



PRÉDICTION DU COÛT HOSPITALIER ANNUEL POUR DES PATIENTS VIVANTS AVEC LE VIH

COMPARAISON DE 14 ALGORITHMES DE MACHINE LEARNING SUR UNE
COHORTE IDENTIFIÉE DANS LES DONNÉES PMSI

DE OLIVEIRA Hugo, PRODEL Martin, VAINCHTOCK Alexandre

Congrès ADEL F EMOIS
30 mars 2018

1/Cas d'étude

Patients vivants
avec le VIH en
France (2013)

30 169
Nb. patients

180M €
Coût total

5 981 €
Coût moyen

HEVA



Sexe



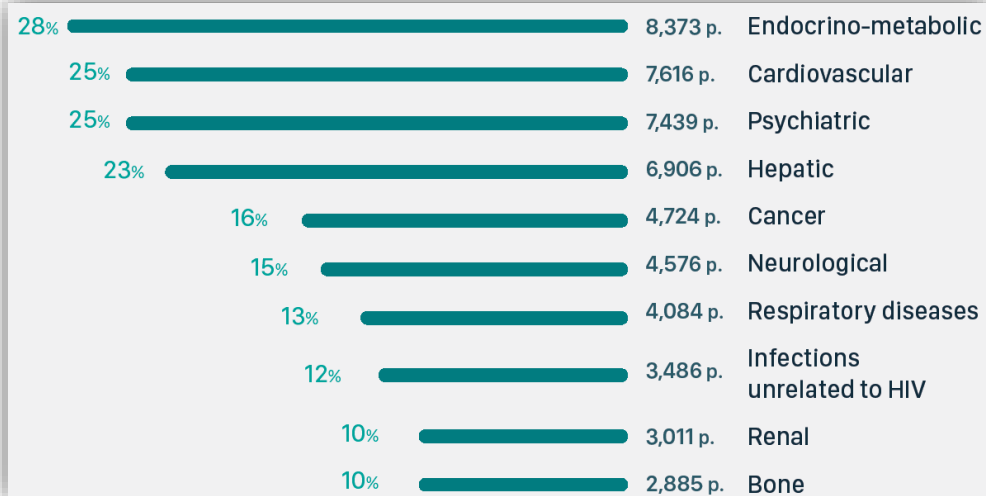
Age moyen



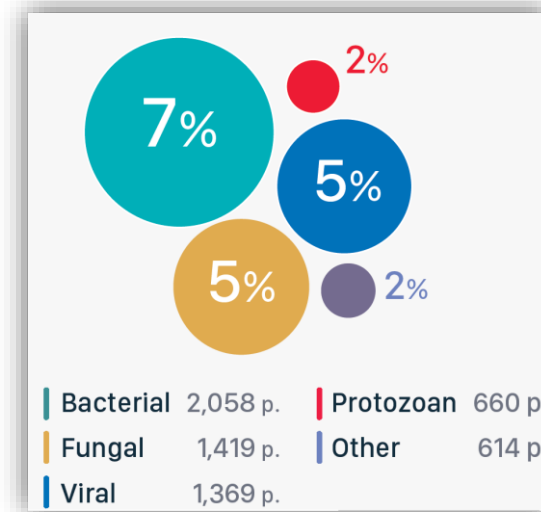
10 comorbidités



5 infections
opportunistes



Détails des comorbidités

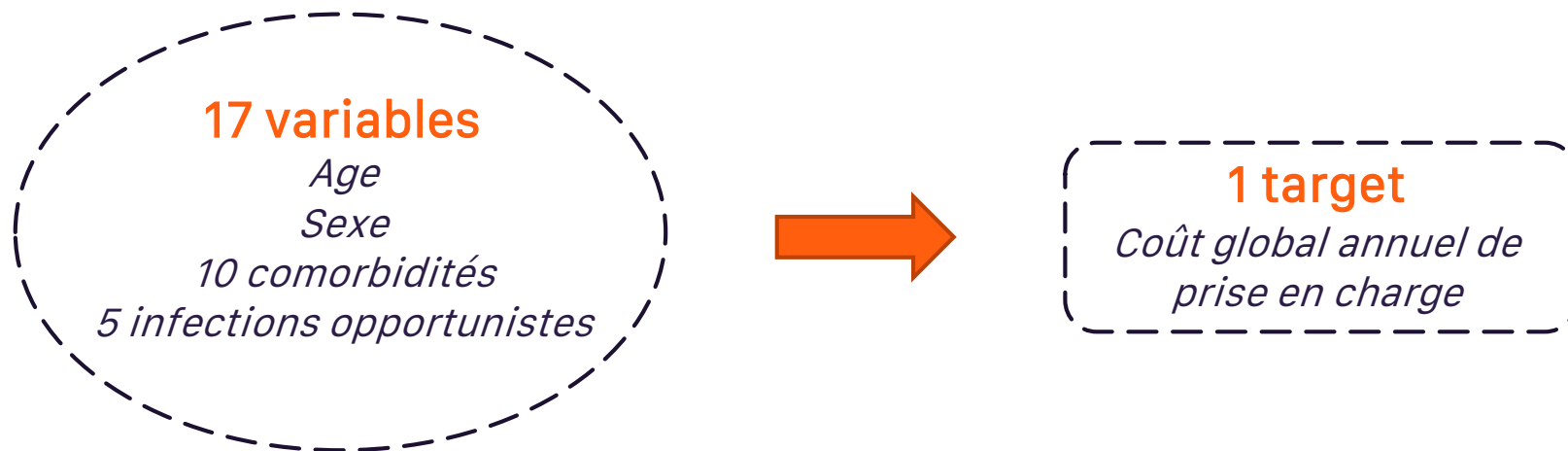


Détails des infections opportunistes

1/Cas d'étude

HEVA

Objectif

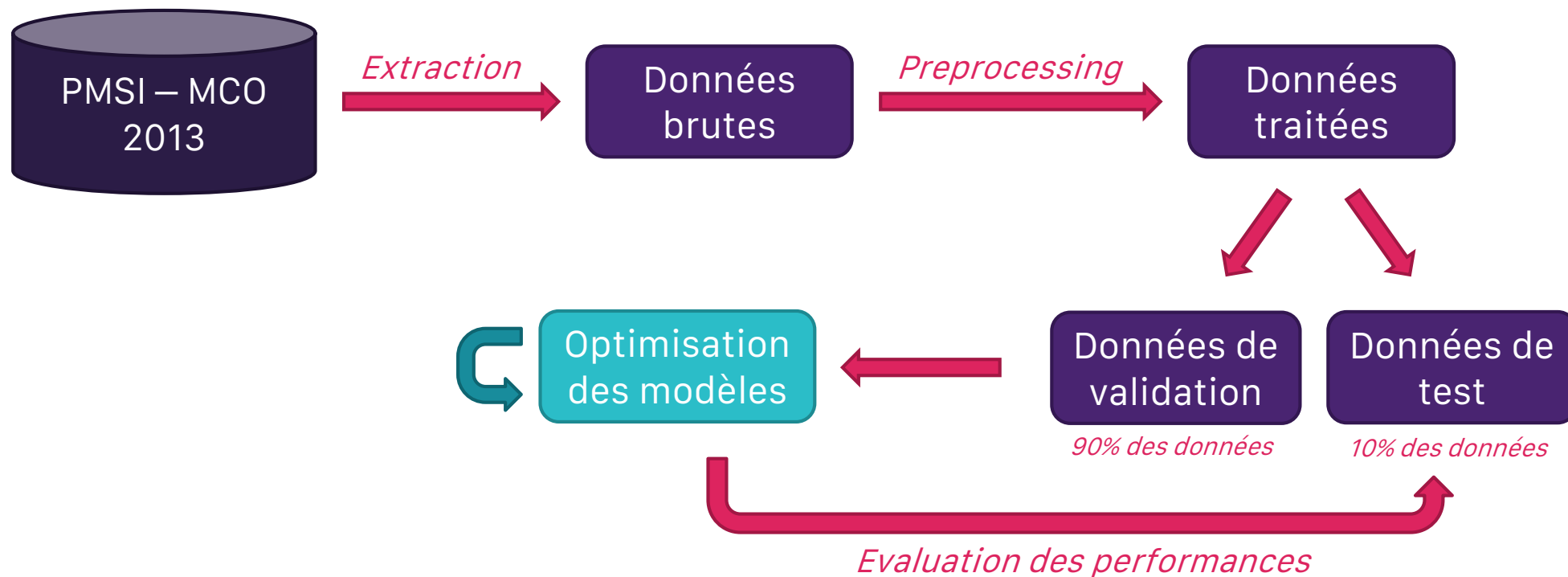


Quel modèle prédictif choisir afin de prédire au mieux le coût annuel d'un patient vivant avec le VIH, en fonction de ses caractéristiques ?

2/La méthodologie

Présentation
du protocole
Machine
Learning
utilisé

HEVA

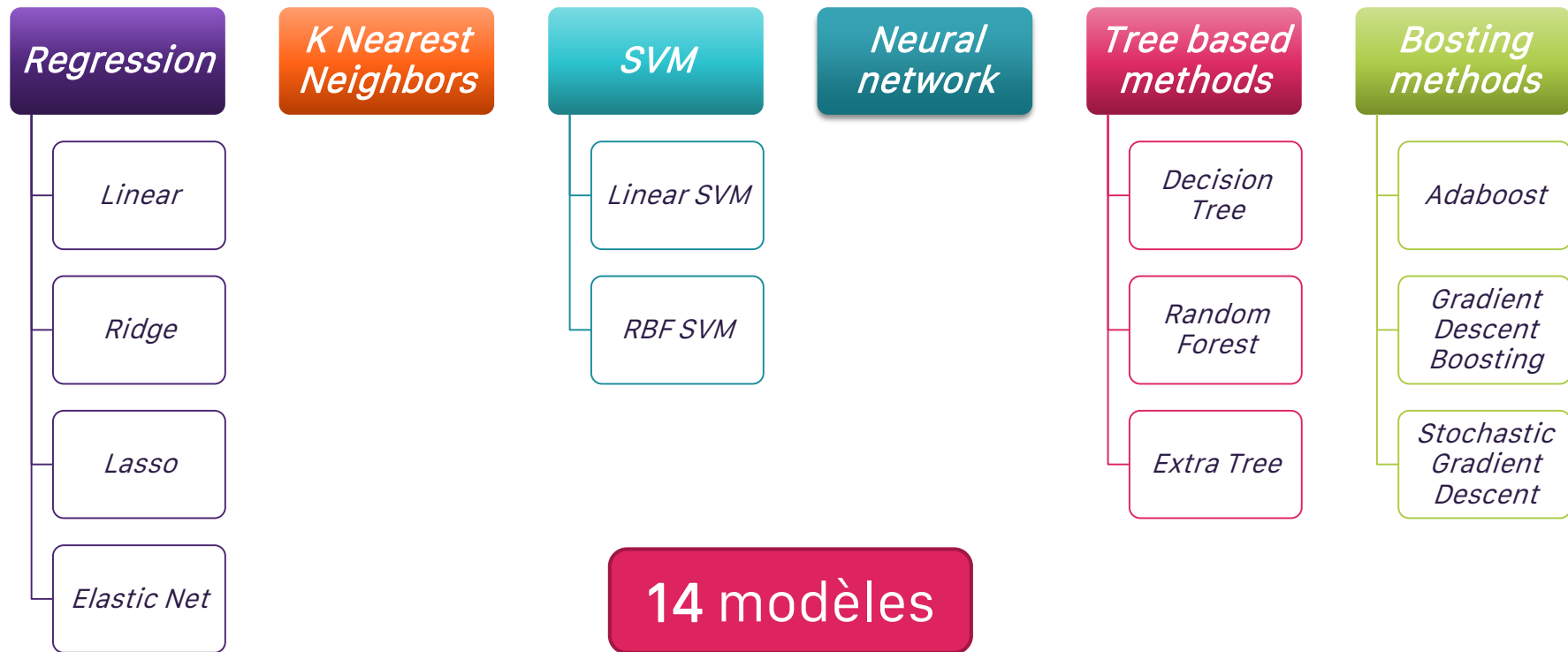


Critère d'inclusion

- *Au moins une hospitalisation en lien avec le VIH en 2013*
- *Suivi des patients concernés pendant 1 an (évaluation des coûts de prise en charge du VIH)*

2/La méthodologie

Présentation du
plan d'expérience



Optimisation des modèles : « Hyperparameter tuning problem »

Evaluation des performances : MAE, MSE, R^2

MAE : mean absolute error
MSE : mean square error
 R^2 : coefficient de détermination

3/Résultats

Analyse des performances
sur le jeu de test après
optimisation des modèles

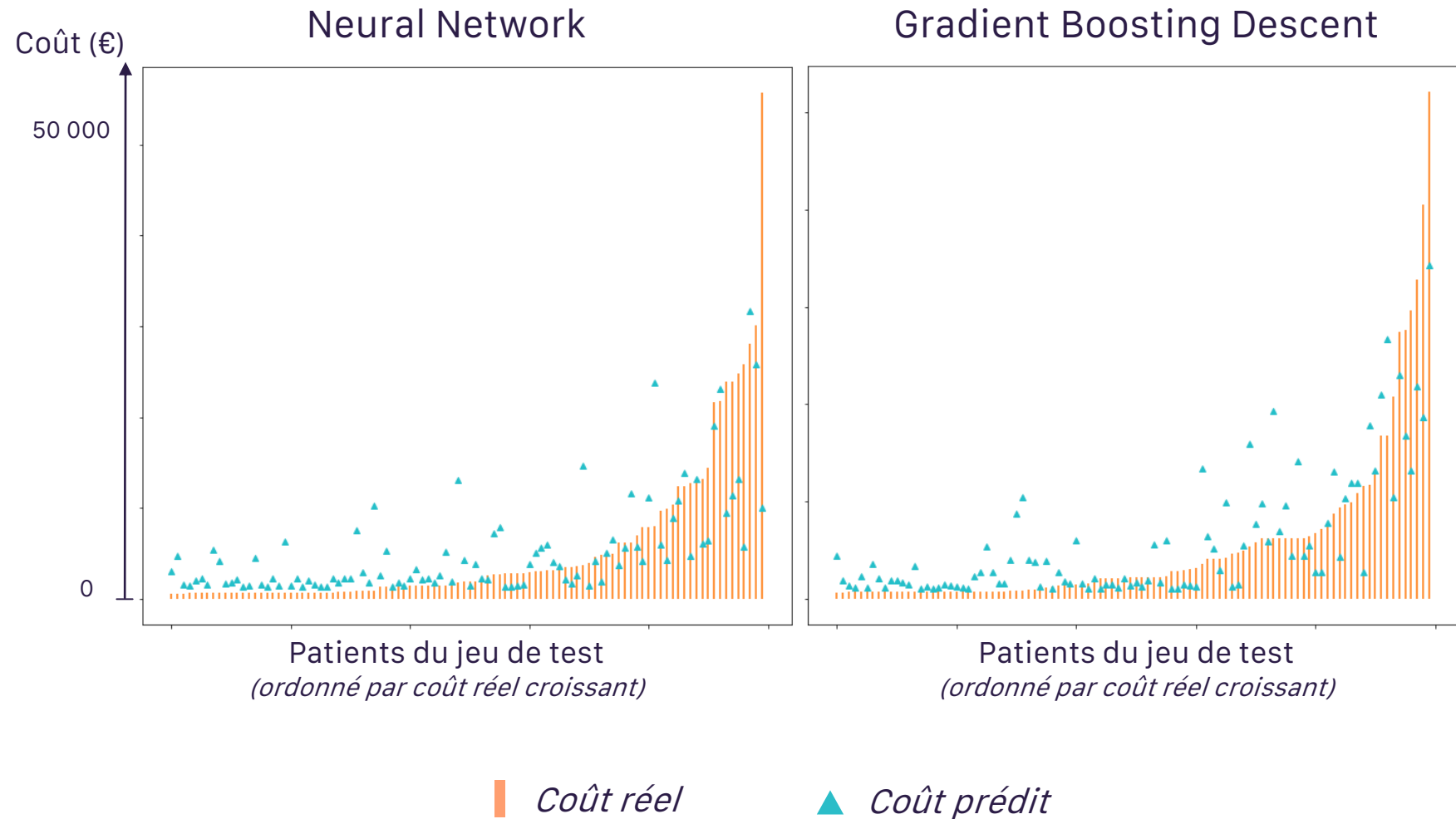
MODEL	MAE/MAX	MSE/MAX	1-R2	SCORE
GDB	0,55	0,54	0,54	0,54
EXT	0,56	0,53	0,54	0,54
RF	0,56	0,54	0,54	0,54
SVM (Lin.)	0,58	0,59	0,59	0,59
NN	0,60	0,48	0,48	0,52
KNN	0,61	0,63	0,63	0,63
DT	0,62	0,67	0,67	0,65
EN	0,65	0,51	0,52	0,56
LASSO	0,65	0,51	0,51	0,56
LR	0,65	0,51	0,51	0,56
SGD	0,66	0,51	0,52	0,56
RIDGE	0,66	0,51	0,51	0,56
SVM (RBF)	0,67	0,86	0,87	0,80
ADB	0,73	0,64	0,65	0,68
Dummy	1,00	1,00	1,00	1,00

- Gradient Descent Boosting est le plus performant au regard de la MAE
- Neural Network obtient le meilleur score (moyenne des 3 critères)

3/Résultats

Illustration des prédictions sur le jeu de test

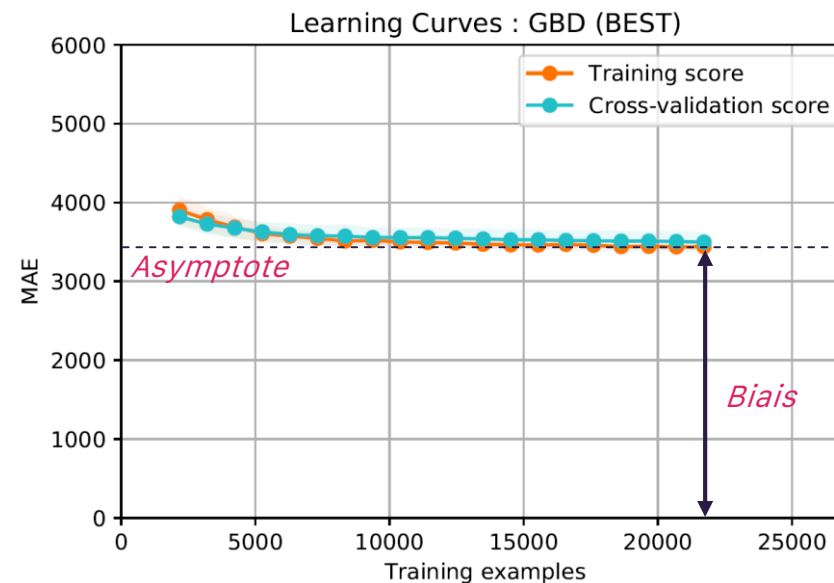
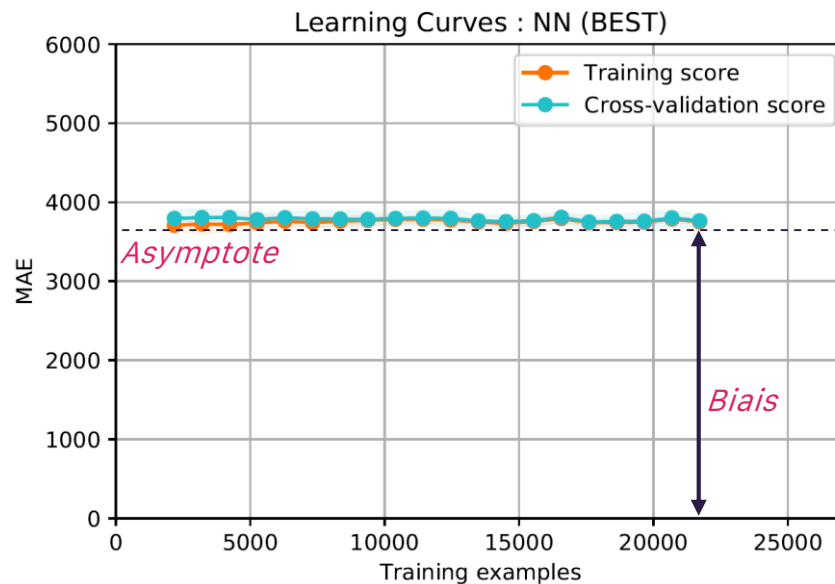
HEVA



3/Résultats

Analyse des courbes d'apprentissage après optimisation

HEVA



- Optimisation réussie (pas d'écart les deux erreurs) : généralisation de l'apprentissage
- Pas d'amélioration possible en augmentant le nombre d'exemples (a priori)

● Erreur d'entraînement

● Erreur de test

Conclusion

- Identification des **algorithmes les plus performants** dans le cadre d'une prédiction de coût (pour un patient donné)
- Exemple **concret** d'application de méthodes **Machine Learning** sur les données **PMSI**
- Application possibles dans d'**autres aires thérapeutiques**
- Amélioration des performances possibles : ajout de **nouvelles variables explicatives** (features) extraites du PMSI
- Piste de recherche : modélisation **complète** des **parcours patients** dans le cadre du développement de modèles prédictifs



Merci de votre attention.

Hugo De Oliveira
HDEOLIVEIRA@hevaweb.com