



كلية العلوم
السفلالية - مراكش
FACULTÉ DES SCIENCES
SEMLALIA - MARRAKECH

Soutenance de projet de fin d'études

Conception et réalisation d'une solution d'analyse des Tweets à l'aide d'une analyse des sentiments basée sur les aspects, appliquée sur les compagnies aériennes.

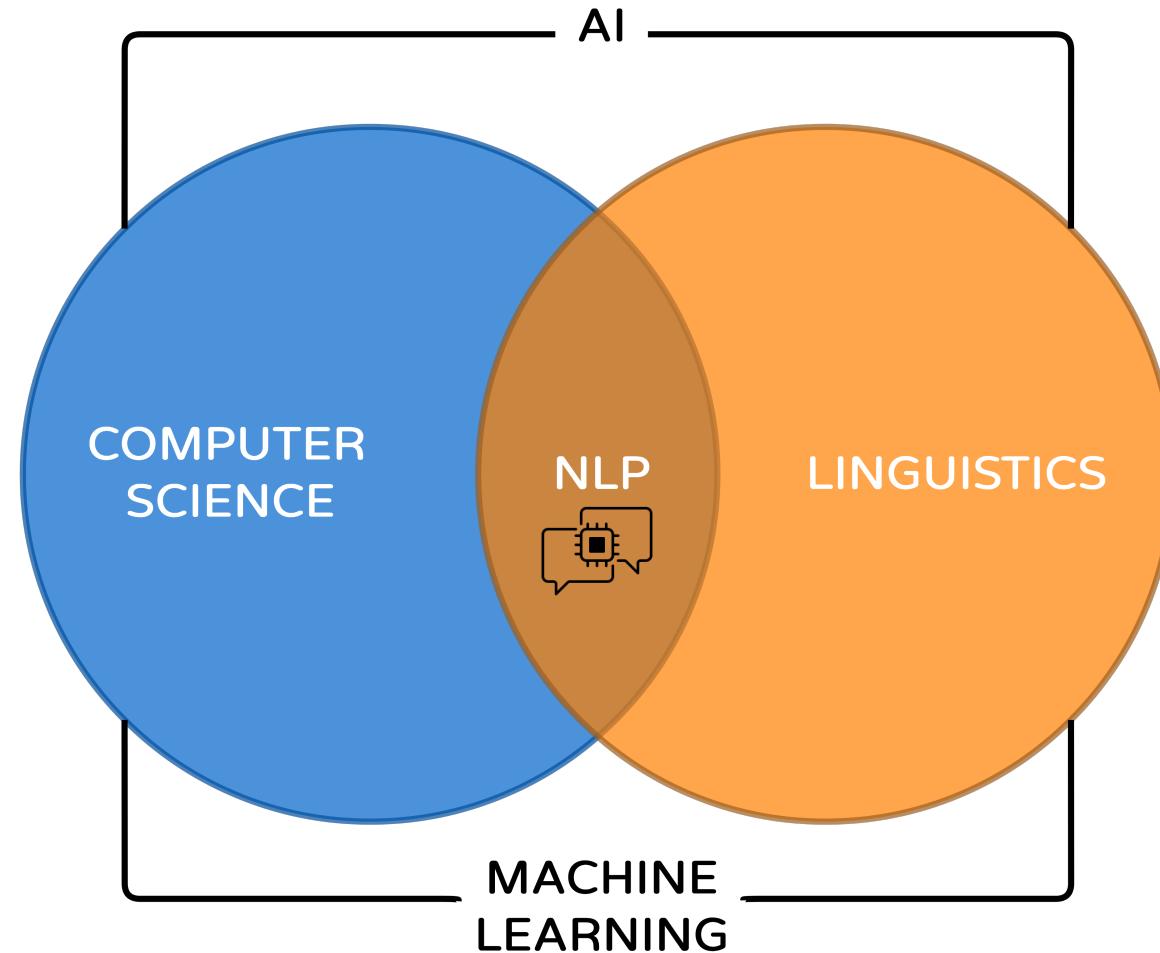
LABIAD Salah Eddine

ZIRY Asmaa

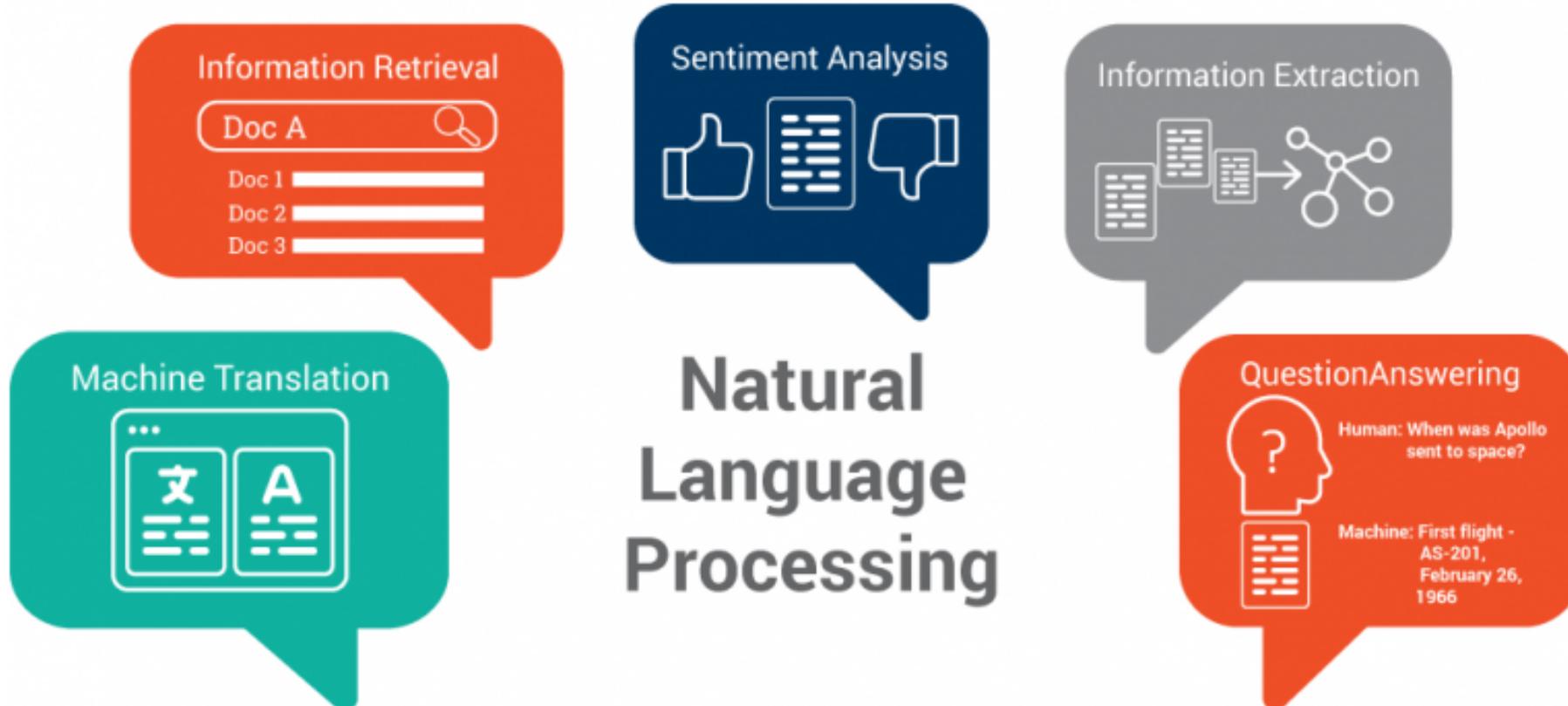
Plan

- I. Introduction
- II. Contexte général du projet
 - i. Problématique
 - ii. Objectif
- III. Analyse et Conception
 - i. Vue d'ensemble
 - ii. Analyse exploratoire des données
- IV. Réalisation et mise en œuvre
 - i. Collecte des données
 - ii. Préparation des données
 - iii. Construction des modèles
 - iv. Évaluation et résultats
- V. L'interface
- VI. Conclusion

Introduction



Introduction



Introduction



Negative

I'm dissatisfied with your customer service.
No one was able to help me with the
problems I had with using your product.



Neutral

The product has multiple features
that are suitable for users with different
levels of experience.

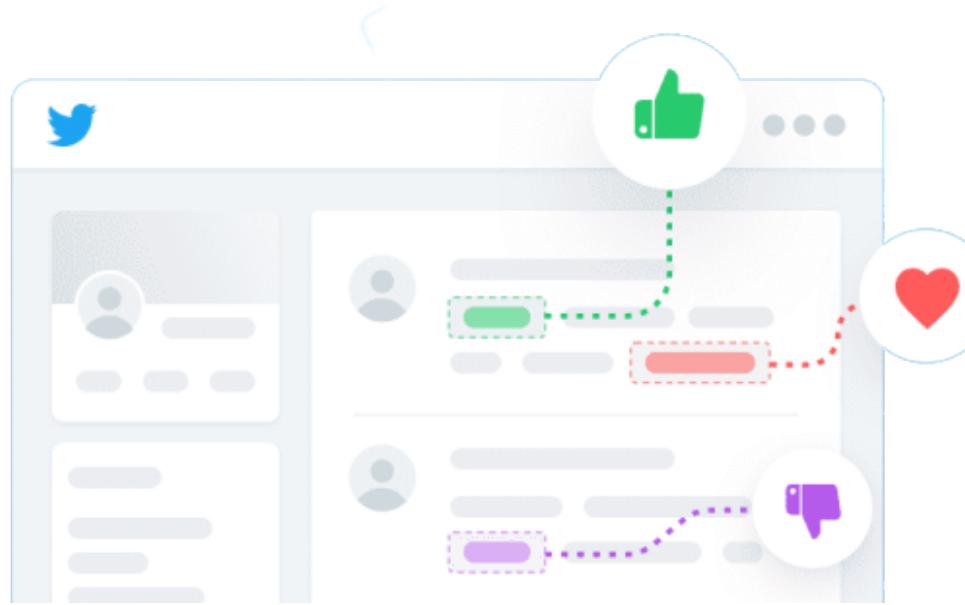


Positive

I really enjoy how easy this product
is to use and how it successfully helps
my team complete their day-to-day tasks.

Contexte général du projet

Problématique



Comment peut-on **collecter** un ensemble de tweets, et par la suite **extraire** les réactions, les opinions, les critiques et les commentaires des gens à l'égard d'un produit ou service spécifique d'une compagnie aérienne?

Objectif

 **Nathalie Smuha**
@NathalieSmuha 

@VirginAmerica What happened 2 ur vegan food options?! At least say on ur site so i know I won't be ...

9:26 AM · May 20, 2020 (i)

 27  See Nathalie Smuha's other Tweets

Sentiment : negative Reason : food & beverage

 Darren O'Keeffe
@darrenjokeeffe



VirginAmerica everything was fine until you lost my bag

10:01 AM · May 27, 2020



 5  See Darren O'Keeffe's other Tweets

Sentiment : negative Reason : ground service

 **Ryan**
@NoMagRyan

It's my lunchbreak which means it's time to see if
Ryanair has tweeted anything today

 Jennifer   
@JacobiteJen

Hi [@VirginAmerica](#) is the best airline I have flown on. Easy to change your reservation, helpful represent...

Analyse et Conception

Vue d'ensemble

- L'approche utilise les avis des passagers recueillis sur le service de micro-blog Twitter.

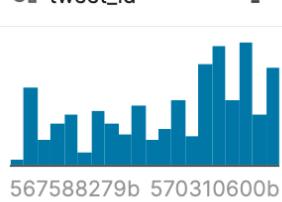
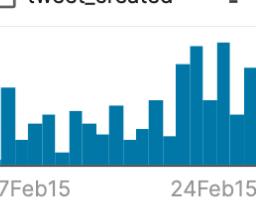
Vue d'ensemble

- L'approche utilise les avis des passagers recueillis sur le service de micro-blog Twitter.
- Élaboration d'une stratégie de prétraitement pour gérer la suppression des informations sans importance.
- La représentation des tokens en tant que sparse matrix et application du pipelining pour améliorer les performances des classifieurs.

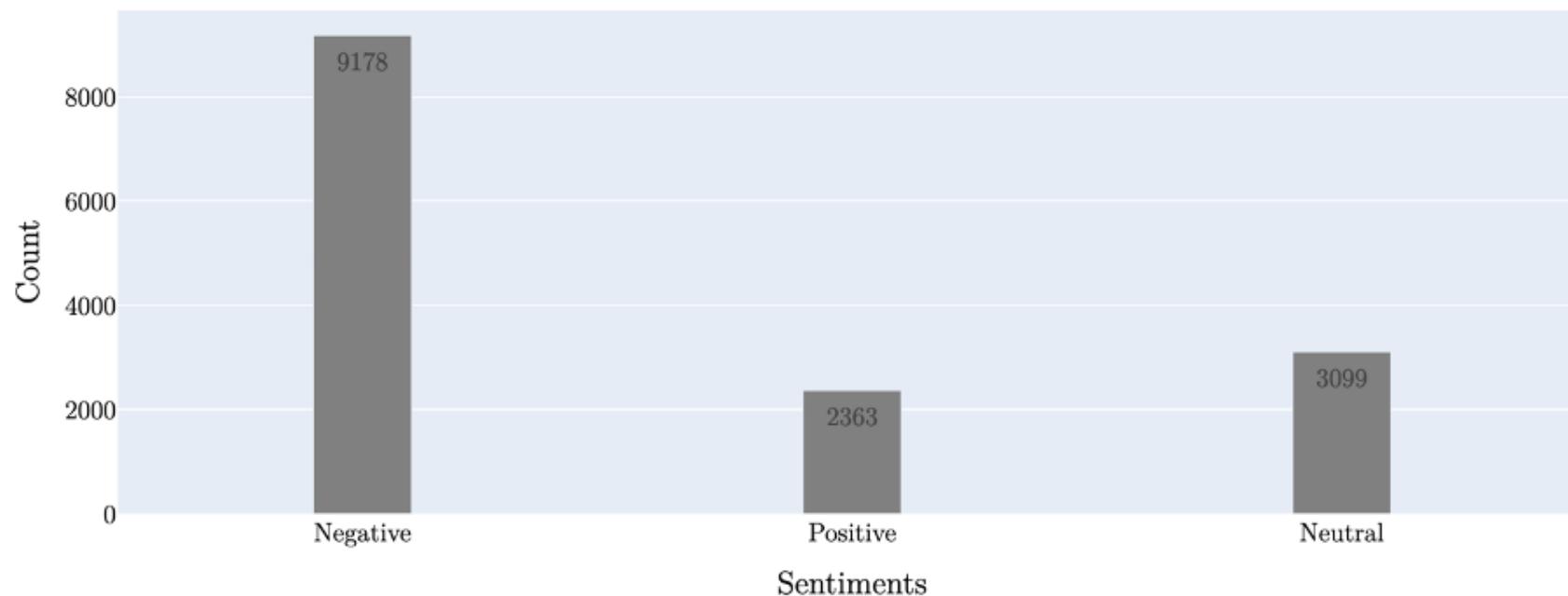
Vue d'ensemble

- L'approche utilise les avis des passagers recueillis sur le service de micro-blog Twitter.
- Élaboration d'une stratégie de prétraitement pour gérer la suppression des informations sans importance.
- La représentation des tokens en tant que sparse matrix et application du pipelining pour améliorer les performances des classifieurs.
- Représentation graphique des données sentimentalement analysées via une application Web qui permet de comparer les compagnies aériennes.

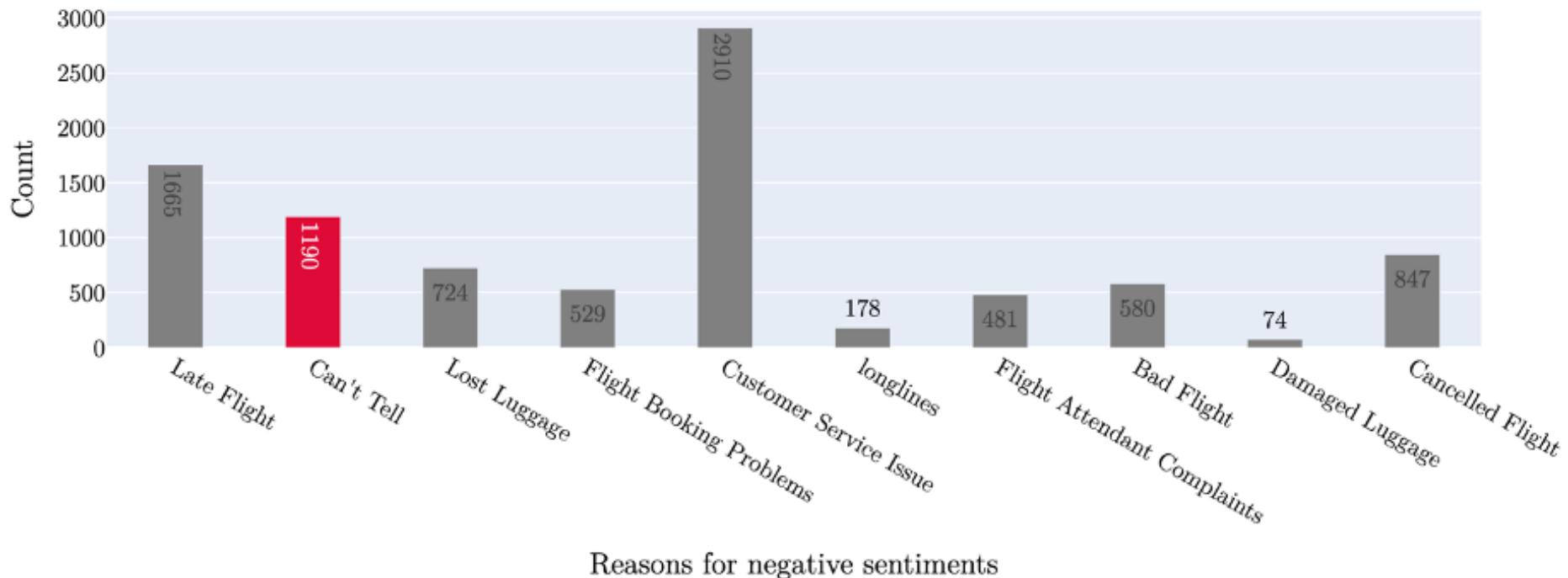
Analyse exploratoire des données

tweet_id	airline_sentiment	negativereason	text	tweet_created
	negative 63%	[null] 37%	14427 unique values	
570088404156698625	negative	Customer Service Issue	@VirginAmerica help, left expensive headphones on flight 89 IAD to LAX today. Seat 2A. No one answer...	2015-02-23 21:10:41 -0800
570084582780899328	negative	Customer Service Issue	@VirginAmerica awaiting my return phone call, just would prefer to use your online self-service opti...	2015-02-23 20:55:30 -0800
570076792993611776	positive		@VirginAmerica this is great news! America could start flights to Hawaii by end of year http://t.co...	2015-02-23 20:24:33 -0800
570051991277342720	neutral		Nice RT @VirginAmerica: Vibe with the moodlight from takeoff to touchdown.	2015-02-23 18:46:00 -0800

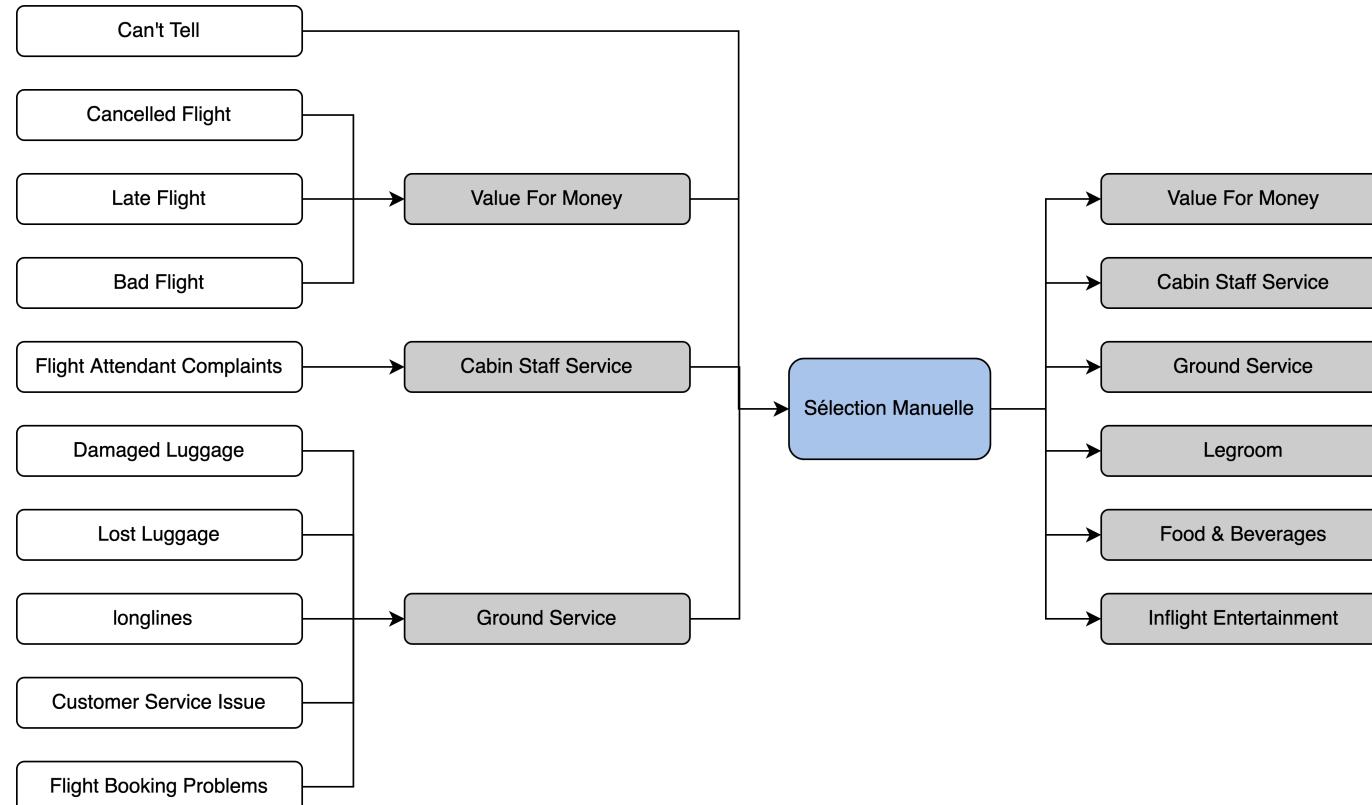
Analyse exploratoire des données



Analyse exploratoire des données



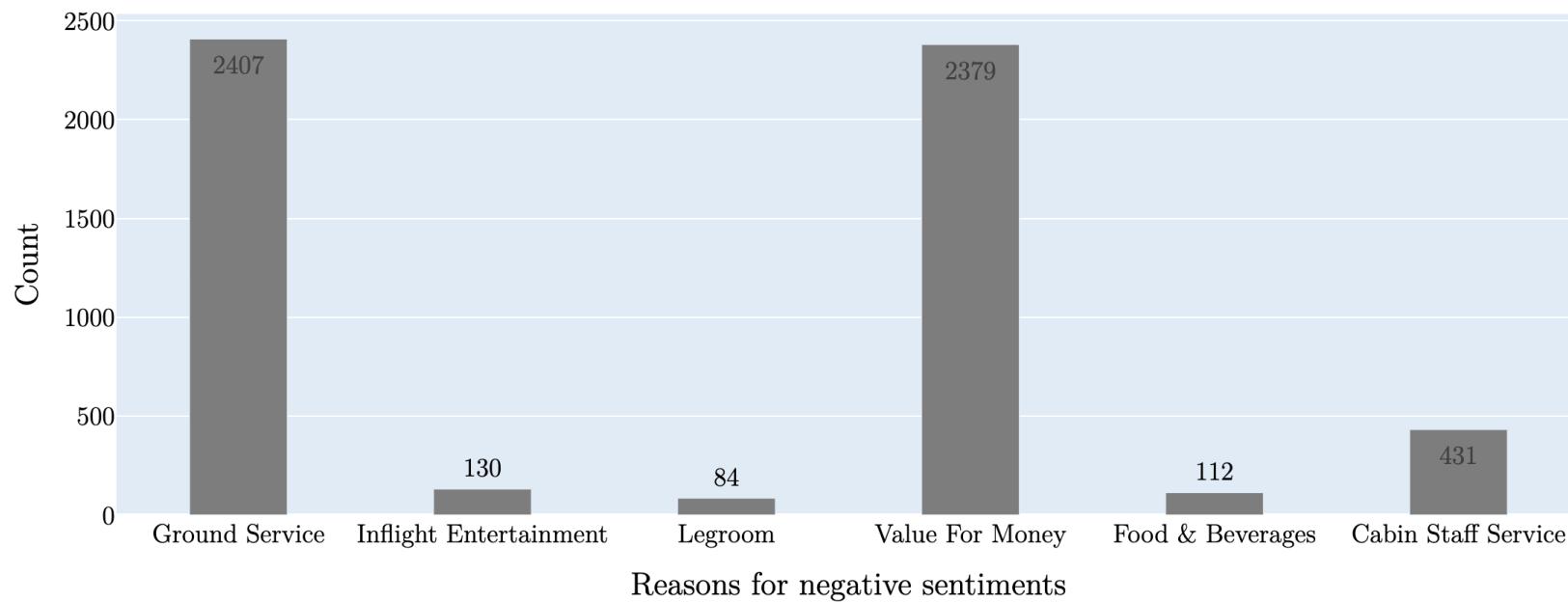
Analyse exploratoire des données



Première sélection à l'aide de Pandas

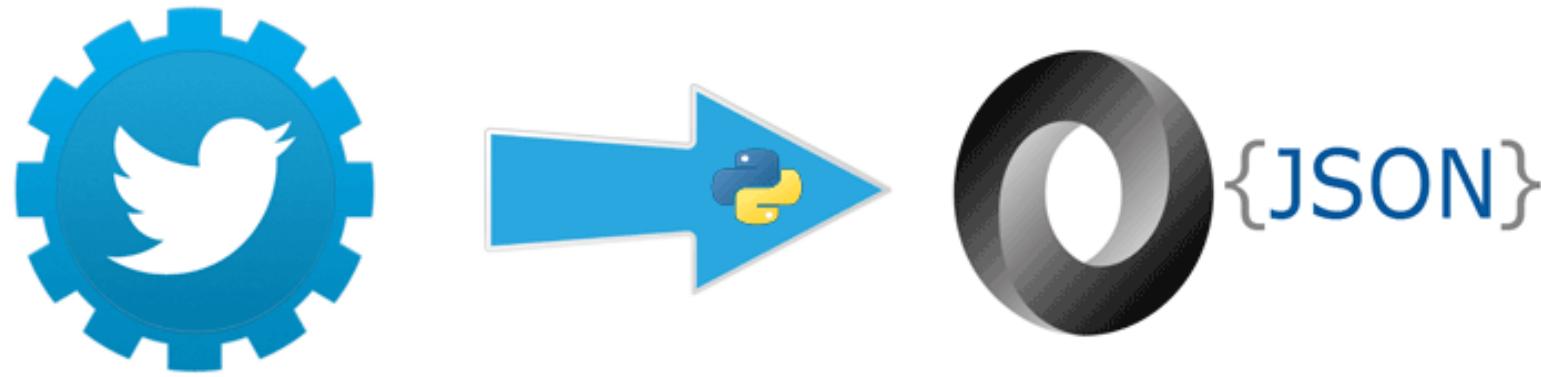
Deuxième sélection en fonction des mots clés

Analyse exploratoire des données



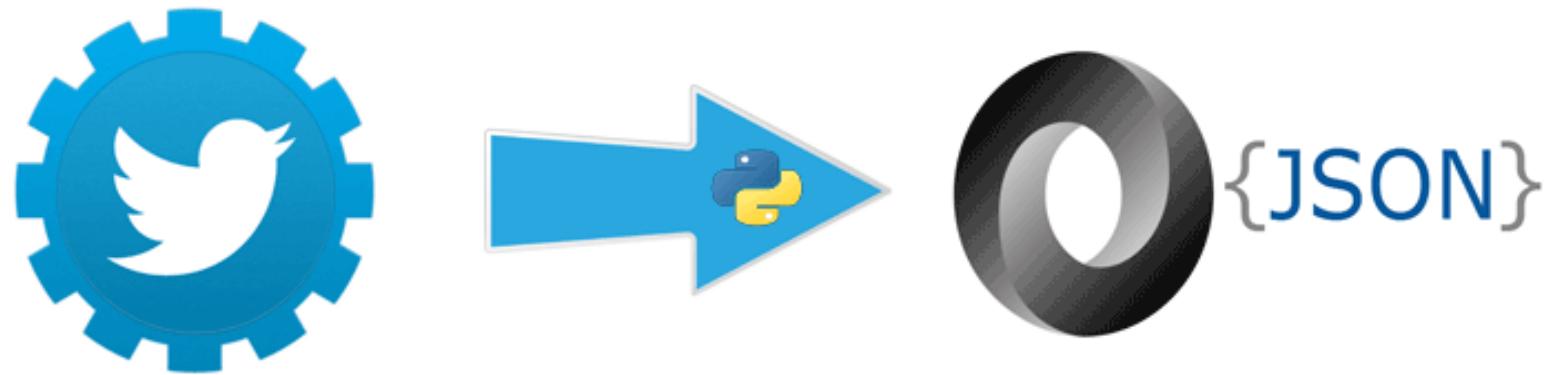
Réalisation et mise en œuvre

Collecte des données



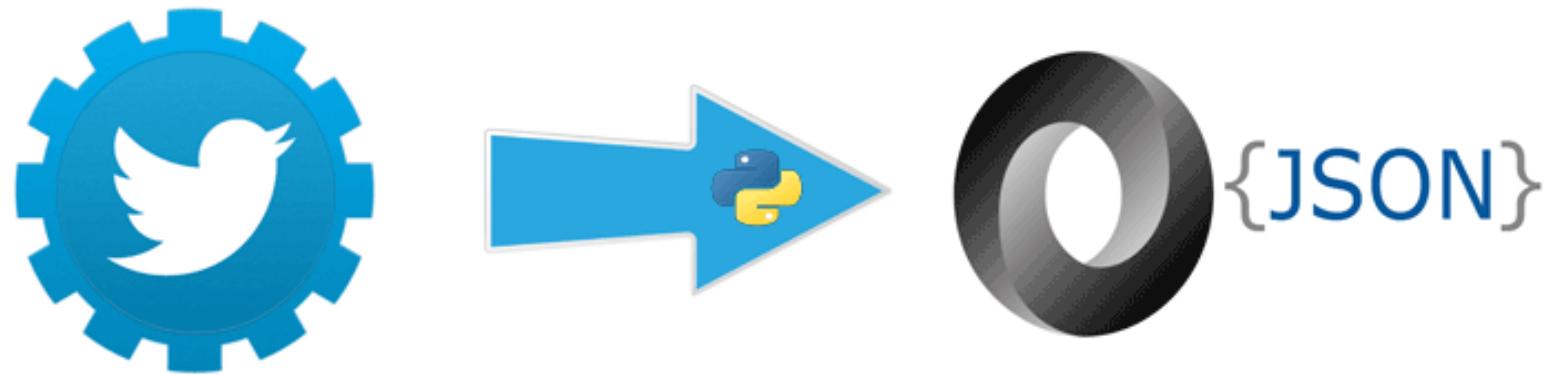
1. Accès à l'API Twitter.

Collecte des données



1. Accès à l'API Twitter.
2. Exploitation de la librairie Tweepy.

Collecte des données



1. Accès à l'API Twitter.
2. Exploitation de la librairie Tweepy.
3. Reformatage du fichier de sortie.

Préparation des données

1. Prétraitement des données

- La mise en minuscule

@VirginAmerica Flight from #BOS > LAS
tomorrow was Cancelled Flightled. No
notification; wait time...



@virginamerica flight from #bos > las
tomorrow was cancelled flightled. no
notification; wait time...

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @

@VirginAmerica @freddieawards Done and
done! Best airline around, hands down!



Done and done! Best airline around, hands
down!

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #

@united #UnitedAirlines Pls Fix
#AspenBaggageFail issues. Had to pay my
own way to #Aspen after flig...



@united UnitedAirlines Pls Fix
AspenBaggageFail issues. Had to pay my
own way to Aspen after flig...

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres

@united yes, a paper voucher that I got on
January 26th 2015

@united yes, a paper voucher that I got on
January th



Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres
- La conversion des emojis et emoticones

@VirginAmerica still waiting to see
@Starryeyes 



@VirginAmerica still waiting to see
@Starryeyes :disappointed:

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres
- La conversion des emojis et emoticones
- La suppression des URLs

@united Thanks for the reminder. It's been a fun ride. <http://t.co/pPVA4Rch9f>



@united Thanks for the reminder. It's been a fun ride.

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres
- La conversion des emojis et emoticones
- La suppression des URLs
- La suppression de la ponctuation

@united wonderful bag claim people _ finally
have my bag. But why did I catch earlier
flight?



@united wonderful bag claim people finally
have my bag But why did I catch earlier flight

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres
- La conversion des emojis et emoticones
- La suppression des URLs
- La suppression de la ponctuation
- La suppression des stop words

@united caught earlier flight to ORD. Gate checked bag, and you've lost it at O'Hare.



@united caught earlier flight ORD. Gate checked bag, lost O'Hare.

Préparation des données

1. Prétraitement des données

- La mise en minuscule
- La suppression des mentions @
- La suppression des hashtags #
- La suppression des chiffres
- La conversion des emojis et emoticones
- La suppression des URLs
- La suppression de la ponctuation
- La suppression des stop words
- Stemming

@United empathizes my disappointment that
1stClass flight nonworking entertainmnt is
worth \$50....



@United empath my disappoint that 1stClass
flight nonwork entertainmnt is worth \$50....

Préparation des données

2. Vectorisation de mots

- CountVectorizer
- TF-IDF Vectorizer

CountVectorizer

	also	love	programming
love programming	0	1	1
programming also love	1	1	1

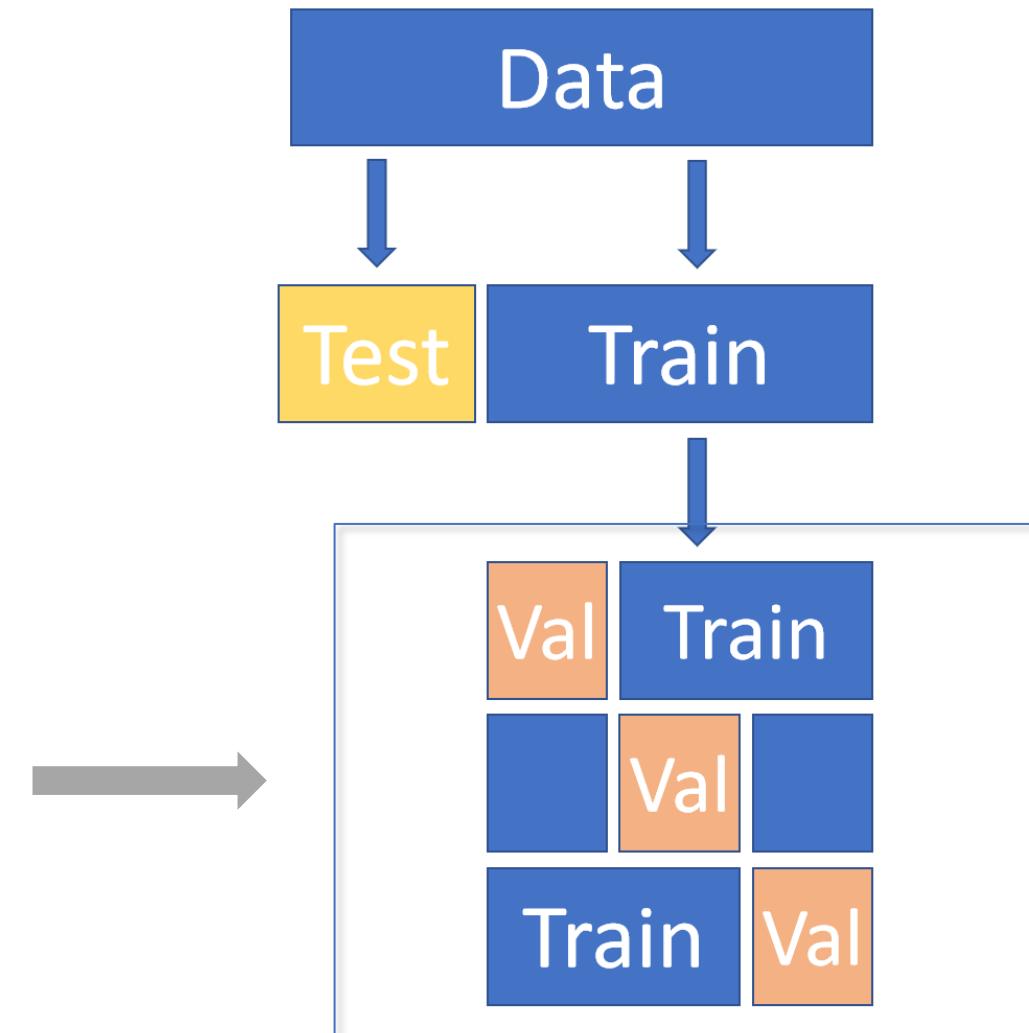
TF-IDF Vectorizer

	also	love	programming
love programming	0.000000	0.707107	0.707107
programming also love	0.704909	0.501549	0.501549

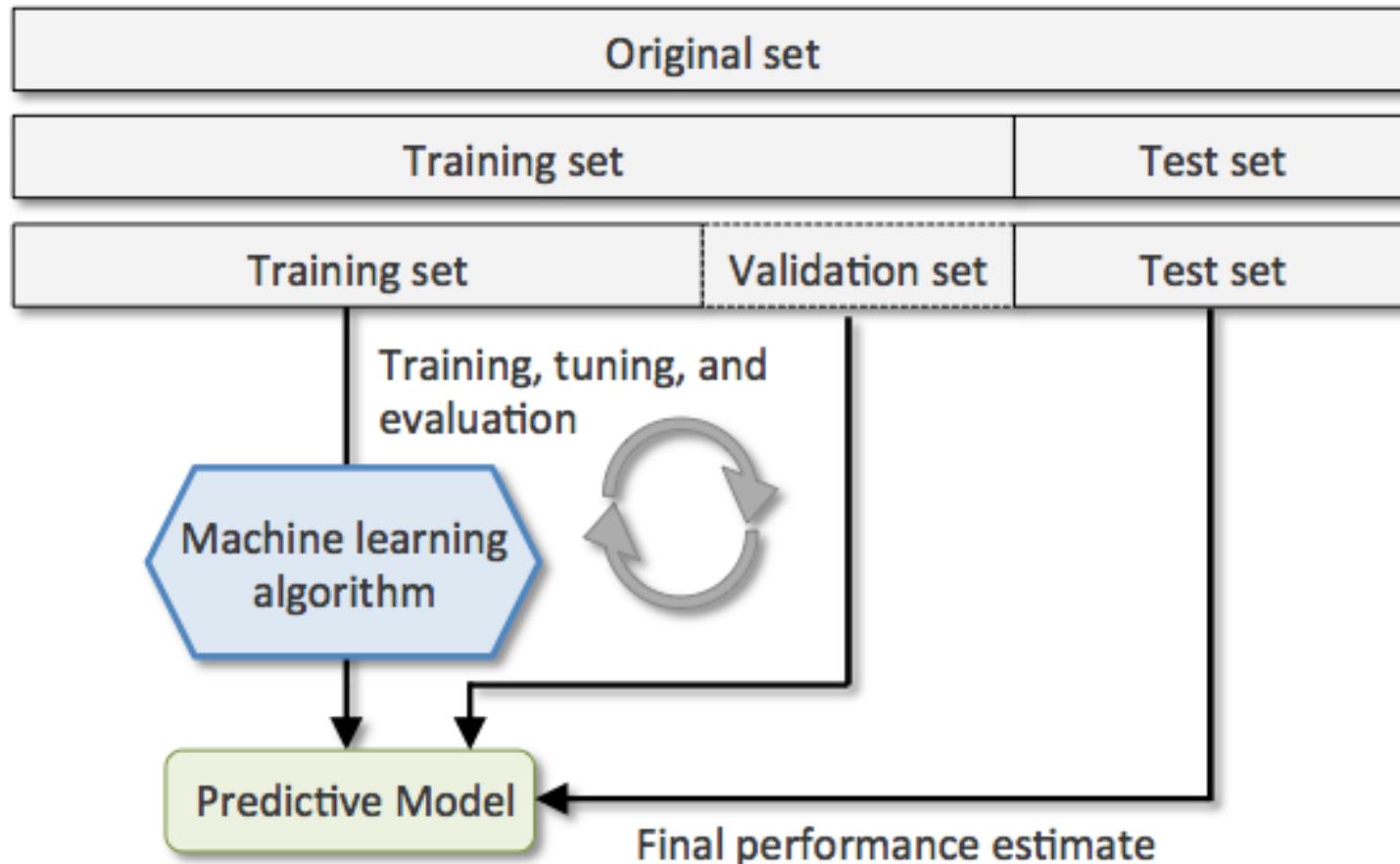
Préparation des données

3. Création de données d'apprentissage/test

- Cross-validation



Construction des modèles



Évaluation et résultats

		CountVectorizer				TfidfVectorizer			
		precision	recall	f-score	accuracy	precision	recall	f-score	accuracy
<i>Modèle du sentiment</i>	MultinomialNB	0.79	0.80	0.79	0.80	0.77	0.77	0.74	0.77
	LogisticRegression	0.80	0.81	0.80	0.81	0.79	0.78	0.79	0.79
<i>Modèle des raisons</i>	MultinomialNB	0.80	0.80	0.80	0.80	0.76	0.77	0.75	0.77
	LogisticRegression	0.78	0.78	0.78	0.78	0.78	0.77	0.76	0.77

L'interface

AIRCOMPARE

AER COMPARE

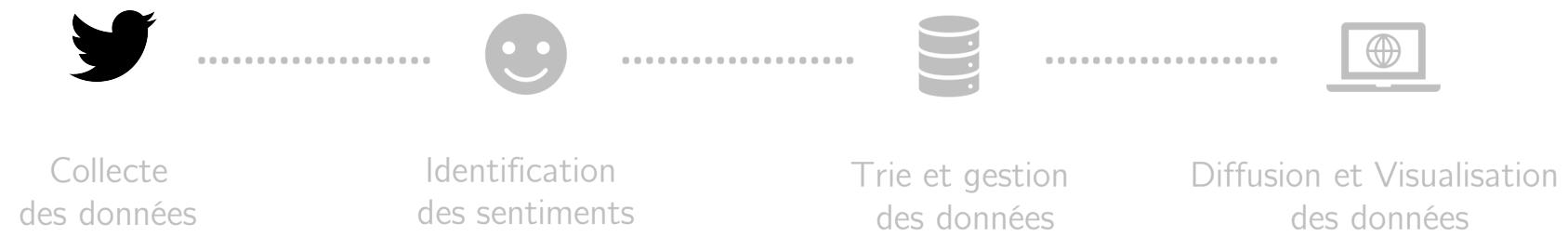
Construction des modèles



Application web



Conclusion



Merci pour votre attention !