



Résultat du pilote Holi.io x Jedha

Florian Guillot - Directeur en chef du succès client

Julien Donche - Scientifique en chef de la donnée

23 août 2021

Aujourd'hui

3 moments

I. Votre besoin, notre démarche

II. La Démonstration

III. Projection dans le déploiement global

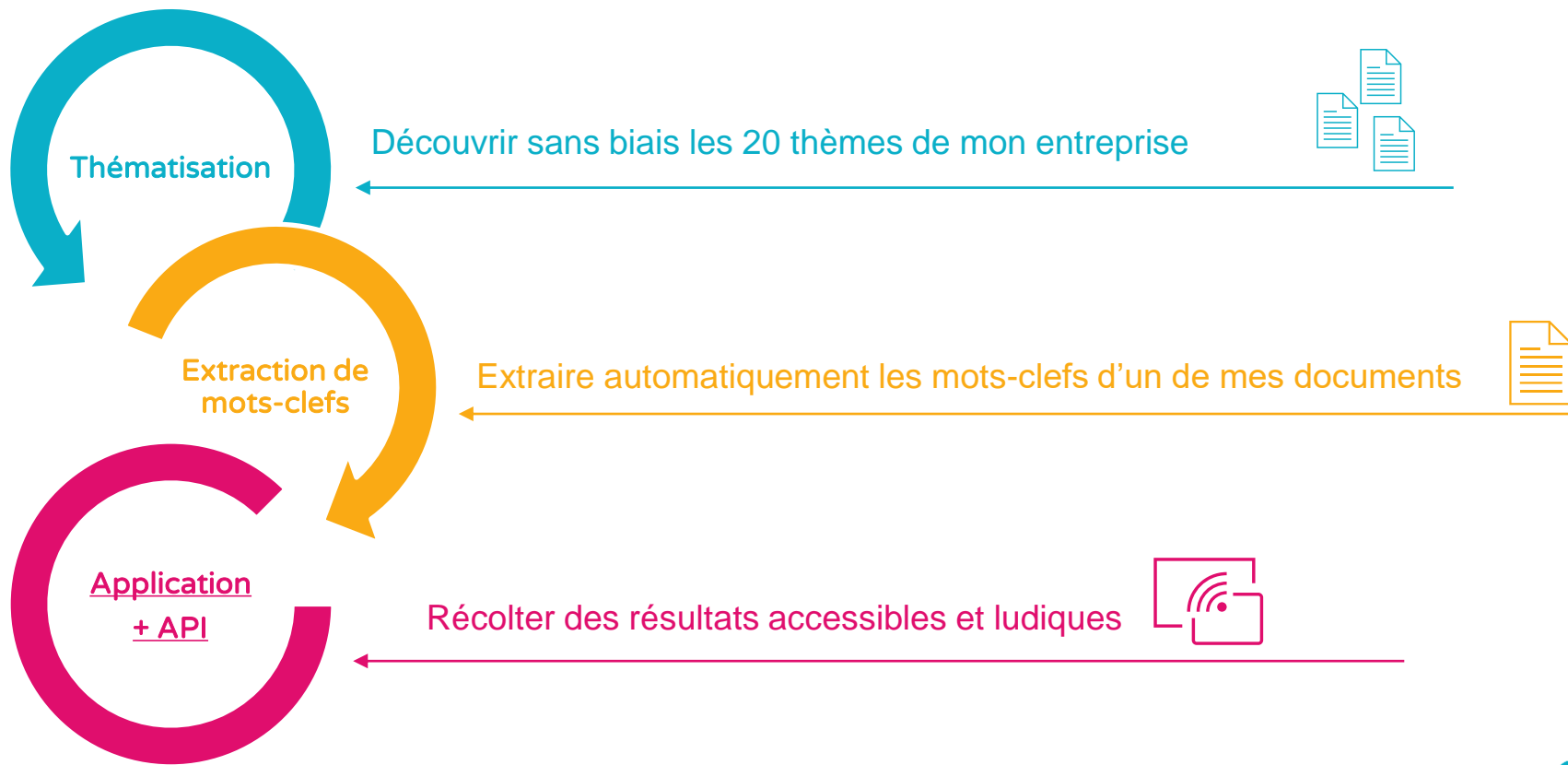
Résultat du pilote Holi.io x Jedha

Florian Guillot - Directeur en chef du succès client

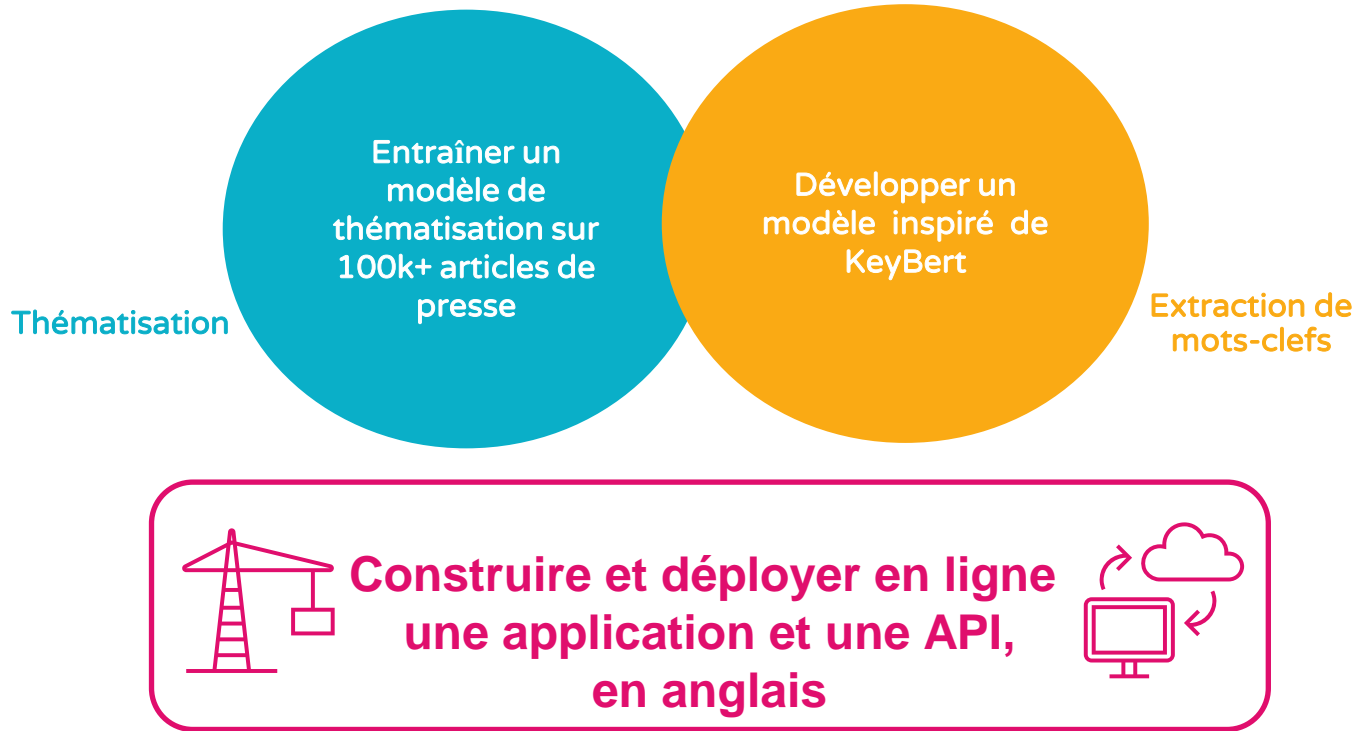
Julien Donche - Scientifique en chef de la donnée

23 août 2021

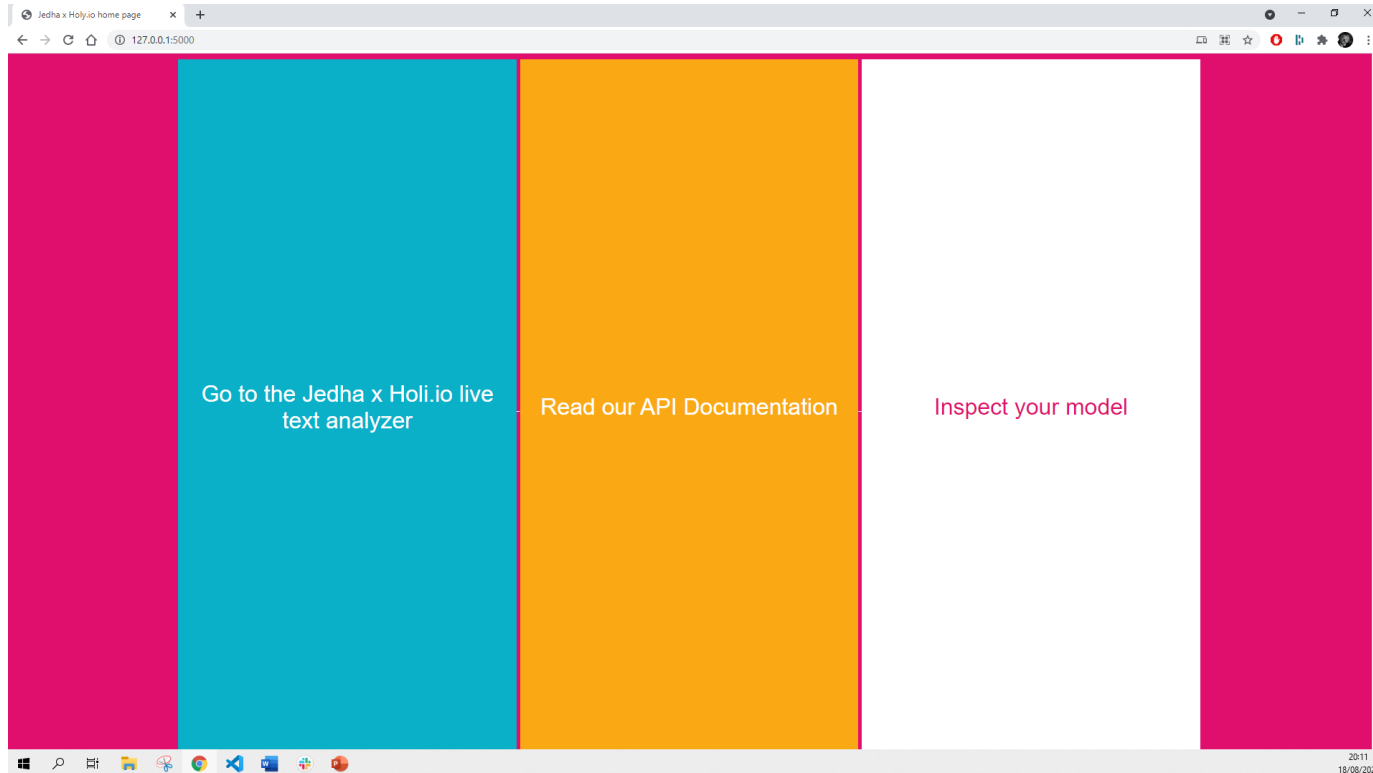
Notre compréhension de votre besoin



Notre démarche lors du pilote



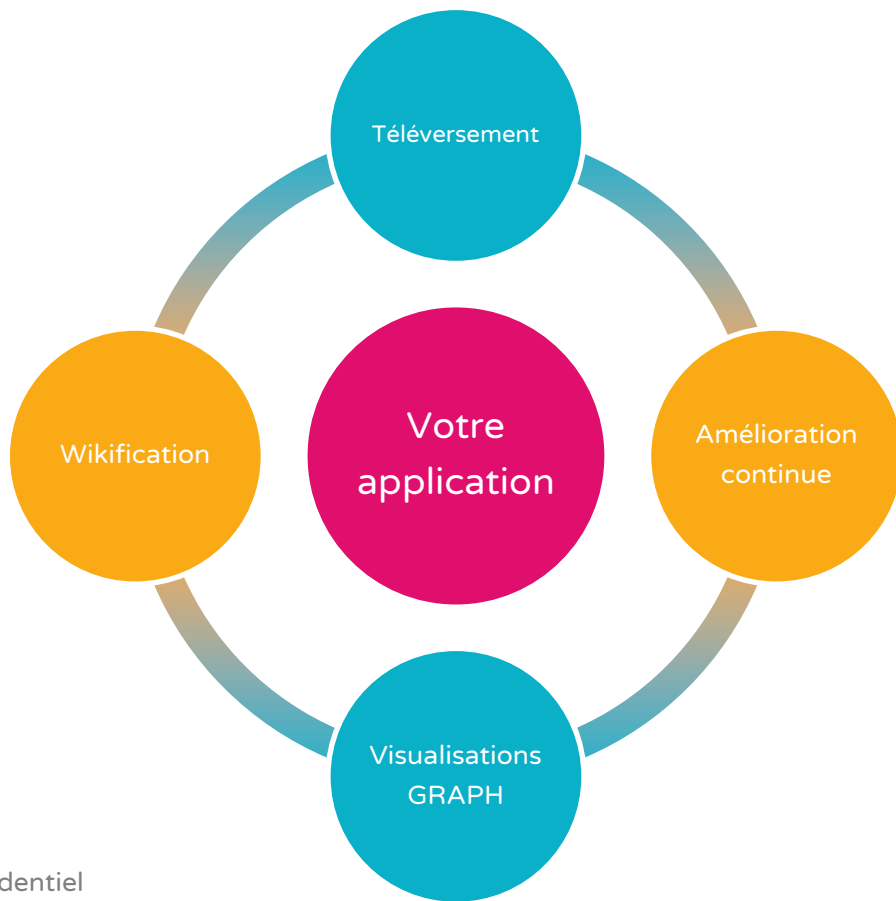
Démonstration



<https://jedha-holi-text-analyzer.herokuapp.com/>



D'un pilote à votre application d'entreprise



✓ Performance optimisée

✓ Modèle personnalisé

✓ Design sur mesure



Nos modèles : LDA & KeyBERT

LDA Latent Dirichlet allocation

Modèle non supervisé génératif probabiliste

« J'assigne au hasard (merci Dirichlet) une probabilité par thème pour chaque mot dans chaque document »

« Je fais l'hypothèse que cette assignation est correcte sauf pour le mot que j'étudie »

« Je calcule pour ce mot la probabilité qu'il soit associé à un thème dans ce document »

« Je répète ce processus pour tous les mots, plusieurs fois, jusqu'à stabilisation »

KeyBERT

BERT

Bidirectional Encoder Representations from Transformers

1. Modèle neuronal pré-entraîné
2. Qui contextualise la phrase
3. En utilisant :

le passé
+
le futur



4. Et transforme chaque mot en vecteur

Similarité Cosinus

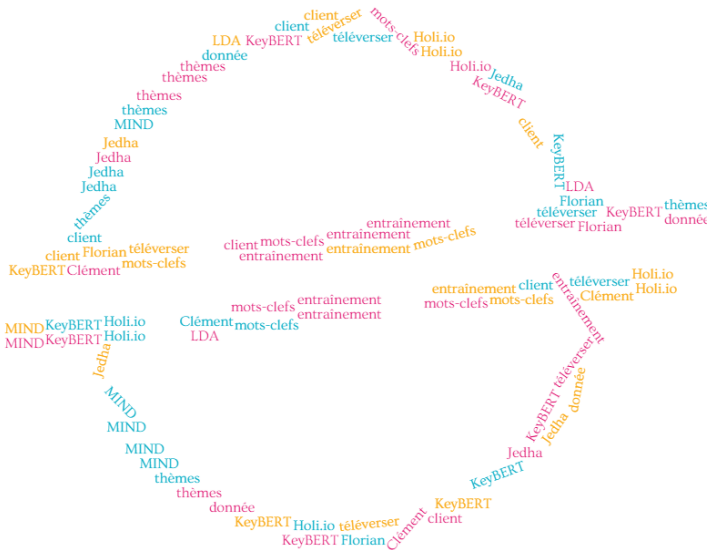
6. Pour trouver les mots les plus importants

'Je'	0,1
'suis'	0,2
'un'	0,01
'joli'	0,5
'hérisson'	0,6

5. Que l'on compare entre eux

→ [-1,2 0,3 -0,4 0,1]





Merci !

FLORIAN GUILLOT



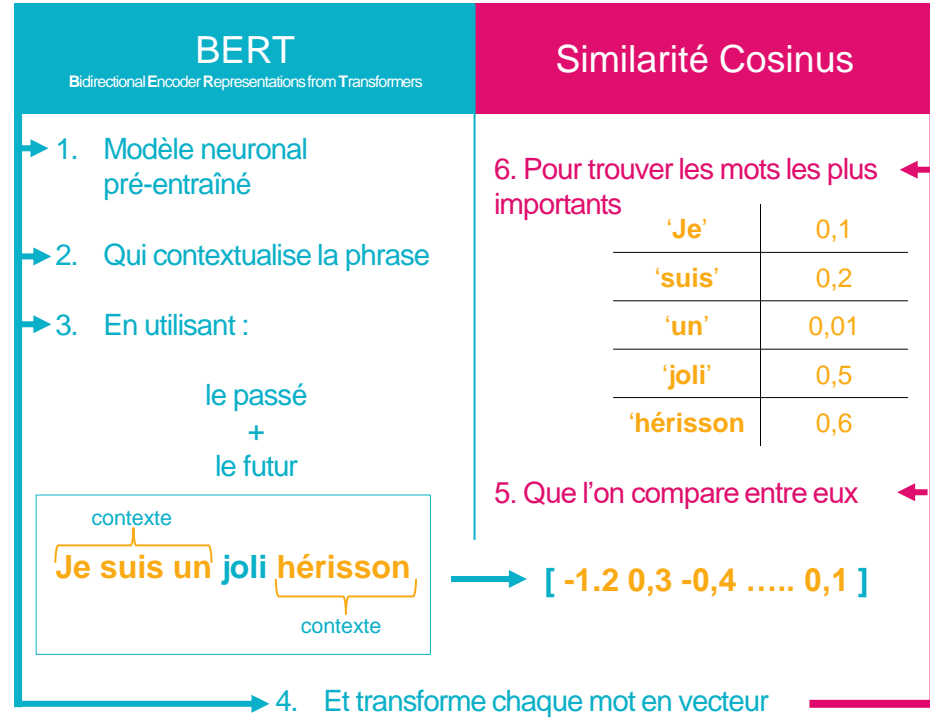
JULIEN DONCHE



Le modèle KeyBERT

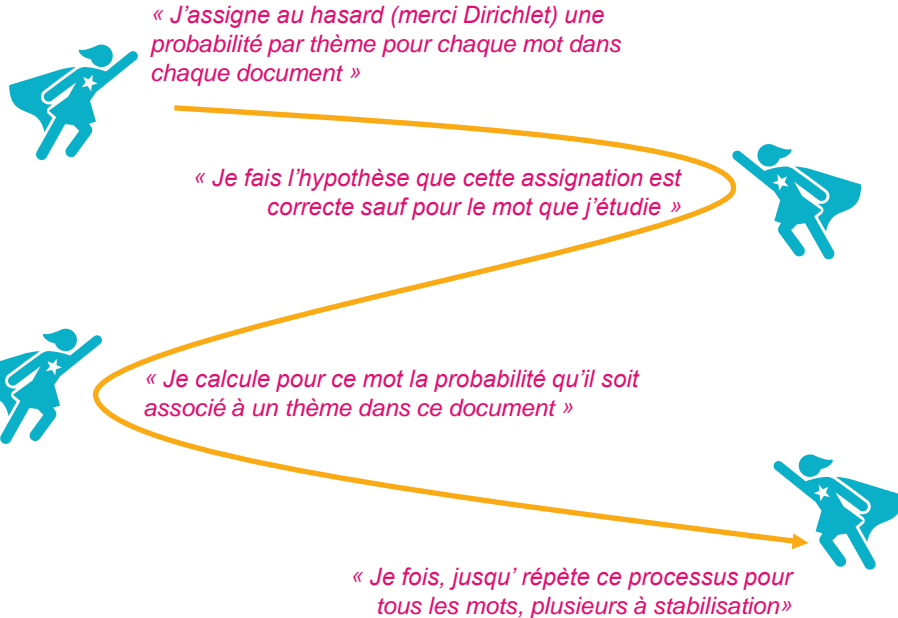
KeyBERT

Extraction de
mots-clefs



Le modèle LDA

Modèle non supervisé génératif probabiliste



LDA
Latent Dirichlet allocation

Thématisation

