

Projet 8: Réaliser un dashboard et assurer une veille technique

Maodo FALL

OpenClassrooms

Soutenance du projet
31 mars 2025



Sommaire

- 1 Dashboard
- 2 Veille technologique
 - Problématique
 - Etat de l'art du modèle DeBERTa
- 3 Conclusion

1 Dashboard

2 Veille technologique

- Problématique
- Etat de l'art du modèle DeBERTa

3 Conclusion

Société financière



Problématique

Nous avons mis en place, pour le compte de la société financière "Prêt à dépenser", un outil de "scoring crédit" qui :

- calcule la probabilité qu'un client rembourse son crédit
- classifie la demande en crédit accordé ou refusé

L'entreprise décide de développer un dashboard interactif pour que les chargés de relation client puissent expliquer aux clients avec transparence les décisions d'octroi de crédit.

Missions

- Visualiser le score pour chaque client
- Visualiser les principales informations descriptives de chaque clients.
- Comparer, à l'aide de graphiques, les principales informations descriptives d'un client à l'ensemble des clients ou à un groupe de clients
- Couvrir des critères d'accessibilité du WCAG.
- Déployer le dashboard dans le cloud
- Mise à jour et/ou ajout de données clients (optionnel)

Le dashboard a été réalisé avec Streamlit et déployé dans [Streamlit Community Cloud](#) pour un accès public.

1 Dashboard

2 Veille technologique

- Problématique
- Etat de l'art du modèle DeBERTa

3 Conclusion

Problématique

Après ce travail de dashboarding, "Prêt à dépenser", soucieux de mettre en œuvre les dernières techniques en data science sur des problématiques de données texte (NLP) et de données d'images, souhaite faire l'état de l'art d'un outil récent de NLP ou de Computer Vision.

Missions :

- Réaliser un état de l'art sur une technique récente de modélisation de données texte ou de données image, l'analyser, la tester et la comparer à une approche plus classique que j'ai réalisée précédemment

Contexte de l'approche précédente

Sur le site d'une entreprise de l'e-commerce nommée "Place de Marché", des vendeurs proposent des articles à des acheteurs en postant une photo et une description.

L'entreprise souhaitait automatiser la catégorisation des articles avec des approches NLP et Computer Vision.

Dans le traitement et la vectorisation des données texte, j'ai appliqué et comparé :

- une approche Word2Vec
- une approche USE
- une approche BERT

L'idée est de faire l'état de l'art DeBERTa le comparer avec avec l'approche BERT

Place de Marché



Concepts de DeBERTa

DeBERTa (Decoding-enhanced BERT with disentangled attention) :

- Proposée par Microsoft en 2020.
- Améliore BERT grâce à deux innovations :
 - Disentangled Attention : qui sépare contenu et position relative dans le mécanisme d'attention.
 - Enhanced Mask Decoder (EMD) : injection des positions absolues uniquement à la fin, avant le softmax.

Avantages :

- Capture plus précise du contexte
- Meilleure prise en compte de la structure syntaxique
- Prédiction plus fine des tokens masqués

Modélisation avec BERT

Le score ARI de mesure de similarité est de : 0.301

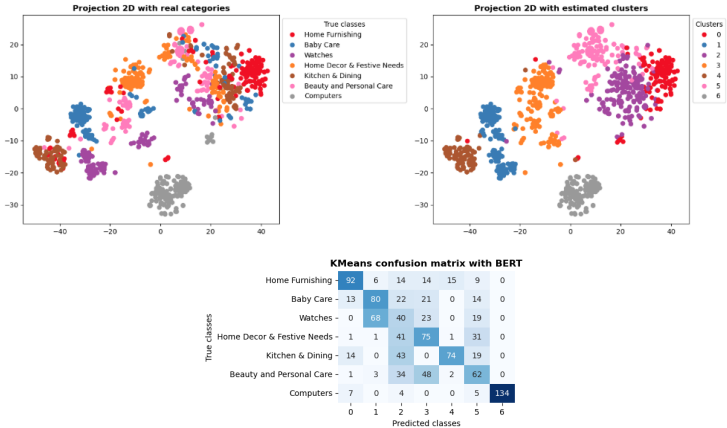


Figure – Résultats avec BERT

DeBERTa-large

Le score ARI de mesure de similarité est de : 0.296

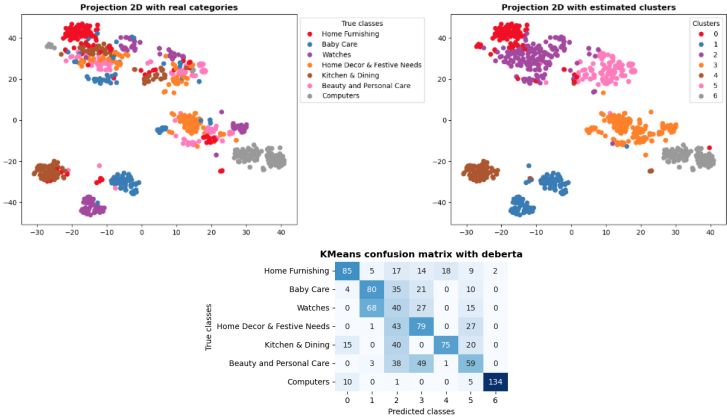


Figure – Résultats avec DeBERTa

DeBERTa-large - max_length optimisé

Le score ARI de mesure de similarité est de : 0.298

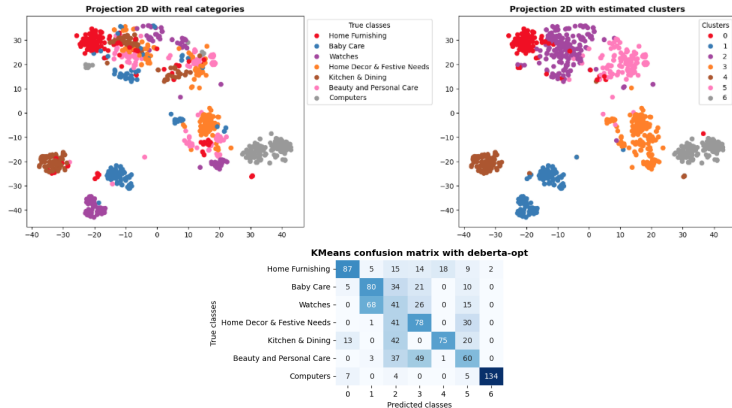


Figure – Résultats de DeBERTa avec optimisation

Analyse comparative

Dans notre cas de figure

- DeBERTa-large n'améliore pas les performances par rapport à BERT.
- BERT obtient un ARI supérieur tout en étant plus léger et plus rapide.
- L'optimisation du max_length n'a pas non plus permis de compenser la baisse de performance vis à vis de BERT.

Feature importance - Modèle proxy

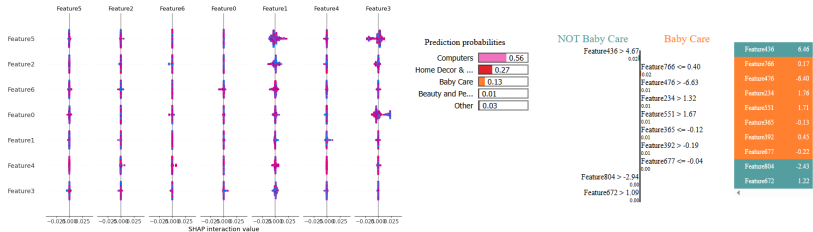


Figure – Globale avec SHAP

Figure – Locale avec LIME

1 Dashboard

2 Veille technologique

- Problématique
- Etat de l'art du modèle DeBERTa

3 Conclusion

Conclusion

Dans cette étude, j'ai

- présenté le dashboard interactif pour l'entreprise "Prêt à dépenser".
- fait l'état de l'art de DeBERTa et comparé ses performances avec BERT.

DeBERTa apporte des innovations intéressantes, notamment le mécanisme d'attention désentrelacée et l'Enhanced Mask Decoder, qui améliorent la représentation du contexte et la prise en compte de la structure syntaxique.

Plusieurs limites ont néanmoins été identifiées avec DeBERTa :

- Temps de calcul élevé
- Absence d'amélioration significative des performances

Pistes d'amélioration :

- Optimisation des hyper-paramètre plus poussée
- Versions plus légères de DeBERTa (ex. DeBERTa-base)
- Enrichir et pré-traiter davantage les données

MERCI BEAUCOUP !