

MACHINE LEARNING :

Un outil de détection de la dépression chez les soignants pendant une crise sanitaire



Table des matières

Introduction

I- Comprendre le problème : l'histoire derrière nos données

II- Choisir les bons outils : quel algorithme pour résoudre ce problème ?

III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?

Conclusion

Introduction

Pendant la pandémie de Covid-19 :
augmentation plus importante des **troubles mentaux** chez les **soignants** (40 %-50 %) que dans la population générale.

En 2020, 56% des soignants français montraient des signes de **détresse psychologique**, 21% des troubles de **stress post-traumatique**

(Données Santé Publique France)

D'où l'importance de prédire et prévenir ces troubles mentaux lors de futures crises sanitaires, à l'aide du machine learning



I- Comprendre le problème : l'histoire derrière nos données

- Étude réalisée sur des **professionnel.le.s de la santé** d'Asie Centrale entre juillet et novembre **2022**
- **2685** réponses à l'enquête
- Types de données :

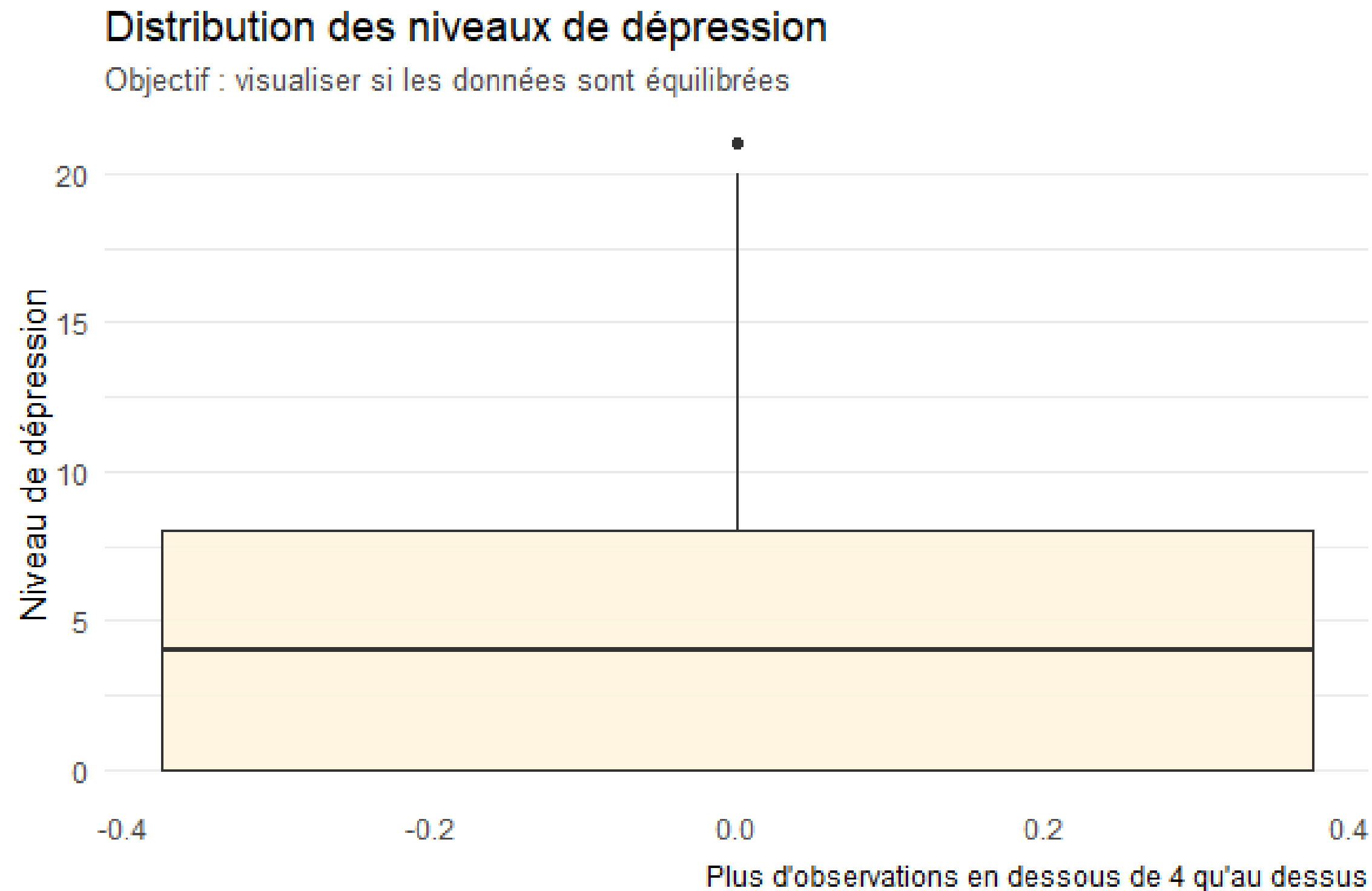
Données socio-démographiques et professionnelles : sexe, âge, emploi (médecins ou infirmières), situation familiale, antécédents de travail en première ligne pendant la pandémie de COVID-19

Données de santé mentale : prévalence, niveau et gravité de la dépression, de l'anxiété et du stress, calculés à partir du questionnaire DASS-21

Qualité de vie, calculée à l'aide de l'indice de bien-être OMS-5.13

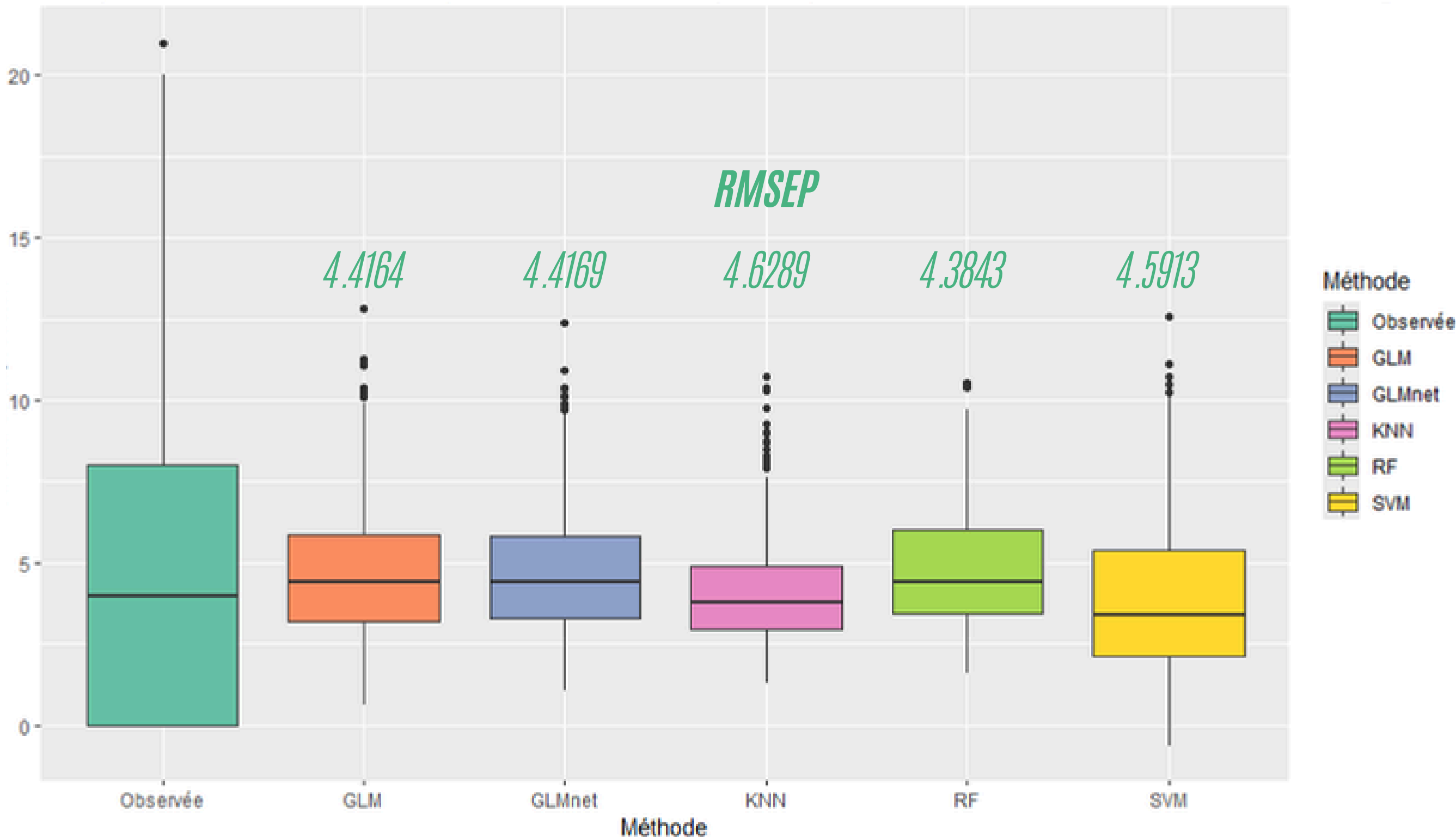
Impact du COVID-19 sur la vie personnelle et professionnelle

I- Comprendre le problème : l'histoire derrière nos données



II- Choisir les bons outils : quel algorithme pour résoudre ce problème ?

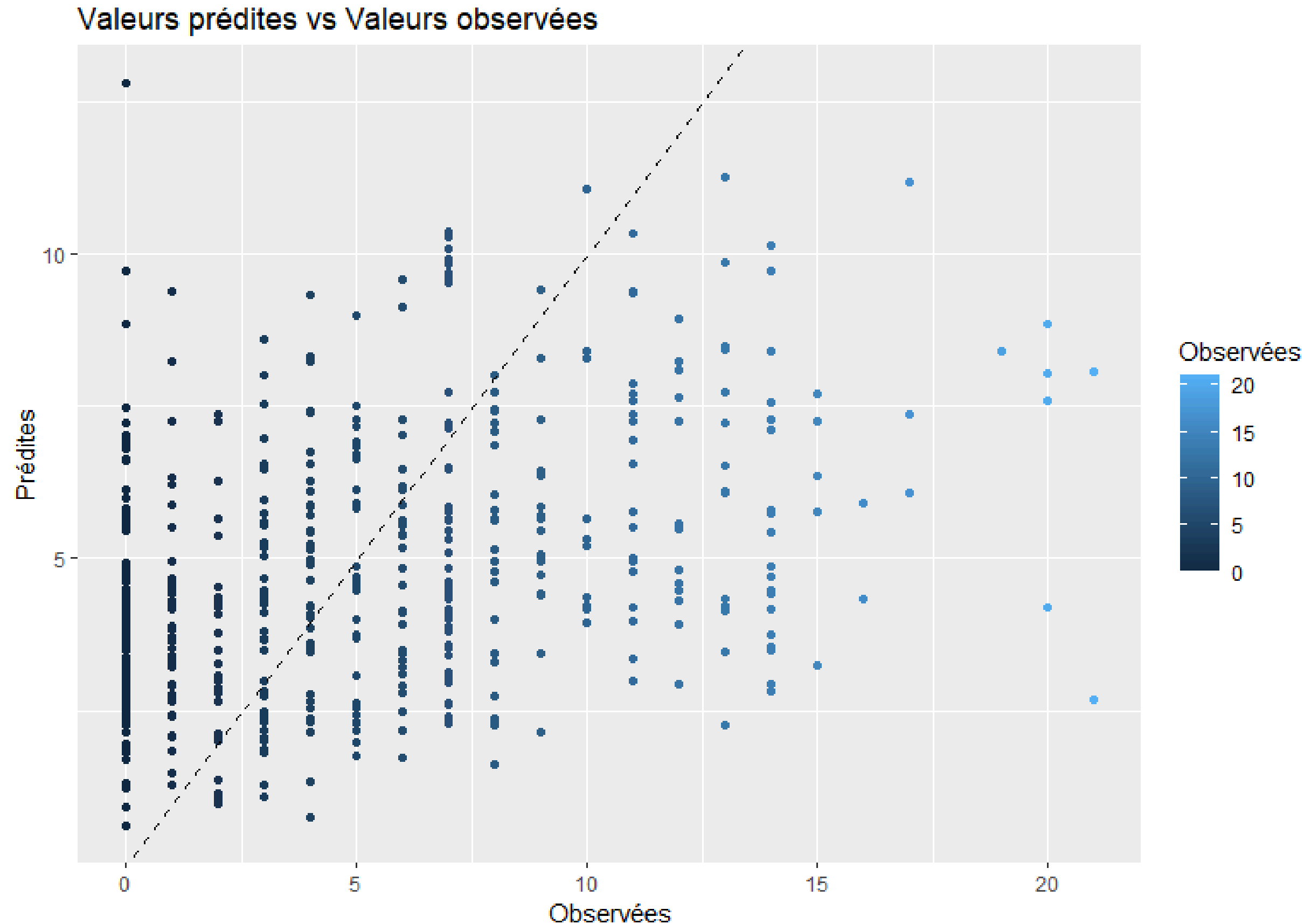
Comparaison des niveaux de dépression observés et prédits par différentes méthodes de Machine Learning



Les RMSEP sont globalement similaires entre les modèles

Choix du **modèle glm**
car il est facile d'expliquer
l'impact de chaque variable
explicative, il est donc plus
adapté à la problématique

II- Choisir les bons outils : quel algorithme pour résoudre ce problème ?



Il y a une sous-estimation du niveau de dépression prédit

III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?

Dans la réalité du terrain : il faudrait avoir accès au **minimum de données** possible sur le soignant (par exemple à travers un **questionnaire rapide**) pour **détecter une dépression éventuelle**.

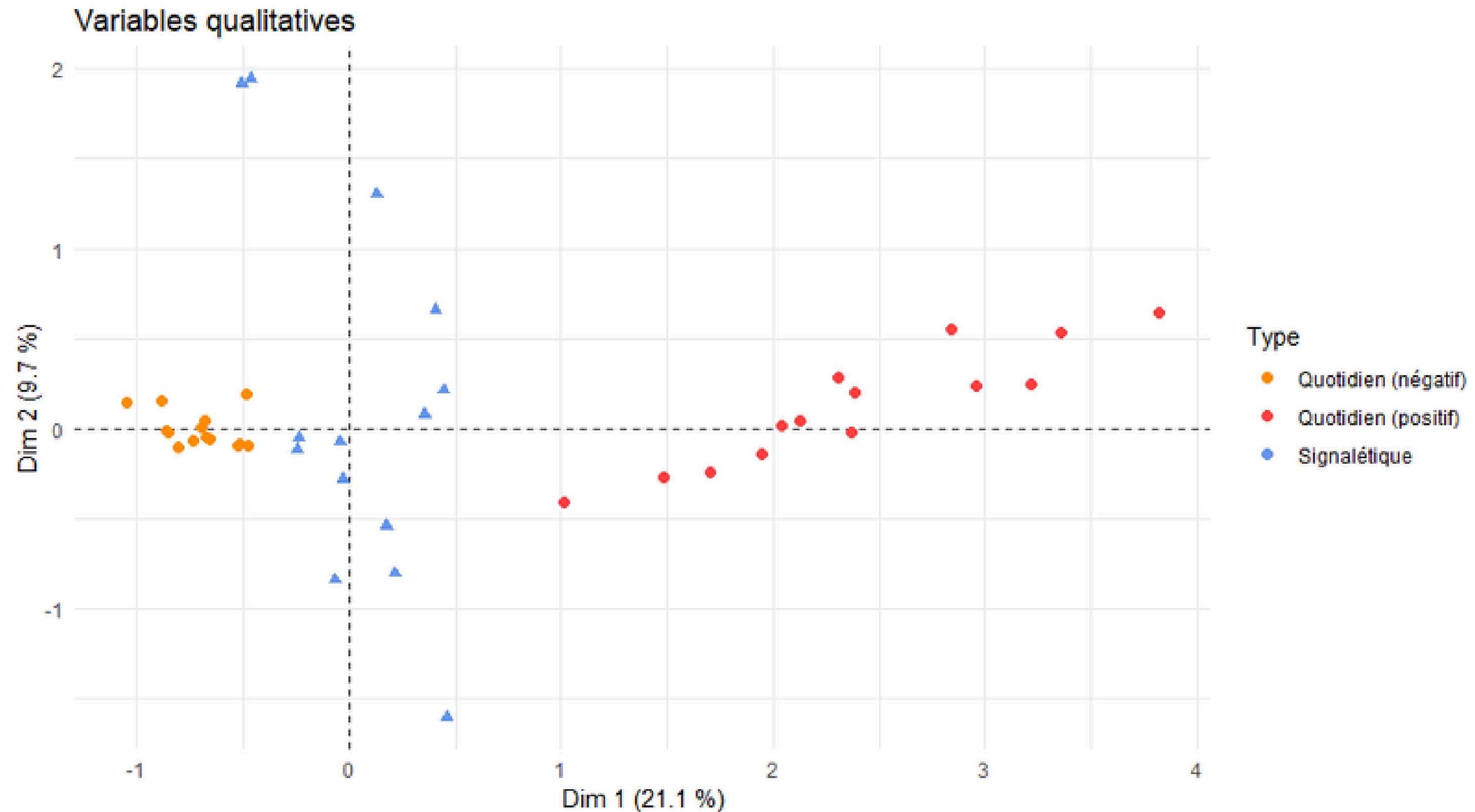


De 21 questions

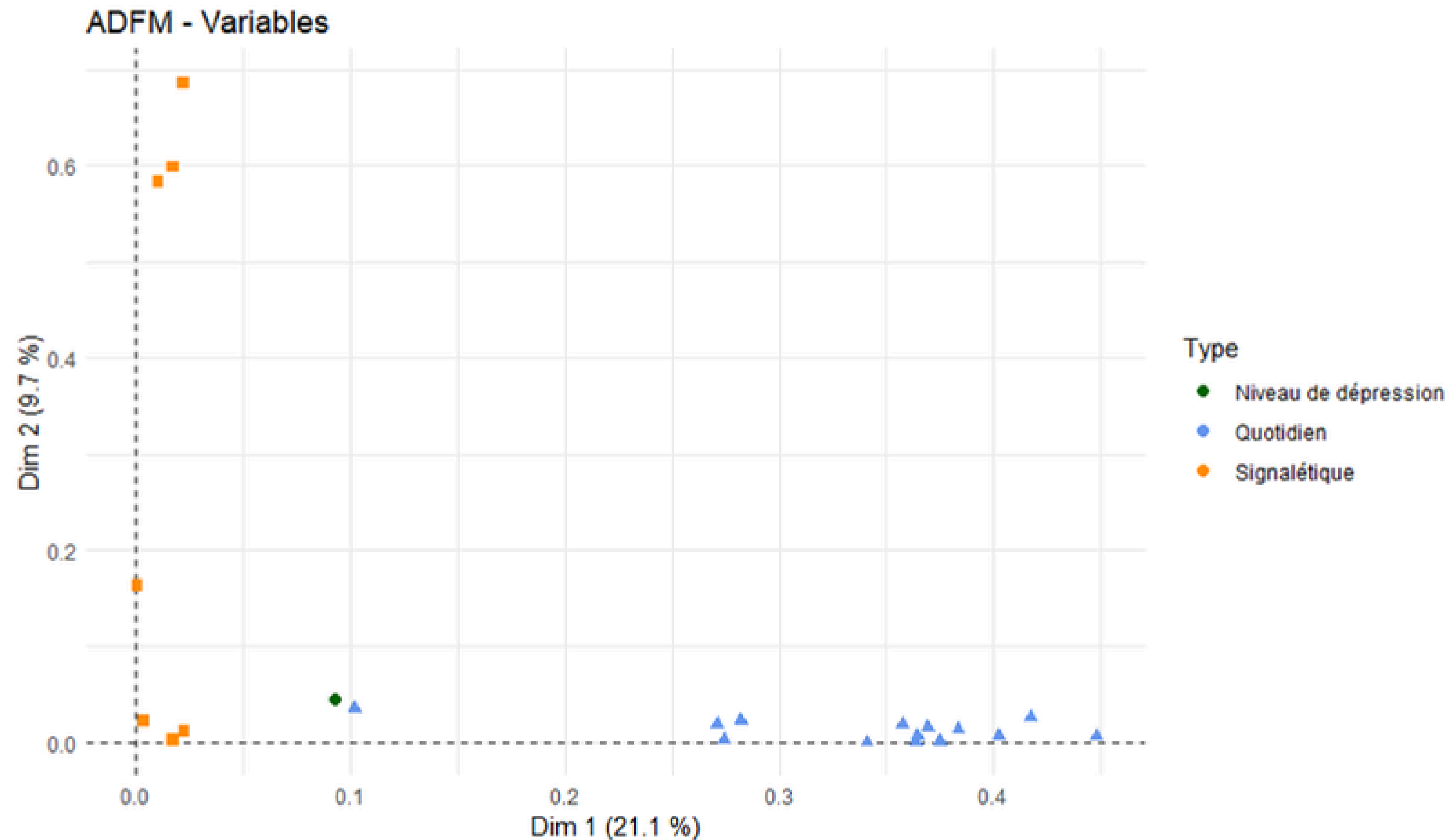


À ? questions

III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?



III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?



III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?

Stepwise Model Path
Analysis of Deviance Table

Initial Model:

Depression_level ~ Gender + Age + Job + Manager_position + COVID_frontline +
Family_status + Children + more_work + additional_workload +
overtime + work_stress + work_conflicts + afraid_family +
people_avoid_me + afraid_others + people_avoid_family + working_attitude +
insufficient_employees + appreciation_employer + appreciation_society +
appreciation_govt

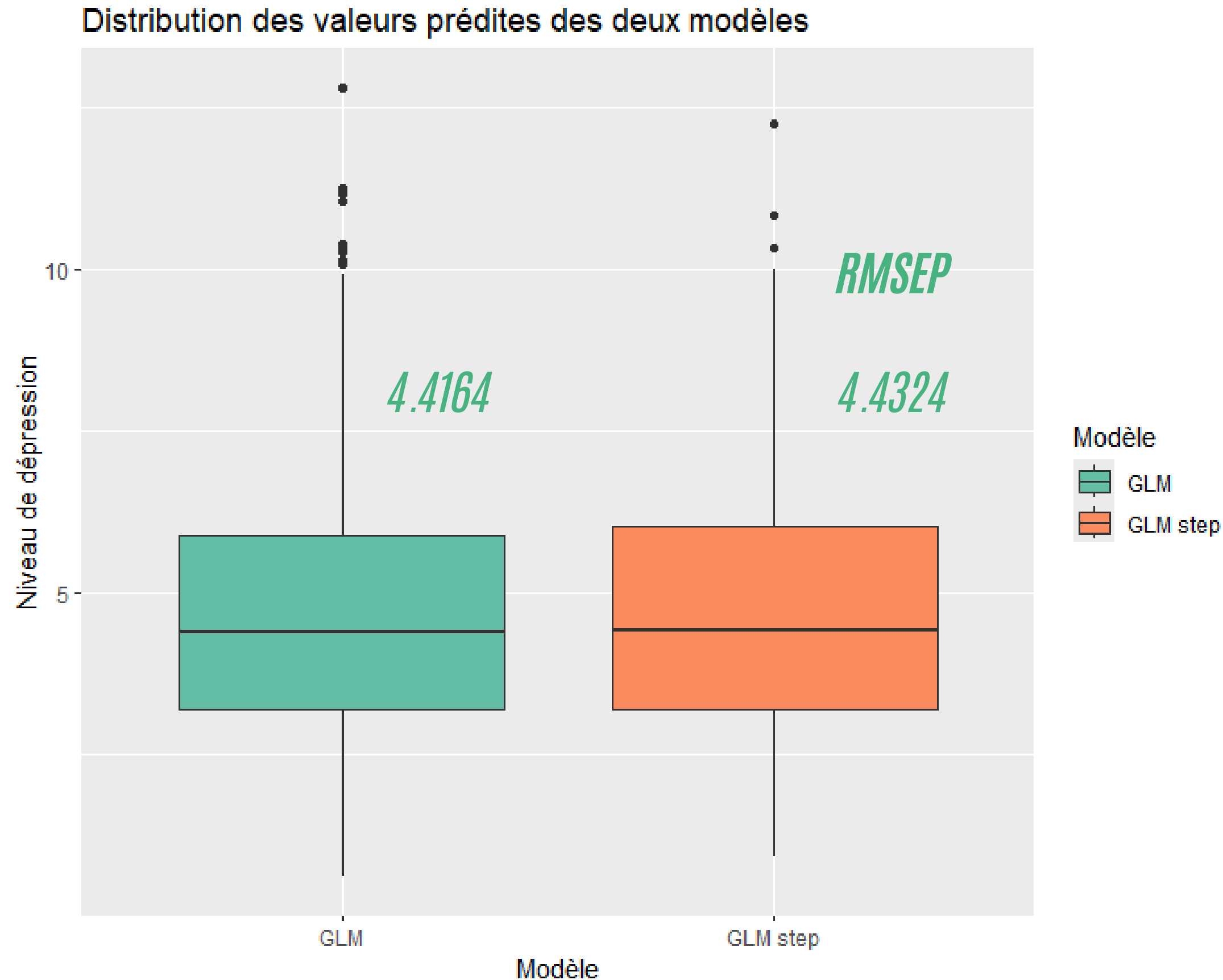
Modèle initial :
21 variables explicatives

Final Model:

Depression_level ~ Age + Job + Manager_position + Children +
more_work + work_stress + work_conflicts + afraid_family +
people_avoid_family + appreciation_employer + appreciation_society

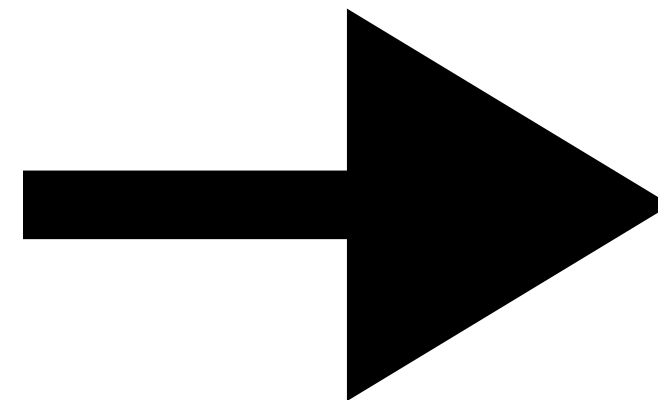
Modèle final :
11 variables explicatives

III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?



III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?

Métrique	Seuil Optimal
Accuracy	5.0
F1 Score	3.0
AUC	4.5

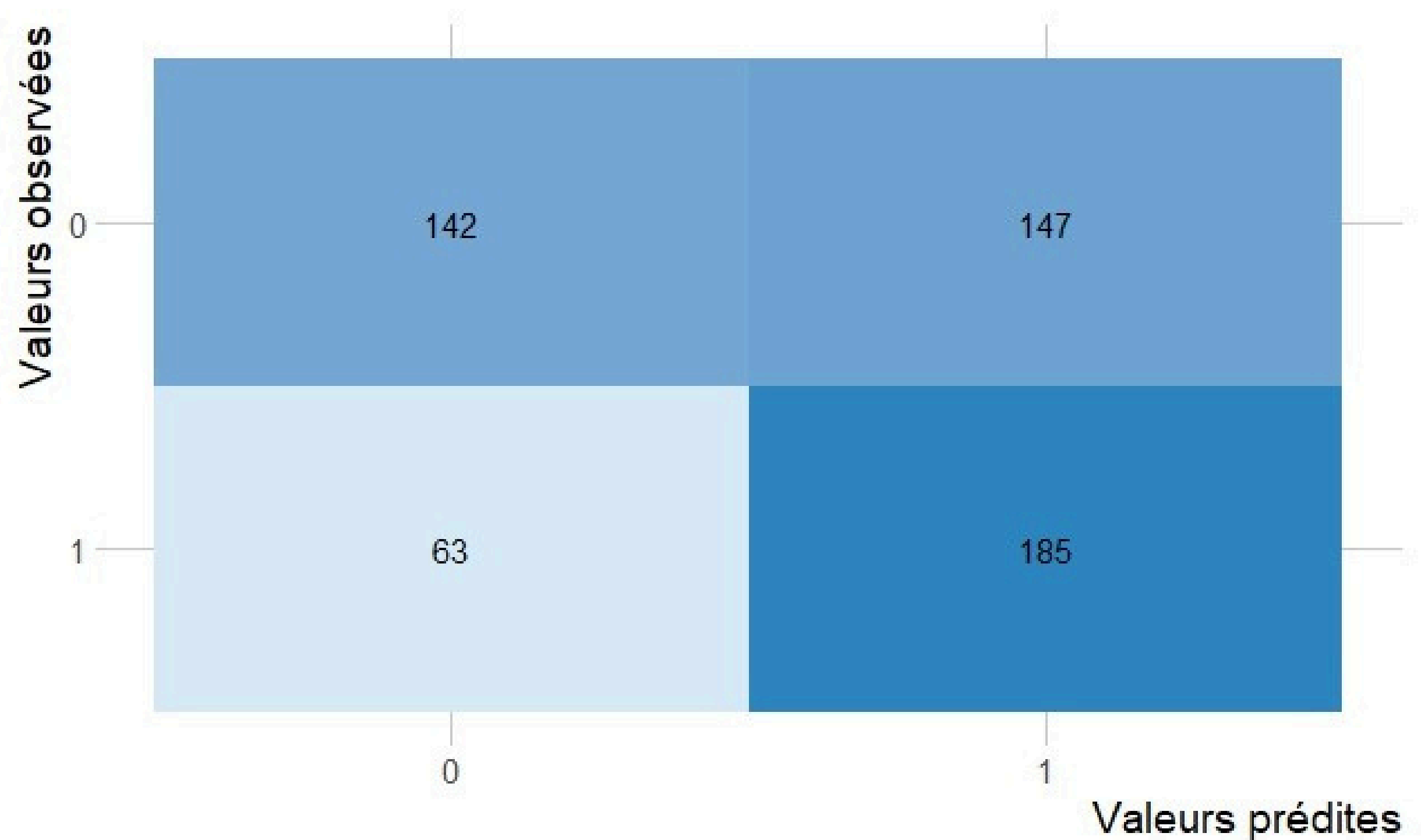


Seuil = 4

(compromis entre les différentes métriques utilisées)

III- Tester et améliorer : peut-on réduire le nombre de variables du modèle ?

Matrice de confusion du modèle final à 11 variables



Conclusion

Développement d'un outil d'aide à la détection permettant de **mettre en alerte** les institutions responsables et de déclencher le **passage de tests plus approfondis**.

Élargissement du jeu de données à d'autres pays et apprentissage pour rendre le modèle plus précis et **généralisable**.



Conclusion

Modèle final	Publication
<div>Âge</div> <div>Métier</div> <div>Rôle</div> <div>Nombre d'enfants</div> <div>Surcharge de travail</div> <div>Stress au travail</div> <div>Conflits au travail</div> <div>Peur de la part de la famille</div> <div>Les autres évitent ma famille</div> <div>Sentiment d'appréciation de la part de l'employeur</div> <div>Sentiment d'appréciation de la part de la société</div>	<div>Âge</div> <div>Métier</div> <div>Rôle</div> <div>Statut marital</div> <div>Nombre d'enfants</div> <div>Surcharge de travail</div> <div>Stress au travail</div> <div>Conflits au travail</div> <div>Sentiment d'appréciation de la part de l'employeur</div> <div>Sentiment d'appréciation de la part de la société</div>

Conclusion

Manque de variables importantes telles que :

- Facteurs environnementaux
- Habitudes de vie
- État émotionnel préexistant
- Génétique