

# IAPAU#3

*Le Festival Machine Learning,  
Data Science et Intelligence Artificielle*

HOUNKONNOU Mehdy  
LABARBARIE Pol  
MENENDEZ Benjamin  
THURIN Gauthier

# Nettoyage des données

## 1) Des doublons qui posent problème

“Homme à Homme”

vs

“homme a homme”

**PALLIER AUX NOMBREUSES ERREURS DE SAISIES**

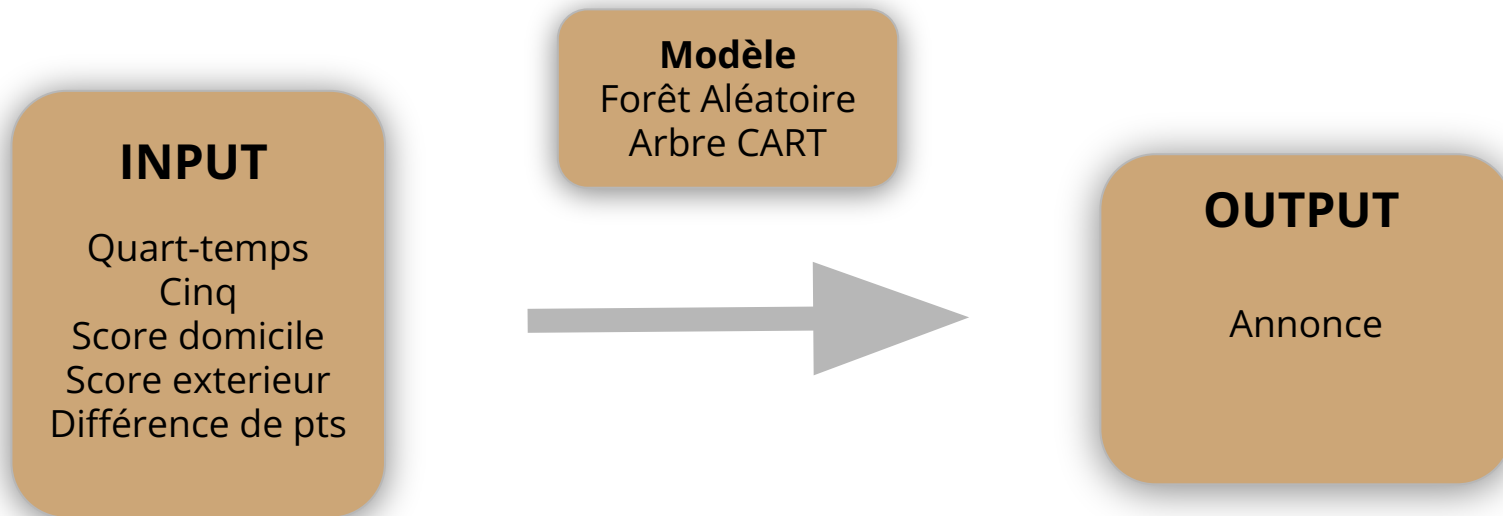
# Nettoyage des données

## 2) Des nouvelles variables, pour mesurer la positivité d'une annonce

	id	quart-temps	cinq_id	annonce_id	resultat_pos	score_home	score_adv	diff
0	64	1	1	8	0	0	0	0
1	65	1	1	34	1	0	0	0
2	66	1	1	169	1	2	0	2
3	67	1	1	169	1	0	0	2
4	68	1	1	200	0	0	2	0
...	...	...	...	...	...	...	...	...
23976	28097	4	817	202	0	0	2	11
23977	28099	4	817	202	0	0	2	9
23978	28100	4	811	80	1	0	0	9
23979	28101	4	811	80	1	1	0	10
23980	28102	4	811	80	1	1	0	11

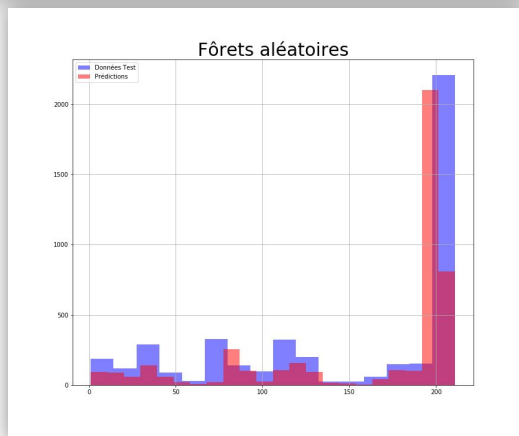
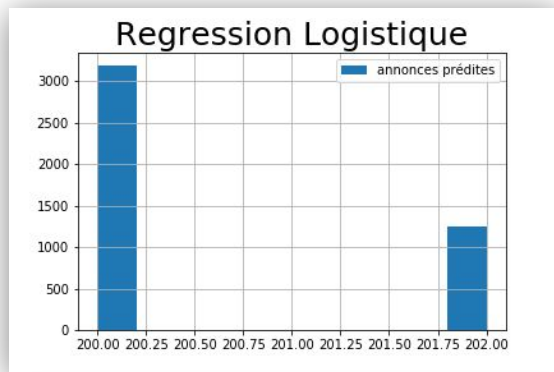
22562 rows x 8 columns

# Modèles établis



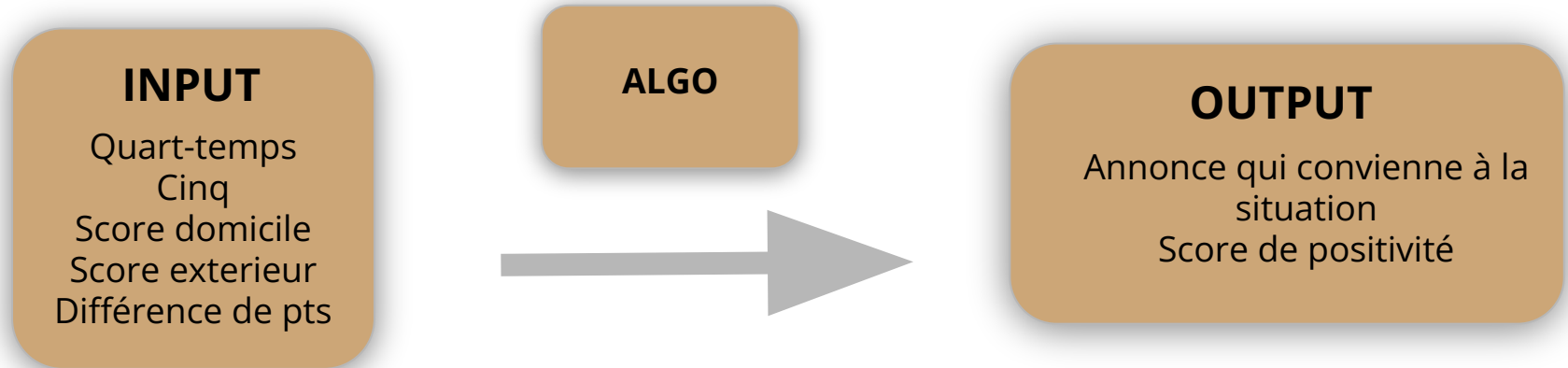
**PB** : sur-représentation des classes d'annonces majoritaire (*ex: id 200 Homme à Homme*)

# Résultats et Interprétation



- Régression logistique : prédiction des classes majoritaires uniquement
- Forêts aléatoires : mieux !
  - Amélioration : prédire un score de positivité associé
  - Nécessite l'avis d'un expert

# Ambition et perspectives



# Ambition et perspectives

- Base de données Annonces à optimiser avec un professionnel
- Intégration de la performance des joueurs
- Algorithme en temps réel

# Qui sommes nous ?

Étudiants MSS (Modélisation Statistiques et Stochastiques) de Bordeaux formés aux métiers de l'ingénierie stochastiques. Passionnés et attentifs à la veille technologique.

**HOUNKONNOU Mehdy**

**LABARBARIE Pol**

**THURIN Gauthier**

**MENENDEZ Benjamin**