

# ***Formation Text Mining***

Module 1 : **Introduction**

# PLAN

---

**Contexte général**

**Définitions**

**Applications**

# Contexte général

## Structured Enterprise Data



Data Warehouse



ERP  
Systems



CRM  
Systems

10%

## Unstructured Enterprise Data



Email Systems



Research Data



Surveys



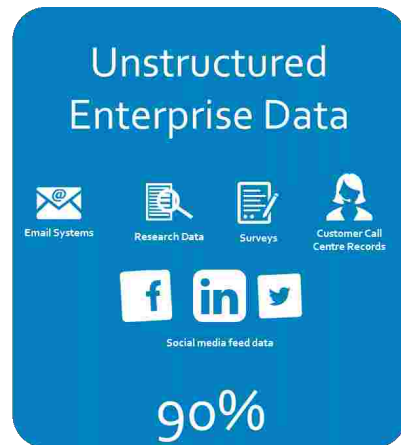
Customer Call  
Centre Records



Social media feed data

90%

# Contexte général



=> Nécessité de tenir en compte ces données dans les projets de data science pour tirer le maximum d'informations et aider à la prise de décision.

# Contexte général

Web 2.0 : Flux de données de plus en plus volumineux et rapide.

La plupart de ces données sont non structurées !



# Contexte général



1. Les outils/techniques d'analyse de la data ne peuvent manipuler que des données structurées.

=> Peut-on transformer les données non structurées en des données structurées ?

# Contexte général



2. Les textes constituent une grande partie des données non structurées.

=> Y'a-t-il des techniques pour manipuler les textes ?

# Définitions

**Mining** : Fouillage, fouiller dans les mines pour chercher les minéraux par exemple.

**Data Mining** : Fouiller la data pour chercher des informations utiles.





# Définitions

**Text Mining** : Fouiller des données textuelles pour extraire des informations utiles.



# Définitions

**NLP** : Ensemble de techniques permettant d'automatiser le traitement des langues.

NLP requiert des BD dites ressources lexicales, comme les dictionnaires.

N.B : Chaque langue a ses propres ressources.



# Applications (1/3)



**Gouvernement** : Identifier les fraudes. Analyser les sentiments/réactions des citoyens à l'égard des nouvelles décisions.



**Finance** : Détecter les fraudes. Comprendre les clients et proposer des offres répondant à leurs besoins.

# Applications (2/3)



**Retail :** Analyser et comprendre les comportements des clients pour être à leur écoute. Améliorer la qualité des services et/ou des produits.



**Santé :** Chercher des patterns similaires dans les rapports d'autres médecins.

# Applications (3/3)

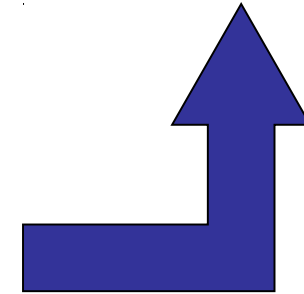
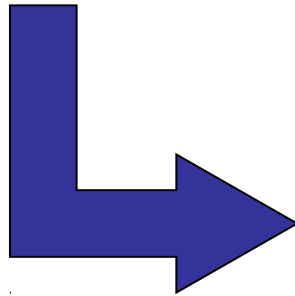


**Télécom** : prévenir l'attrition des clients.  
Proposer des offres de cross-selling ou up-selling à travers l'analyse des commentaires des clients.

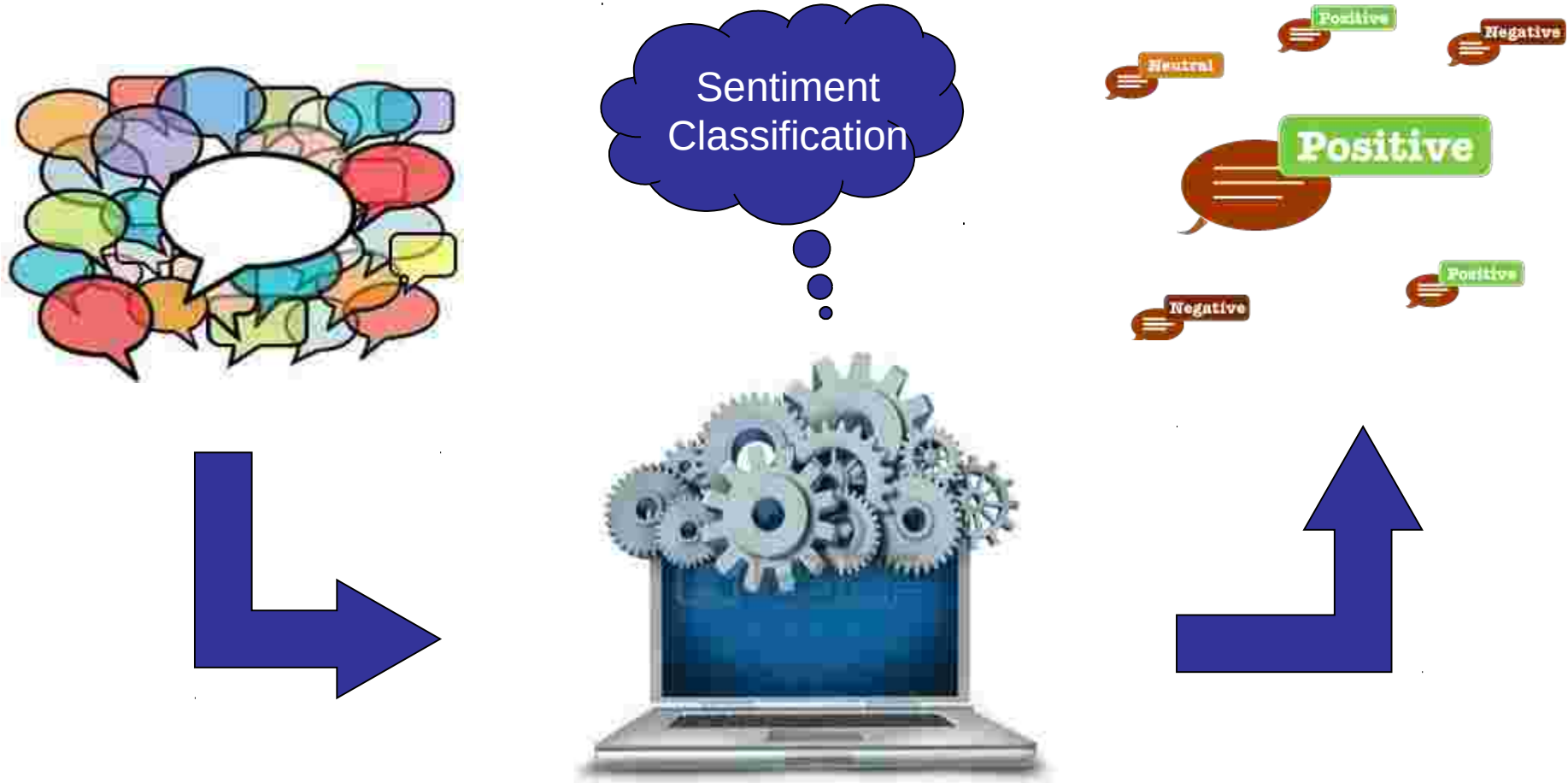


**Assurances** : Détection de fraudes.  
Intelligence compétitive.

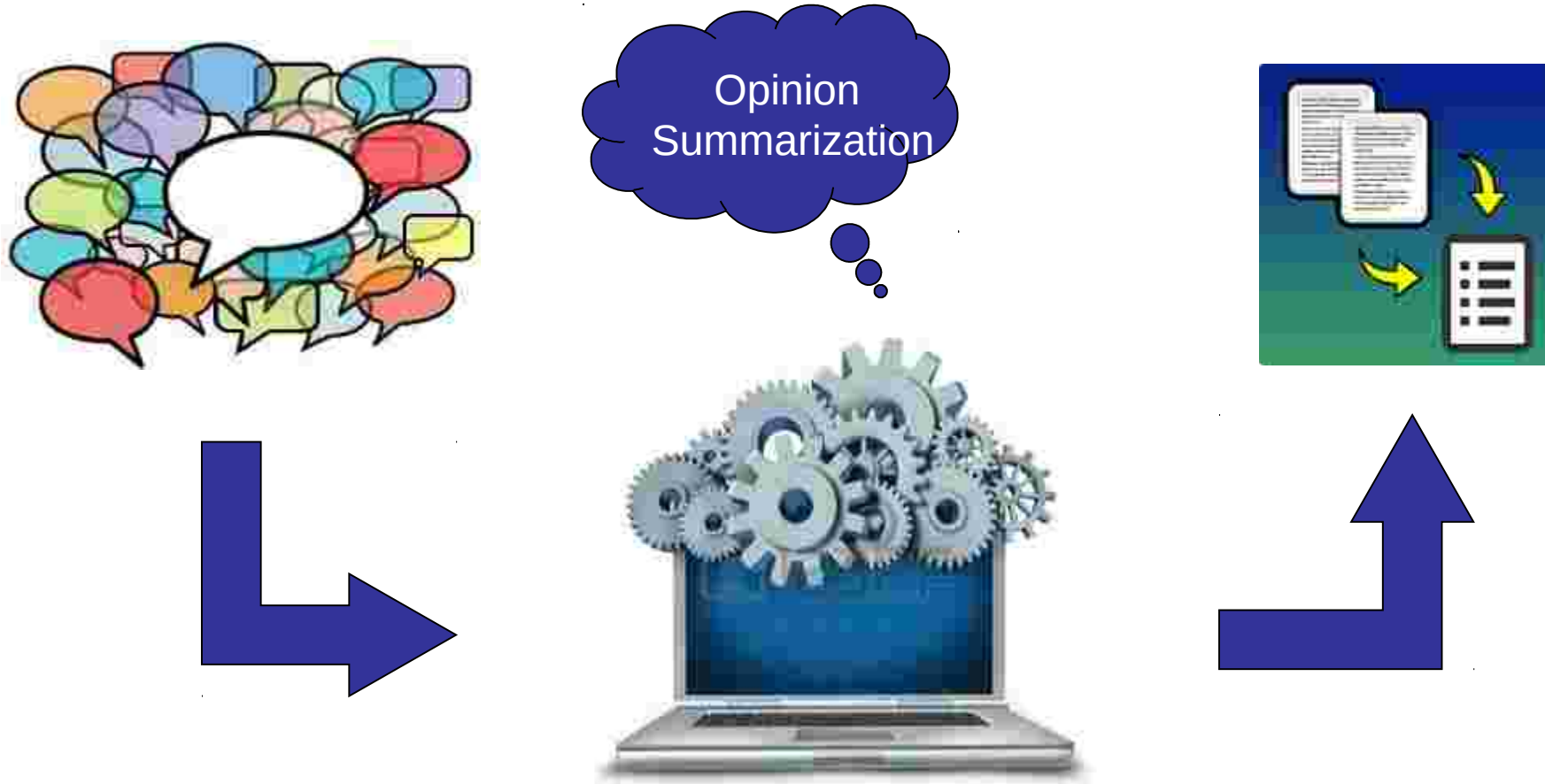
# Tâches



# Tâches

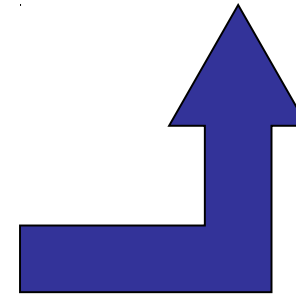
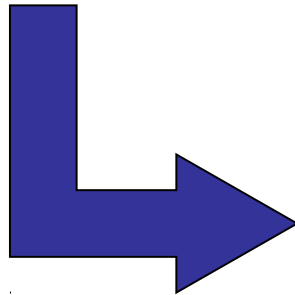


# Tâches





# Tâches



# Tâches

Traduction automatique

Résumé de textes

Transcription

Détection de polarité

Information Retrieval

Information Extraction

**TEXT  
MINING**

Etc..