

Objectif

"75 cl de vin rouge le soir pour le antioxydants très bon pour la santé (ils l'ont dit à la télé) 25 mg d'Anafranil le soir 1 ou 2 comprimés de xanax 0,5mg/jour Parfois du Tercian le soir pour dormir. Est-ce que vous pensez que c'est un bon traitement ?" → alcool contrindiqué + dosage

"je devais commencer mon utrogestran 200 a j16 bien sur j'ai oublier! donc je l'ai pris ce soir!!!!" → oubli de prise

"ma petite soeur de 7 ans a bu pas tout le flacon mais beaucoup est-ce que c'est grave ??????" → âge contrindiqué + dosage

L'objectif est de catégoriser automatiquement des messages de forum selon qu'ils contiennent un mésusage de médicament ou pas.

Mésusage : Utilisation d'un médicament d'une façon ou pour un objectif différents des recommandations médicales



"moi ma psy m'a bien dit, pas d'alcool, mais j'en bois quand même en week-end quand y'a des occasions" → alcool contrindiqué

"Je vous demande donc si vous en avez en trop de me vendre de la sertraline 50mg et de l'alprazolam à fort dosage" → si ces médicaments sont sur ordonnance, c'est un mésusage

"j'ai peur je le prends pas" → ne pas prendre un médicament qui a été prescrit est aussi un mésusage

"j'ai déconné avec des somnifères pour me défoncer en gros" → usage psychotrope

Tâche

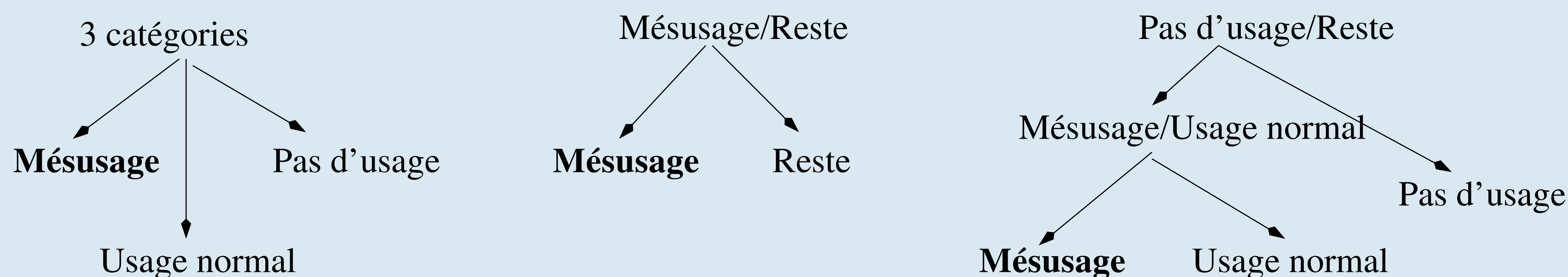
Classification en trois catégories :

pas d'usage : "il ne t'a pas prescrit d'aspegic nourisson ?"

usage normal : "est ce que c'est normal que le loxapac que je prends met des heures à agir ?"

mésusage : "je me suis trompée j'ai pris 3mg au lieu d'1mg est-ce que c'est déjà arrivé à qqun ?"

Différents chemins pour isoler la catégorie **mésusage** :



Paramètres

► Texte

► Formes

"Je suis sous seroplex et pour l'instant à 15 ans, les effets secondaires se font beaucoup trop fort"

► Lemmes

"je suivre | etre sous seroplex et pour le instant avoir @card@ an , le effet secondaire se faire beaucoup trop fort"

► Lemmes Lexicaux

"suivre|etre seroplex instant avoir an effet secondaire faire beaucoup trop fort"

► Médicaments/maladies

On remplace les noms de médicaments et de maladies pour évaluer leur impact sur la classification

► normal, texte original

"je suis sous seroplex"

► code, remplacé par un code où toutes les appellations d'un même médicament/maladie sont réduites à un code unique

"je suis sous N06"

► normal + code, texte original avec ajout du code

"N06 | je suis sous seroplex"

► substitution, remplacé par une marque unique

"je suis sous \$drug\$"

► supprimé

"je suis sous"

► Algorithme (implémentation Weka)

► NaiveBayes

► NaiveBayes Multinomial

► RandomForest

► SimpleLogistic

Corpus

- messages des forums santé de *Doctissimo*
- contient un nom de médicament
- < 2500 caractères
plus longs : difficiles à annoter, contiennent rarement plus d'information, ou sont des articles copiés/collés

total : 1 850 messages
 pas d'usage : 600 messages
 usage normal : 1 117 messages
 mésusage : 133 messages

Dans les expériences les classes sont équilibrées

Résultats

Meilleur classifieur (Expérience *mésusage / reste* avec *nom + code des médicaments*)

	NaiveBayes		
	Précision	Rappel	F-mesure
formes	0,735	0,734	0,733
lemmes	0,778	0,772	0,773
lemmes lexicaux	0,821	0,810	0,809
NaiveBayesMultinomial			
formes	0,669	0,646	0,627
lemmes	0,608	0,582	0,579
lemmes lexicaux	0,812	0,810	0,810

Expériences sur l'impact des médicaments et des maladies. Avec texte lemmatisé, algorithme de type NaiveBayes, exprimé en F-mesure.

	3 classes	mésusage/reste	pas d'usage/reste	usage/mésusage
normal	0,544	0,734	0,713	0,741
code	0,546	0,728	0,700	0,739
normal + code	0,563	0,740	0,684	0,758
substitution	0,540	0,731	0,702	0,721
supprimé	0,554	0,731	0,694	0,721
3 classes mésusage/reste pas d'usage/reste usage/mésusage				
normal	0,577	0,763	0,768	0,763
code	0,542	0,720	0,750	0,651
normal + code	0,579	0,793	0,749	0,763
substitution	0,554	0,734	0,755	0,661
supprimé	0,544	0,730	0,751	0,676

Limites

- Petit corpus peu varié
- Médicaments peu variés
- Corpus de test équilibré entre les classes

Conserver le nom du médicament/maladie (*normal* ou *normal+code*) améliore légèrement les résultats