

Hotel Online Reviews for Customer Satisfaction's Analysis

Asma Ameer

Encadré par :
Mr Dhafer Malouche

Colloque **CiDE.21**:
04- 06- Avril 2019 – Djerba

Ecole Polytechnique de Tunisie
(EPT)

Unité MASE- Ecole Supérieure
de la Statistique et Analyse de
l'Information





Contexte du projet
Méthodes de gestion de Mining
Conclusion et perspectives

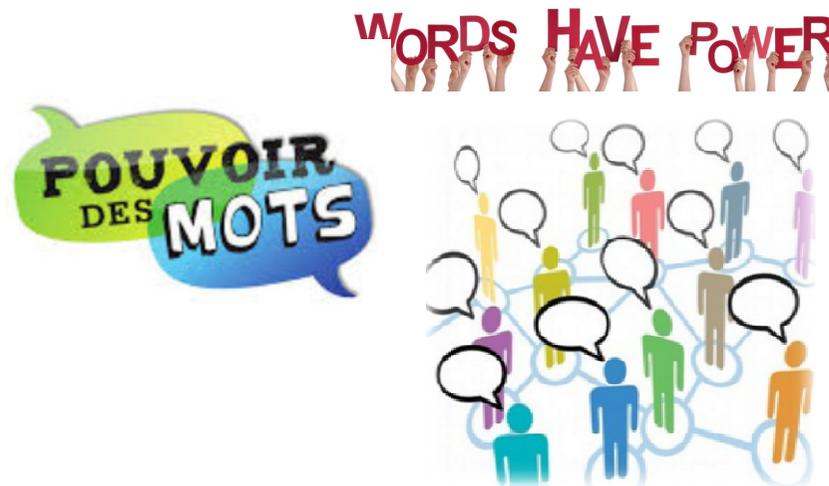
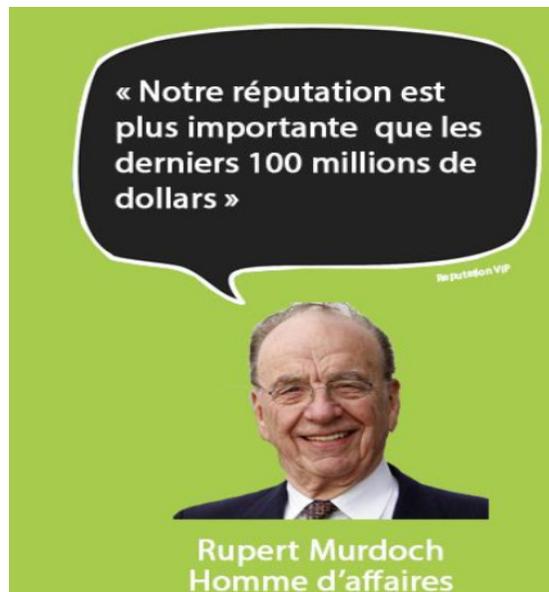


Introduction générale



La réputation

- ❖ Le mot réputation provient du latin « reputatio » qui veut dire **réflexion, examen ou considération**.
- ❖ Selon le dictionnaire Larousse : « Opinion **favorable ou défavorable** attachée à quelqu'un ou à quelque chose », « *manière dont quelqu'un, quelque chose est **connu, considéré dans un public*** »



Réputation vers E-réputation

- ❖ Avec internet, les réseaux sociaux, les téléphones portables, la réputation sort du **monde réel** pour **s'installer sur la toile**.



Développement continu des réseaux sociaux



Apparition de nouvelles formes de medias

Transformation de l'information, de sa digestion par les internautes et donc de son traitement par les services d'E-réputation



- ❖ La réputation est **peu contrôlable** à l'ère du **Web** et des médias sociaux.



E-réputation

E-réputation = réputation numérique

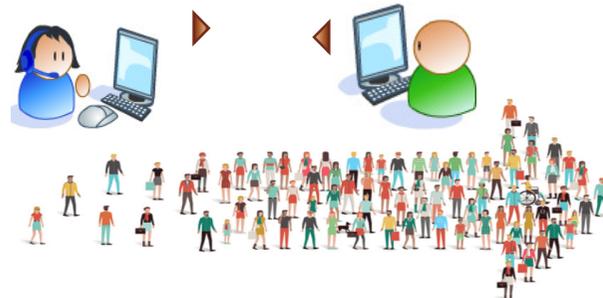
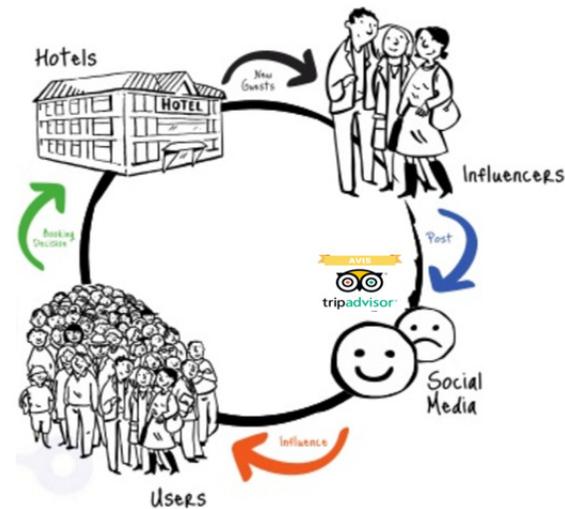
- ❖ C'est l'image que les internautes se font d'une marque, d'un établissement touristique à partir de ce qu'ils lisent / écrivent **sur Internet**.
- ❖ E-réputation **multiplie les informations**, leurs canaux de diffusion, leurs sources et leurs producteurs.
- ❖ Une accélération de la distribution des commentaires **en termes de temps et d'espace**.



Contexte du projet



E-réputation dans le domaine hôtelier



Importance de l'e-réputation d'un hôtel



- Pouvoir **différencier les concurrents d'un hôtel** et son positionnement dans le secteur hôtelier



- Pour le marketing on permet l'hôtelier de connaître de son hôtel dans le marché

Les concurrents sont en ligne !!
Les consommateurs sont en ligne !!



« contrôler » et de gérer
on et donc d'éviter
avec une mauvaise



- Pour l'internaute: Permet d'exprimer son opinion et d'obtenir une idée sur les avis des autres internautes envers un hôtel.

Optimiser la gestion d'image de l'hôtel sur Internet: Il suffit d'un client insatisfait pour que toute la planète, ou presque, le sache.

Sites d'opinions en hôtellerie



A collage of negative user reviews and ratings. It includes:

- A review with 4 empty green stars and the text «Do not recommend».
- A review with 1 star and the text «miserable!».
- A review with 4 empty green stars and the text «Minable...».
- A review with 4 empty green stars and the text «Honteux».
- A review with 1 star and the text «Nul».
- A review with 1 star and the text «Pas pour moi».
- A review with 1 star and the text «A EVITER ABSOLUMENT!».

Other elements include a yellow owl icon, a thumbs up icon, and a red star icon.



Problématique



Comment peut-on évaluer ce que disent les internautes et mesurer leurs satisfactions envers un hôtel?

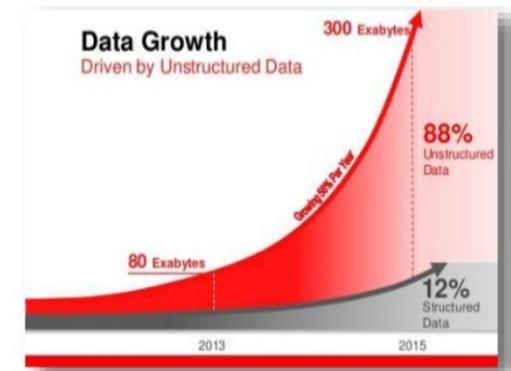
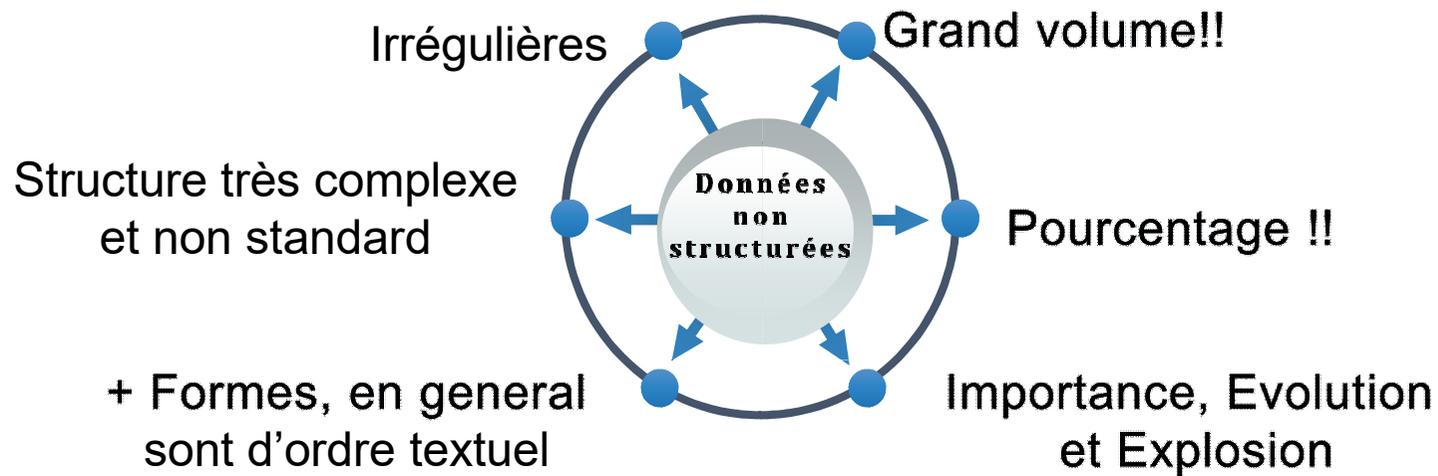


Comment peut-on déterminer la position d'un hôtel dans le marché à partir des commentaires des internautes?

Techniques de Text Mining



Données non structurées



Data Mining et Text Mining



a Mining

Analyse des données numériques et catégoriques

Données structurées et simples



● **Text Mining = Etude statistique des données textuelles**

est l'ensemble
Technologies
et méthodes:

- * ...destinées au traitement automatique des **données textuelles**
- * ...en vue d'en structurer et analyser ce contenu
- * ...dans une perspective de découverte d'informations cachées et alors de prise automatique de décision



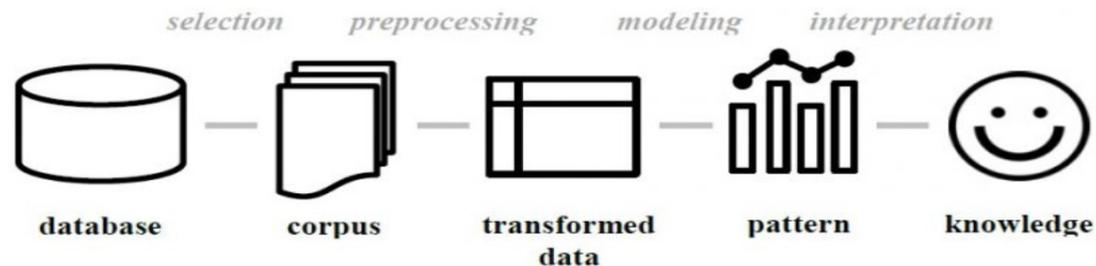
Solution

Opinion Mining

Opinion Mining

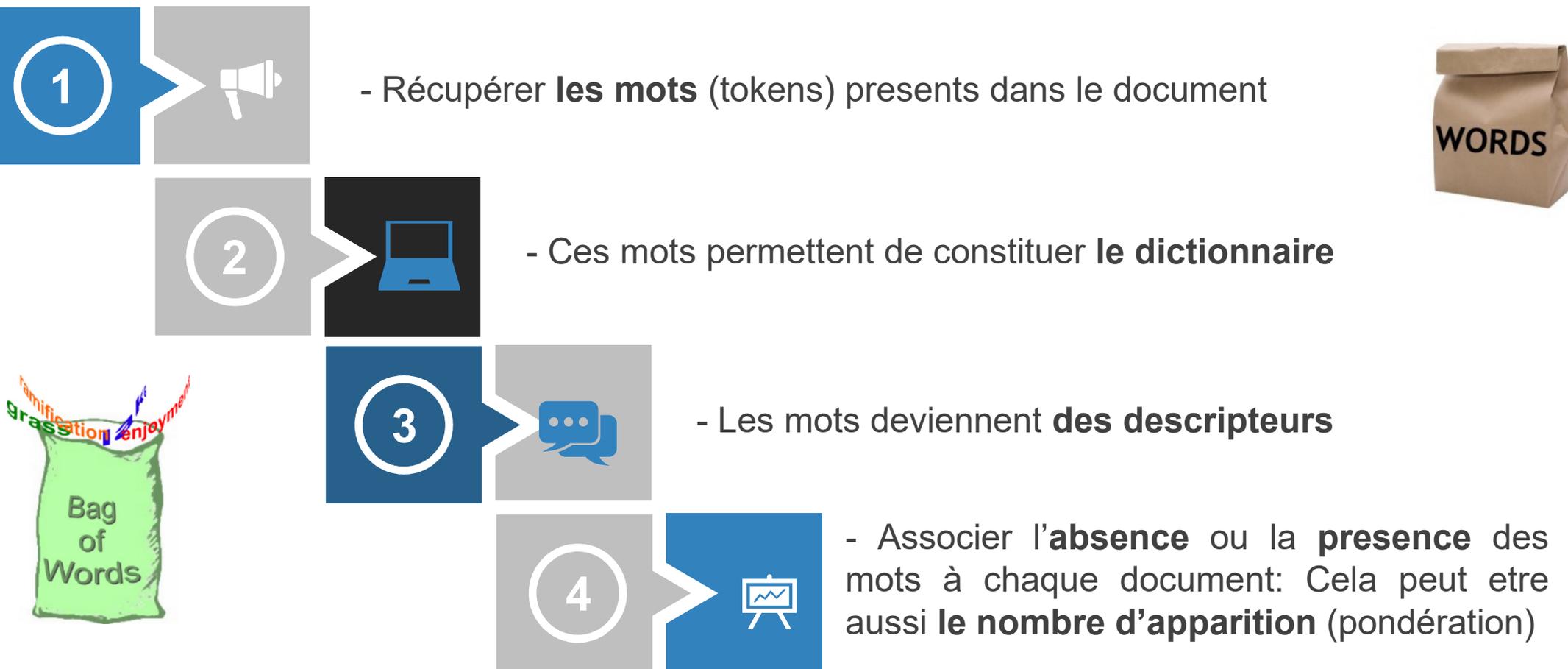
Un sous-domaine de la fouille de textes (Text Mining)

Permet d'analyser les textes afin d'en extraire des informations liées aux opinions et aux sentiments.



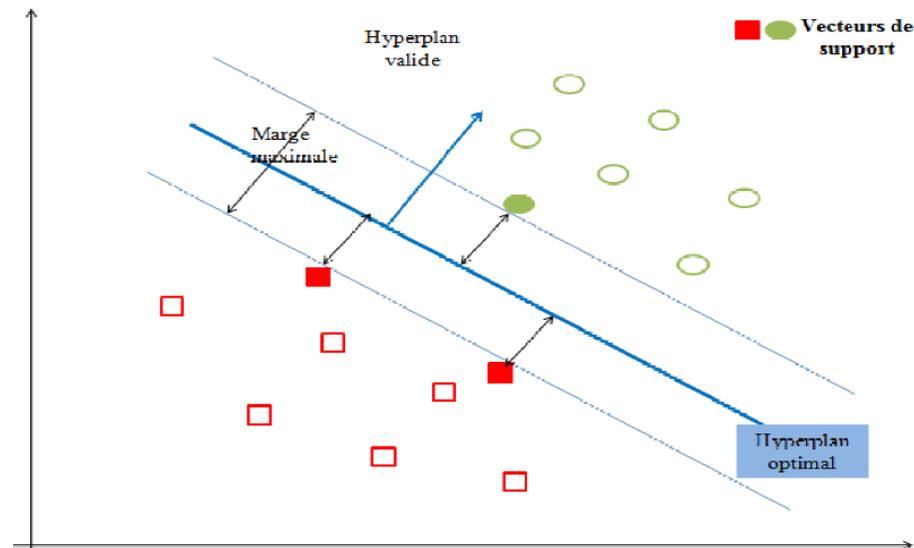
Approche1: Bag of Words (Bow)

- ❖ Une technique permettant **d'extraire l'information du texte pour une modélisation** statistique
- ❖ **Un processus de « tokenization »**: découpage d'un flux de caractère en phrases, symboles, mots/(plus particulièrement aux mots.) → Méthode sac de mots



Approche 2: Support Vector Machine

- Support Vector Machine (SVM) est un **algorithme d'apprentissage automatique supervisé**.
- L'objectif de SVM est **de trouver un classificateur** (une frontière de décision) pour séparer les données en **maximisant la marge** entre les classes.
- Le classificateur linéaire est appelé **hyperplan**. Il permet de séparer l'espace en deux régions.



Application



Présentation du logiciel R

- Un **logiciel de développement scientifique** spécialisé dans le calcul et l'**analyse statistique**
- Un **langage de programmation** et de traitement de données qui est **de plus en plus répandu**
- Gratuit et libre (**open source**)



- Performant pour la **manipulation des données**, le calcul et l'**affichage des graphiques**.
- **Très puissant** et **multiplateforme** (Linux, Mac, Windows)
- Riche en **modules statistiques** et de procédures d'analyse



Packages
utilisés



Exemples

- rvest: Outil facile de scraping
- tm: Un framework pour les applications de text mining
- openNLP: Collection d'outils de traitement du langage naturel
- ggplot2 : Système de création de graphiques

Extraction des données avec TripAdvisor

Hôtels à Hammamet

Travel feed | Hammamet | Hôtels | Activités | Restaurants | Vols | Forfaits touristiques

Économisez de l'argent. Nous interrogeons plus de 200 sites pour trouver les prix d'hôtels les plus bas.

Hotel Diar Lemdina

1 683 avis | N°10 sur 115 Hôtels à Hammamet
Rue de la Medina | Yasmine Hammamet, Hammamet 8050, Tunisie

Infos pratiques | Emplacement | **Avis et plus**

Bad service. Scattered hotel

Google Traduction

«Scattered hotel. Not secure. Bad food. Bad service. You seem to be lost and no directions to get to the reception. Situated in heart of a souk. I didn't even find the covered pool during my stay. The rooms are excellent. Clean and you find a small kitchen. It's like an s3...»

Plus

Date du séjour : janvier 2019

Utiliser | Republier | Partager

Réponse de MedinaHotels, Responsable relations publiques de Hotel Diar Lemdina
A répondu il y a 3 semaines

Google Traduction

Dear Guest, Thank you for your feedback. First of all we are glad to know that you enjoyed our rooms. Indeed they are conceived as apartments and they are unique in their kind. In fact, like you noticed, our entire complex is unique, being situated in the heart of the Medina of Yasmine Hammamet. It is quite an

Celia918 a écrit un avis le févr. 2019
Londres, Royaume-Uni • 34 contributions • 18 votes utiles

Lovely Hotel

Google Traduction

«PLEASE read about buying rugs/carpets below*** A lovely welcoming hotel. The complex includes a theme park, restaurants, museum, shisha lounges, clothes, gift shops and more. People are extremely friendly, and every point of entry into or off complex had security to ensure your...»



Prétraitement des données

**Nettoyage
!!!!!!**



**Données
Hétérogènes
!!!!**

**Corpus
non
nettoyé**

```
reviews
55 Stayed here back in May and I must say I would>
56 I'm cam to hotel Diar medina<U+2661>&hotel so>
57 Hotel was ok, Reception needs to be more organ>
58 Stayed 5 nights in July.. Upgraded to a one be>
59 Been to the hotel with my friends it was reall>
60 Me and boyfriend had an excellent week spent a>
61 I have booked for a family stay in Diar Lemdin>
62 A friend and I stayed here a few weeks ago and>
63 Its our second time in this hotel we really lo>
64 I know I am late putting this on here but bett>
65 We arrived at the hotel on sunday 21st june an>
66 We have just returned from a wonderful 2 week >
67 We stayed here 12th-26th June and absolutely l>
68 We have just returned from two wonderful weeks>
69 This hotel can accommodate for every possible >
70 This was the first all inclusive holiday I had>
71 Visited this hotel with my boyfriend for a wee>
72 The food here is great, lovely entertainment an>
73 We been To this hotel twice we really enjoyed >
74 I totally agree with other reviews that this i>
75 We are just back from our 4 weeks stays at Dia>
76 Food was excellent and most staff were very he>
77 Myself and my partner went for a week all incl>
78 We had such a good time we stayed for a week a>
```





> stopwords("english")

[1]	"i"	"me"	"my"	"myself"	"we"	"our"	"ours"	"ourselves"	"you"
[10]	"your"	"yours"	"yourself"	"yourselves"	"he"	"him"	"his"	"himself"	"she"
[19]	"her"	"hers"	"herself"	"it"	"its"	"itself"	"they"	"them"	"their"
[28]	"theirs"	"themselves"	"what"	"which"	"who"	"whom"	"this"	"that"	"these"
[37]	"those"	"am"	"is"	"are"	"was"	"were"	"be"	"been"	"being"
[46]	"have"	"has"	"had"	"having"	"do"	"does"	"did"	"doing"	"would"
[55]	"should"	"could"	"ought"	"i'm"	"you're"	"he's"	"she's"	"it's"	"we're"
[64]	"they're"	"i've"	"you've"	"we've"	"they've"	"i'd"	"you'd"	"he'd"	"she'd"
[73]	"we'd"	"they'd"	"i'll"	"you'll"	"he'll"	"she'll"	"we'll"	"they'll"	"isn't"
[82]	"aren't"	"wasn't"	"weren't"	"hasn't"	"haven't"	"hadn't"	"doesn't"	"don't"	"didn't"
[91]	"won't"	"wouldn't"	"shan't"	"shouldn't"	"can't"	"cannot"	"couldn't"	"mustn't"	"let's"
[100]	"that's"	"who's"	"what's"	"here's"	"there's"	"when's"	"where's"	"why's"	"how's"
[109]	"a"	"an"	"the"	"and"	"but"	"if"	"or"	"because"	"as"
[118]	"until"	"while"	"of"	"at"	"by"	"for"	"with"	"about"	"against"
[127]	"between"	"into"	"through"	"during"	"before"	"after"	"above"	"to"	"under"
[136]	"from"	"up"	"down"	"in"	"out"	"on"	"off"	"when"	"why"
[145]	"again"	"further"	"then"	"once"	"here"	"there"	"when"	"more"	"other"
[154]	"how"	"all"	"any"	"both"	"each"	"few"	"more"	"own"	"so"
[163]	"some"	"such"	"no"	"nor"	"not"	"only"			
[172]	"than"	"too"	"very"						

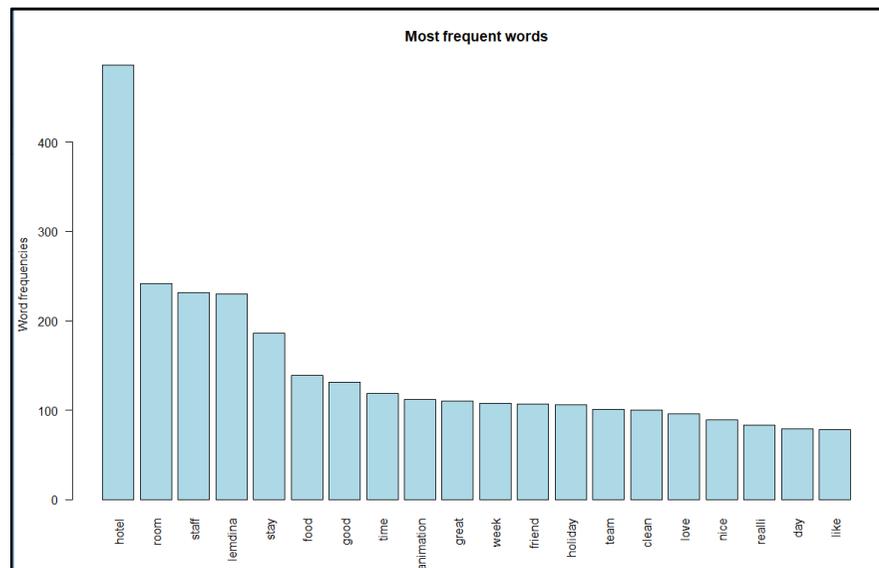


DTM et fréquences des mots

DTM =
Document
Terme
Matrice

- C'est une matrice Mathématique qui décrit la fréquence des termes existants dans le corpus.

	helpful	house	lemdina	appear	consist
128	0	0	0	0	0
129	0	0	0	0	0
130	0	0	0	0	0
131	1	1	1	0	0
132	0	0	0	1	1
133	0	0	0	0	0
134	0	0	0	0	0
135	0	0	0	0	0
136	0	0	0	0	0



word	freq
hotel	486
room	242
staff	232
lemdina	230
stay	186
food	139
good	131
time	119
animation	112
great	110
week	108
friend	107
holiday	106
team	101
clean	100
love	96
nice	89
realli	83
day	79
like	78
arriv	76
pool	70
back	70
first	61

Visualisations graphiques

Nuage de mots



Co-occurrence



Une façon simple de présenter les idées **en utilisant DTM**

Visualisation graphique du contenu des documents avec une **graduation de couleurs** et une **différenciation de polices**

- Présenter les **associations** entre deux unités linguistiques.
- Mettre en valeur le **présence simultanée de deux ou de plusieurs** mots dans le même contexte.

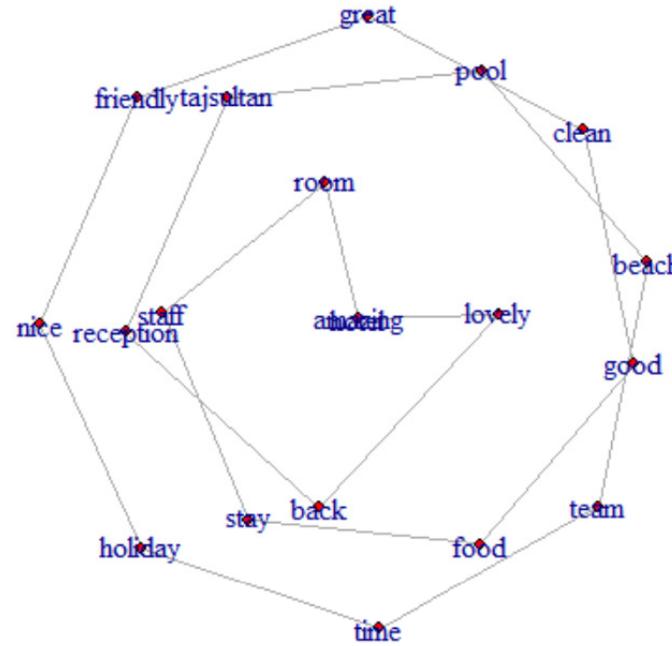
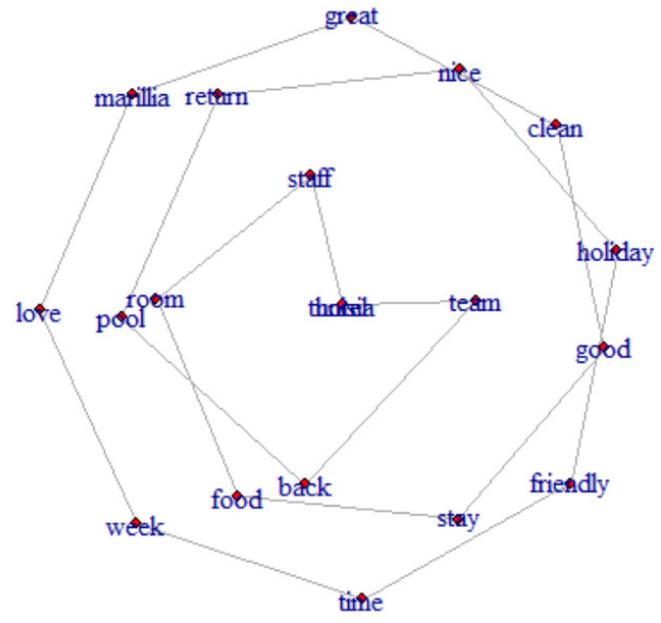
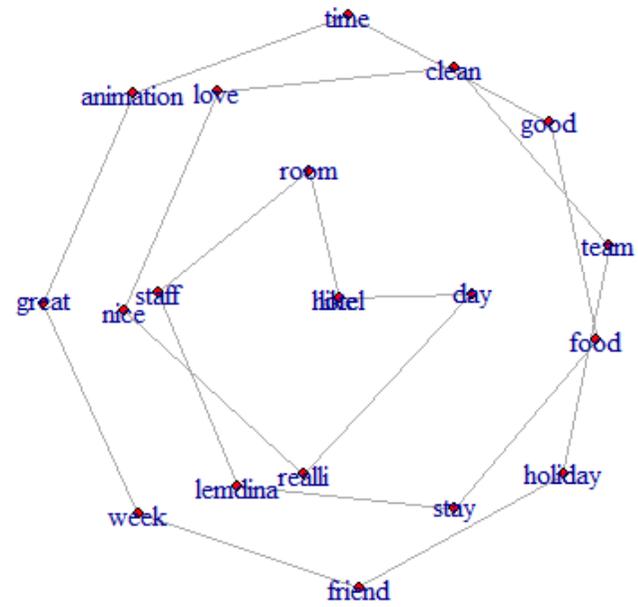


Co-occurrence

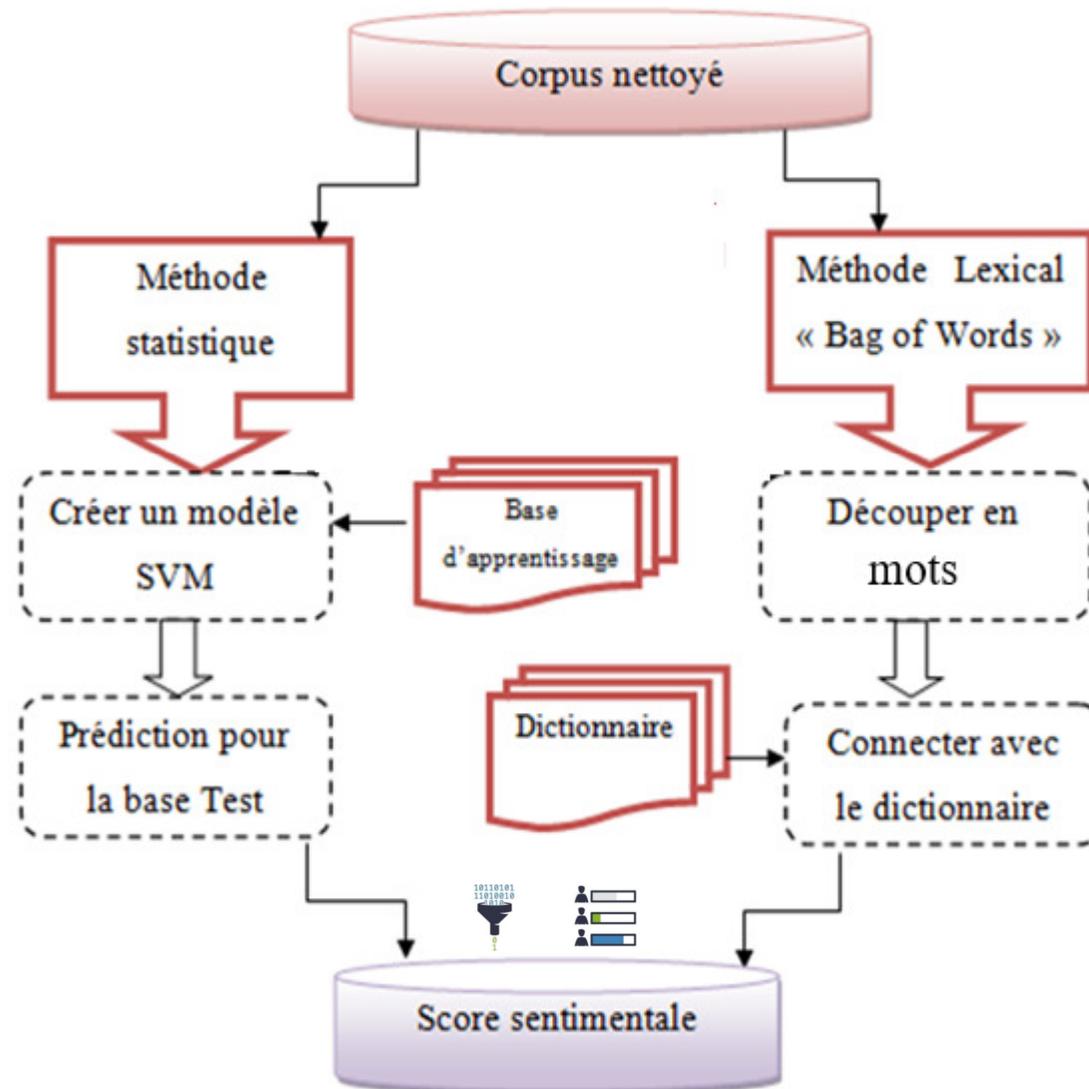
```

> assocs
      [,1]      [,2]
[1,] "hotel"    "room"
[2,] "room"     "staff"
[3,] "staff"    "stay"
[4,] "stay"     "food"
[5,] "food"     "good"
[6,] "good"     "clean"
[7,] "clean"    "great"
[8,] "great"    "friendly"
[9,] "friendly" "nice"
[10,] "nice"    "holiday"
[11,] "holiday" "time"
[12,] "time"    "team"
[13,] "team"    "beach"
[14,] "beach"   "pool"
[15,] "pool"    "tajsultan"
[16,] "tajsultan" "reception"
[17,] "reception" "back"
[18,] "back"    "lovely"
[19,] "lovely"  "amazing"

```



Méthodologie du score de satisfaction



Calcul du score avec la méthode Bow

Considerer le corpus comme un “ sac de mots”!!

Découper le corpus en mots



Construire un dictionnaire lexical

Vérifier si le matching entre les termes et le dictionnaire est vrai

Tenir compte du nombre d'occurrence de chaque mot

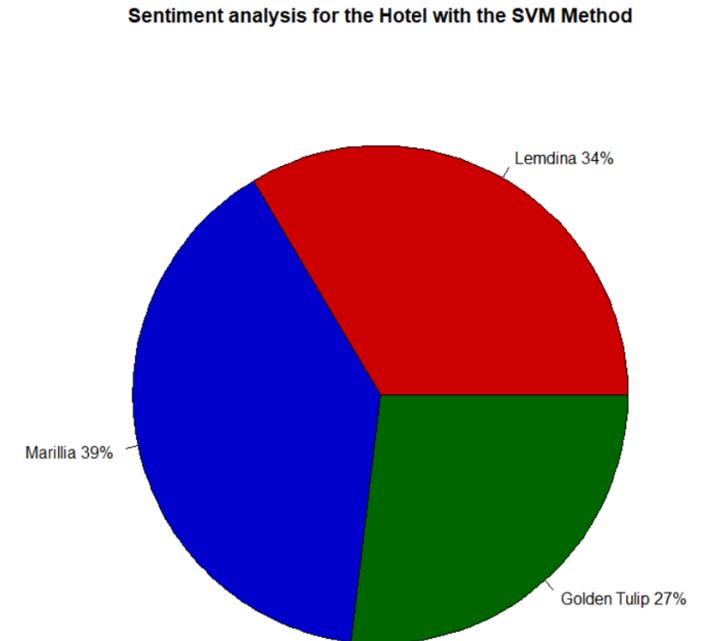
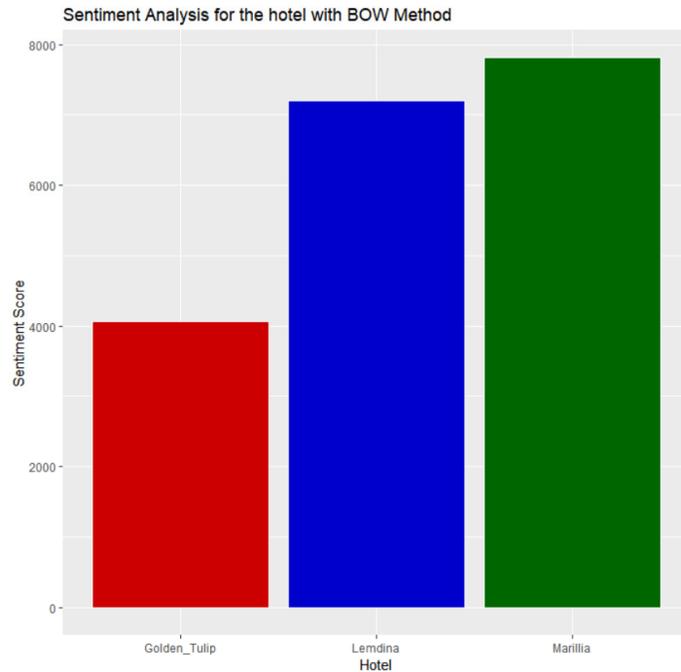
Calculer le score pour chaque hôtel



super	4
top	5
accomodation	1
delicious	2
enjoyable	3
equipped	1
trouble	-2
poor	-3
hard	-2



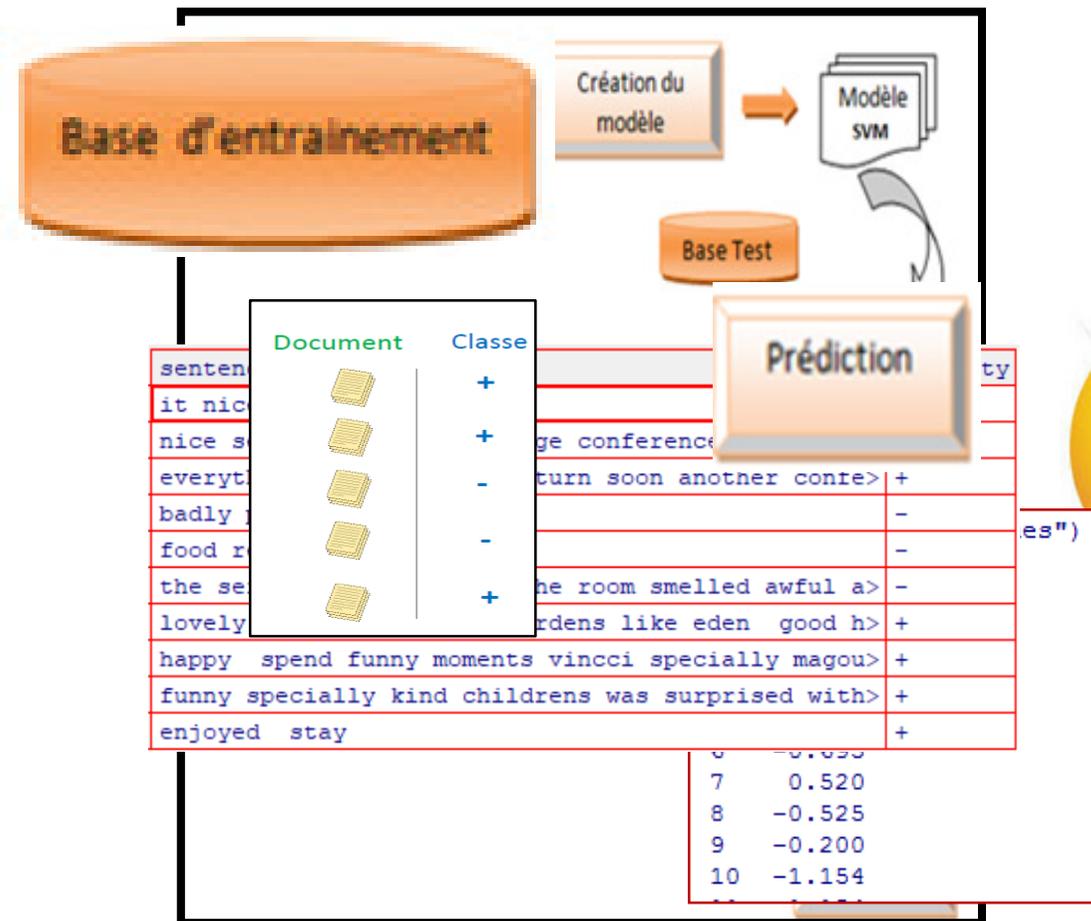
Les résultats du Bow



$$Score_{BOW} = \sum_{i=1}^{i=n} f_i * p_i$$

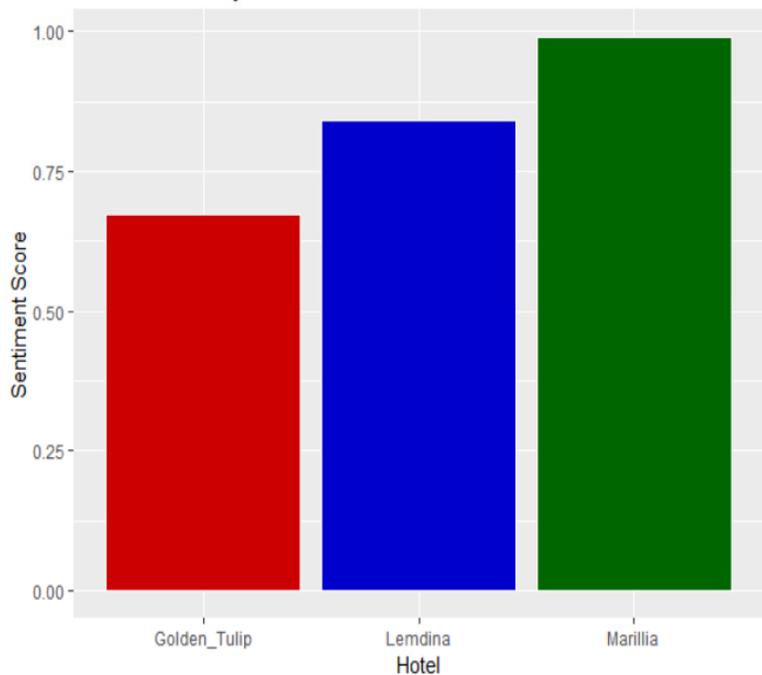
Avec: n = le nombre des mots dans le corpus
 p_i est la pondération des mots dans le dictionnaire
 f_i est la fréquence des mots dans le corpus

Calcul du score avec la méthode SVM

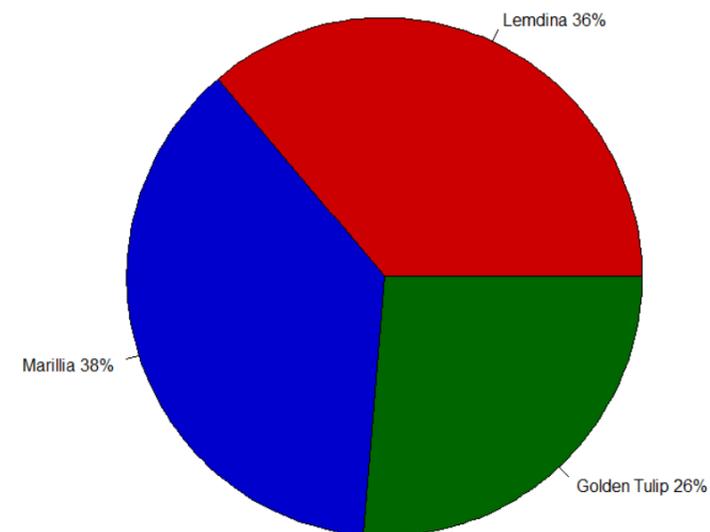


Les résultats de SVM

Sentiment Analysis for the hotel with SVM Method



Sentiment analysis for the Hotel with the SVM Method



$$Score_{SVM} = \sum_{i=1}^{i=k} score \text{ du text } j$$

avec i = le texte utilisé dans une base test avec k documents

Validation du modèle

La phase de validation est une étape primordiale pour déterminer la **qualité d'un modèle**.

En s'appuyant sur la **matrice de confusion** qui permet de confronter les valeurs **réelles avec celles prédites**, on a calculé ces indicateurs suivants:

Indicateurs	Pour la base test
Taux de bon classement	81%
Rappel	75%
Précision	79%
F_{mesure}	77%

 **Validation**



Conclusion

Ce travail sert à:

- ❖ Présenter **l'importance de l'e-réputation** dans le domaine hôtelier.
- ❖ Fournir une introduction pour certains **techniques** utilisées pour une **analyse textuelle des commentaires** des internautes.
- ❖ Mettre en place **un outil d'aide à la décision** permettant de comparer la **satisfaction des internautes** envers un hôtel donné.



Perspectives



- ❖ Evaluer les **commentaires** des internautes pour **plusieurs langues**
- ❖ Elargir le volume des données exploités: **différents zones géographiques** d'hôtel
- ❖ Exploiter d'autres **techniques d'Opinion Mining** et comparer les résultats
- ❖ Modéliser les commentaires **filtrés par type de clientèle**
- ❖ Traduire **les émoticons des sentiments** au sein des commentaires en un score





Merci pour votre
attention





➤ *Taux de bon classement* = $\frac{\text{Vrai positif} + \text{vrai négatif}}{\text{Total de l'échantillon}}$

➤ *Précision* = $\frac{\text{Vrai positif}}{\text{Vrai positif} + \text{Faux positif}}$

➤ *Rappel* = $\frac{\text{Vrai positif}}{\text{Vrai négative} + \text{Vrai positif}}$

➤ $F_{\text{mesure}} = \frac{2 * \text{Rappel} * \text{Précision}}{\text{Rappel} + \text{Précision}} = \frac{2 * VP}{2 * VP + FP + FN}$

Observation en ligne d'hôtels p l'analyse de la satisfaction de clientèle



Asma Ameer

Encadré par :
Mr Dhafer Malouche

Colloque **CiDE.21**:
04- 06- Avril 2019 – Djerba

Ecole Polytechnique de Tunisie
(EPT)

Unité MASE- Ecole Supérieure
de la Statistique et Analyse de
l'Information





Hotel Online Reviews for Customer Satisfaction's Analysis

Colloque **CiDE.21**:
04- 06- Avril 2019 – Djerba

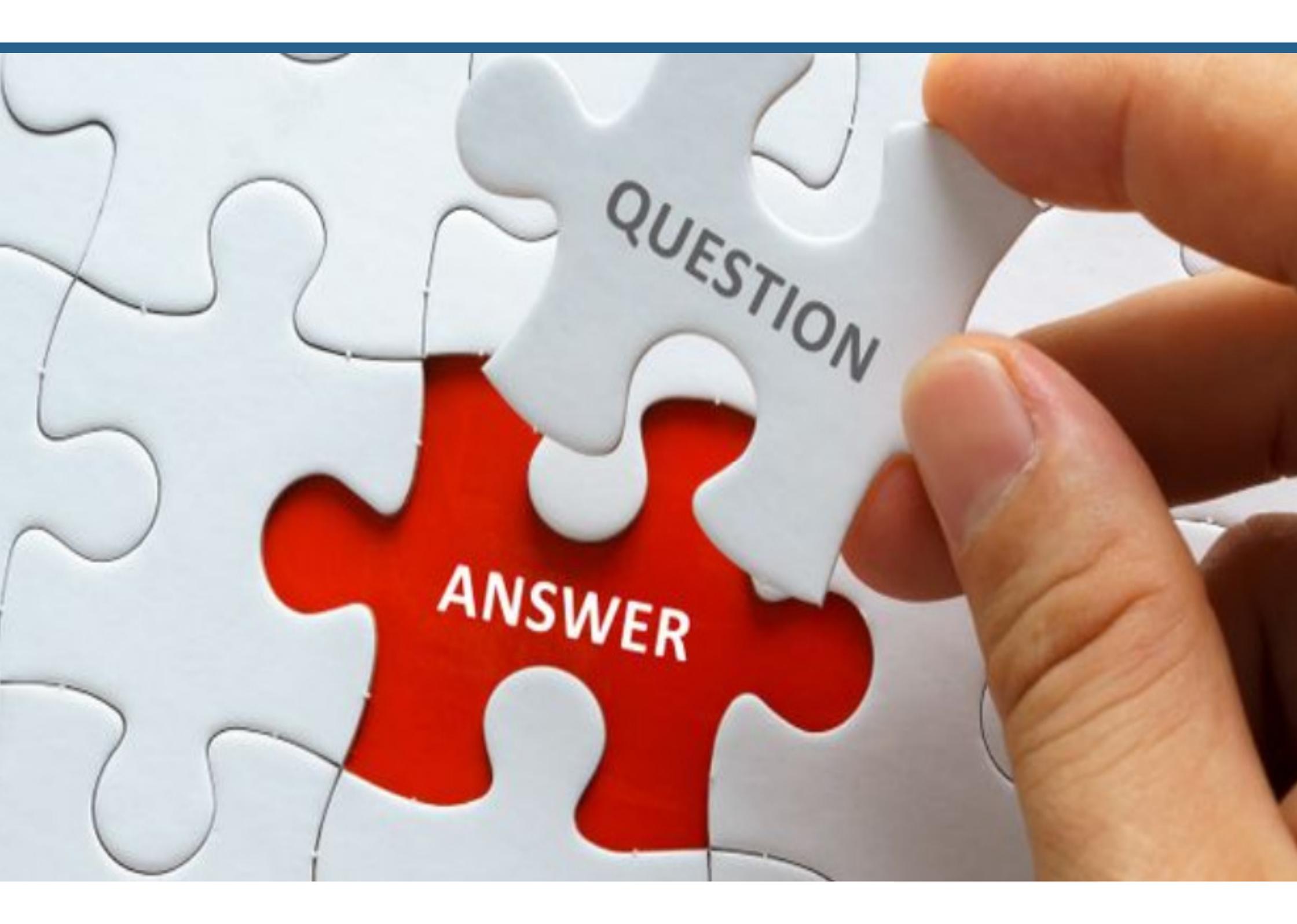
Asma Ameer

Encadré par :
Mr Dhafer Malouche

Ecole Polytechnique de Tunisie
(EPT)

Unité MASE- Ecole Supérieure
de la Statistique et Analyse de
l'Information





QUESTION

ANSWER