

Projet 5 : Catégorisez automatiquement des questions



Claire Gayral

Décembre 2021 - Janvier 2022

Introduction

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering


Classification Non
Supervisée





Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Products



[Join this community](#)

[Home](#)

[PUBLIC](#)

[Questions](#)

[Tags](#)

[Users](#)

[COLLECTIVES](#)

[Explore Collectives](#)


[FIND A JOB](#)

[Jobs](#)

[Companies](#)

[TEAMS](#)

Stack Overflow for Teams - Collaborate and share knowledge with a private group.



[Create a free Team](#)

What is Teams?



PHP mail function doesn't complete sending of e-mail


Asked 7 years, 6 months ago


Active 2 months ago

Viewed 453k times

533




135



```
<?php
$name = $_POST['name'];
$email = $_POST['email'];
$message = $_POST['message'];
$from = 'From: yoursite.com';
$to = 'contact@yoursite.com';
$subject = 'Customer Inquiry';
$body = "From: $name\n E-Mail: $email\n Message:\n $message";

if ($_POST['submit']) {
    if (mail($to, $subject, $body, $from)) {
        echo '<p>Your message has been sent!</p>';
    } else {
        echo '<p>Something went wrong, go back and try again!</p>';
    }
}
?>
```

I've tried creating a simple mail form. The form itself is on my `index.html` page, but it submits to a separate "thank you for your submission" page, `thankyou.php`, where the above PHP code is embedded. The code submits perfectly, but never sends an email. How can I fix this?

php


html

email


Share

Follow

edited Nov 8 '19 at 11:28

 Peter Mortensen
29.3k • 21 • 97 • 124

asked Jul 9 '14 at 2:18

 user3818620
5,373 • 3 • 10 • 3

The Overflow Blog

- Favor real dependencies for unit testing
- Podcast 403: Professional ethics and phantom braking

Featured on Meta

- Providing a JavaScript API for userscripts
- Congratulations to the 59 sites that just left Beta

Linked

20

[PHP mail\(\) doesn't work](#)

11

[Troubleshooting PHP Mail](#)

9

[Mail function is not working in PHP](#)

4

[Send mail by php mail\(\) function](#)

2

[PHP mail function is not working on Hostgator](#)

2

[PHP 'mail\(\)' Function Not Sending Email](#)

4

[PHP mail\(\) Not Functioning](#)

Suivi du travail sur [github](#)

Introduction - Plan

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

Import des données

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Les données d'entrée :

- Publications sur StackOverFlow
- 3 parties : titre, corps et tags
- Sélection des publications avec tags parmi les 10 000 premières

Requête SQL sur <https://data.stackexchange.com>

```
SELECT Id, Title, Tags, Body
FROM posts
WHERE Id < 100000 AND Tags <> ''
```

Les données - Prétraitements

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

`<p>How do I forcefully unload a <code>ByteArray</code> from memory using ActionScript 3?</p>`

- 1 Format, ponctuation, filtre versions des langages de programmation
how, do, i, forcefully, unload, a, bytearray, from, memory, using, actionscript'
- 2 Stop words
forcefully, unload, bytearray, memory, actionscript
- 3 Lemmatisation
forc, unload, bytearray, memori, actionscript

Les données textuelles - Répartition des mots

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

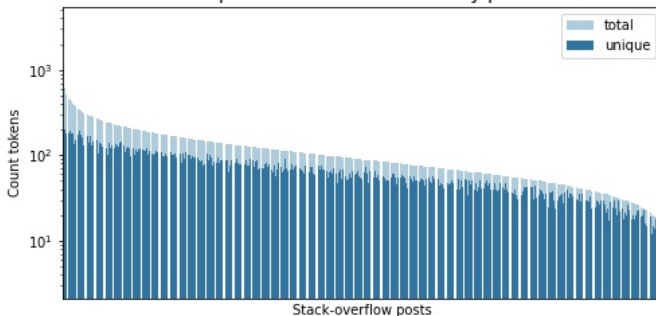
Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Barplot of tokens used in every posts



Les données textuelles - Représentation des tokens

Projet 5 : Catégorisez automatiquement des questions

Claire Gayral

Les données textuelles

Pré-traitements données textuelles

Exploration sur les tags

Réduction de dimension - NMF

Réduction de dimension - NMF

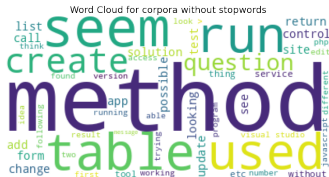
Réduction de dimension - clustering

Classification Non Supervisée

Classification Supervisée

Modèles de classification

Résultats des classifications



- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

Les tags - Pré-traitements

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

- `<c#><.net><wcf><web-services><soa>`
- Filtre nom de langages :
`c#, C#, c#-2.0, c#-3.0, c#-4.0` → `csharp`
- Data Frame en one hot encoding

Les tags - distribution

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

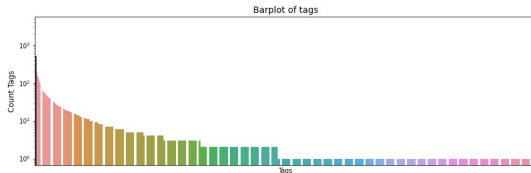
Classification
Supervisée

Modèles de
classification

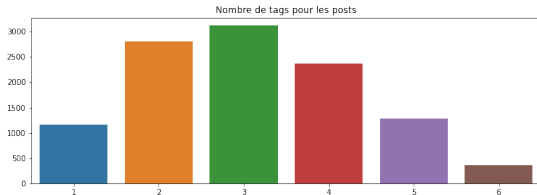
Résultats des
classifications

Conclusion

Combien de fois apparaissent chaque tags ?



Nombre de tags par publication



Tags - Création d'une variable univariée

Projet 5 :

Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

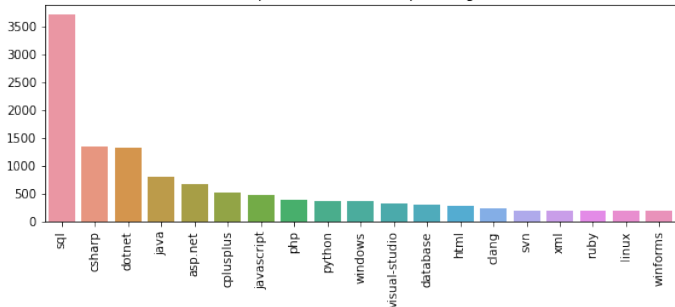
Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Repartition of most frequent tags



$\hookrightarrow y = \text{tags}["\text{csharp}"]$

Tags - NMF 1

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

La NMF :

$$\begin{matrix} p & & n \\ \boxed{X} & = & \boxed{W} \begin{matrix} K \\ \boxed{H} \\ n \end{matrix} \end{matrix} \quad \begin{matrix} X = WH \\ W_{jk} \geq 0 \quad H_{ki} \geq 0 \end{matrix}$$

[source](#)

Modélisation :

- Sur d'autres tags (Id > 10 000)
- Le choix des hyper-paramètres :
 - Séparation en train - validation
 - NMF en changeant : `n_components`, `alpha`, `l1_ratio`
 - Choix des meilleurs paramètres (minimisent le score)

Tags - NMF 2

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

**Réduction de
dimension - NMF**

Réduction de
dimension -
clustering

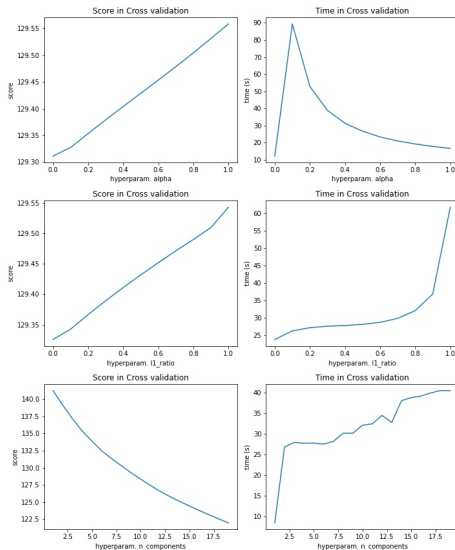
Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion



Les premiers topics de la NMF :

Conclusion

15 / 31

Tags - Clustering hiérarchique

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

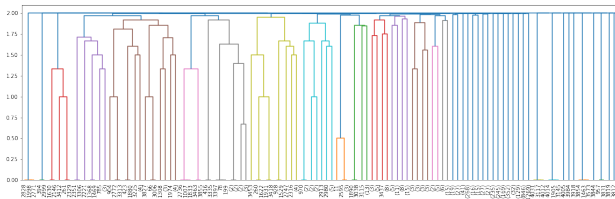
Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion



↪ 12 clusters nommés à partir des tag :

linux, language, microsoft, micro_service, create_website,
python_website, ruby, tests, python, computer_architecture,
multimedia, object_oriented

Tags - Répartition des clusters

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

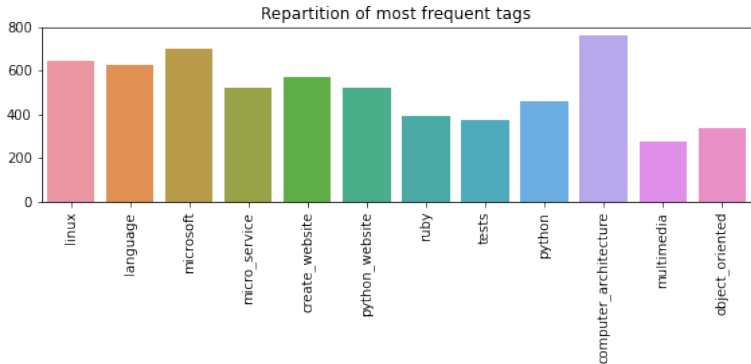
Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion



- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

LDA et NMF

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

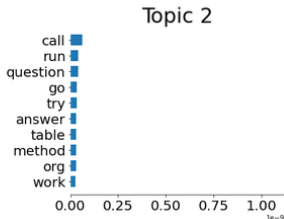
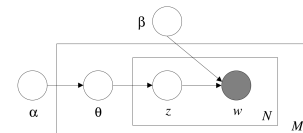
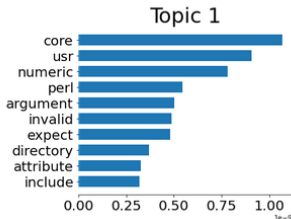
Résultats des
classifications

Conclusion

NMF :

$$X = W \times H$$

Résultats de la NMF :



LDA sur le corpus

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

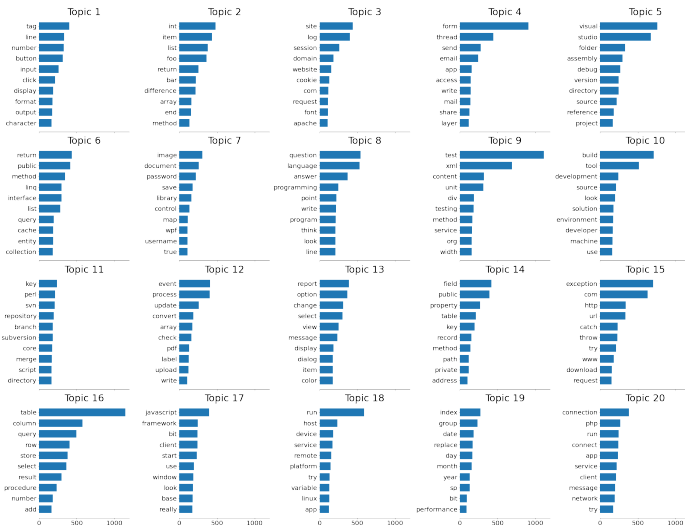
Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Topics in LDA model



LDA sur le corpus

Projet 5 : Catégorisez automatiquement des questions

Claire Gayral

Les données textuelles

Pré-traitements
données textuelles
Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non Supervisée

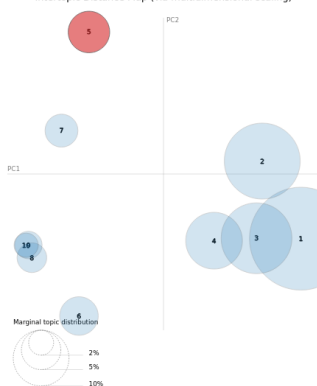
Classification Supervisée

Modèles de
classification
Résultats des
classifications

Conclusion

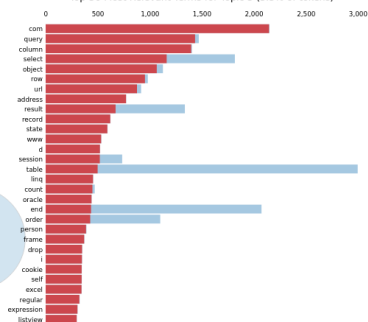
Out[179]: Selected Topic: 5 Previous Topic Next Topic Clear Topic

Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance
metric: (2) $\lambda = 1$

Top-30 Most Relevant Terms for Topic 5 (5.5% of tokens)



Overall term frequency
Estimated term frequency within the selected topic

1. $\text{saliency}(\text{term } w) = \text{frequency}(w) - \sum_t \text{p}(t | w) * \log(\text{p}(t | w) / \text{p}(t))$ for topics t : see Chuang et al.
2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * \text{p}(w | t) + (1 - \lambda) * \text{p}(w | t) / \text{p}(w)$: see Sievert & Shirley (2014)

- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

Les différents modèles utilisés

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

- Naive Bayes
- Gradient Boosting
- Random Forest

Différentes modélisations :

- Classification binaire
 - $y = \text{tag le plus courant } csharp$
- Classification multi-class :
 - $Y = \text{méta-tags issus de la classification hiérarchique}$
 - ou $Y = \text{les 20 tags les plus courants (à généraliser)}$
 - uni-label/multi-label

Mesure des performances des modèles de classification

Projet 5 :

Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

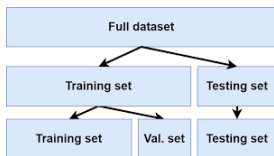
Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

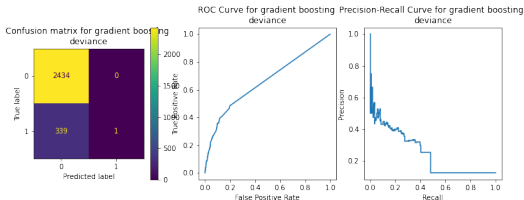


Séparation des publications

- train/test
- Puis train = train/validation par validation croisée
- Corpus pré-traité

Les métriques de classification

- Accuracy
- AUC
- Log-loss



- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

Résultats de la classification binaire :

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Classification binaire Naive Bayes (NB) :

Loi	α	Accuracy	log-loss	AUC	Temps
Bernoulli	100	0.876	4.27	0.49	0.12s
Complement	125	0.862	4.78	0.52	0.05s
Multinomial	600	0.877	4.25	0.5	0.05s

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Classification binaire Gradient Boosting :

Loss	n estimators	Accuracy	log-loss	AUC	Temps
Deviance	100	0.88	4.16	0.524	13.9s
Exponential	100	0.878	4.22	0.501	13.5s

Résultats de la classification multi-classe :

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Modèle	Accuracy	log-loss	AUC	Temps
Complement NB	0.615	2.33	0.574	0.25s
Arbre décision				
gini	0.615	3.1	0.71	1.14s
entropie	0.609	3.07	0.73	1.05s
Gradient Boosting				
deviance	0.685	1.137	0.84	3.5min

↪ meilleur modèle = Gradient Boosting en multi-class

Résultats de la classification multi-classe :

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

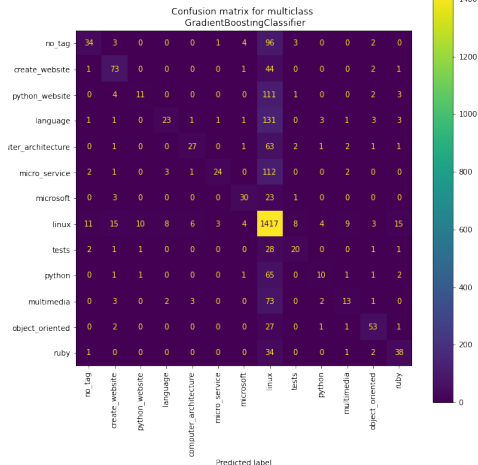
Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion



Meilleur modèle et API

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles
Exploration sur les
tags

Réduction de
dimension - NMF
Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification
Résultats des
classifications

Conclusion



monAPIflask Create

Index de l'API

Présentation

Cette API permet de prédire les tags les plus probables associés à un texte.

Utilisation

Pour l'utiliser, il suffit de coller le contenu de la publication à tagguer dans le lien suivant. Elle peut être avec balise (format html par exemple) ou juste les mots.

C'est parti !



monAPIflask Create

Traiter une nouvelle publication

Le texte de la publication :

```
operations: AXI AX2
To enable them in other operations, rebuild TensorFlow with
the appropriate compiler flags.
hello, [[4,]]

From the message, it seems that the installation was
installed successfully. But what does This TensorFlow binary
is optimized with google Deep Neural Network Library
[100000] to use the following CPU instructions in
performance-critical operations: AVX AVX2 mean exactly?

Am I using a tensorflow version with some limited features?
Any side effects?

I am using Windows 10.
```

Prédire les tags



monAPIflask Create

Résultats pour la publication

Le texte de la publication :

```
I just installed tensorflow v2.3 on anaconda python. I tried to test out the installation using the
python command below: $ python -c "import tensorflow as tf; x = [[2,]]; print(tensorflow
version', tf._version_); print('hello, {}' .format(tf.matmul(x, x)))" I got the following message:
2020-12-15 07:59:12.411852: I tensorflow/core/platform/cpu_feature_guard.cc:142] This
TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use
the following CPU instructions in performance-critical operations: AVX AVX2 To enable them
in other operations, rebuild TensorFlow with the appropriate compiler flags. hello, [[4,]] From
the message, it seems that the installation was installed successfully. But what does This
TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use
the following CPU instructions in performance-critical operations: AVX AVX2 mean exactly?
Am I using a tensorflow version with some limited features? Any side effects? I am using
Windows 10.
```

Les tags associés :

['linux']

Nouvelle publication à traiter

- 1 Les données textuelles
 - Pré-traitements données textuelles
 - Exploration sur les tags
- 2 Classification Non Supervisée
- 3 Classification Supervisée
 - Modèles de classification
 - Résultats des classifications
- 4 Conclusion

Conclusion

Projet 5 :
Catégorisez
automatiquement
des questions

Claire Gayral

Les données
textuelles

Pré-traitements
données textuelles

Exploration sur les
tags

Réduction de
dimension - NMF

Réduction de
dimension - NMF

Réduction de
dimension -
clustering

Classification Non
Supervisée

Classification
Supervisée

Modèles de
classification

Résultats des
classifications

Conclusion

Résumé

- Dimension énorme, beaucoup de mots
- Deux façons de modéliser le problème : supervisée, non supervisée
- Un modèle déployé sur une API flask

Améliorations et suite :

- Comparer avec et sans pré-traitements, les différentes représentations des mots (tf-idf, word2vec, ...)
- Améliorer le modèle en combinant les analyses (ex : LDA comme réduction de dimension)

Merci pour votre écoute !