

# Projet UE-4 : Identification d'interlocuteurs

GUNDUZ Maxime  
AIT SAID Hicham

6 Janvier 2025

Master 2  
Informatique  
Médical  
2024-2025



Introduction et présentation

Conception de notre jeu de données et représentation bag-of-word (TF-IDF)

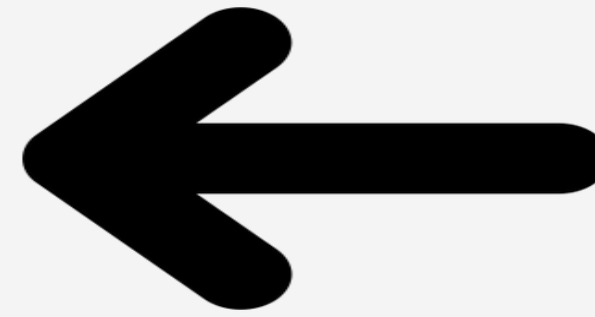
Réalisation de nos modeles

Présentation des résultats et discussions des performances

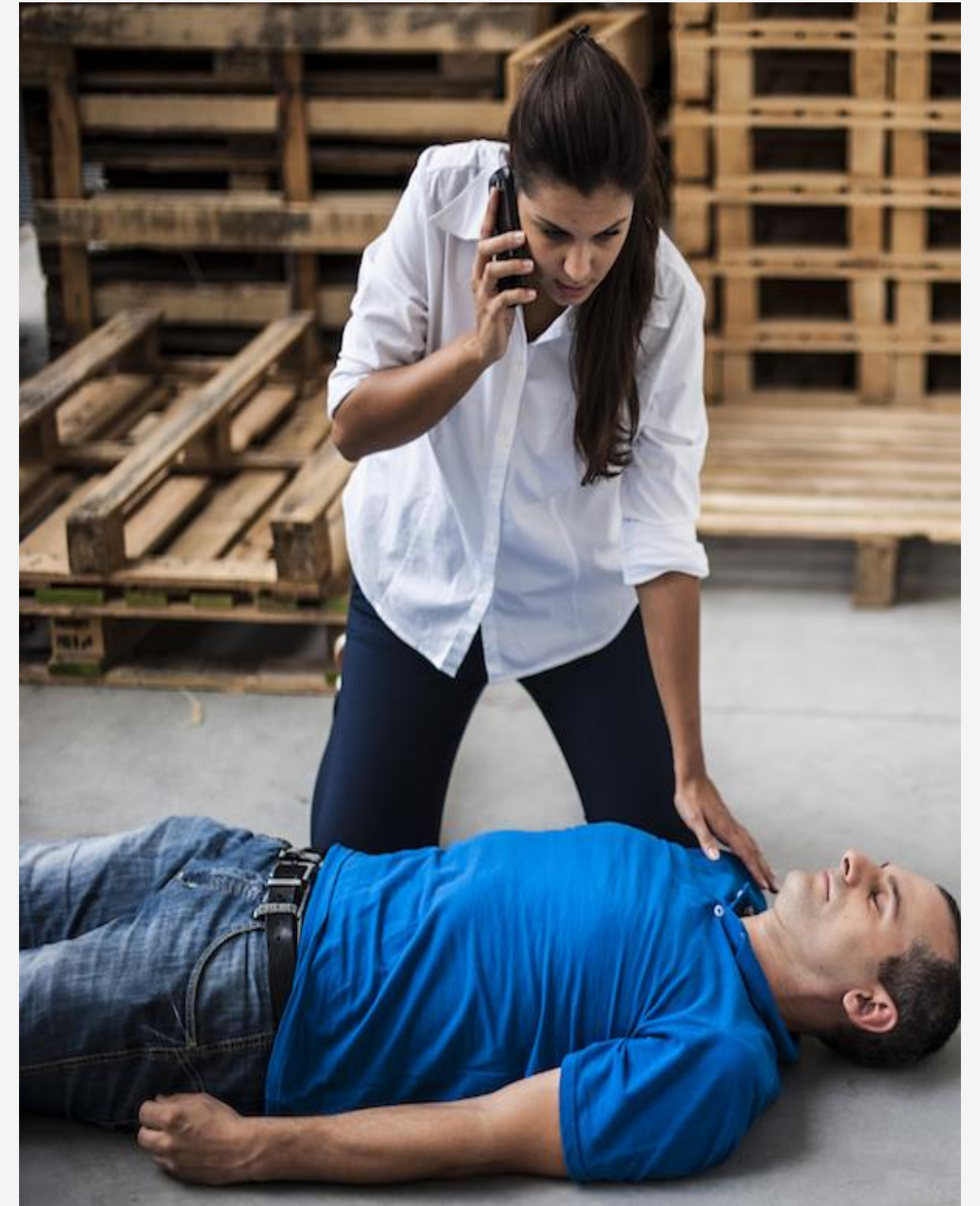
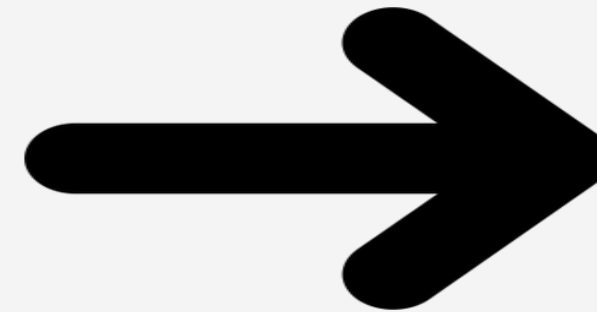
Conclusion et avenir du projet



**SAMU (Service d'Aide Médicale Urgente)**



**Échange  
téléphonique  
“Patient”  
“Médecin”**

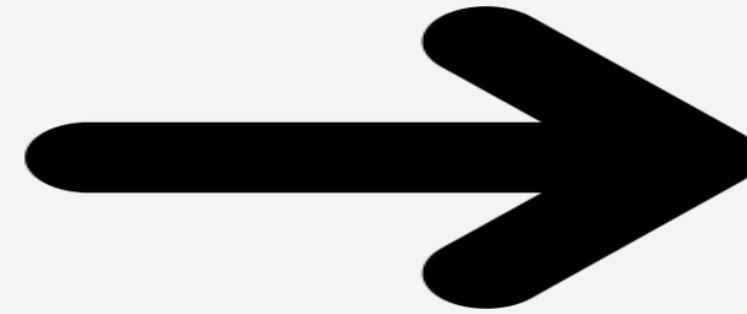


**2021 :**

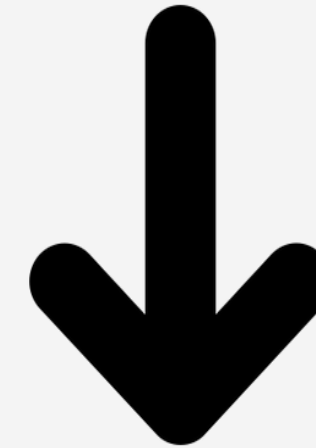
- 32,3 millions d'appel, selon sfmu (société française de médecine d'urgence)**
- 2-3 minutes par appels en moyennes**



Dataset : medkit/simsamu



Importation du jeu de données



```
[  
  {  
    'patient' : ['toute les phase du patient dans le cas'],  
    'medecin' : ['toute les phase du médecin dans le cas']  
  },  
  {Cas 2} , {cas 61}  
]
```



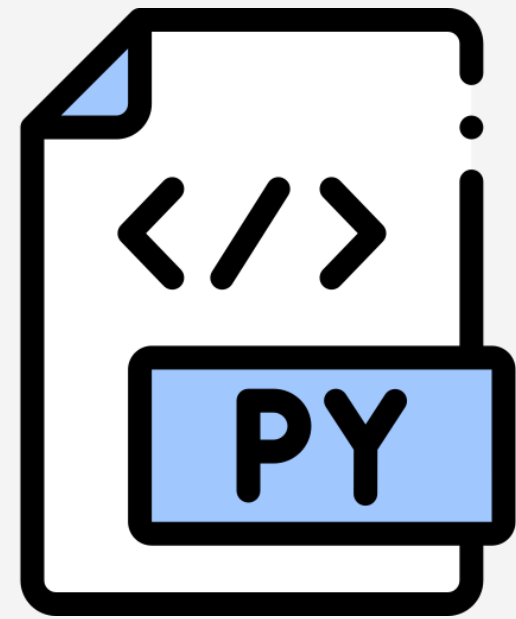
Fichier srt



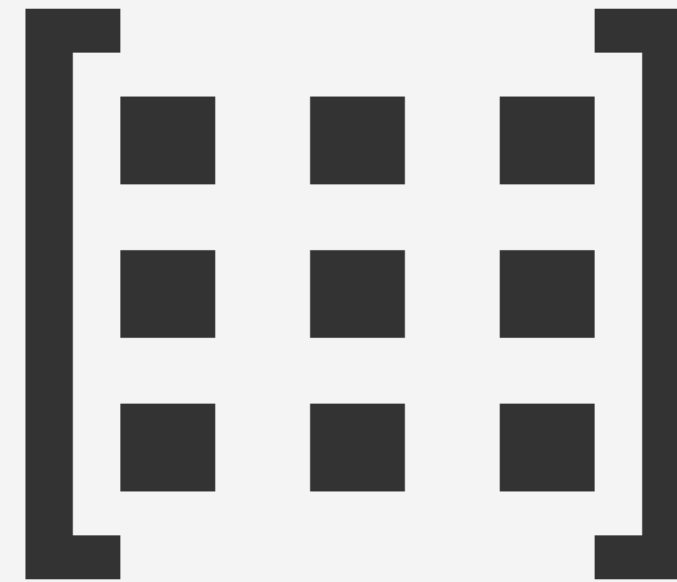
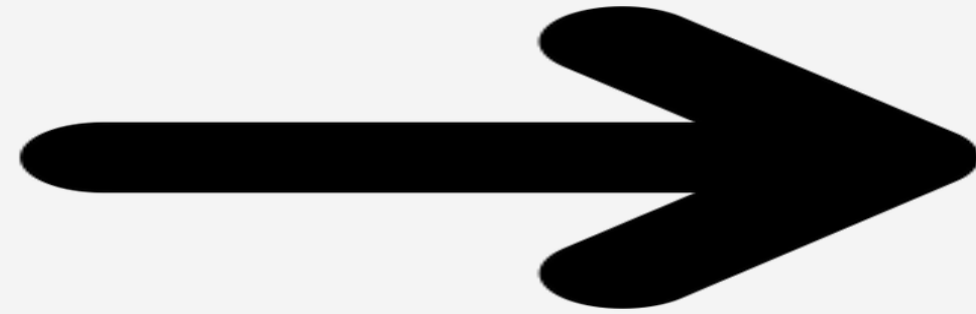
Fichier rttm



# Représentation bag-of-words (TF-IDF)



Dataset



Matrice (3061x2625 termes)

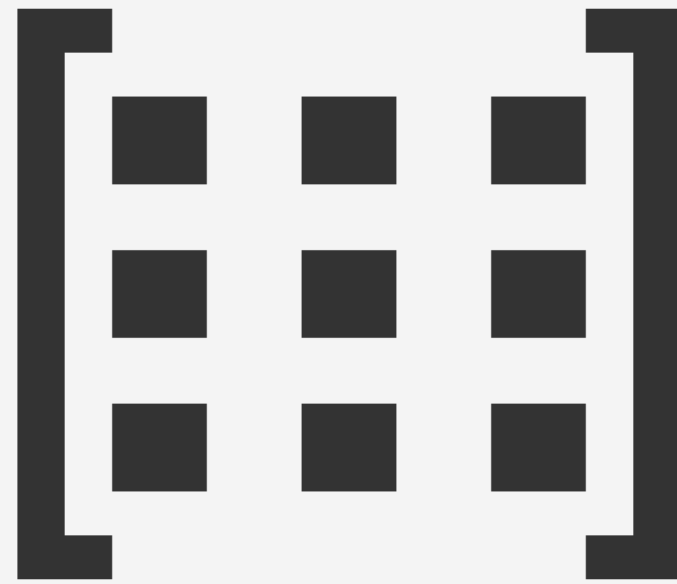
- Conversion en minuscule
- Suppression de mot sans valeur ('heu', 'euh', 'hum', 'bah')

3 Mot les plus importants :

- Medecin : [accord, madame, ça]
- Patient : [non, oui, là]

3061 phrases :

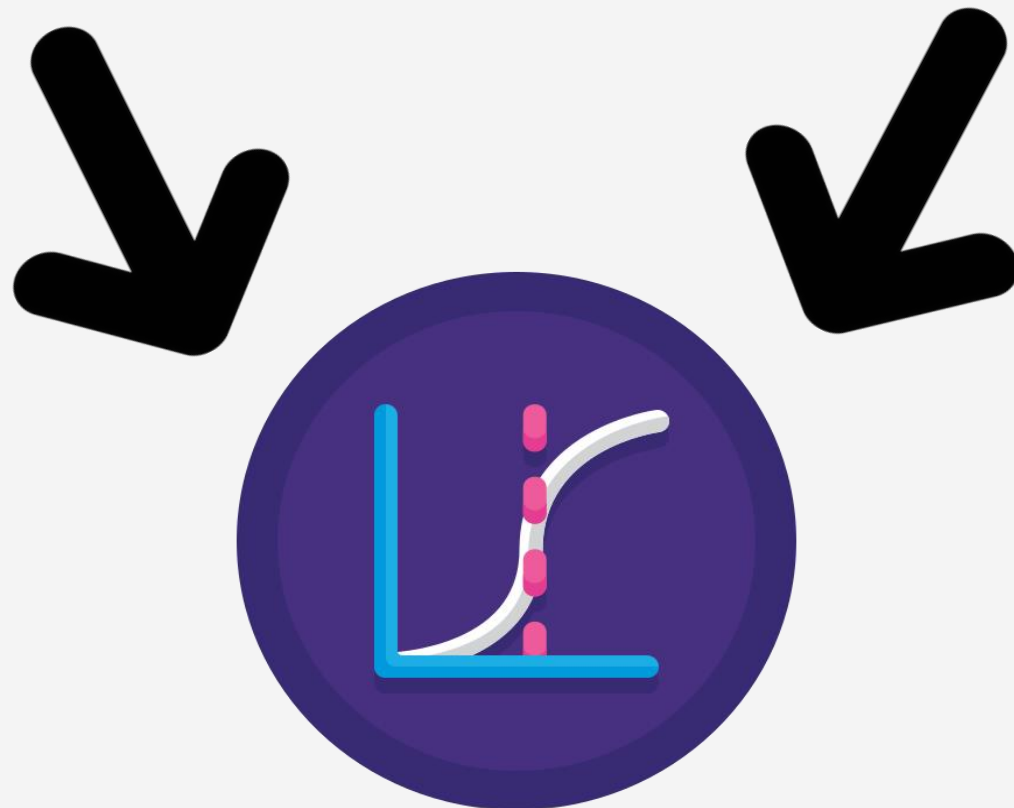
- Médecin : 1594 phrases
- Patient : 1467 phrases



Matrice TF-IDF



Importation de modele de sklearn



Modèle méthode Classique Régression logistique

## - Feature Selection :

Réduire le nombre de caractéristiques en ne gardant que les plus pertinentes.

## - Grid Search :

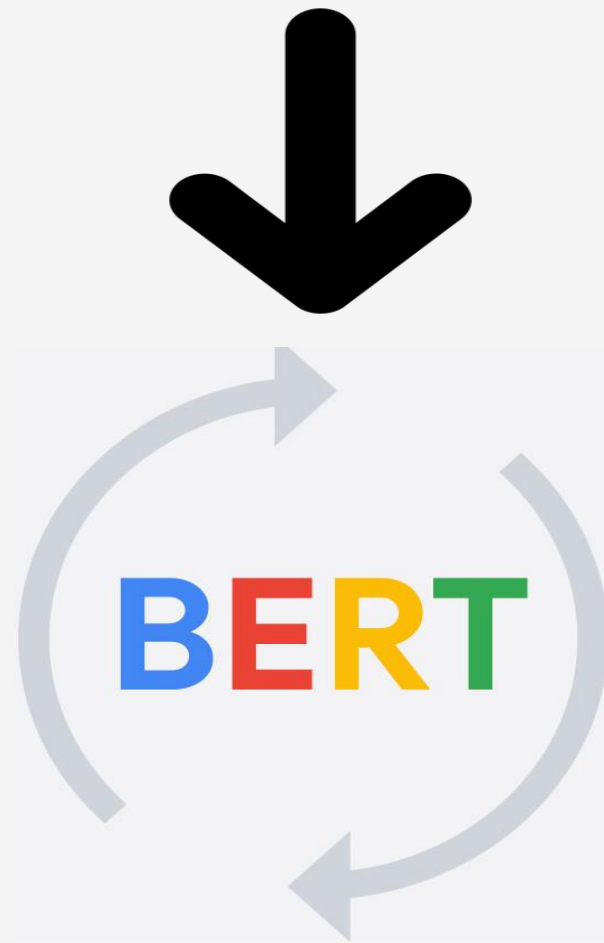
Trouver les meilleurs paramètres pour optimiser les performances du modèle

## - Hyperparameter Tuning

Ajuster les paramètres du modèle pour un équilibre optimal entre précision et généralisation.



Importation de modele Bert (CamemBERT)



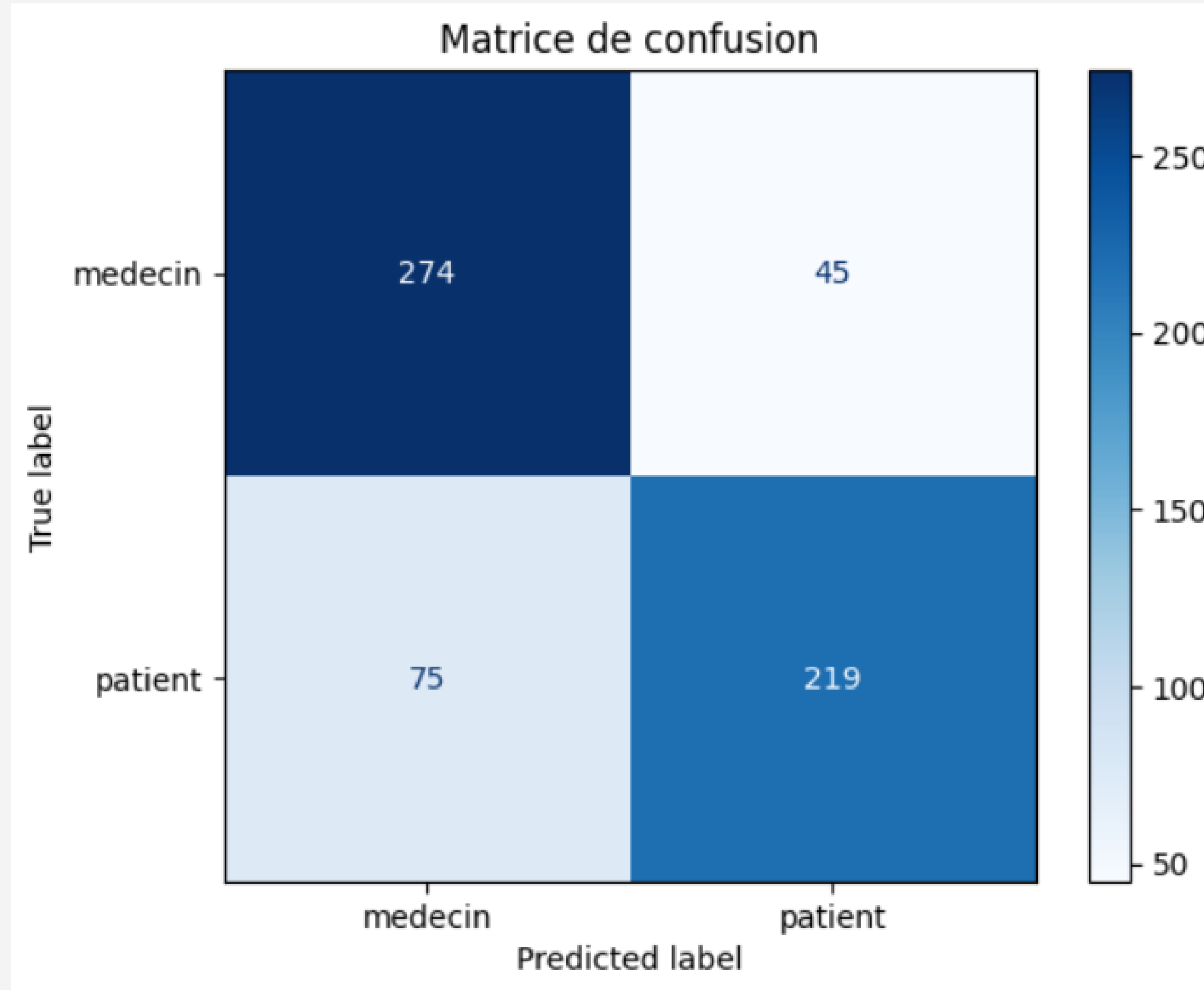
Modele Bert (CamemBERT)

Test de Prompt :

- **Défaut** : la phrase est donnée telle quelle
- **Question** : la phrase est présentée sous forme de question, comme "Est-ce que cette phrase vient d'un médecin ou d'un patient ?«
- **Context** : la phrase est introduite avec un contexte, comme "Dans le contexte d'un appel au SAMU, analyser qui parle."
- **Role** : La tâche est explicitement décrite, par exemple "Identifier le rôle du locuteur (médecin/patient) : [texte]."

Test de température : [0.5, 1.0, 2.0]

# Analyse Résultat de la Régression logistique



**Précisions : 80,42%**

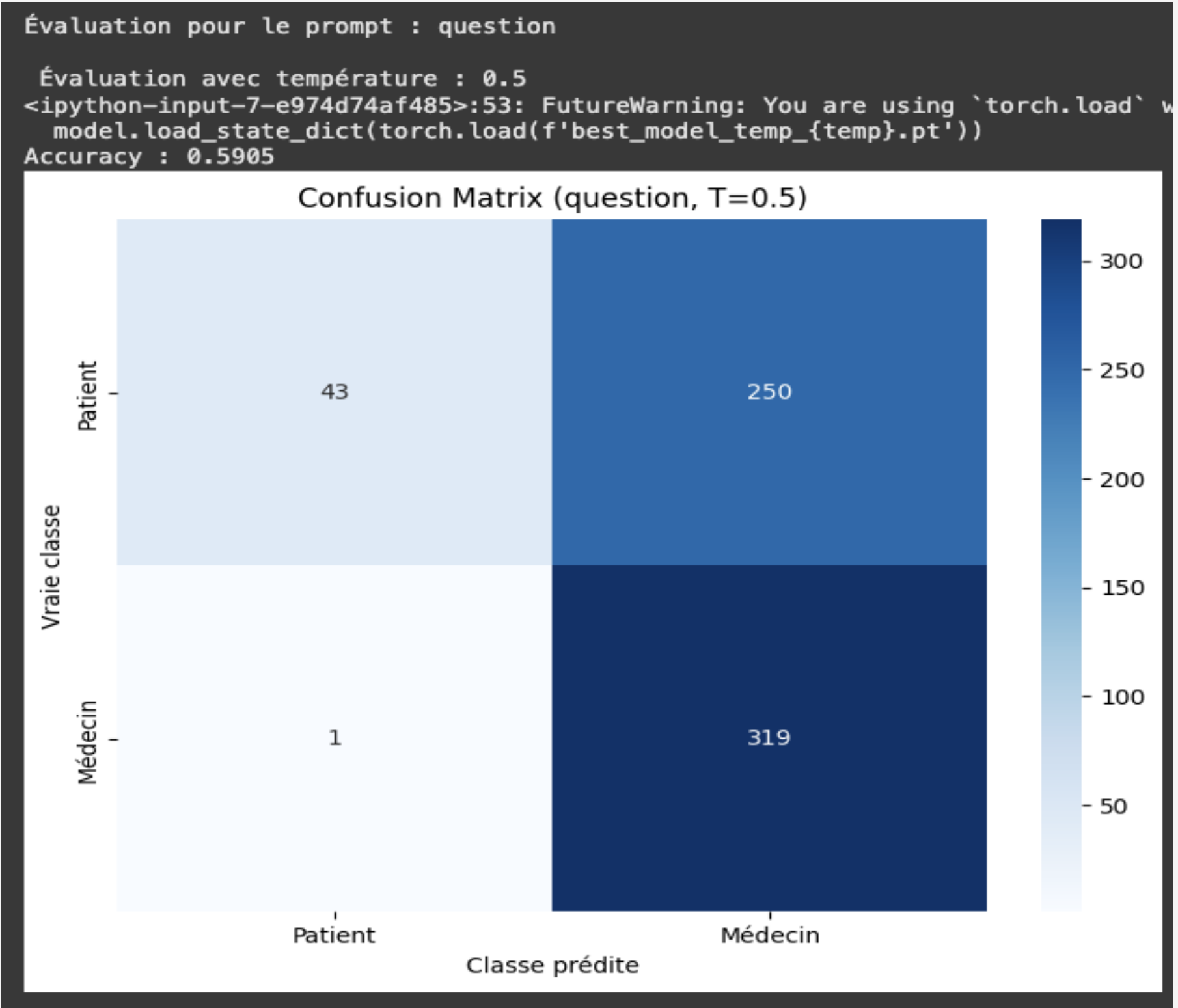
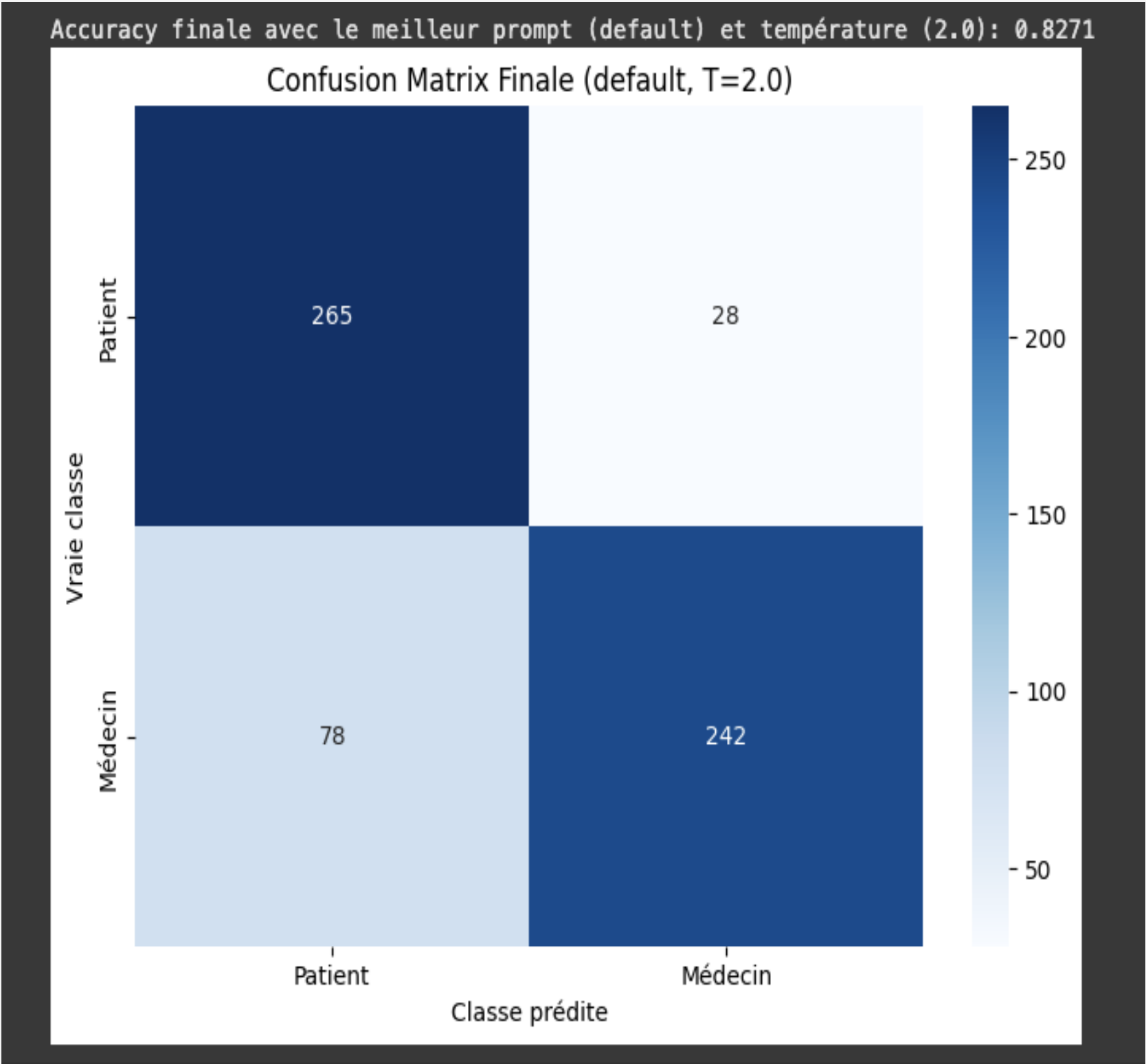
**Medecin (319) : 79%**

**patient (294) : 83%**

**Surclassification en  
médecin**

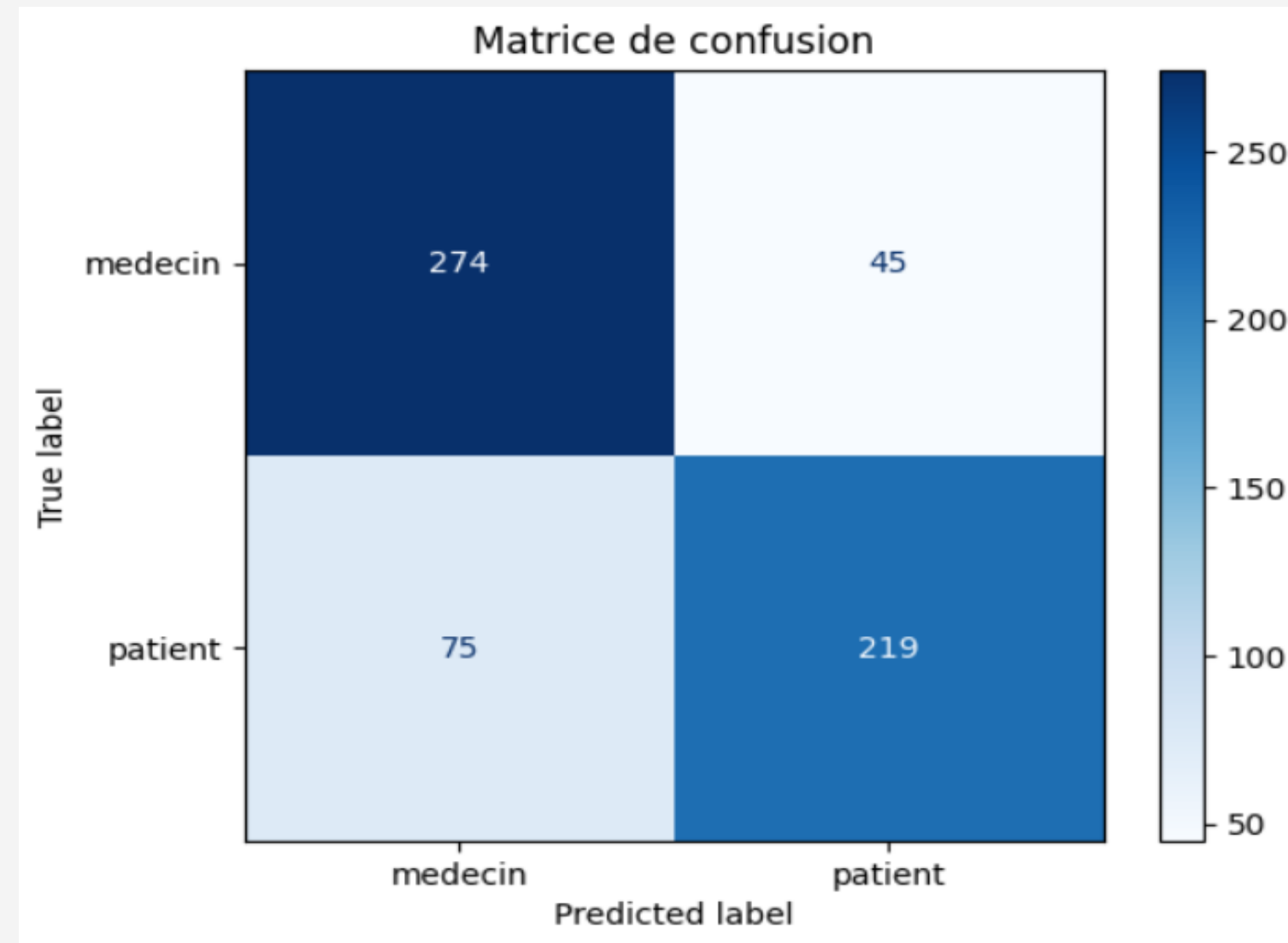


# Analyse Résultat de BERT (CamemBERT)



# Comparaison de nos 2 modèles

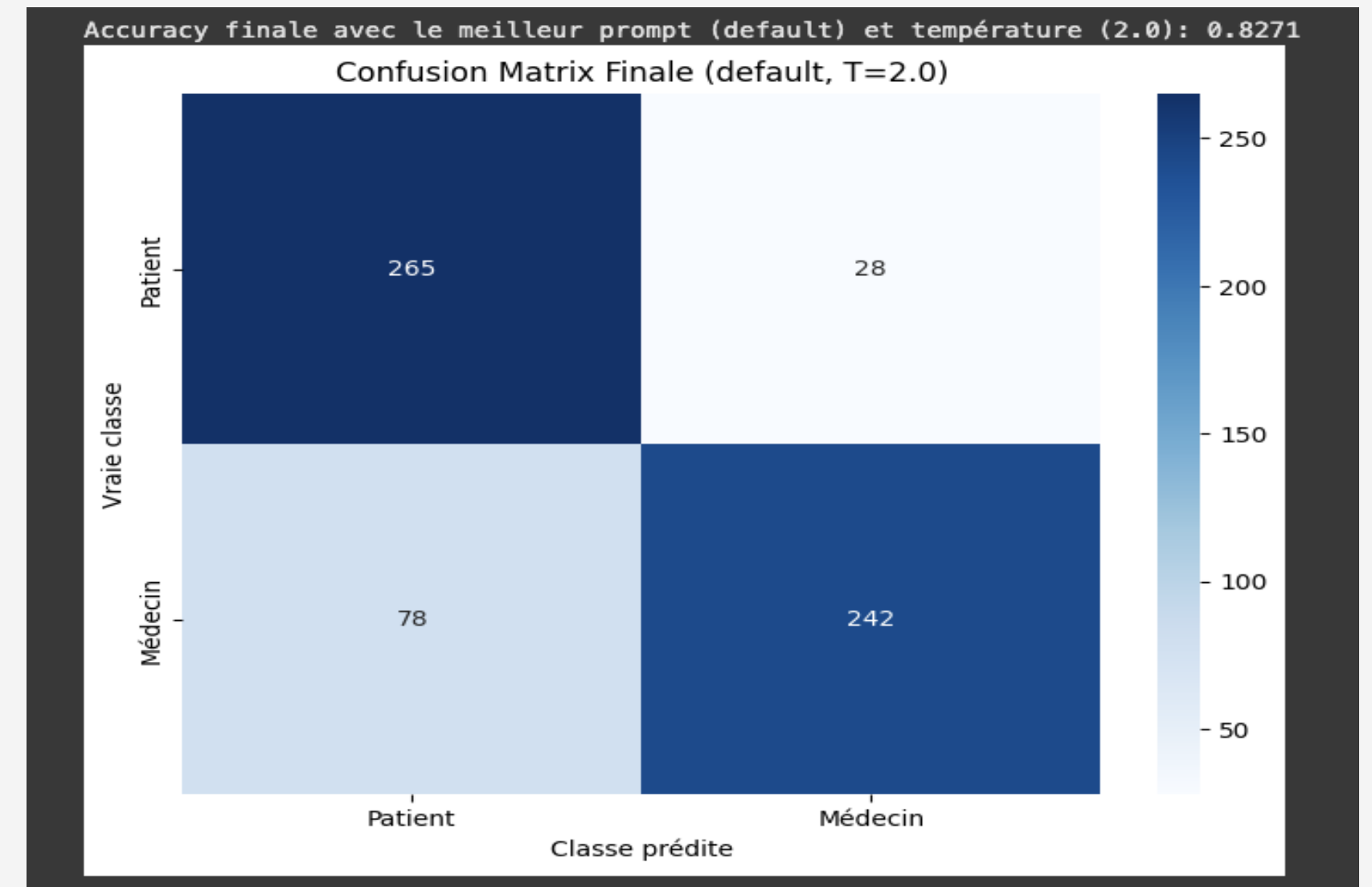
## Régression logistique



**Précisions : 80,42%**

**75 faux négatifs "patient"**

## CamemBERT



**Précisions : 82,71%**

**78 faux négatifs "médecin"**





**- Préparation d'un jeu de données et conception de modele**

**- Augmenté la taille du jeu de données**

**- Entraîné CamemBert sur des textes avec exemple médical**

**- Ajuster l'entraînement pour plus d'attention aux phrase des patients**

**- Ajout de sous-classes pour l'interlocuteur "patient"**