

# PYTHON FOR DATA ANALYSIS

Par Samuel Pariente et Marius Ortega

# SOMMAIRE

**1** Contexte et Problématique

**2** Preprocessing

**3** Optimisation du succès

**4** Prédire votre succès

**5** Conclusion

61 variable  
(initialement)

58 variables  
prédictives

1 target (les  
partages)

Variables

39 797 articles

Provenant de  
Mashable.com

Publiés entre 2013  
et 2014

Articles



## ONLINE NEWS POPULARITY

*présentation du dataset*

Sources

INESC TEC

ALGORITMI  
Research Centre

Universidade de  
Aveiro

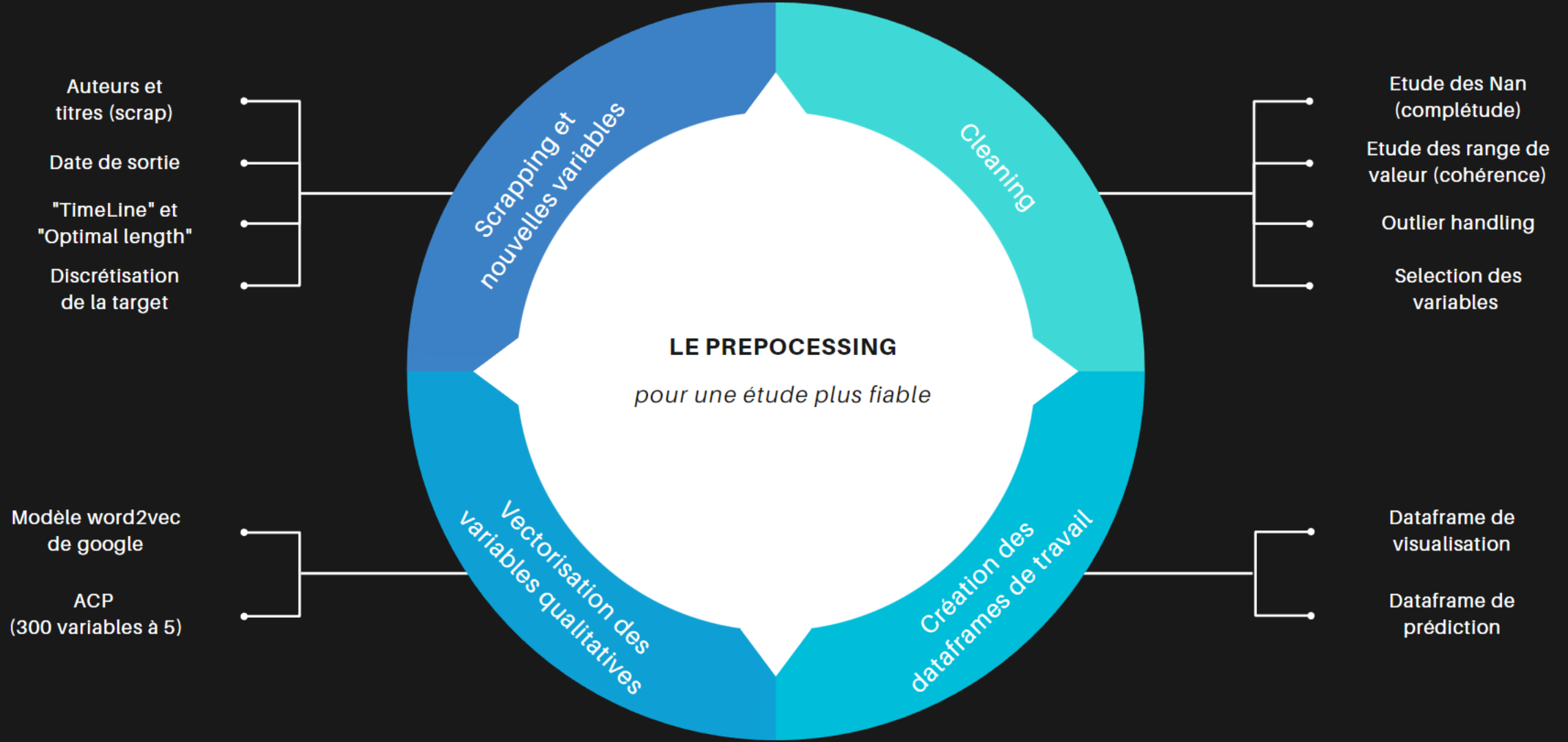
Specs

acquisition le  
08/01/2015

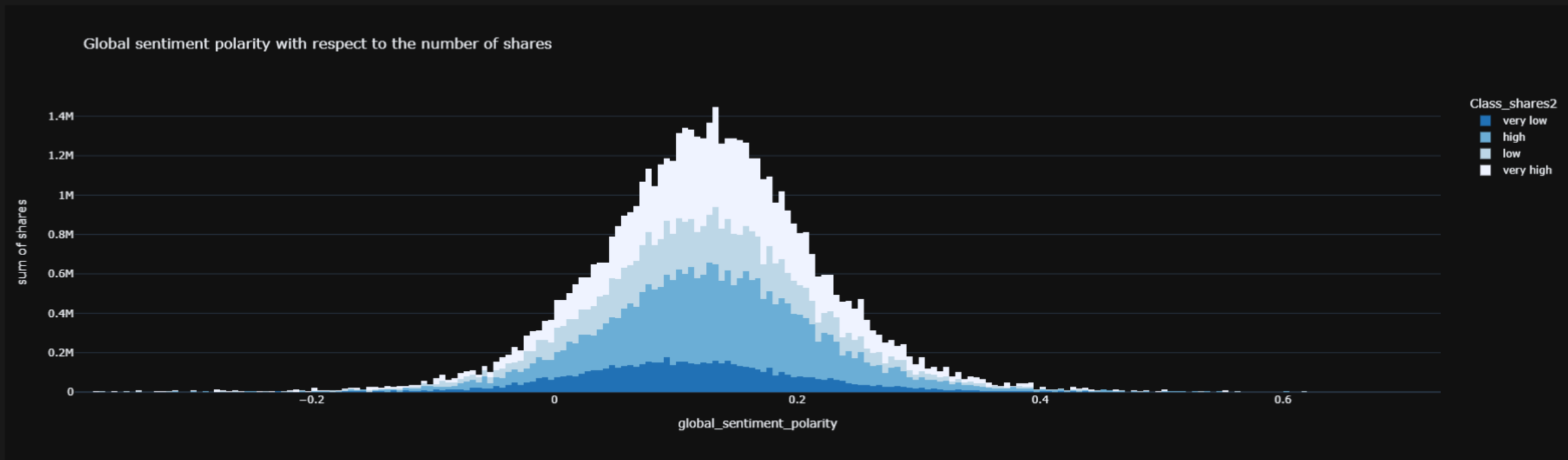
Pas de Nan

Attributs réels  
ou integer

### 1. CONTEXTE ET PROBLÉMATIQUE



## 2. PREPROCESSING



**SOYEZ POSITIF**

Polarité moyenne positive

Quel que soit le palier de nombre de partages

Number of shares per author and types of articles written



## DIVERSIFIEZ VOTRE CONTENU

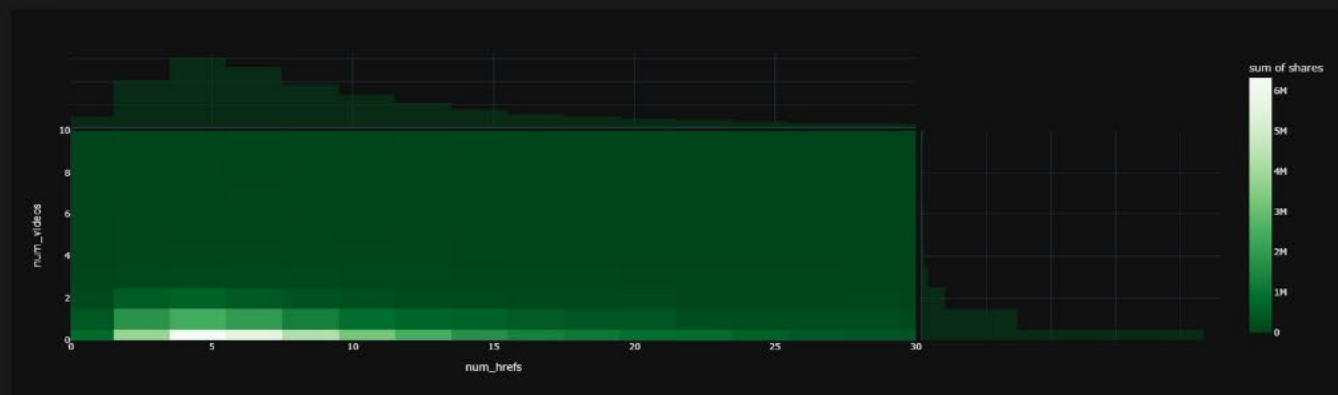
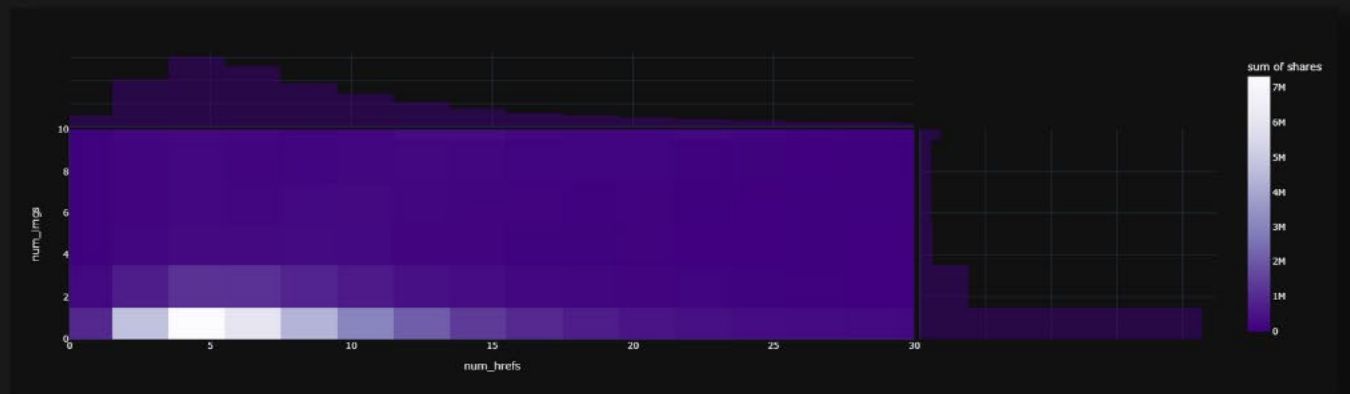
Inspirez-vous des plus grands

Publiez sur plusieurs topics différents



4-6 liens

Citez vos sources



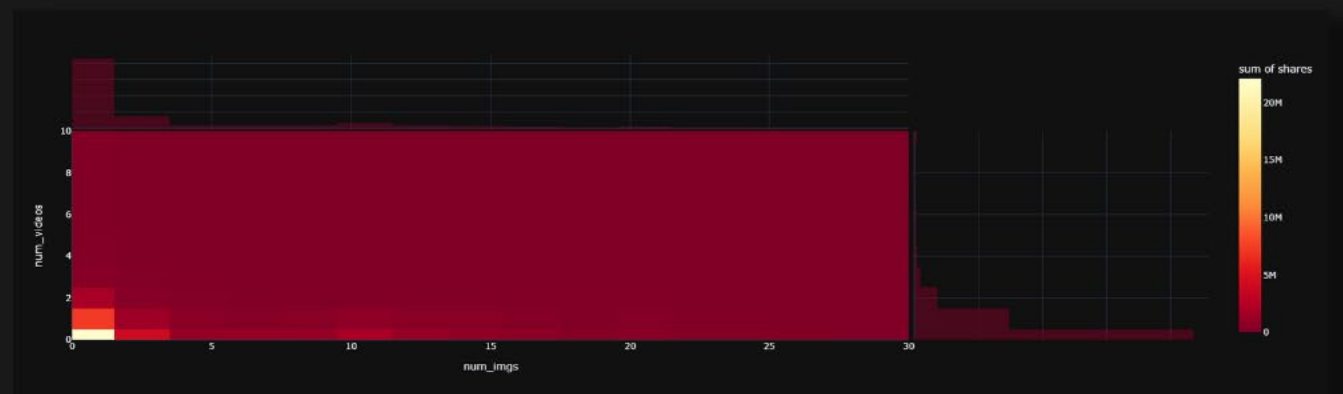
0 à 3 images

N'en faites pas trop



0 à 2 vidéos

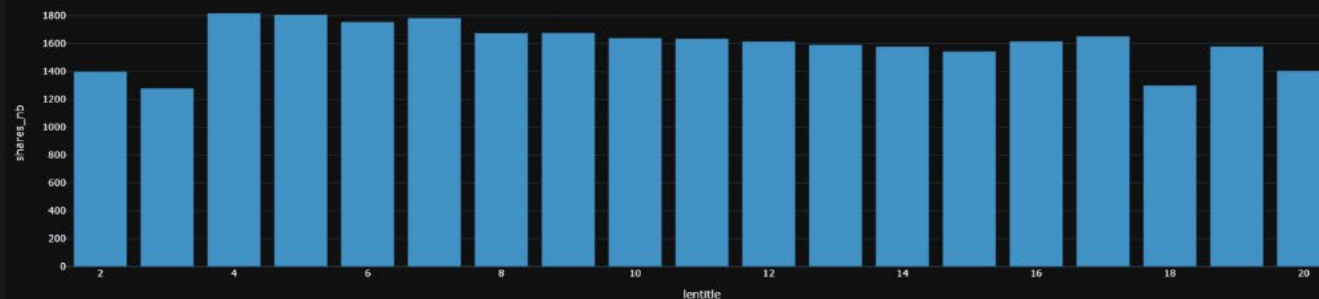
Insigh potentiellement biaisé



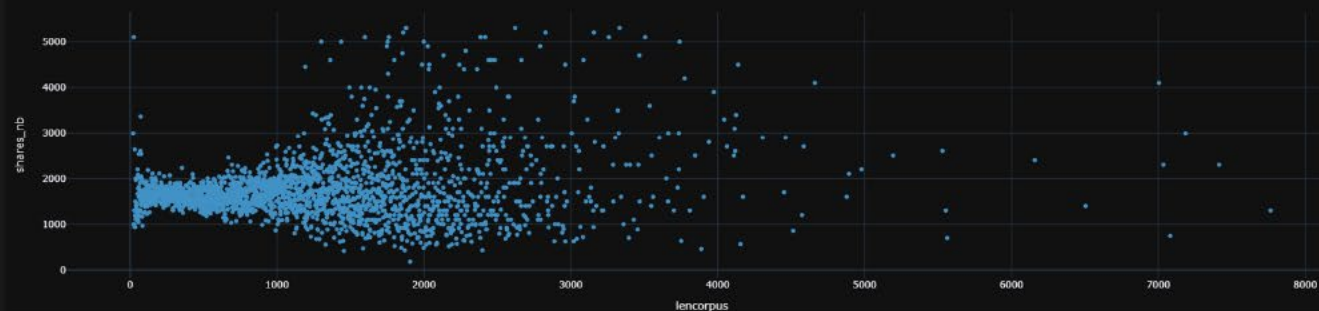
## Votre titre

- Soyez impactant : 4 à 7 mots
- Soyez précis : 16 à 20 mots
- Faites votre choix

Shares with respect to the Bar plot of the length of Titles



Scatter plot of the length of the body with respect to shares



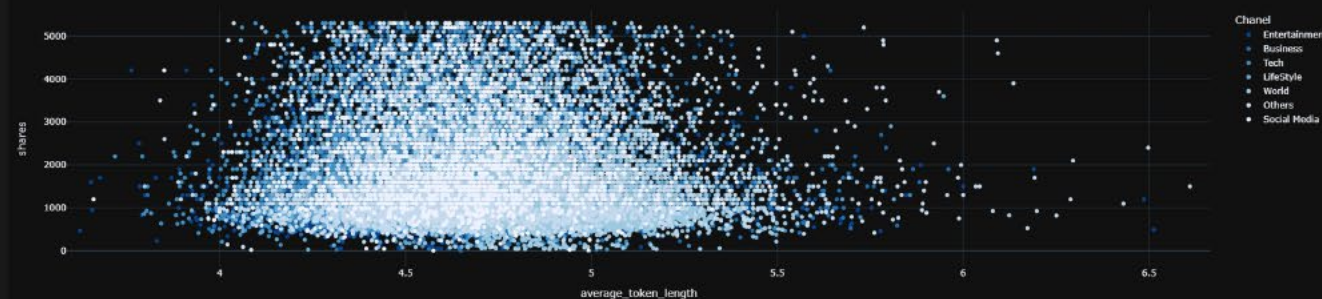
## Votre article

- Longueur optimale : 1000 à 3000 mots
- Trouvez votre équilibre

## Vos mots

- Taille optimale : 4 à 5 caractères
- Restez accessible

Complexity of words with respect to channels and shares



## 3. OPTIMISATION DU SUCCÈS



## Apprenez des meilleurs

Diversifiez votre contenu

## Restez accessible

Réfléchissez à votre format



## Pensez au lecteur

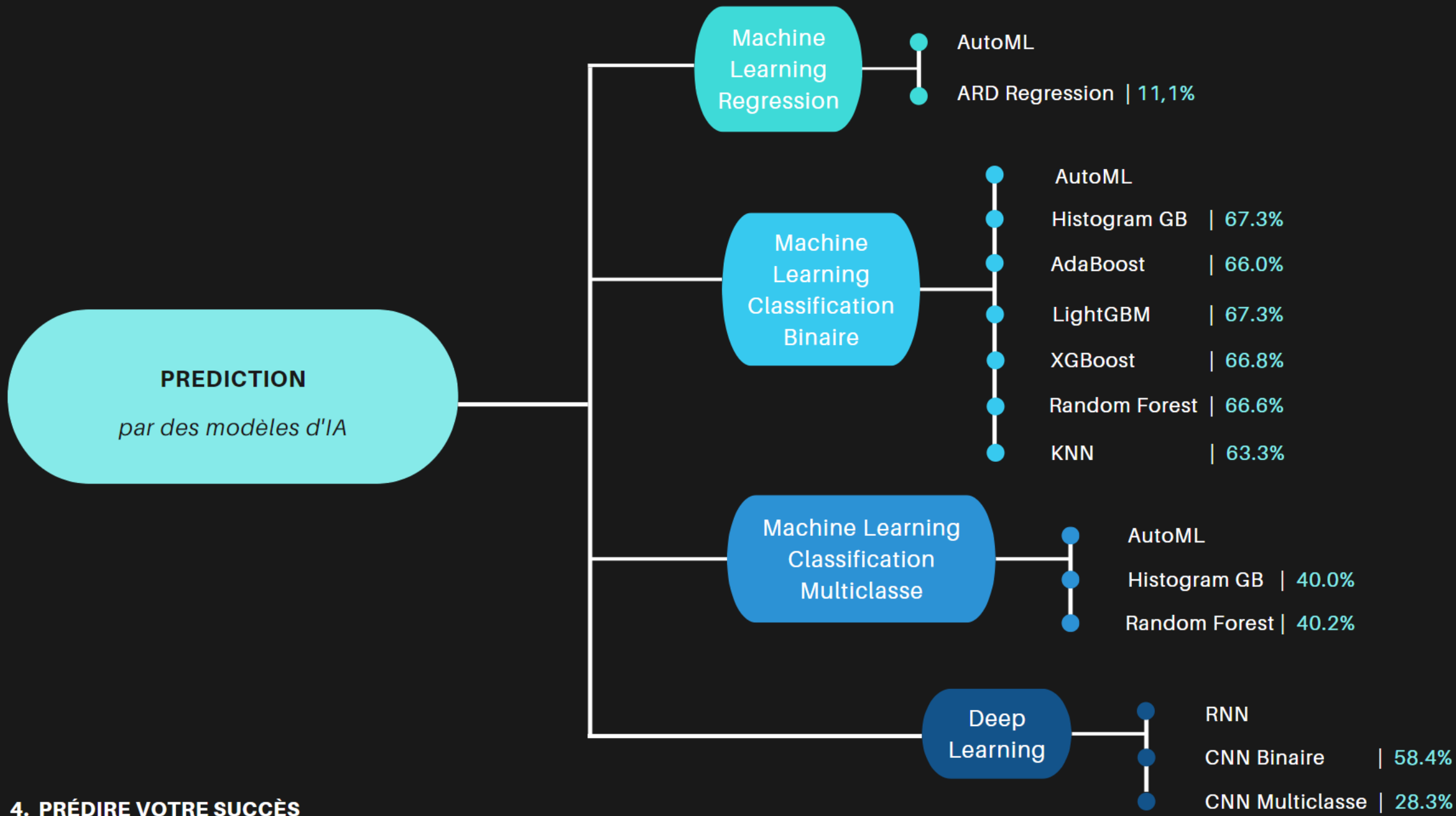
Soyez positif

## Soyez lisible

Choisissez vos médias

## Choisissez votre moment

Planifiez votre jour de sortie



#### 4. PRÉDIRE VOTRE SUCCÈS

## LIGHT GBM

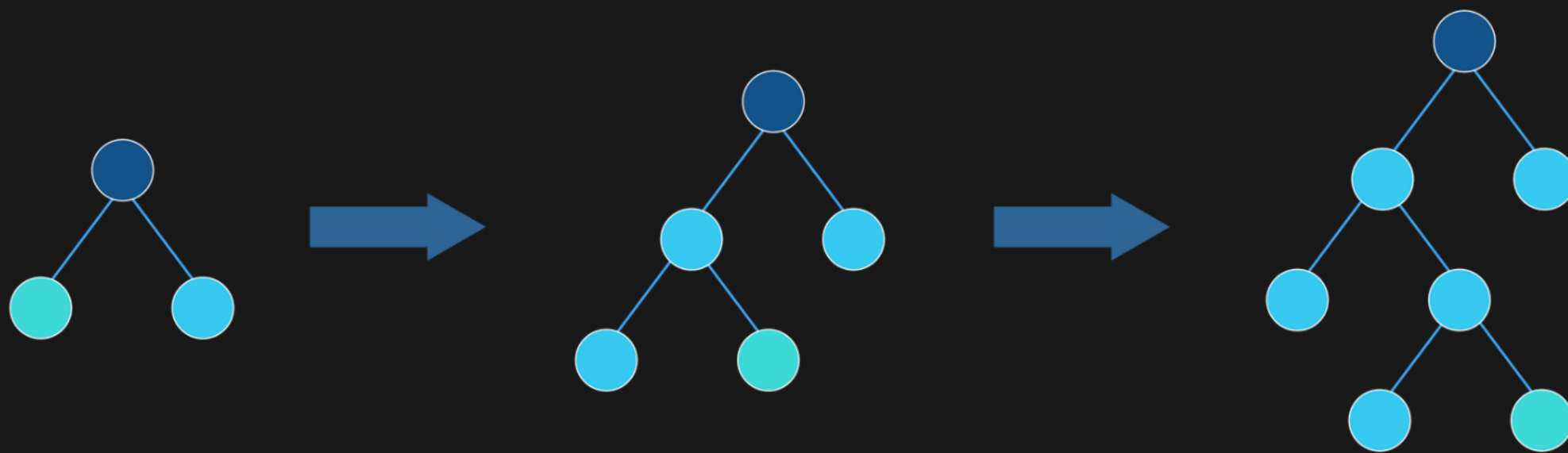
Open source

Developed by Microsoft

Gradient boosting framework

Base sur un accroissement  
basé sur les feuilles  
réduisant la perte

Possède les avantages du XGBoost



## Keras Sequential Neural Network

Sequential API

Sélectionnez le nombre  
de neurones par couche

Définition des fonctions  
d'activation (Relu, Softmax...)

Flexible

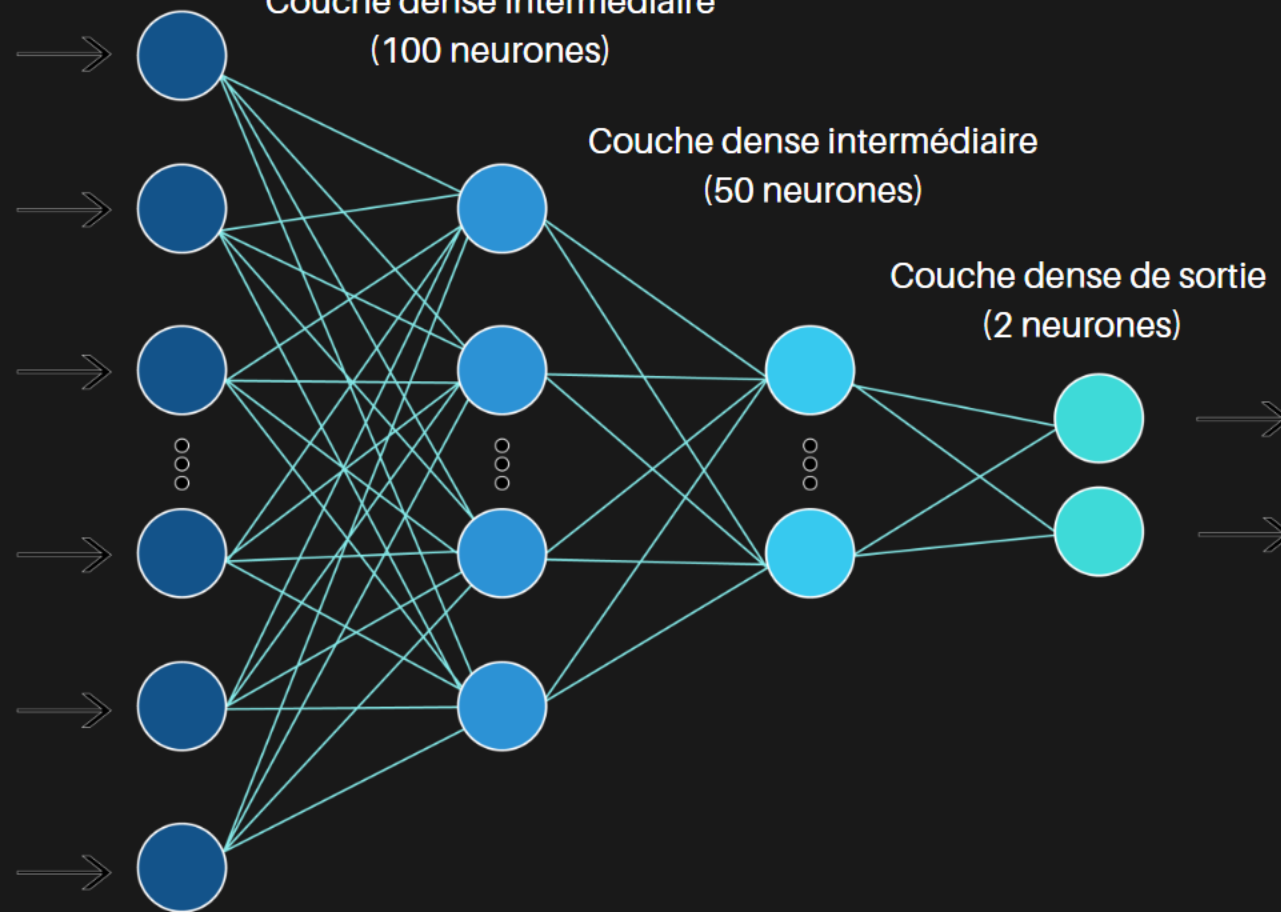
Intuitif

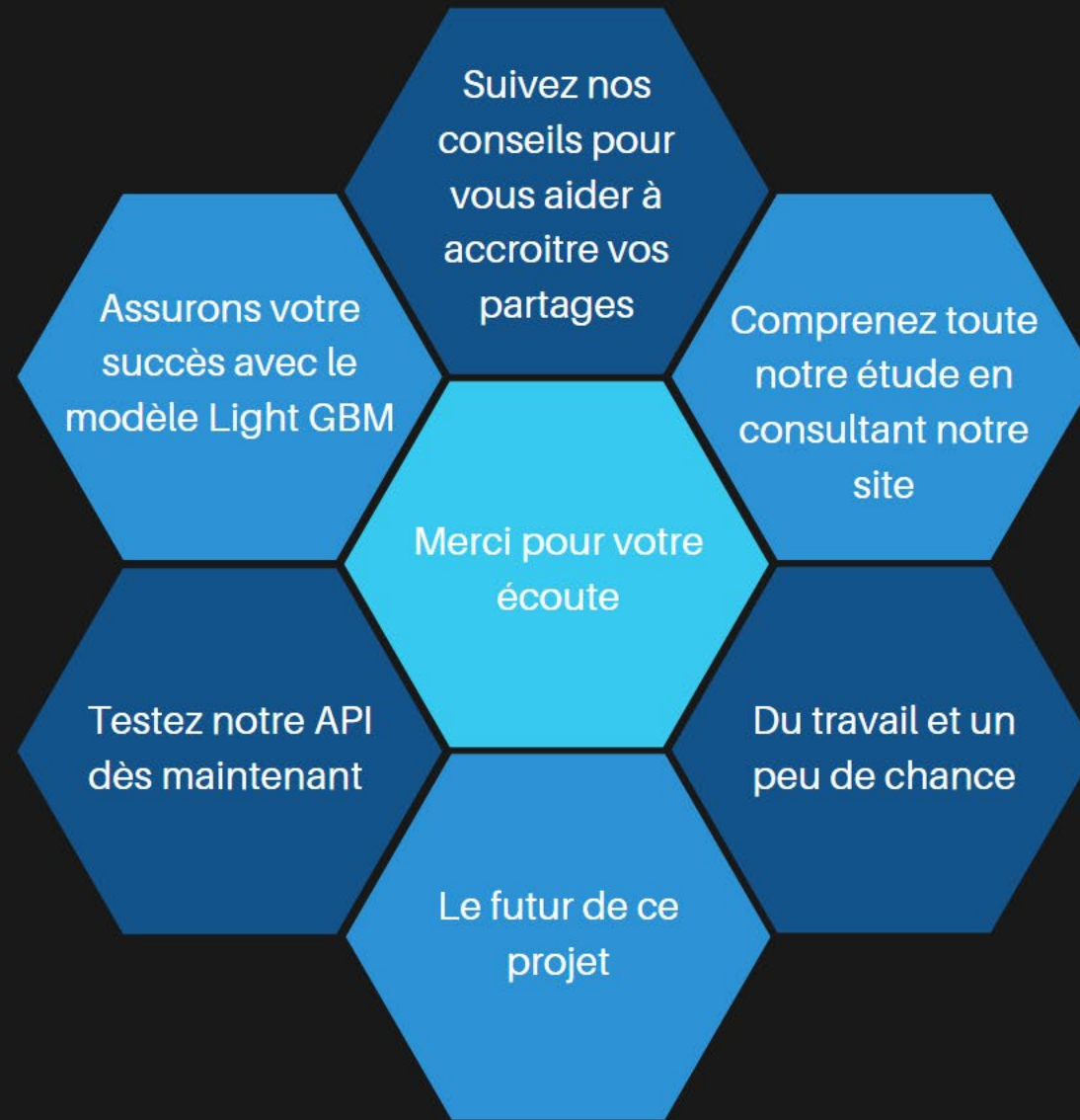
Couche Dense d'entrée  
(500 neurones)

Couche dense intermédiaire  
(100 neurones)

Couche dense intermédiaire  
(50 neurones)

Couche dense de sortie  
(2 neurones)





**ESSAYEZ PAR VOUS MÊME**

