

Medical FAQ Chatbot



Sommaire

- Sujet
- Présentation dataset
- Processing
- Model
- Demo




DATASET

- Lien : Kaggle Diseases and Symptoms Dataset (<https://www.kaggle.com/datasets/sanskrutik/diseases-and-symptoms>)
- Description : Cet ensemble de données contient des transcriptions médicales, qui peuvent inclure des informations sur les maladies, les symptômes et les traitements. Bien que ce ne soit pas spécifiquement un ensemble de données axé sur les maladies et les symptômes, il peut contenir des informations utiles pour l'analyse de ces aspects.

Détails sur l'ensemble de données

Code	Name	Symptoms	Treatments
1	Panic disorder	Palpitations, Sweating, Trembling, Shortness o...	Antidepressant medications, Cognitive Behavior...
2	Vocal cord polyp	Hoarseness, Vocal Changes, Vocal Fatigue	Voice Rest, Speech Therapy, Surgical Removal
3	Turner syndrome	Short stature, Gonadal dysgenesis, Webbed neck...	Growth hormone therapy, Estrogen replacement t...
4	Cryptorchidism	Absence or undescended testicle(s), empty scro...	Observation and monitoring (in cases of mild o...
5	Ethylene glycol poisoning-1	Nausea, vomiting, abdominal pain, General mala...	Supportive Measures, Gastric Decontamination, ...
6	Ethylene glycol poisoning-2	Metabolic acidosis, apid breathing, rapid hear...	Blood tests, Supportive Measures, Gastric Deco...
7	Ethylene glycol poisoning-3	Decreased urine output, swelling in the legs a...	Supportive Measures, Gastric Decontamination, ...
8	Atrophic vaginitis	Vaginal dryness, Vaginal burning, frequent uri...	Vaginal moisturizers, Vaginal estrogen therapy...
9	Fracture	Pain, Swelling, Bruising, Deformity, Difficult...	Immobilization, Surgery, Rehabilitation
10	Cellulitis	Redness, Pain, tenderness, Swelling, Skin chan...	Antibiotics, Warm compresses, immobilization, ...

The background of the slide is a dense field of dark, 3D-rendered dollar signs (\$). In the center of the image, there is a single, prominent orange question mark (?).

problème
sur
données

Tokenisation

- **Diapositive 1 : Introduction**

- Contexte : Expliquez brièvement pourquoi le prétraitement du texte est nécessaire dans le traitement des données textuelles.
- Objectif : Présentez l'objectif de votre code de prétraitement.

- **Diapositive 2 : Fonction de prétraitement**

- Code : Affichez le code de la fonction `preprocess_text()` que vous avez écrit.
- Description : Expliquez chaque étape du prétraitement, notamment la tokenisation, la suppression des stopwords et la lemmatisation.

- **Diapositive 3 : Application de la fonction**

- Code : Montrez comment vous appliquez la fonction de prétraitement aux données.
- Explication : Expliquez comment vous utilisez la fonction `apply()` pour appliquer le prétraitement à vos données.

- **Diapositive 4 : Tokenisation**

- Code : Affichez le code de tokenisation utilisant `Tokenizer` et `text_to_sequences()`.
- Détails : Expliquez le processus de tokenisation et comment il convertit les mots en séquences numériques.

- **Diapositive 5 : Remarques finales**

- Récapitulation : Résumez les principales étapes de votre prétraitement du texte.
- Résultats attendus : Évoquez brièvement ce à quoi ressembleraient les données après le prétraitement.

RNN : Entraînement du modèle

```
model.add(Dense(400, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\patel\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\patel\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]   C:\Users\patel\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
WARNING:tensorflow:From C:\Users\patel\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packa
WARNING:tensorflow:From C:\Users\patel\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packa
```

```
Epoch 1/10
WARNING:tensorflow:From C:\Users\patel\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalC
WARNING:tensorflow:From C:\Users\patel\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalC

10/10 - 13s - loss: 5.9947 - accuracy: 0.0000e+00 - val_loss: 6.0004 - val_accuracy: 0.0000e+00 - 13s/epoch - 1s/step
Epoch 2/10
10/10 - 4s - loss: 5.9800 - accuracy: 0.0250 - val_loss: 6.0166 - val_accuracy: 0.0000e+00 - 4s/epoch - 428ms/step
Epoch 3/10
10/10 - 4s - loss: 5.9681 - accuracy: 0.0688 - val_loss: 6.4464 - val_accuracy: 0.0000e+00 - 4s/epoch - 427ms/step
Epoch 4/10
10/10 - 4s - loss: 5.9088 - accuracy: 0.0063 - val_loss: 6.1176 - val_accuracy: 0.0000e+00 - 4s/epoch - 404ms/step
Epoch 5/10
10/10 - 4s - loss: 5.8780 - accuracy: 0.0125 - val_loss: 7.0172 - val_accuracy: 0.0000e+00 - 4s/epoch - 410ms/step
Epoch 6/10
10/10 - 4s - loss: 5.8077 - accuracy: 0.0063 - val_loss: 6.8610 - val_accuracy: 0.0000e+00 - 4s/epoch - 417ms/step
Epoch 7/10
10/10 - 5s - loss: 5.7282 - accuracy: 0.0094 - val_loss: 7.7180 - val_accuracy: 0.0000e+00 - 5s/epoch - 488ms/step
Epoch 8/10
10/10 - 5s - loss: 5.5704 - accuracy: 0.0094 - val_loss: 8.6622 - val_accuracy: 0.0000e+00 - 5s/epoch - 528ms/step
Epoch 9/10
10/10 - 7s - loss: 5.4072 - accuracy: 0.0094 - val_loss: 8.9091 - val_accuracy: 0.0000e+00 - 7s/epoch - 722ms/step
Epoch 10/10
10/10 - 6s - loss: 5.2188 - accuracy: 0.0063 - val_loss: 9.8046 - val_accuracy: 0.0000e+00 - 6s/epoch - 600ms/step
```


DEMO

