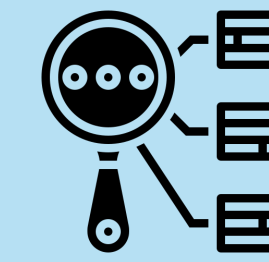
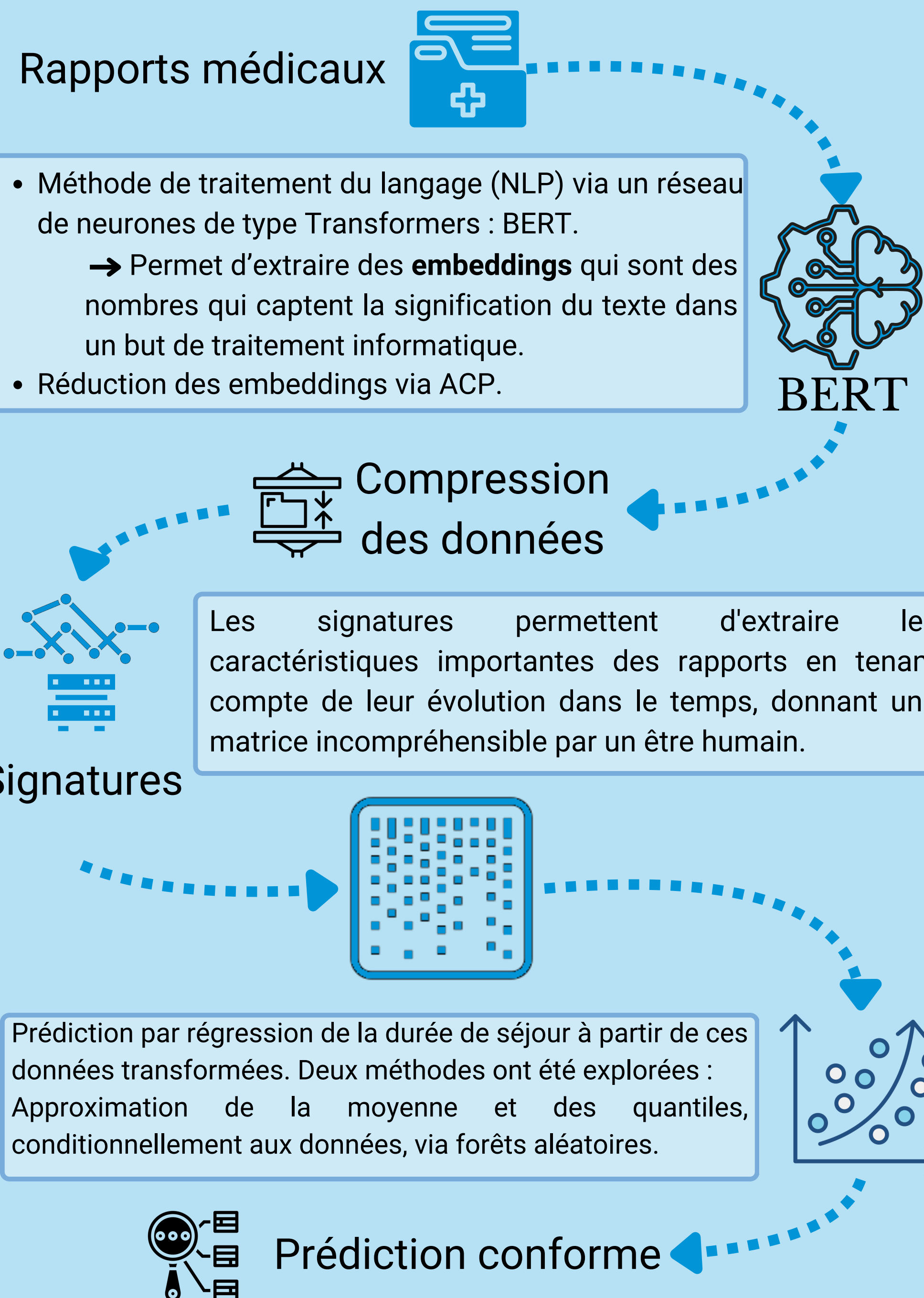


Contexte

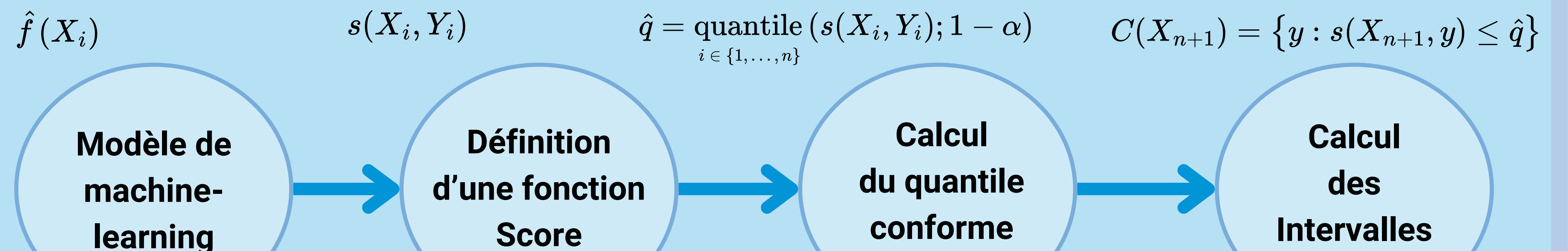
A chaque admission, les médecins rédigent de nombreux rapports médicaux, générant beaucoup de documents à lire. Dans le cadre d'une analyse préalable à un projet ambitieux du Centre Léon Bérard qui vise à libérer du temps aux médecins et à fournir une aide précise à la décision, nous partons de ces rapports médicaux pour prédire le temps de séjour à l'hôpital d'un patient.

Données et traitements

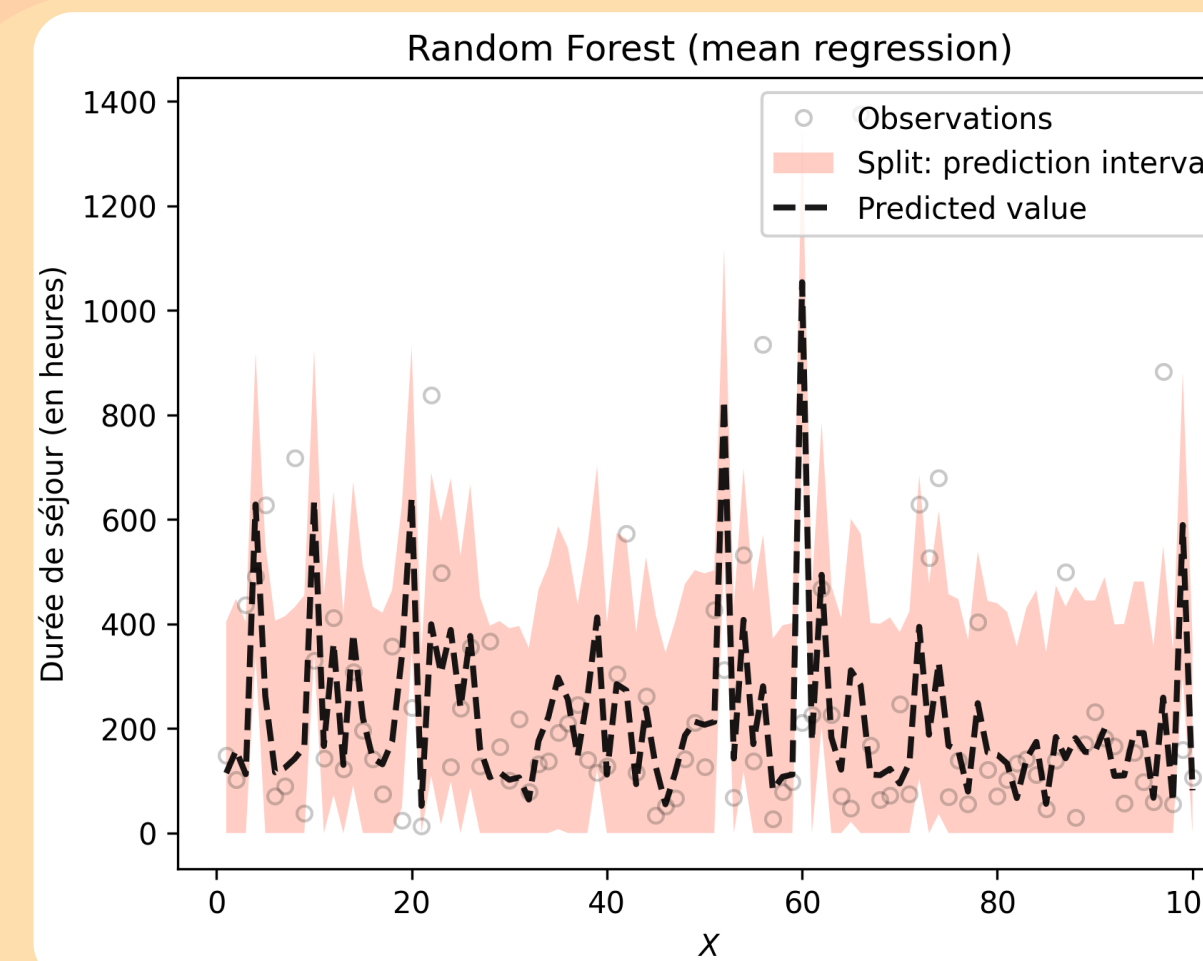


Prédiction Conforme

La prédiction conforme est une technique statistique qui génère des intervalles de prédiction selon un niveau de confiance $1 - \alpha$ désiré. Basée sur le principe d'échangeabilité des données, elle peut être utilisée à la fois dans le cadre de la classification ou de la régression. Plusieurs variantes de cette méthode existent (Full, Split, Cross Validation +, ...) et nécessitent parfois un n-échantillon supplémentaire de "calibration" lors du calcul du quantile conforme. L'approche générale est décrite ci-dessous :



Résultats et Conclusion



Le graphique présente les observations réelles de la durée de séjour en heures (points gris) par rapport aux valeurs prédites par la forêt aléatoire (ligne noire en pointillés). La zone en rose représente les intervalles de prédiction fournis par la méthode de prédiction conforme Split, capturant l'essentiel des observations mal prédites du modèle.

Méthode	Couverture	Médiane	Ecart-type
Split	87.96%	461	77
CV+	90.48%	491	82
CQR	88.55%	413	467

Les méthodes Split, CV+, et CQR offrent toutes des taux de couverture proches des 90 % désirés. Cependant, la méthode CV+ montre une meilleure adaptativité que Split, qui tend à être moins flexible en raison d'une taille d'intervalle fixe. La CQR se démarque par sa capacité d'adaptation élevée, illustrée par un écart type élevé et une médiane plus performante. Cela souligne sa supériorité en terme de précision lorsque le modèle le permet.

Les limites du projet résident dans la complexité des données médicales et les temps de calcul élevés. De plus, une adaptation pour le format des données du centre Léon Bérard reste nécessaire.