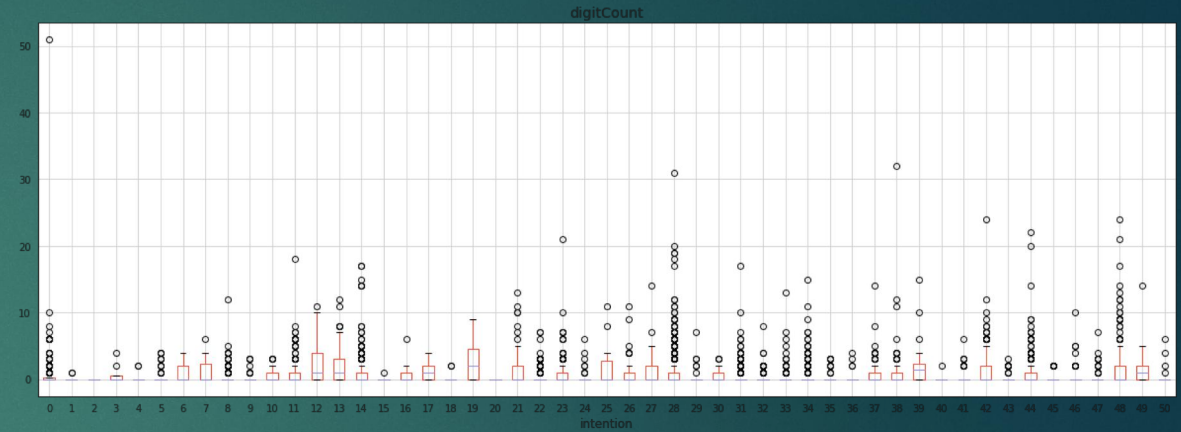
Analyse des données

POSOS

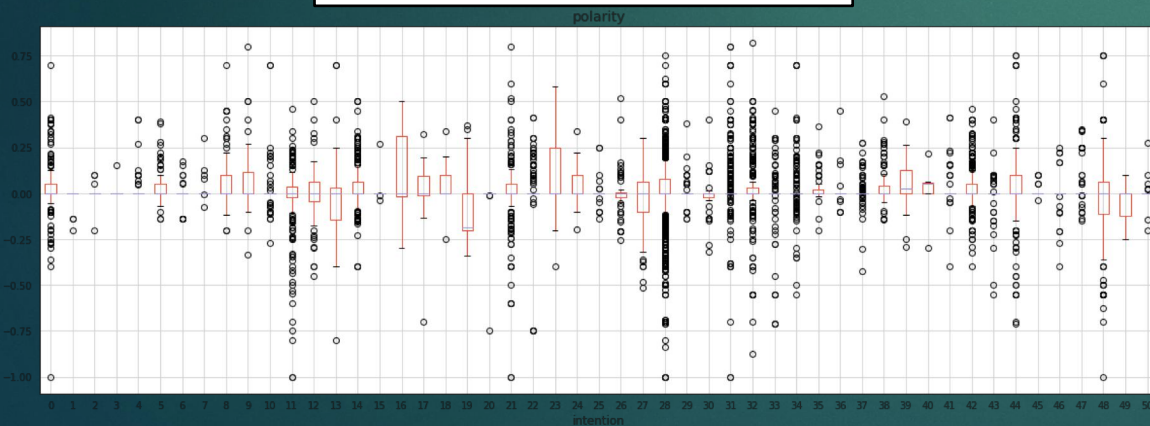
Nombres de mots



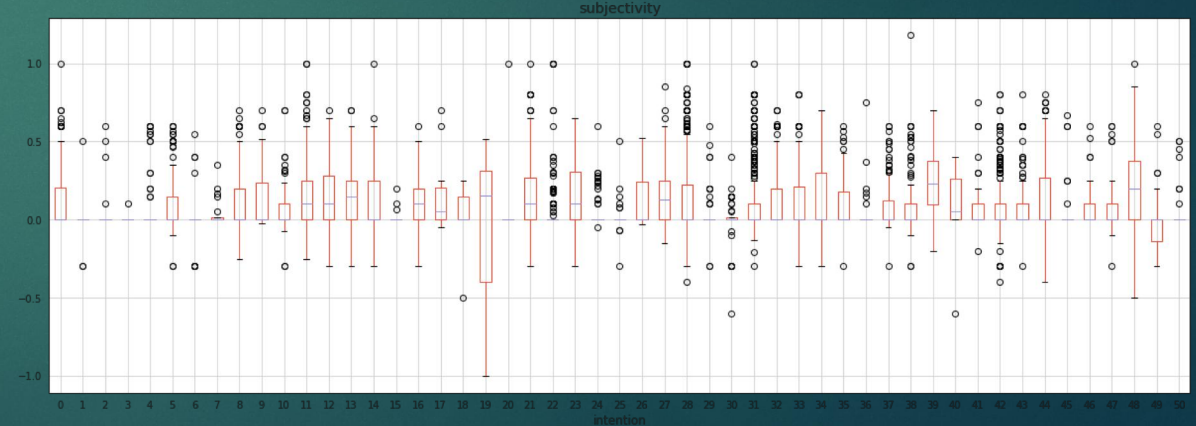
Nombres de chiffres



Polarité



Subjectivité



Analyse des données

POSOS

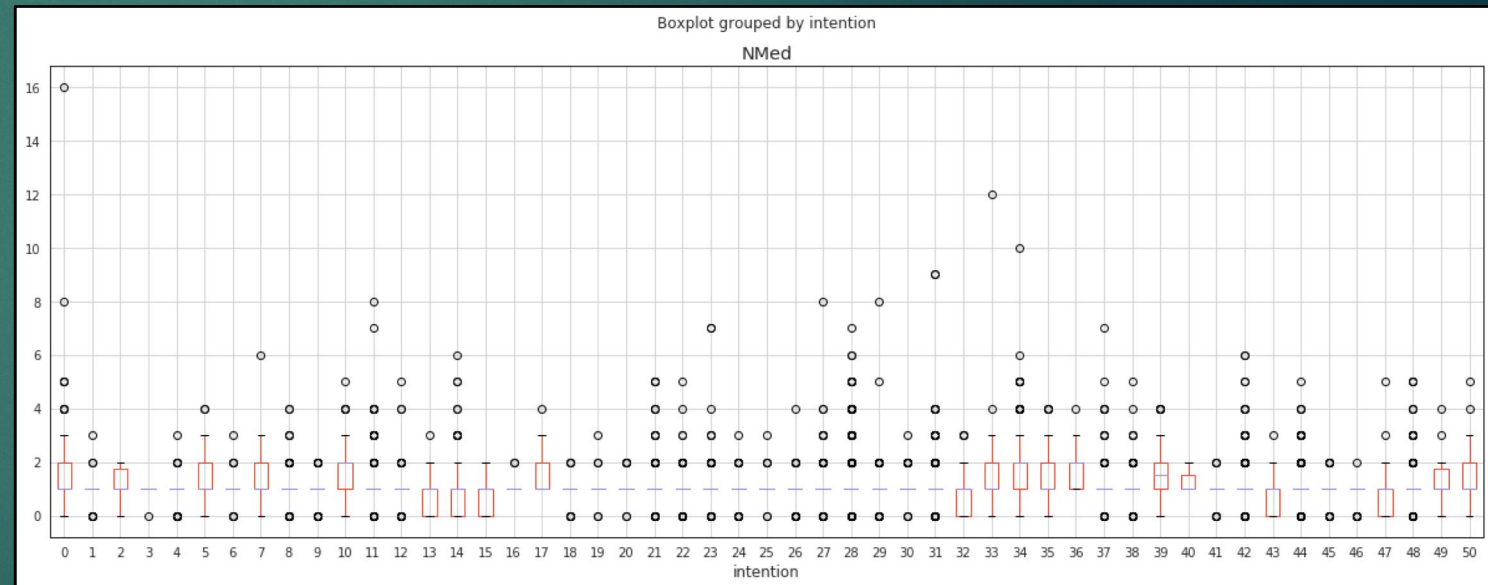
NOMBRE DE MÉDICAMENTS PAR CATÉGORIE

CIS.txt

62869109	A 313 50 000 U.I., capsule molle	capsule molle	orale	Autorisation active	Procédure nationale
61855042	A B I, Émulsion pour application locale	émulsion pour application	cutanée	Autorisation abrogée	Procédure nationale
61679174	ABACAVIR DEXTREG 300 mg, comprimé pelliculé sécable	comprimé pelliculé sécable	orale	Autorisation active	Procédure nationale
62481060	ABACAVIR WYLAN 300 mg, comprimé pelliculé sécable	comprimé pelliculé sécable	orale	Autorisation active	Procédure d'centralisation
63797011	ABACAVIR SANDOZ 300 mg, comprimé pelliculé sécable	comprimé pelliculé sécable	orale	Autorisation active	Procédure d'centralisation
62828870	ABACAVIR/LAMIVUDINE ARROW 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation
63431640	ABACAVIR/LAMIVUDINE BIOGARAN 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure nationale
65196479	ABACAVIR/LAMIVUDINE EG 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation
62170486	ABACAVIR/LAMIVUDINE WYLAN PHARMA 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure nationale
63011387	ABACAVIR/LAMIVUDINE PHAROS 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation
67720261	ABACAVIR/LAMIVUDINE SANDOZ 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure nationale
65857902	ABACAVIR/LAMIVUDINE TEVA 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation
67567004	ABACAVIR/LAMIVUDINE VALE 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation
62202005	ABACAVIR/LAMIVUDINE ZENTIVA 600 mg/300 mg, comprimé pelliculé	comprimé pelliculé	orale	Autorisation active	Procédure d'centralisation



['novodil', 'carbosomag', 'metfogamma', 'haxifal', 'diarsed', 'aspip', 'zymaduo', 'cocculus', 'polyionique', 'homeoptic', 'gammatetanos', 'c', 'lavement', 'acal', 'formoterol', 'psyllium', 'cypropharm', 'phyccitux', 'buspar', 'kapanol', 'panzyga', 'hirucreme', 'cliavist', 'gristam', 'vegetal', 'variquel', 'dalmene', 'clonazone', 'durogesic', 'rmgel', 'lyodenal', 'lytos', 'hantrux', 'teicoplanine', 'omix', 'ger', 'bronchyteine', 'cinnabaris', 'thuya', 'valcote', 'vegetoserum', 'ar', 'betahistine', 'gelaspan', 'osmogel', 'miconazole', 'gastric', 'fluc', 'e', 'ionoka', 'wakix', 'gaosedal', 'mimpara', 'oxinate', 'cancidas', 'diafusor', 'aldyzine', 'isovist', 'zinc-nickel-cobalt', 'rimcure', 'e', 'tussilene', 'noctium', 'domistan', 'seropram', 'tetanea', 'fer', 'l', 'azonutril', 'elvorine', 'indometan', 'theodel', 'viread', 'gere', 'xe', 'cynomel', 'synacthene', 'mucocil', 'ozurdex', 'mintexed', 'hes', 'lycobyl', 'blincyto', 'proglcem', 'isudrine', 'suxamethonium', 'sar', 'gine', 'citrocholine', 'nibiol', 'sargemag', 'solute', 'kispplx', 'i



Analyse des données

COMPRENDRE LE CONTENU DES CATÉGORIES

```
Cluster_2 = ['bilan', 'glycemie', 'glycémie', 'biologique', 'serologie', 'suivi', 'sérologie'] ## Attention to Sang
Cluster_3 = ['rupture', 'stock', 'fabricant']
Cluster_4 = ['être', 'appartenir', 'famille', 'apparenter', 'classe', 'pharmacologique', 'categorie']
Cluster_5 = ['compose', 'composer', 'contenir', 'composition', 'difference']
Cluster_6 = ['disponible', 'a-t-il', 'exister', 'savoir', 'prendre']
Cluster_7 = ['garder', 'durer', 'duree', 'conservation', 'conserver', 'frigo', 'frais', 'reutilisable', 'utilisable',
Cluster_8 = ['prescription', 'ordonnance', 'posologie', 'obtenir', 'legislation', 'delivrer']
Cluster_9 = ['prescrire', 'prescription', 'faire', 'changer', 'pharmacien', 'infirmier', 'medecin']
Cluster_10 = ['notice', 'enchaîner', 'passer']
Cluster_11 = ['help', 'suite', 'boutons', 'rouge', 'gaz', 'reaction', 'frissons', 'depuis', 'gauche', 'saignement', 'c
Cluster_12 = ['depasser', 'oublier', 'oubli', 'exces', 'prendre', 'manquer', 'avalér', 'regurgiter', 'tromper', 'surde
Cluster_13 = ['risque', 'oublier', 'oubli', 'protéger', 'dangereux']
Cluster_14 = ['arreter', 'plaquette', 'est-ce-que', 'devoir', 'faire', 'pouvoir']
Cluster_15 = ['patch']
Cluster_16 = ['perime', 'conservation', 'ouverture', 'utilisable', 'depuis']
Cluster_17 = ['vaccin', 'inutilisable', 'perime', 'refrigerateur']
Cluster_18 = ['commercialiser', 'commercialisation', 'accessible', 'quel', 'disponible', 'existe', 'apparition', 'rece
Cluster_19 = ['comsotion', 'prendre', 'perimer', 'peremption', 'utiliser', 'consomer', 'date', 'validite', 'expirer']
Cluster_20 = ['version', 'marque', 'nom', 'commercial', 'specialité'] # Generique supprimé par Sofiane
Cluster_21 = ['Arret', 'sevrage', 'effet', 'reaction', 'reaction', 'stopper', 'arreter']
Cluster_22 = ['question', 'cest', 'quoi', 'que', 'connaitre', 'info', 'renseignement', 'renseigner']
Cluster_23 = ['premiere', 'combien', 'temps', 'marche', 'efficace', 'effets', 'attendre', 'action', 'bout']
Cluster_24 = ['disponibilite', 'procurer', 'trouver', 'acheter', 'retire', 'retour']
Cluster_25 = ['dosage', 'combien', 'unidose', 'doser', 'injection']
Cluster_26 = ['efficace', 'effet', 'efficacite', 'protege', 'agir']
Cluster_27 = Cluster_effet_second + Cluster_temps
Cluster_28 = ['trouver', 'France', 'france', 'exister', 'vente', 'commercialisé', 'disponible', 'où', 'ou', 'marché']
Cluster_30 = ['couper', 'diluer', 'mélanger', 'lait', 'sirop', 'réduire', 'doses', 'dose', 'sécable', 'sécable',
'secable', 'réduire']
Cluster_31 = ['indications', 'indication', 'role', 'rôle', 'roles', 'rôles'] + Cluster_verbe_utiliser

Cluster_33 = ['Alcool', 'alcool', 'fumer', 'boire', 'boit', 'cigarette', 'cigarettes', 'café', 'cafe', 'drogue', 'soleil
Cluster_34 = ['mélanger', 'associer', 'en même temps']
Cluster_35 = ['au lieu de', 'moyen', 'moyens']
Cluster_38 = ['moment', 'soir', 'matin', 'heure', 'midi', 'réveil', 'quand', 'moment'] + ['prendre']

Cluster_40 = ['origine', 'base', 'vient', 'où', 'demande', 'demander', 'souhaite', 'souhaiter']
Cluster_41 = ['éliminer', 'ressort', 'demi-vie', 'métabolisation', 'métaboliser', 'hépatique', 'absorption', 'éliminat
Cluster_42 = ['mg', 'dose', 'combien']

Cluster_44 = ['enfants', 'bébé', 'enceinte', 'grossesse', 'nourisson', 'bebe', 'enfant']
Cluster_45 = ['version', 'générique', 'DCI']
Cluster_46 = ['prix', '€', 'couter', 'coûte', 'cout', 'coûter', 'payer', 'tarif', 'coute']
Cluster_47 = ['Remboursée', 'remboursement', 'sécurité', 'securite', 'sociale', 'déremboursé',
'remboursable', 'charge', 'payant', 'rembourser', 'derembourser', 'rembourse']
Cluster_48 = ['dangereux', 'risques', 'risque', 'risquer']
Cluster_49 = ['dissoudre', 'découpables', 'sécable', 'casser', 'couper', 'coupé', 'divise']
Cluster_50 = ['substituable', 'non-substituable', 'substituté', 'substitué']
```

Prétraitement des questions

Dictionary.txt

(Nom) |décadence|diminution|humiliation|affaiblissement|amollissement|amoindrissement|abattement|affaires
déperissement|déclin|baisse|décrépiteude|agenouillement|genuflexion|prosternement|inclinaison|aplatissement
aveulissement|avachissement|avilissement|réduction|soustraction|décroissance|retranchement|ab
dévaluation|amenuisement|amincissement|amaigrissement|dépréciation|fléchissement|courbure|fle
dégradation|vexation|honte|confusion|affront|avanie|prosternation|rabaissement|rétrogradation
servitude|esclavage|dépendance|subordination|sujétion|assujettissement|joug|vassalité|asservi
abaisser|1
(Verbe) |avilir|diminuer|humilier|baisser|abattre|anéantir|assommer|démolir|décourager|renvers
descendre|pencher|incliner|courber|rabattre|faiblir|décroître|dévaluer|dévaloriser|déprécier|
mortifier|vexer|blesser|confondre|cingler|gifler|inférioriser|rabaisser|prosterner|retrancher
nettoyer|ravilir
abajoue|1
(Nom) |abajoue
abandon|1
(Nom) |renoncement|cessation|délaissement|abdication|capitulation|démission|désistement|renonc
liquidation|arrêt|fin|interruption|suspension|répit|rémission|relâche|apaisement|suppression|
acceptation|déterminisme|passivité|renoncement|résignation|infidélité|déloyauté|trahison|inco
tromperie|erreur|inexactitude|isolement|isolation|solitude|claustration|exil|séquestration|qui
négligence|incurie|désinvolture|désordre|relâchement|débraillé|largage|lâchage|parachutage|pe
plaquage|placage|reddition|refus|négation|dénégation|récusation|inacceptation|rejet|veto|renv
dédit|démenti|résistance|reniement|désintéressement|soumission|abnégation|constance|désespér
abandonne|1
(Verbe) |planter|végéter|graminée|arbuste|sème|campe
abandonné|1
(Adjectif Nom Verbe) |infréquenté|délaisse|dépeuplé|désert|désolé|écarté|inhabité|perdu|retiré
déserté|solitaire|nu|orphelin|pupille|seul|reculé|isolé|farouche|fauve|indompté|craintif|brut
simple|esseulé|reclus|vacant|libre|disponible
abandonner|1
(Verbe) |abdiquer|cesser|délaisser|quitter|renoncer|capituler|démissionner|se démettre|se dési
renier|rétracter|désavouer|choir|tomber|plaquer|dévouer|dédier|consacrer|offrir|sacrifier|vou
faiblir|reculer|mollir|laisser|négliger|omettre|consentir|confier|remettre|léguer|transmettre
larguer|parachuter|renvoyer|pâmer|défaillir|évanouir|admirer|enthousiasmer|partir|émigrer|fui
déguerpir|disparaître|appareiller|démarrer|planter|semer|camper|coller|immoler|résigner|retom
foncer|succomber
abaque|1
(Nom) |anamorphe|boulrier|calculateur|compteur
abasourdi|1
(Adjectif Verbe) |coi|tranquille|muet|sidéré|stupéfait|pantois|époustouflé|ébahi|ahuri|déconcerté|décont
éberlué|ébouiffé|émerveillé|étonné|fasciné|interdit|médusé|pétrifié|effondré|consterné|étourdi|atterré
chagriné|confondu|suffoqué|surpris|frappé|hébété|abêti|abruti|troubé|stupide|bête|engourdi|ensommeillé
abasourdie|1
(Adjectif Verbe) |coite|tranquille|muette|sidérée|stupéfaite|pantoise|époustouflée

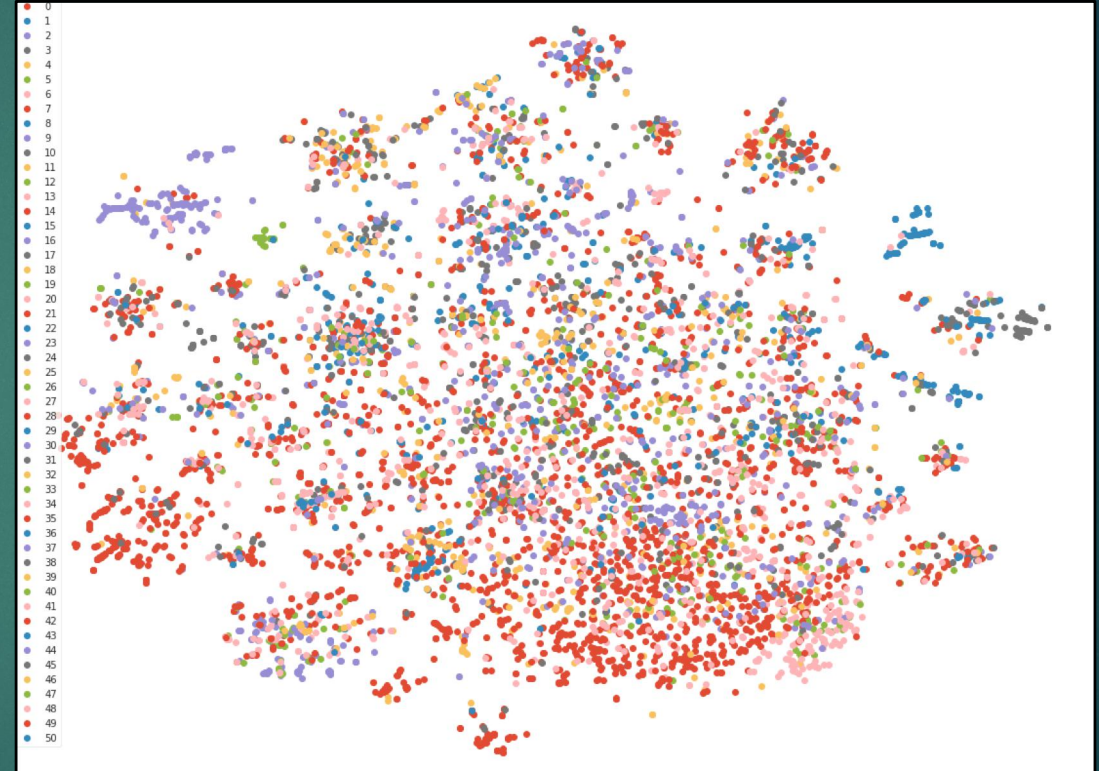
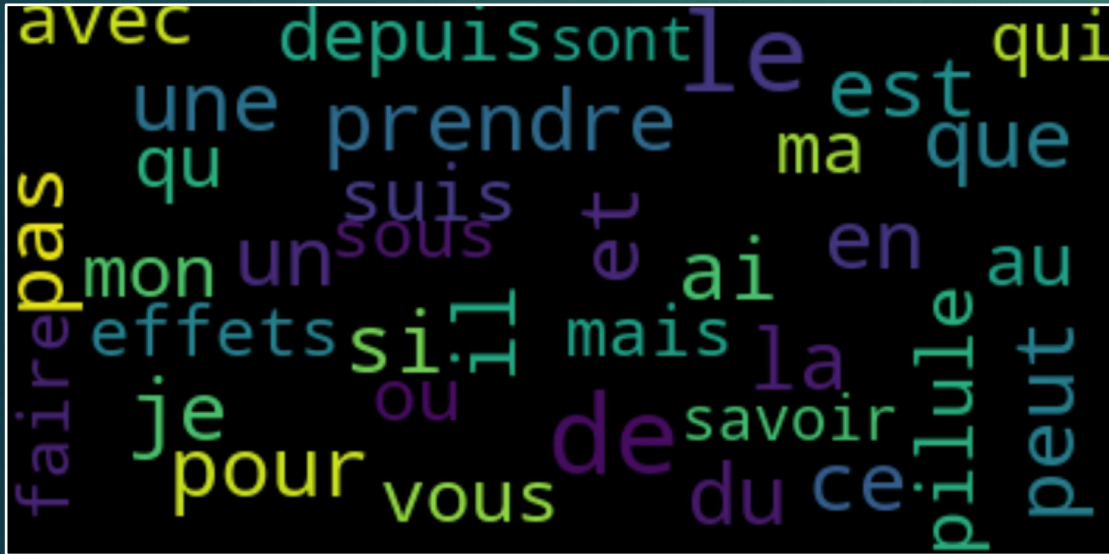
```
def tokenizel(text):  
    text = string.lower()  
    text = unicode.unidecode(string)  
    text = re.sub(r'[^a-z]', ' ', string)
```

spacy

```
def normalisel(words_list):  
    words_list_normalised = words_list  
    return words_list_normalised  
  
def text_preprocessing1(text):  
    return normalisel(remove_noisel(tokenizel(text)))
```


Analyse des données

TOPIC MODELING



Features Engineering

GÉNÉRÉES MANUELLEMENT

- Nombre de mots
- Pourcentage :
 - Nombre de chiffres
 - Nombre de médicaments
 - Nombre de mots clés présents (265)

```
Cluster_2 = ['bilan', 'glycemie', 'glycémie', 'biologique', 'serologie', 'suivi', 'sérologie'] ## Attention to Sang
Cluster_3 = ['rupture', 'stock', 'fabricant']
Cluster_4 = ['être', 'appartenir', 'famille', 'apparenter', 'classe', 'pharmacologique', 'categorie']
Cluster_5 = ['composer', 'composer', 'contenir', 'composition', 'difference']
Cluster_6 = ['disponible', 'a-t-il', 'exister', 'savoir', 'prendre']
Cluster_7 = ['garder', 'durer', 'duree', 'conservation', 'conserver', 'frigo', 'frais', 'reutilisable', 'utilisable',
Cluster_8 = ['prescription', 'ordonnance', 'posologie', 'obtenir', 'legislation', 'delivrer']
Cluster_9 = ['prescrire', 'prescription', 'faire', 'changer', 'pharmacien', 'infirmier', 'medecin']
Cluster_10 = ['notice', 'enchaîner', 'passer']
Cluster_11 = ['help', 'suite', 'boutons', 'rouge', 'gaz', 'reaction', 'frissons', 'depuis', 'gauche', 'saignement', 'c
Cluster_12 = ['depasser', 'oublier', 'oubli', 'exces', 'prendre', 'manquer', 'aval', 'regurgiter', 'tromper', 'surde
Cluster_13 = ['risque', 'oublier', 'oubli', 'protéger', 'dangereux']
Cluster_14 = ['arreter', 'plaquette', 'est-ce-que', 'devoir', 'faire', 'pouvoir']
Cluster_15 = ['patch']
Cluster_16 = ['perime', 'conservation', 'ouverture', 'utilisable', 'depuis']
Cluster_17 = ['vaccin', 'inutilisable', 'perime', 'refrigerateur']
```



GÉNÉRÉES PAR MODÈLES

- Topic modeling
- Random Forest
- Prédiction de probabilités

UN SEUL MODEL

- Naive Bayes, Random Forest, XGBoost appliqués sur counterVect et TF-IDF—> reduction de dimension.
- Deep Learning

DEUX MODELS

- Déséquilibre aussi présent en test set → Classes distingués --> RF
- Model séparé qui contient les features TF-IDF et R

A faire :

PROJECTIONS D'IDÉE À FAIRE :

- Confusion MATRIX
 - Ensembling
 - Stacking
- Ajout des features du topic modeling

MERCI DE VOTRE ATTENTION