

DEFI-IA 2022

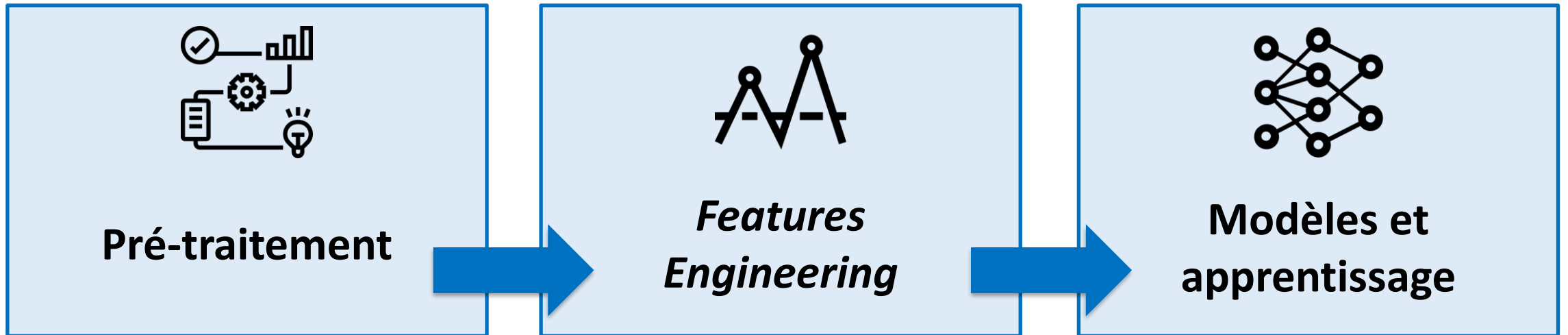
Prédire les précipitations cumulées en des stations météorologiques

Tuteurs : David Bertoin
Adil Zouitine

Quentin DOUZERY
Alexia GHOZLAND
Dario MOED

5^{ème} année – Mathématiques appliquées

Vendredi 7 Janvier 2022
Année Académique 2021/2022



0. Jeux de données

$X_{station}$



Observations météorologiques au jour **J-1** :
*wind speed, wind dir., temperature,
dew point temp., humidity, precip*

$X_{forecast}$



Prévisions de *METEO FRANCE* pour le jour **J**

Y



Précipitations cumulées pour le jour **J**

I. Pré-traitement

Ajout des coordonnées géographiques

Latitude, longitude, altitude

Alignement des échantillons entre eux

Alignement de X_{train} et y_{train}
Alignement de X_{test} sur X_{train}

Gestion des données manquantes

Suppression des *NaNs*
ou Méthode *backfill*

Suppression des outliers

Suppression des valeurs impossibles
Suppression des valeurs extrêmes
Suppression de certaines stations

Normalisation des données

Normalisation de X_{test} avec μ_{train} et σ_{train}

II. *Features engineering*

Transformation des variables initiales

**Somme des précipitations sur la
journée**

**Moyenne des variables
(excepté *precip*)**

- Journée entière
- Distinction nuit/journée/soir
- 2 méthodes: *all/just24*

**Récupérer la dernière valeur
enregistrée de la journée**

- Uniquement si la journée comporte 24 heures
- Capturer une certaine temporalité

II. *Features engineering*

Ajout de nouvelles variables

Récupération du mois et de la saison

- Saisons définies en fonction des précipitations

Calcul de la *smooth mean* associée

- Estimateur Bayésien
- Calculée sur la variable *precip*
- Éviter le surapprentissage

$$\mu = \frac{n \times \bar{x} + m \times w}{n + m}$$

n : nombre de valeurs
 \bar{x} : estimation de la moyenne
 w : globale moyenne
 m : poids attribué à la moyenne globale

II. *Features engineering*

Conversion de la variable *month*

Expression de *month* sous la forme d'une combinaison d'un sinus et d'un cosinus

- Fonctions cycliques (2π -périodiques)
- Récupérer l'information temporelle

Observation	<i>month</i>
x_0	1
x_1	1
...	...
x_n	12



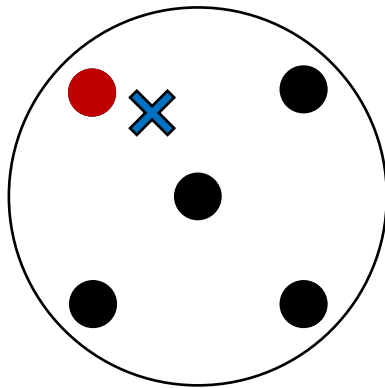
Observation	<i>cos_month</i>	<i>sin_month</i>
x_0	0.87	0.50
x_1	0.87	0.50
...
x_n	1	0

II. *Features engineering*

Incorporation du modèle physique

Ajout de la prévision météo

- Utilisation de la *baseline forecast*
- Récupération au point le plus proche de la station

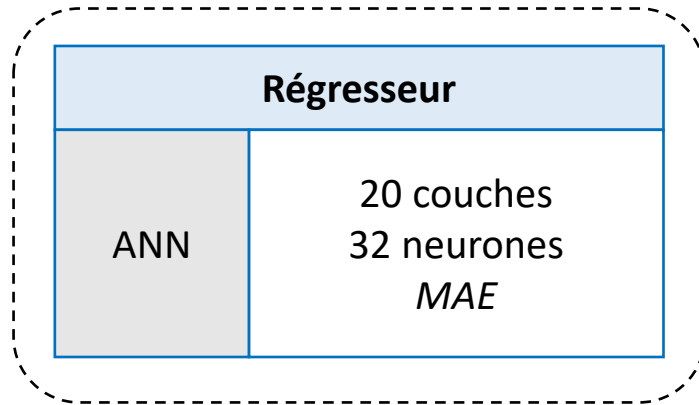


- Prévisions du modèle physique
- × Station
- Prévision attribuée à la station

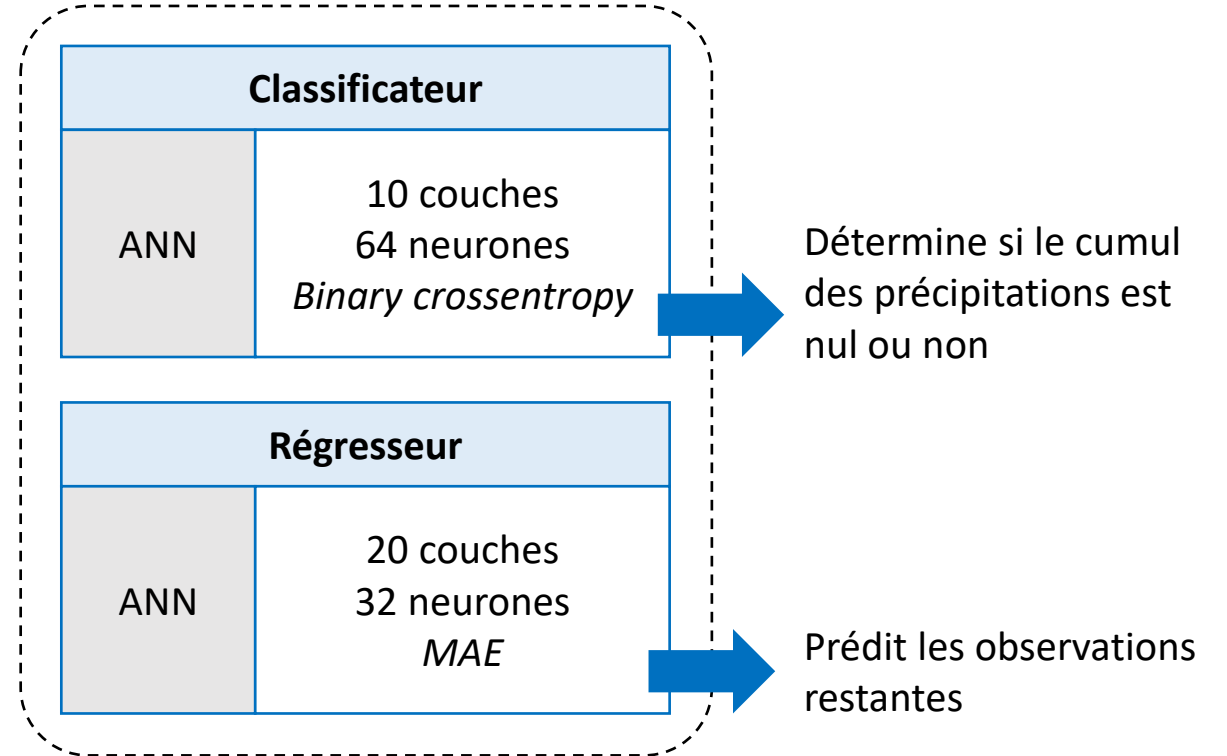
III. Modèles et apprentissage

Les différents modèles implémentés

Régresseur simple

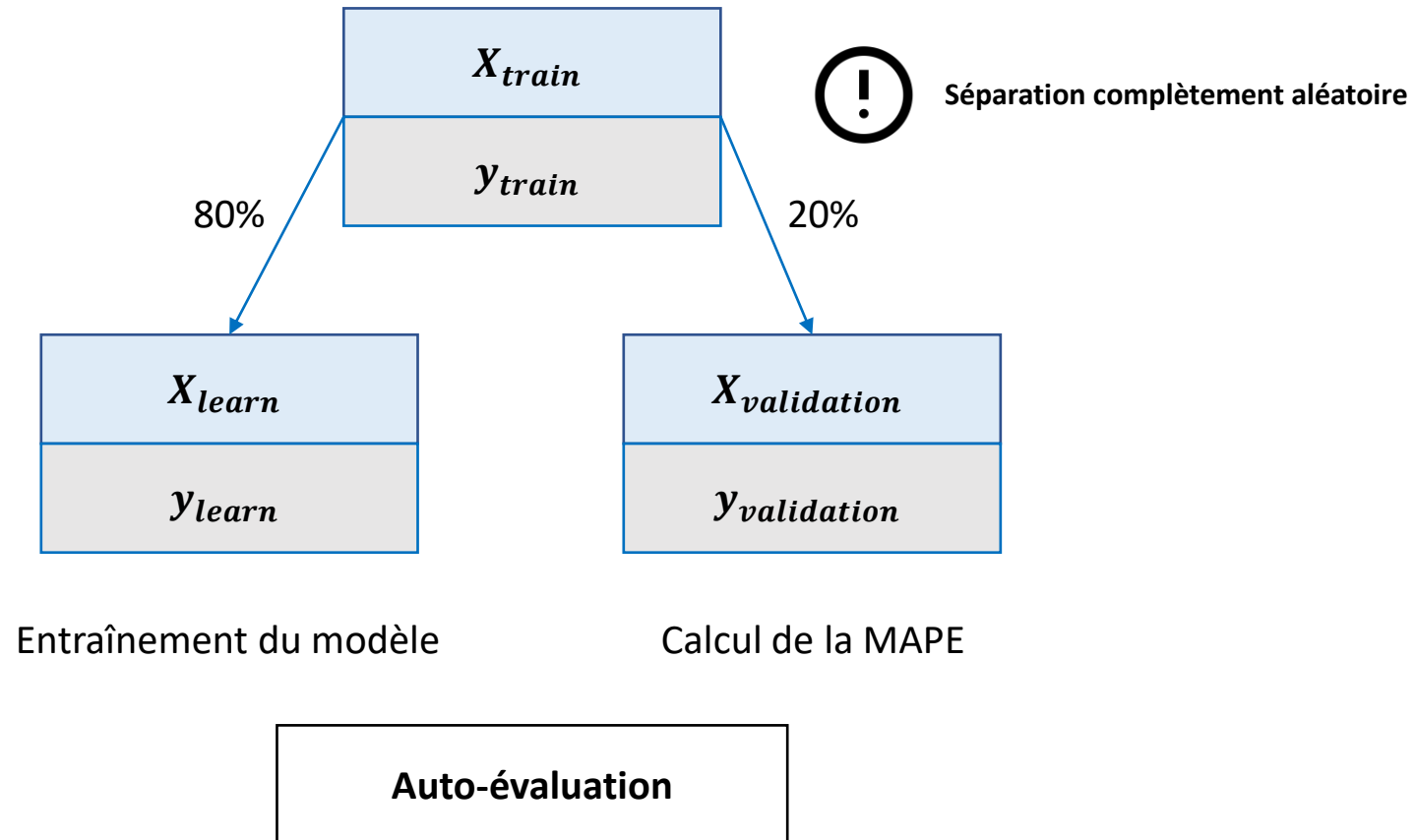


Classificateur + Régresseur



III. Modèles et apprentissage

Méthode de validation et paramètres d'apprentissage



Hyperparamètres

Époques = 20
Batch size = 200

« Optimisation » à la main, en tâtonnant

Discussion

K-means

Partitionnement des stations en clusters
Entraînement/Test sur chacun des clusters

$X_{forecast}$

Prise en compte des prévisions météo du jour J
Considérer les **N** prévisions **les** plus proches

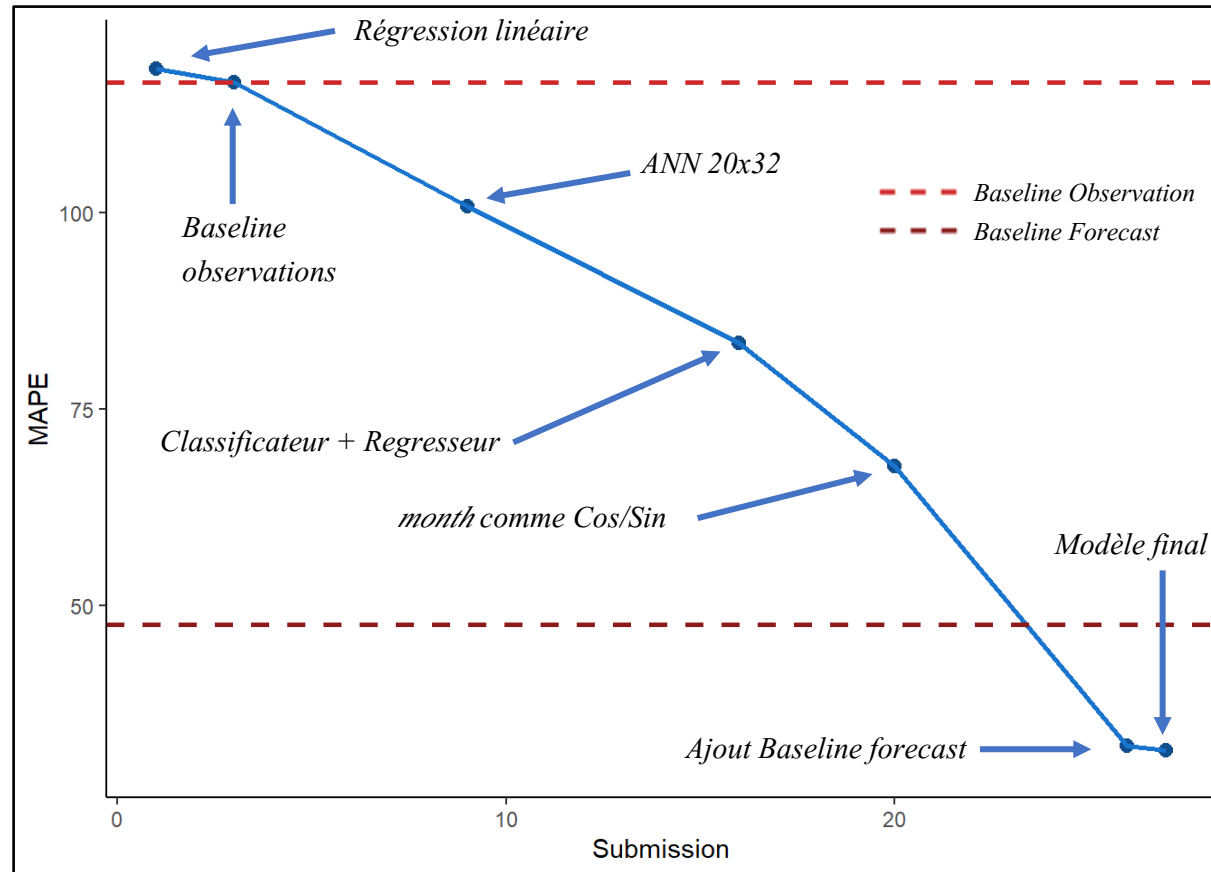
Distance [station – mer]

Ajout d'une *feature* pour considérer l'influence maritime

**MAPE comme fonction
coût**

La MAPE pénalise surtout les surestimations
Intérêt de n'avoir que des petites prédictions

Évolution de la MAPE en fonction des méthodes utilisées



Merci pour votre attention

Quentin DOUZERY
Alexia GHOZLAND
Dario MOED

5^{ème} année – Mathématiques appliquées

Tuteurs : David Bertoin
Adil Zouitine

Vendredi 7 Janvier 2022
Année Académique 2021/2022