

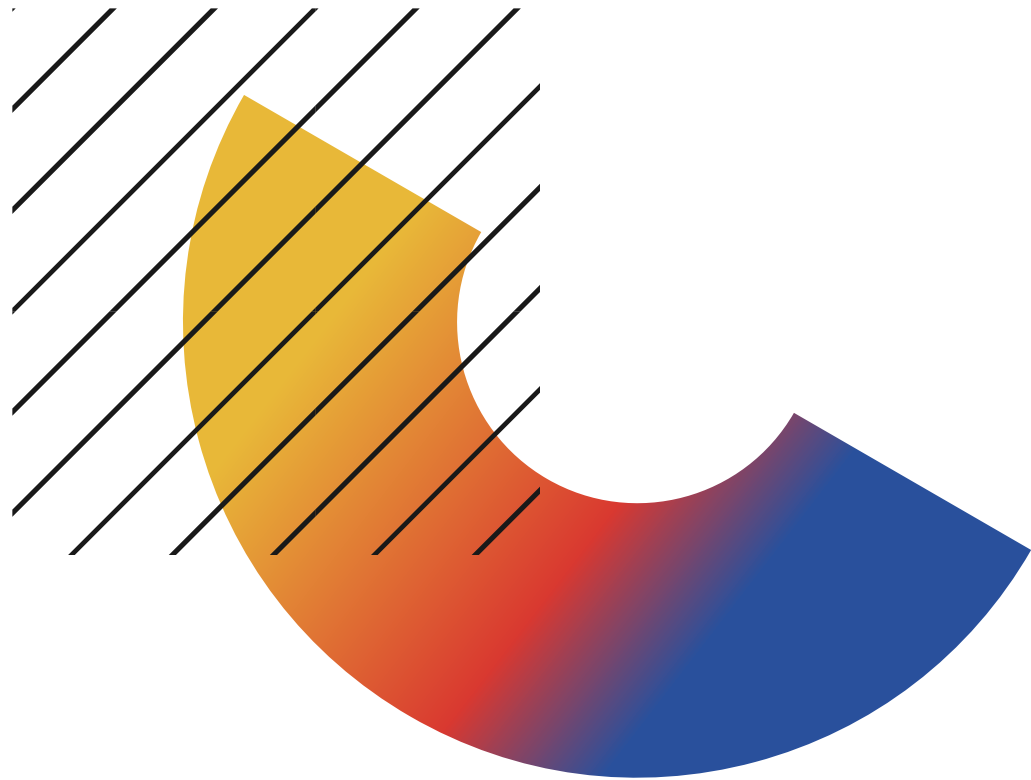


BUREAU D'ÉTUDE S&S

# Projet MRR

Par Naïm SOUNI et Abdoulaye SAKHO  
(Groupe 10)

# Introduction



## Mise en contexte

- Qui sommes-nous ?
- Client
- Objectif

# Base de Données



## Présentation

- Observations:  $n = 10\,308$
- Variables explicatives:  $p = 376$  ( dont 1 qualitative )



## Restructuration

- Compression des images et éliminations des observations aberrantes:  $n' = 9596$  et  $p' = 29$



## Cross Validation

# Les modèles utilisés



Type de modèle 1

Persistence

$Y = aC(\text{windspeed})$

Référence

Type 2

Régression  
linéaire simple

Type 3

Méthodes  
incrémentales

Forward

Bacward

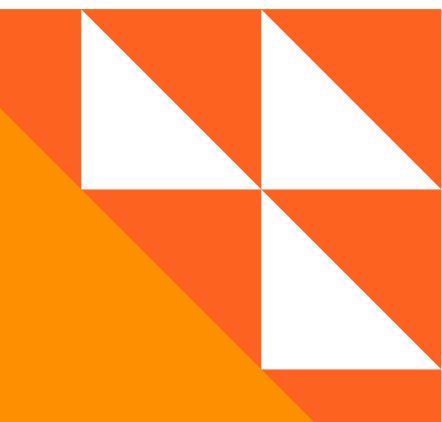
Bothward

Type 4

Lasso

Type 5

Ridge



# Les résultats obtenus:

Sous Cross validation (spéciale)

Hypothèses des modèles:

- Gaussianité des résidus
- Homoscédalité des résidus

Compression:

- Résultats très proches de ceux de la non compression .  
( à 0.5 près pour le RMSE et 0.1 pour R-carré )
- A permis convergences des méthodes incrémentales en un temps décent

RMSE moyen:

- Persitence: 25
- Autres : 12

R-carré moyen:

- Persitence: 0.05
- Autres : 0.74

Variance RMSE:

- Persitence: 7
- Autres : 0.4

Variance R-carré:

- Persitence: 0.02
- Autres : 0.003

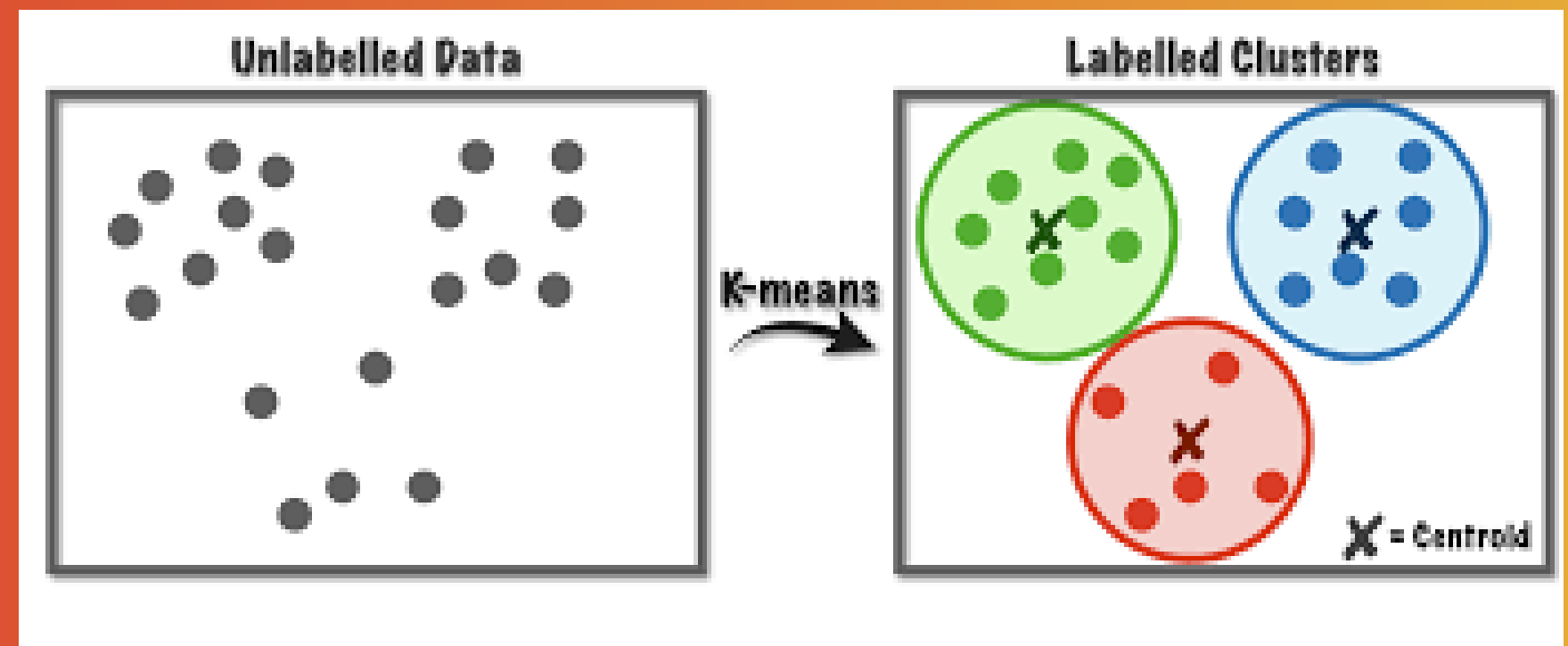
# Pour aller plus loin:

Limites des modèles dit "strictement linéaire"

Un modèle pour aller plus loin:

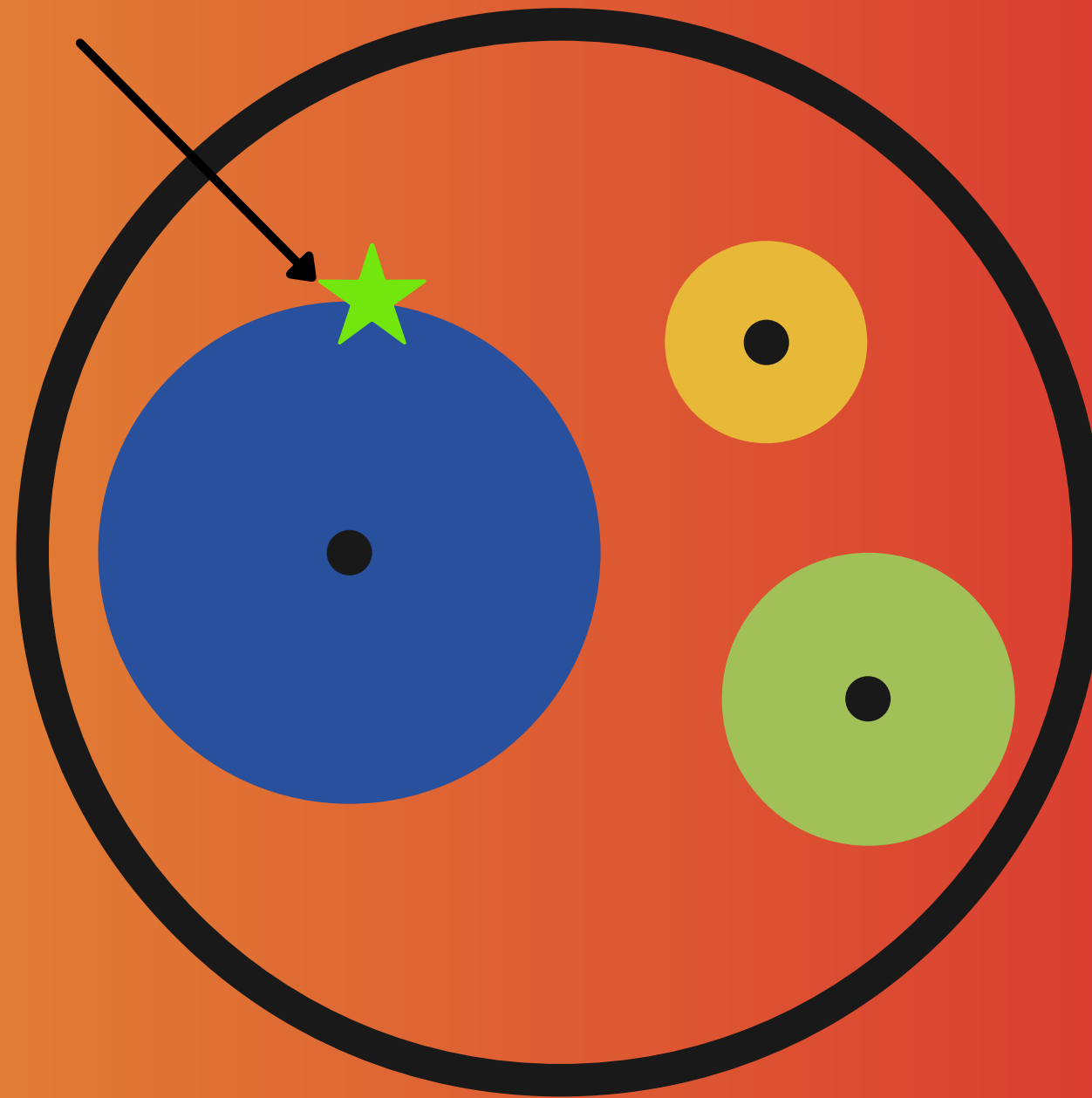
- Clustering des données d'apprentissages en K cluster à partir des variables explicatives
- Construction d'un modèle linéaire pour chaque cluster
- Prédictions

## 1. Description des k-means:



## 2. Fonctionnement du nouveau modèle:

I - Une observation de test est affecté au cluster dont il est le plus proche (par rapport aux barycentres)



II- Le modèle linéaire du cluster en question est utilisé pour prédire la valeur cible d'étoile

# Résultats de ce type de modèle:

- Dans l'ensemble, similaire à ceux des modèles dits "strictement linéaires"



# Conclusions sur le projet:

- Construction de meilleurs modèles que celui de Persitance
- Insuffisance des modèles pour atteindre notre objectif:

catégorie	vent moyen en km/h
1	118 - 153
2	154 - 177
3	178 - 210
4	210 - 249
5	+ 249

Échelle de Saffir-Simpson

1 noeud = 1,85km/h

12 noeuds = 22,52km/h

- Pousser plus loin la recherche d'un modèle adéquat



# Merci !

N'hésitez pas à nous contacter  
si vous avez des questions.



BUREAU D'ÉTUDE S&S



# Les résultats détaillés

(présentation indigeste)

A l'aide de la compression:

	mean(RMSE) <dbl>	var(RMSE) <dbl>	mean(R-squared) <dbl>	var(R-squared) <dbl>
simple	24.83048	7.3622599	0.04231985	1.777095e-02
lm	12.64951	0.4171164	0.74046021	3.198731e-03
Bacward	12.66731	0.4157671	0.73955886	3.155247e-03
Forward	12.66446	0.4197694	0.73920921	1.616284e-05
Both	12.66731	0.4157671	0.73921691	1.594405e-05
Lasso	12.64963	0.4164492	0.73952146	3.201413e-03
Ridge	12.64851	0.4140842	0.73715697	3.162626e-03

Sans la compression:

	mean(RMSE) <dbl>	var(RMSE) <dbl>	mean(R-squared) <dbl>	var(R-squared) <dbl>
lm	12.69944	0.4240635	0.9644352	0.020434849
Bacward	0.00000	0.0000000	0.0000000	0.0000000000
Forward	0.00000	0.0000000	0.0000000	0.0000000000
Both	0.00000	0.0000000	0.0000000	0.0000000000
Lasso	12.45893	0.3537939	0.7499010	0.003697000
Ridge	12.47618	0.3777993	0.7369234	0.003642341