

Lucas Artaud
Iswarya Sivasubramaniam

DIA1

Analyse des avis clients en assurance



Sommaire

1. **Contexte et problématique**
2. **Import de données**
3. **Nettoyage de la donnée**
 - Mise en évidence des mots (et n-grammes) fréquents
 - Correction orthographique
4. **Résumé et traduction**
 - Résumé
 - Traduction
5. **Détection des sentiments**
 - Prédiction positive ou négative
 - Prédiction du nombre d'étoiles
6. **Embeddings pour identifier des mots similaires et enrichir liste des thèmes**
 - Vectorisation des avis
 - Entraînement Word2Vec
 - Visualisation des embeddings avec Matplotlib
 - Implémentation de la distance euclidienne et de la similarité cosinus
 - Réponse aux questions avec recherche sémantique
7. **Apprentissage supervisé**
 - TF-IDF
 - Modèle basique avec une couche d'embedding
 - USE (Universal Sentence Embedding)
8. **Création de l'application Streamlit**
 - Résumé
 - Prédiction des étoiles
 - Recherche d'informations

Lien GitHub

[https://github.com/lucasartaud/Machine Learning for NLP Project2](https://github.com/lucasartaud/Machine_Learning_for_NLP_Project2)

Lien Streamlit

<https://artaud-sivasubramaniam-dia1-nlp-project2.streamlit.app/>

Contexte et problématique

Des enquêtes de satisfaction, sur des assurances, ont été réalisées auprès des clients.

Notre problématique est de prédire le nombre d'étoiles (de 1 à 5) des avis sur des contrats d'assurance.

Import de données

Pour la réalisation de ce projet nous avons pris le fichier CSV fournis avec avis sur des assurances.

La donnée importée comporte 34435 lignes, répartie en 36 fichiers CSV avec 11 colonnes.

Le fichier initial comportait ces colonnes:

- Note
- Auteur
- Avis
- Assureur
- Produit
- Type
- Date_publication
- Date_exp
- Avis_en
- Avis_cor
- Avis_cor_en

Nettoyage de la donnée

Nettoyage Initial

- Dans un premier temps, les colonnes 'type', 'avis_en', 'avis_cor' et 'avis_cor_en' ont été enlevées de notre base de données.
- Nous avons enlevé les 10 331 notes manquantes de notre dataframe.

Mise en évidence des mots (et n- grammes) fréquents

- En premier lieu, nous avons utilisé `word_tokenize` afin de diviser le texte en mots individuels. Nous avons enlevé toute la ponctuation de la liste.
- À partir des tokens, les 10 mots les plus fréquents sont des mots courants.
- À partir des tokens, nous avons aussi déterminé les 10 bigrammes les plus fréquents.

10 most frequent words:

```
[('de', 58163),  
 ('et', 34910),  
 ('je', 31451),  
 ('la', 25634),  
 ('le', 25032),  
 ('à', 23218),  
 ('pour', 20070),  
 ('que', 19123),  
 ('en', 19095),  
 ('les', 18985)]
```

10 most frequent bigrams:

```
[(('je', 'suis'), 8159),  
 (('de', 'la'), 5231),  
 (('que', 'je'), 3572),  
 (('je', 'ne'), 2955),  
 (('à', 'la'), 2568),  
 (('de', 'mon'), 2546),  
 (('pas', 'de'), 2305),  
 (('m', 'a'), 2171),  
 (('du', 'service'), 2145),  
 (('suis', 'satisfait'), 2065)]
```

Correction orthographique

- La classe 'LanguageTool' de la bibliothèque 'language_tool_python' a permis d'effectuer une vérification orthographique et grammaticale des avis.
- Les avis vérifiés ont été ajoutés dans une nouvelle colonne nommée 'avis_corriges' :

note	auteur	avis	assureur	produit	date_publication	date_exp	avis_corriges
0	5	brahim--k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021 01/09/2021	Meilleures assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractifs et services de qualité et rapidité
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet ,impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021 01/05/2021	Je suis globalement satisfait, sauf que vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé, interlocutrice compétente et très agréable
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boitier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021 01/03/2021	Prix très abordable plusieurs options s'offrent à nous comme le boîtier connecter à la voiture, l'option tranquillité et zéro franchise ce qui est tout à fait plaisant
3	4	boulain-f-116580	je satisfait du service, une réponse très rapide de votre service je vous en remercie, vous êtes une assurance la moins cher sur le marche, Cordialement.	L'olivier Assurance	auto	10/06/2021 01/06/2021	Je satisfais du service, une réponse très rapide de votre service je vous en remercie, vous êtes une assurance le moins cher sur la marche, Cordialement.
4	1	ouaille31-51798	Client depuis plus de 25 ans, très déçu de cette "mutuelle" qui n'a plus rien d'une Mutuelle, la recherche du profit immédiat est devenu leur priorité. à l'agence on ne sait que essayer de vous fourguer des contrats inutiles, on vous fait payer une protection juridique sur chaque contrat mais vo...	Matmut	auto	29/01/2017 01/01/2017	Client depuis plus de 25 ans, très déçu de cette "mutuelle" qui n'a plus rien d'une Mutuelle, la recherche du profit immédiat est devenu leur priorité. À l'agence, on ne sait qu'essayer de vous fourguer des contrats inutiles, on vous fait payer une protection juridique sur chaque contrat, mais v...

Résumé et traduction

Résumé

- La bibliothèque Hugging Face Transformers a été utilisée pour créer un pipeline de résumé automatique en utilisant le modèle t5.
- Les résumés ont été rajoutés dans une nouvelle colonne nommée 'résumé' :

	note	auteur	avis	assureur	produit	date_publication	date_exp	avis_corriges	résumé
0	5	brahim-k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021	01/09/2021	Meilleures assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractifs et services de qualité et rapidité	Je recommande cette compagnie pour vous des prix attractifs et des services de qualité. Vous pouvez obtenir des assurances, des prix, des solutions, une écoute, une rapidité, des services rapides
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet ,impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021	01/05/2021	Je suis globalement satisfait, sauf que vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé, interlocutrice compétente et très agréable	Je suis globalement satisfait, sauf si vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boîtier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021	01/03/2021	Prix très abordable plusieurs options s'offrent à nous comme le boîtier connecter à la voiture, l'option tranquillité et zéro franchise ce qui est tout à fait plaisant	Le boîtier connecté à la voiture, l'option tranquillité et la franchise zéro sont parmi les options qui s'offrent à nous. Le prix est très abordable
3	4	boulain-f-116580	je satisfait du service, une réponse très rapide de votre service .je vous en remercie, vous êtes une assurance la moins cher sur le marche, Cordialement.	L'olivier Assurance	auto	10/06/2021	01/06/2021	Je satisfais du service, une réponse très rapide de votre service .je vous en remercie, vous êtes une assurance le moins cher sur la marche, Cordialement.	Je vous en remercie, vous êtes une assurance le moins cher sur la marche, cordialement. Je vous remercie de votre réponse très rapide de votre service
4	1	ouaille31-51798	Client depuis plus de 25 ans, très déçu de cette "mutuelle" qui n'a plus rien d'une Mutuelle, la recherche du profit immédiat est devenu leur priorité. à l'agence on ne sait que essayer de vous fourguer des contrats inutiles, on vous fait payer une protection juridique sur chaque contrat mais vo...	Matmut	auto	29/01/2017	01/01/2017	Client depuis plus de 25 ans, très déçu de cette "mutuelle" qui n'a plus rien d'une Mutuelle, la recherche du profit immédiat est devenu leur priorité. À l'agence, on ne sait qu'essayer de vous fourguer des contrats inutiles, on vous fait payer une protection juridique sur chaque contrat, mais v...	La recherche du profit immédiat est devenu la priorité de tous les assureurs depuis plus de 25 ans. Vous ne serez pas défendu qu'une fois et vous n'aurez plus qu'un seul mot à dire

Traduction

- Nous utilisons également la bibliothèque Hugging Face Transformers pour créer un pipeline de traduction automatique du français vers l'anglais.
- Les traductions ont été rajoutées dans une nouvelle colonne nommée 'traduction' :

note	auteur	avis	assureur	produit	date_publication	date_exp	avis_corriges	résumé	traduction	
0	5	brahim--k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021	01/09/2021	Meilleures assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractifs et services de qualité et rapidité	Je recommande cette compagnie pour vous des prix attractifs et des services de qualité. Vous pouvez obtenir des assurances, des prix, des solutions, une écoute, une rapidité, des services rapides.	Best insurance, prices, solutions, listening, speed, and I recommend this company for you Attractive prices and services of quality and speed
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet ,impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021	01/05/2021	Je suis globalement satisfait, sauf que vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé, interlocutrice compétente et très agréable	Je suis globalement satisfait, sauf si vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé.	I am generally satisfied, except that you have a problem with your website, impossible to report a disaster online after several attempts made by phone or everything went very well, competent interlocutor and very pleasant
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boitier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021	01/03/2021	Prix très abordable plusieurs options s'offrent à nous comme le boitier connecter à la voiture, l'option tranquillité et zéro franchise ce qui est tout à fait plaisant	Le boîtier connecté à la voiture, l'option tranquillité et la franchise zéro sont parmi les options qui s'offrent à nous. Le prix est très abordable.	Very affordable price several options are available to us like the case connect to the car, the option tranquility and zero franchise what is quite pleasant

Détection des sentiments

Prédiction positive ou négative

- Nous avons utilisé le modèle par défaut de la bibliothèque Transformers afin de créer un pipeline de prédiction de sentiment.
- Pour chaque avis, nous retournons si celui-ci est positif ou négatif dans une nouvelle colonne nommée 'pos_neg_pred' :

	note	auteur	avis	assureur	produit	date_publication	date_exp	avis_corriges	résumé	traduction	pos_neg_pred
0	5	brahim--k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021	01/09/2021	Meilleures assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractifs et services de qualité et rapidité	Je recommande cette compagnie pour vous des prix attractifs et des services de qualité. Vous pouvez obtenir des assurances, des prix, des solutions, une écoute, une rapidité, des services rapides.	Best insurance, prices, solutions, listening, speed, and I recommend this company for you Attractive prices and services of quality and speed	POSITIVE
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet ,impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021	01/05/2021	Je suis globalement satisfait, sauf que vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé, interlocutrice compétente et très agréable	Je suis globalement satisfait, sauf si vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé.	I am generally satisfied, except that you have a problem with your website, impossible to report a disaster online after several attempts made by phone or everything went very well, competent interlocutor and very pleasant	NEGATIVE
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boîtier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021	01/03/2021	Prix très abordable plusieurs options s'offrent à nous comme le boîtier connecter à la voiture, l'option tranquillité et zéro franchise ce qui est tout à fait plaisant	Le boîtier connecté à la voiture, l'option tranquillité et la franchise zéro sont parmi les options qui s'offrent à nous. Le prix est très abordable.	Very affordable price several options are available to us like the case connect to the car, the option tranquility and zero franchise what is quite pleasant	NEGATIVE

Prédiction du nombre d'étoiles

- Afin d'effectuer la prédiction de sentiment avec le nombre d'étoiles, nous utilisons le modèle 'BERT base multilingual uncased sentiment'.
- Pour chaque avis, nous stockons les résultats dans la colonne 'stars_pred' :

	note	auteur	avis	assureur	produit	date_publication	date_exp	avis_corriges	résumé	traduction	pos_neg_pred	stars_pred
0	5	brahim--k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021	01/09/2021	Meilleures assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractifs et services de qualité et rapidité	Je recommande cette compagnie pour vous des prix attractifs et des services de qualité. Vous pouvez obtenir des assurances, des prix, des solutions, une écoute, une rapidité, des services rapides.	Best insurance, prices, solutions, listening, speed, and I recommend this company for you Attractive prices and services of quality and speed	POSITIVE	5
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet , impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021	01/05/2021	Je suis globalement satisfait, sauf que vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé, interlocutrice compétente et très agréable	Je suis globalement satisfait, sauf si vous avez un problème avec votre site internet, impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé.	I am generally satisfied, except that you have a problem with your website, impossible to report a disaster online after several attempts made by phone or everything went very well, competent interlocutor and very pleasant	NEGATIVE	4
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boitier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021	01/03/2021	Prix très abordable plusieurs options s'offrent à nous comme le boitier connecter à la voiture, l'option tranquillité et zéro franchise ce qui est tout à fait plaisant	Le boîtier connecté à la voiture, l'option tranquillité et la franchise zéro sont parmi les options qui s'offrent à nous. Le prix est très abordable.	Very affordable price several options are available to us like the case connect to the car, the option tranquility and zero franchise what is quite pleasant	NEGATIVE	4

Embeddings pour identifier
des mots similaires et
enrichir liste des thèmes

Vectorisations des avis

- Le modèle Word2Vec pré-entraîné en français a été chargé depuis Hugging Face et utilisé afin de vectoriser les avis de notre base de données.
- Pour chaque avis, on découpe puis on vectorise chaque mot à l'aide de notre modèle pour ensuite les combiner.
- Nous stockons les résultats dans la colonne 'avis_vectorisé' :

note	auteur	avis	assureur	produit	date_publication	date_exp	avis_vectorisé
0	5	brahim--k-131532	Meilleurs assurances, prix, solutions, écoute, rapidité, et je recommande cette compagnie pour vous Des prix attractif et services de qualité et rapidité	Direct Assurance	auto	06/09/2021 01/09/2021	[0.24681999, 0.16044001, 0.06285334, -0.20820668, -0.047286663, 0.055593334, -0.0342, -0.024773335, -0.012233335, 0.18871333, -0.018540004, -0.022406666, -0.20176001, 0.012426669, 0.124753326, 0.15680668, 0.11167333, 0.14868668, 0.032006666, -0.09978666, -0.15604666, -0.19735332, 0.08191333, -0....
1	4	bernard-g-112497	je suis globalement satisfait , sauf que vous avez un problème avec votre site internet ,impossible de déclarer un sinistre en ligne après plusieurs tentatives déclaration faite par téléphone ou tout c'est très bien passé , interlocutrice compétente et très agréable	Direct Assurance	auto	03/05/2021 01/05/2021	[0.23545946, 0.0059973006, 0.07420811, -0.12728918, -0.04100541, 0.1200162, 0.03873243, -0.012500001, -0.043240547, 0.11099732, 0.047227032, -0.14077027, -0.19022162, 0.0057081087, 0.17322434, 0.14932163, 0.114981085, 0.07671892, 0.024351351, -0.10270002, -0.063910805, -0.27562165, 0.06892161, -...
2	5	virginie-t-107352	Prix tres abordable plusieurs options s'offrent a nous comme le boitier connecter à la voiture, l'option tranquiliter et zero franchise ce qui est tout a fait plaisant	Direct Assurance	auto	21/03/2021 01/03/2021	[0.2317636, 0.08315454, 0.08915909, -0.14952275, -0.053027272, 0.15733637, 0.0069772736, -0.011186368, 0.002445454, 0.12730455, 0.09446363, -0.107200004, -0.14418635, 0.06645909, 0.22583637, 0.12176817, 0.12100908, 0.093022734, 0.039404545, -0.070640914, -0.1143409, -0.24184546, 0.048531808, -0....

Entraînement Word2Vec

- Feature : Avis vectorisé
- Target : Nombre d'étoiles
- Modèle : Régression linéaire

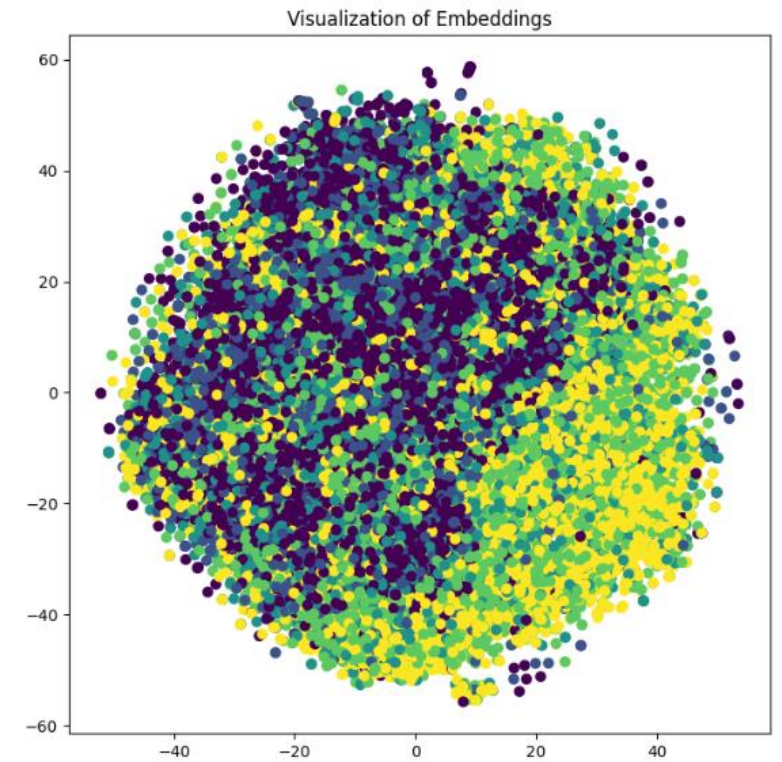
Résultats :

- RMSE = 1,09 ce qui signifie qu'en moyenne, le modèle se trompe d'une étoile
- $R^2 = 0,49$ ce qui signifie que la moitié de la variabilité de la variable prédite est capturée par le modèle

Visualisation des embeddings avec Matplotlib

Nous avons créé une matrice à partir des embeddings de la colonne 'avis_vectorisé'. Puis grâce au t-SNE, nous avons réduit la dimension des embeddings à 2 dimensions.

Voici une représentation 2D des embeddings:



Les couleurs représentent les notes de l'avis. Nous pouvons observer que les points les plus clairs et les plus sombres sont proches. Ce qui signifie qu'il y a une corrélation entre les embeddings et les notes des avis.

Implémentation de la distance euclidienne et de la similarité cosinus

- Nous avons calculé la distance euclidienne et la similarité cosinus entre les 1000 premiers embeddings.
- La valeur moyenne de la distance euclidienne entre les paires d'embeddings est de 41,26. On peut dire qu'ils sont éloignés les uns des autres.
- La valeur moyenne de la similarité cosinus entre les paires d'embeddings est d'environ - 0,000336. Cela signifie de l'alignement n'est pas particulièrement prononcé.
- En conclusion, nous avons une dispersion relativement uniforme des 1000 premiers embeddings dans l'espace.

Réponse aux questions avec recherche sémantique

- Pour chaque question en entrée, nous la vectorisons et calculons la similarité entre la question posée et les avis vectorisés.
- Puis, nous trouvons l'indice de la phrase qui a la similarité cosinus maximal avec la question et nous la retournons.

Apprentissage supervisé

TF-IDF

- Nous utilisons le module 'TfidfVectorizer' afin de calculer la matrice TF-IDF.
- Nous avons ignoré les mots qui apparaissent dans moins de 10% des avis afin de garder uniquement les mots les plus représentatifs.
- Nous avons également ignoré les mots qui apparaissent dans plus de 30% des avis afin d'ignorer les mots trop communs de la langue française qui sont peu intéressants.

	ans	après	au	avec	avoir	bien	car	ce	cette	chez	...	service	si	sinistre	sont	sur	toujours	tout	très	vous	été
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.463420	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.409916	0.000000
1	0.000000	0.353899	0.000000	0.275163	0.000000	0.299268	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.347201	0.000000	0.000000	0.0	0.308504	0.502697	0.278332	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.446994	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.474637	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.589438	0.000000	0.000000	0.000000	0.309236	0.0	0.000000	0.273858	0.606516	0.000000
4	0.125882	0.000000	0.090823	0.000000	0.120062	0.000000	0.000000	0.000000	0.112058	0.000000	...	0.000000	0.000000	0.123646	0.106267	0.101074	0.0	0.000000	0.089511	0.594721	0.000000
...
24098	0.000000	0.000000	0.267761	0.288896	0.000000	0.000000	0.000000	0.000000	0.000000	0.343382	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
24099	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.358224	0.000000	0.000000	...	0.000000	0.000000	0.428090	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000
24100	0.000000	0.000000	0.175249	0.283623	0.000000	0.308469	0.122427	0.199646	0.216224	0.000000	...	0.000000	0.117110	0.000000	0.102525	0.195030	0.0	0.105996	0.000000	0.382519	0.345765
24101	0.123220	0.000000	0.000000	0.095920	0.000000	0.104323	0.000000	0.000000	0.109688	0.000000	...	0.000000	0.118818	0.000000	0.000000	0.000000	0.0	0.107542	0.000000	0.194049	0.350807
24102	0.000000	0.000000	0.216590	0.000000	0.000000	0.254158	0.302615	0.246742	0.267230	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.284887

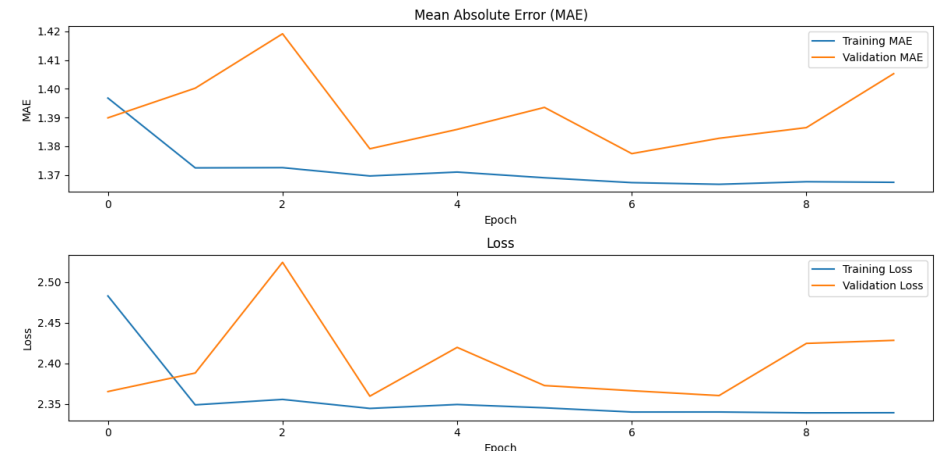
24103 rows × 46 columns

Modèle basique avec une couche d'embedding

Couches du modèle :

1. Embedding Layer : Transforme les données catégorielles en vecteurs d'incrustation.
2. Flatten Layer : Aplatit les vecteurs d'incrustation en un vecteur 1D.
3. Dense Layer (64, 'relu') : Couche dense avec 64 neurones et activation 'relu'.
4. Dense Layer (1, 'linear') : Couche dense avec 1 neurone de sortie et activation 'linear'.

Les métriques MAE (Mean Absolute Error) et loss sont élevées et très volatiles ce qui montre un modèle non satisfaisant.

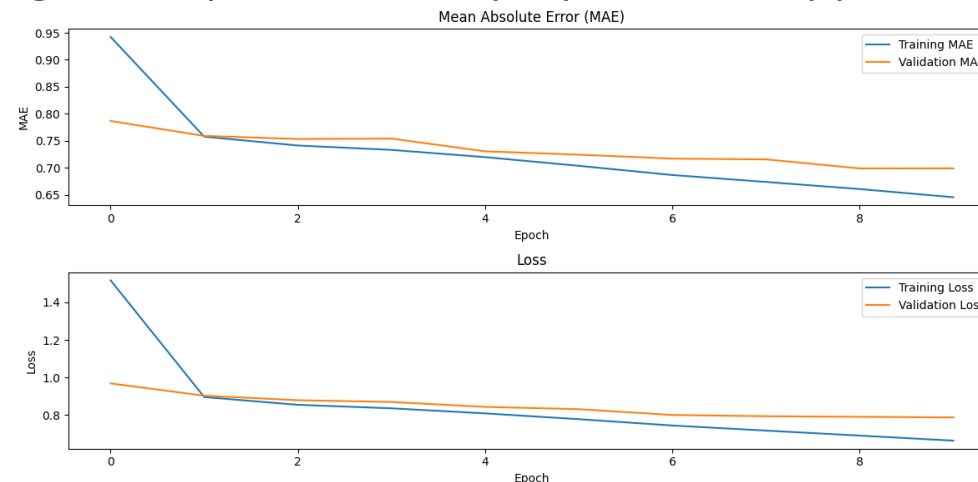


USE (Universal Sentence Embedding)

USE (Universal Sentence Embedding) est un modèle de Google qui permet de convertir du texte (ici les avis) en vecteurs. Ce modèle est pré-entraîné sur un large ensemble de données et fonctionne dans de nombreuses langues.

Par rapport au modèle basique avec une couche d'embedding, la MAE (Mean Absolute Error) est plus faible ce qui signifie que le modèle se trompe moins sur la note des avis.

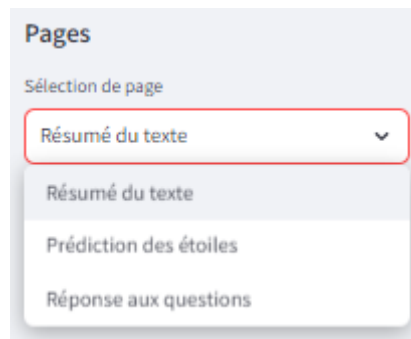
Le Loss est également plus faible ce qui montre que le modèle se trompe moins. La validation loss n'augmente pas donc il n'y a pas de surapprentissage (overfitting).



Création de l'application Streamlit

Présentation

- Nous avons créé une application Streamlit qui contient 3 pages.
- La première page permet de résumer des avis.
- La deuxième page permet de prédire le nombre d'étoiles d'un avis.
- La troisième page permet de rechercher des informations, c'est-à-dire trouver l'avis le plus proche possible de l'entrée de l'utilisateur.



Résumé

Résumé du texte

Entrez votre texte ici :

Un assureur m'a contacté par telephone pour me vendre cette assurance pour mon chiot en me vendant que les jolis détails, j'ai donc souscrit à l'assurance pour etre remboursée à 80%.

Je vais ensuite chez le vétérinaire pour vacciner mon chiot, et on me prescrit une pommade pour des boutons qu'il a, santé vet ne me la rembourse pas car il y a une carrence de 45 jours dont on ne m'avait pas parlé, ce détail passe encore.

1 mois plus tard je fais le rappel du vaccin pour mon chiot (comme pour tous les chiots et achète des vermifuges), j'envoie donc une facture de 118 euros et la on me dit que l'on ne me rembourse pas car je n'ai droit qu'à 50 euros de frais par an (que l'on ne m'avait pas dis non plus sinon je n'aurais jamais souscrit, le but de l'assurance est d'etre remboursé)

J'appelle l'assurance pour résilier et la bien entendu on me dit que je dois attendre 1 an pour me rétracter chose qu'on ne m'avait jamais dite.

Fuyez cette assurance!!

Résumer

Résumé du texte :

Un assureur a contacté par téléphone pour me vendre cette assurance pour mon chiot en vendant que les jolis détails, j'ai donc souscrit à l'assurance pour etre remboursée à 80 %. Je vais ensuite chez le vétérinaire pour vacciner mon Chiot, et on me prescrit une pommade pour des boutons qu'il a, il y a une carrence de 45 jours dont on ne m'a pas parlé, ce détail passe encore. 1 mois plus tard je fais le rappel du vaccin pour mon Chizot (comme pour tous les chits eet achète des vermifuges).

Prédiction du nombre d'étoiles

Prédiction des étoiles

Entrez votre texte ici :

Un assureur m'a contacté par telephone pour me vendre cette assurance pour mon chiot en me vendant que les jolis détails, j'ai donc souscrit à l'assurance pour etre remboursée à 80%.
Je vais ensuite chez le vétérinaire pour vacciner mon chiot, et on me prescrit une pommade pour des boutons qu'il a, santé vet ne me la rembourse pas car il y a une carrence de 45 jours dont on ne m'avait pas parlé, ce détail passe encore.
1 mois plus tard je fais le rappel du vaccin pour mon chiot (comme pour tous les chiots et achète des vermifuges), j'envoie donc une facture de 118 euros et la on me dit que l'on ne me rembourse pas car je n'ai droit qu'à 50 euros de frais par an (que l'on ne m'avait pas dis non plus sinon je n'aurais jamais souscrit, le but de l'assurance est d'etre remboursé)
J'appelle l'assurance pour résilier et la bien entendu on me dit que je dois attendre 1 an pour me rétracter chose qu'on ne m'avait jamais dite.
Fuyez cette assurance!!

Prédire

Nombre d'étoiles prédit :

1

Prédiction des étoiles

Entrez votre texte ici :

Service impeccable, a la hauteur de nos attentes et plus encore.
Prix honnête, difficile de trouver mieux.
Je recommande vivement a toutes et a tous.

Prédire

Nombre d'étoiles prédit :

5

Recherche d'informations

Notre page 'Recherche d'informations' met du temps car un modèle de 6 Go doit être téléchargé pour vectoriser les avis.

Réponse aux questions

Entrez votre texte ici :

Je souhaite une assurance avec des conseillers efficaces et disponibles en agence ou au téléphone.

Soumettre

Réponse trouvée :

Note : 4

Avis : Tres bonne assurance avec des conseillers à l'écoute et disponibles. Je recommande cette assurance. Il n'y a pas de surprises dans les garanties couvertes

Assureur : MAIF

Produit : auto

Date de publication : 20/07/2020