

12/11/2024



# DEMI-JOURNÉE DATA SCIENCE

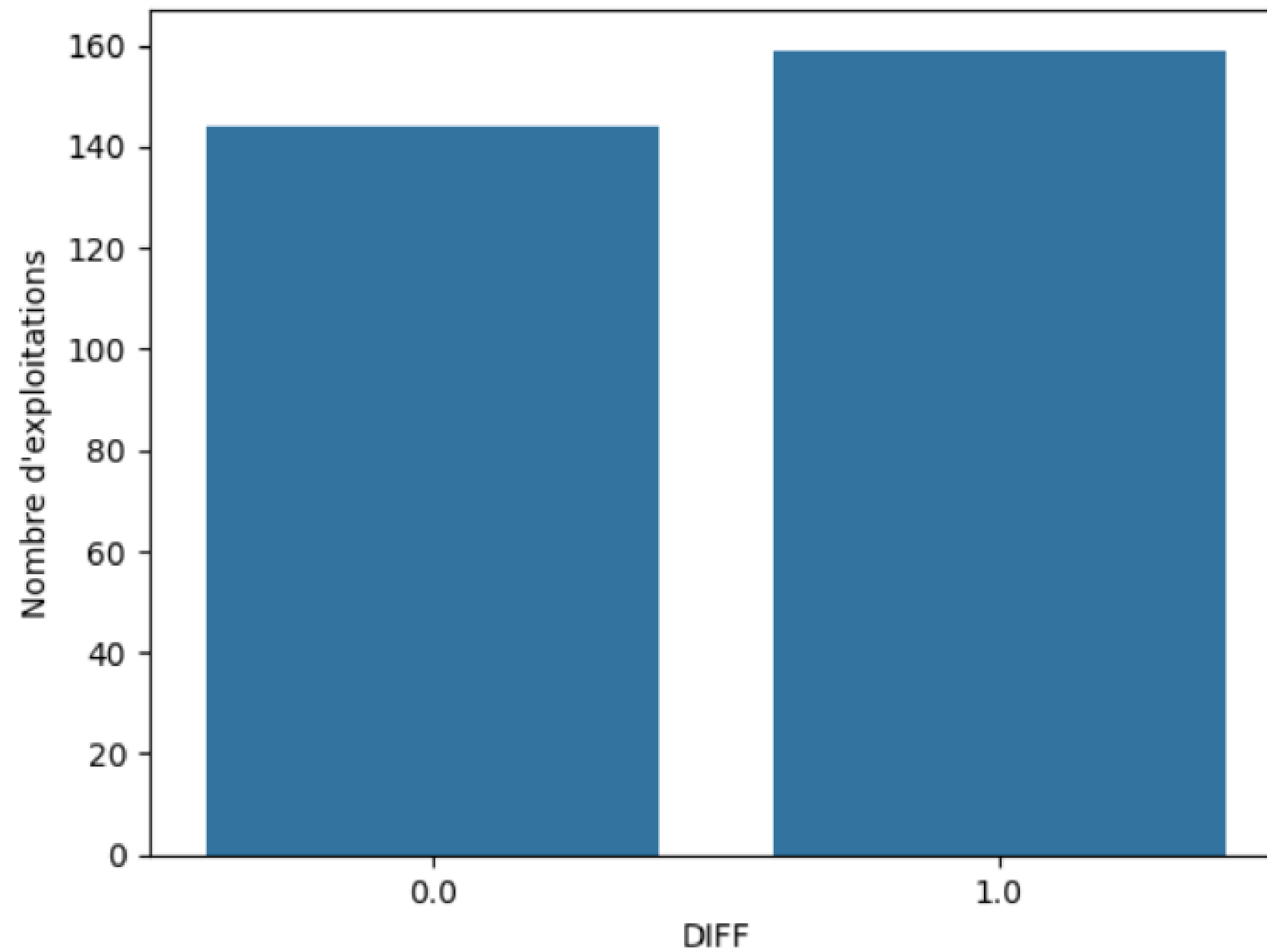
Groupe 5

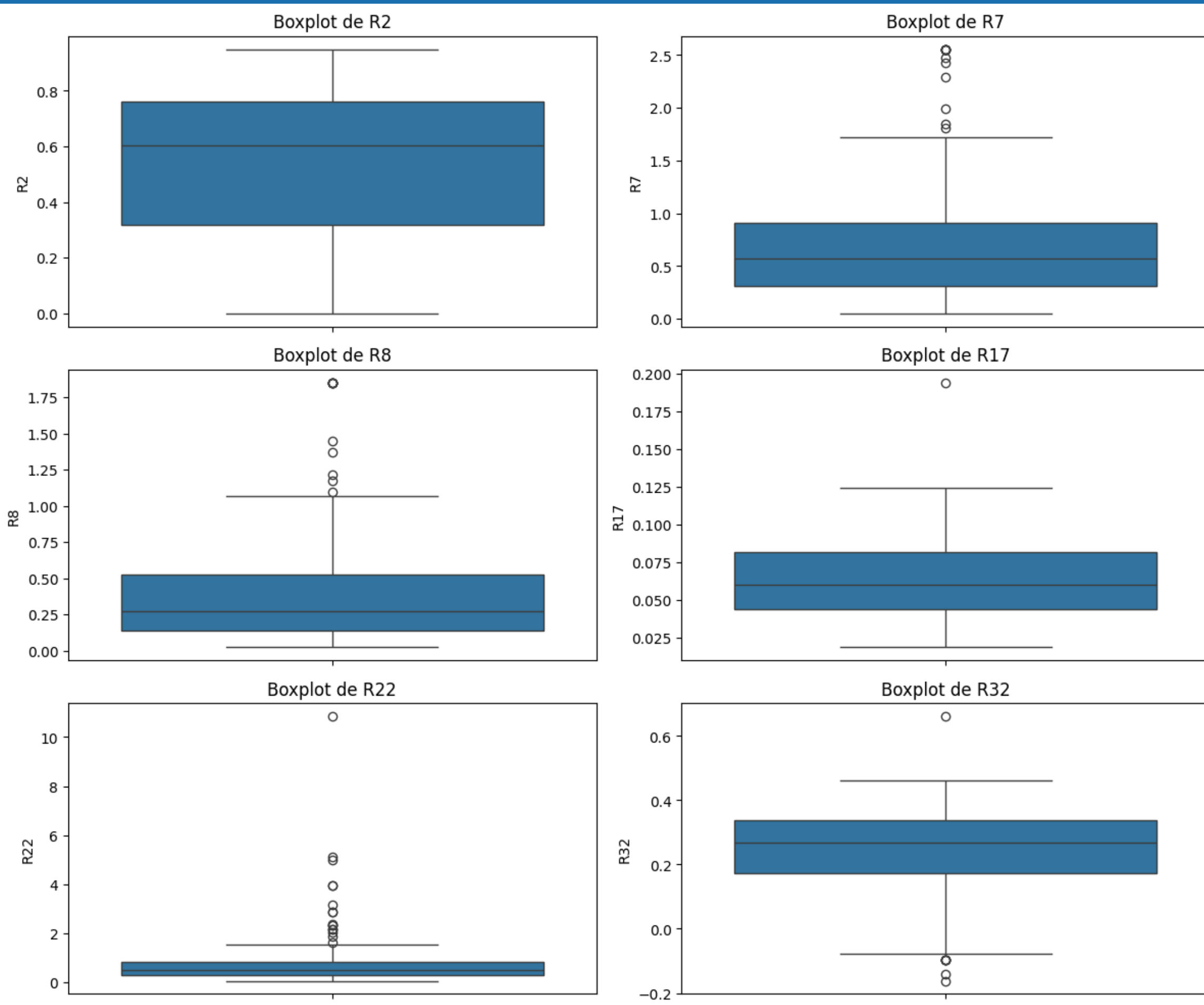
SAOUD Yasmina - EMANUEL Justin - EL OUALYDY Mohamed-Amine



# EXPLORATION DES DONNÉES

Distribution de la variable cible (DIFF)



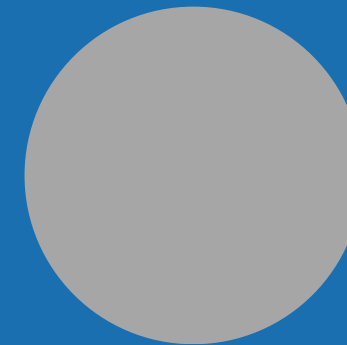


# MODELE

**REGRESSION  
LOGISTIQUE**

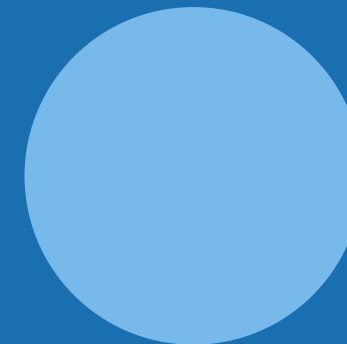
**SVM**

**ARBRE DE  
DECISION**



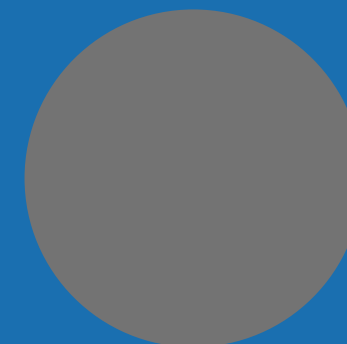
## **REGRESSION LOGISTIQUE**

Ce modèle est souvent performant pour des données financières et est facilement interprétable grâce aux coefficients de chaque variable.



## **ARBRE DE DE DECISION**

Choisi pour sa simplicité d'interprétation. Il nous permet de visualiser comment les décisions sont prises et quelles variables influencent le plus la classification.

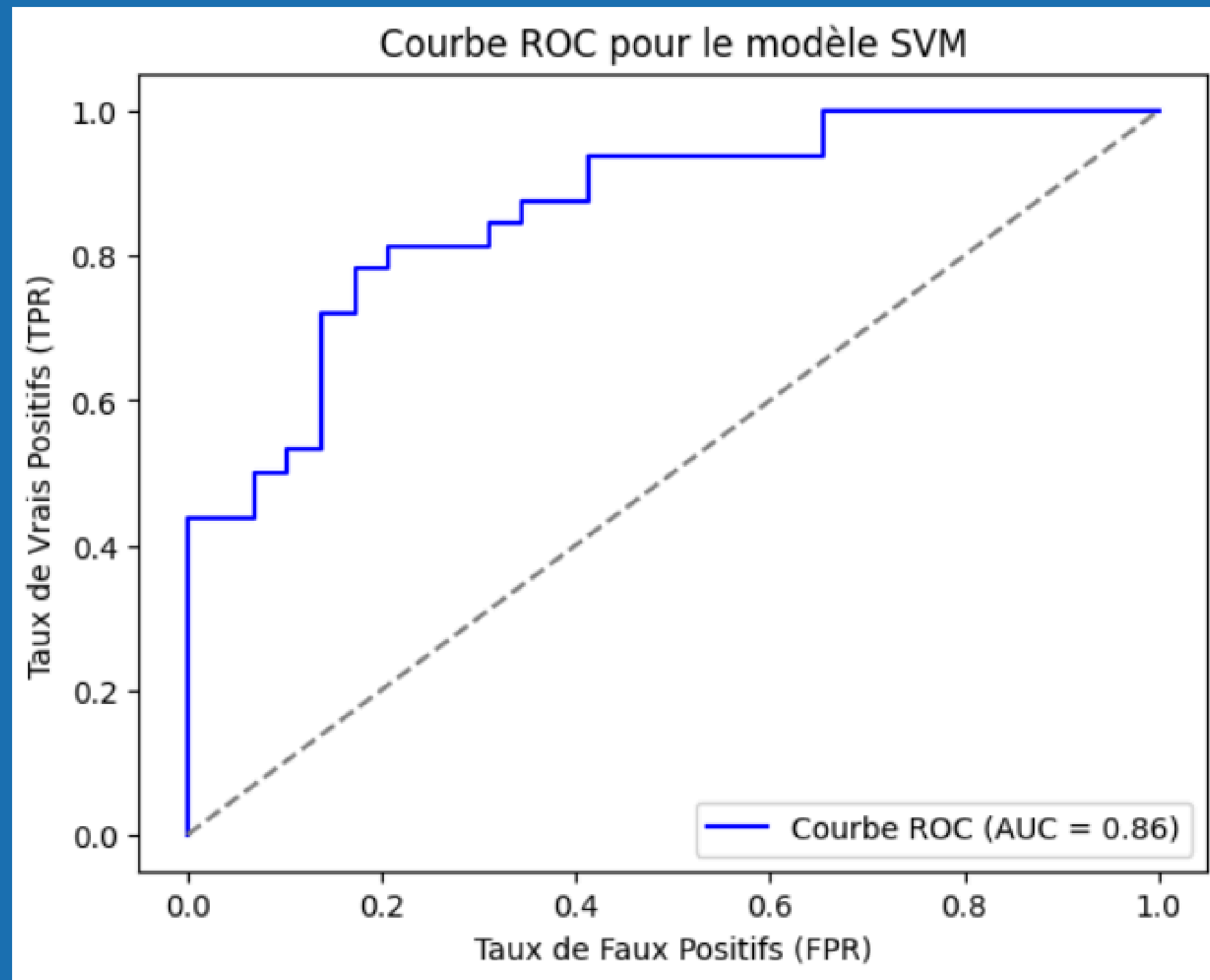


## **SVM**

Ce modèle est bien adapté pour des données où les classes sont bien séparées et peut offrir de bonnes performances en termes de classification.

# **ÉVALUATION DES RÉSULTATS ET VISUALISATION**

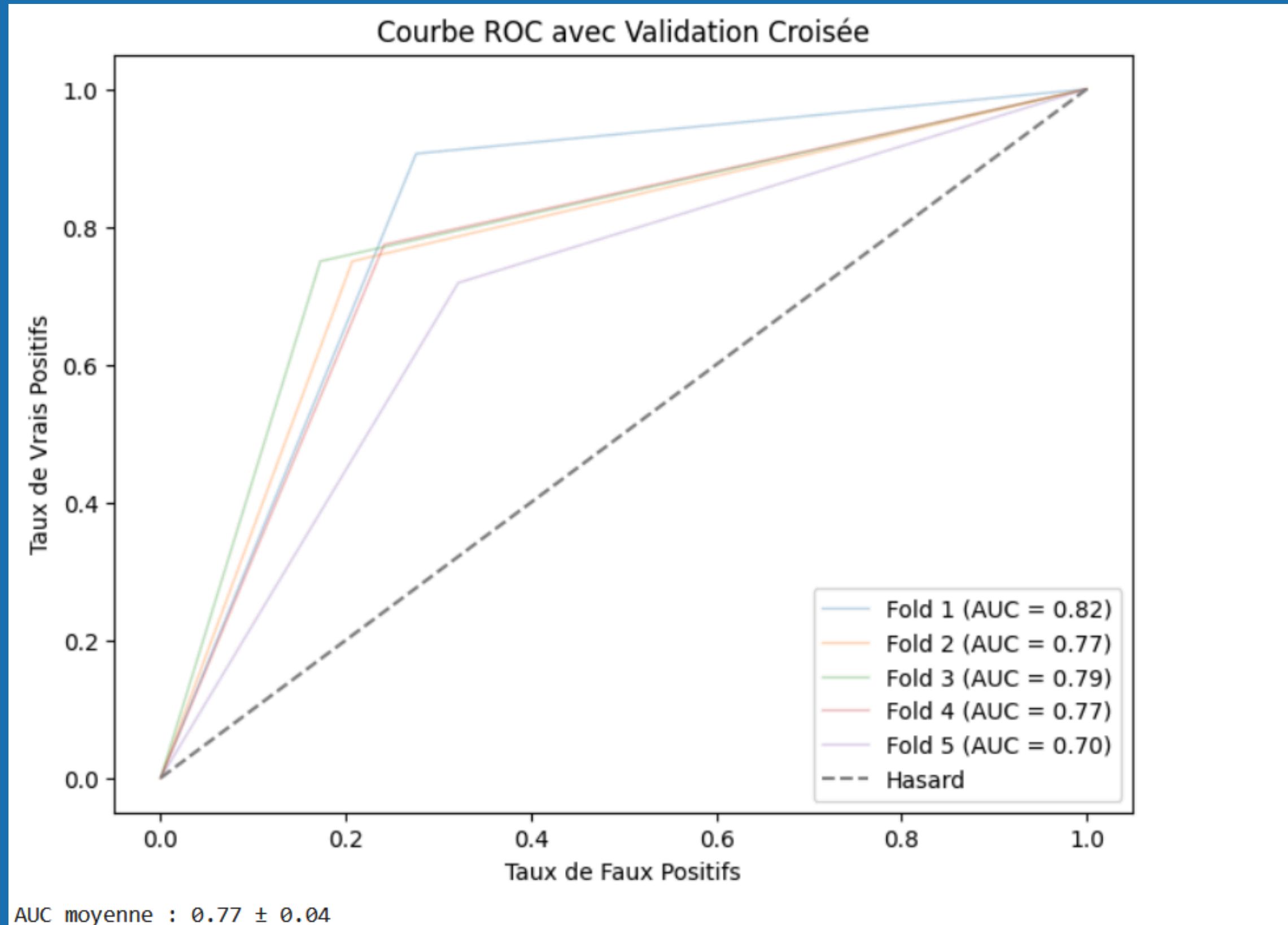
# MODELE SVM



AUC : 0.86

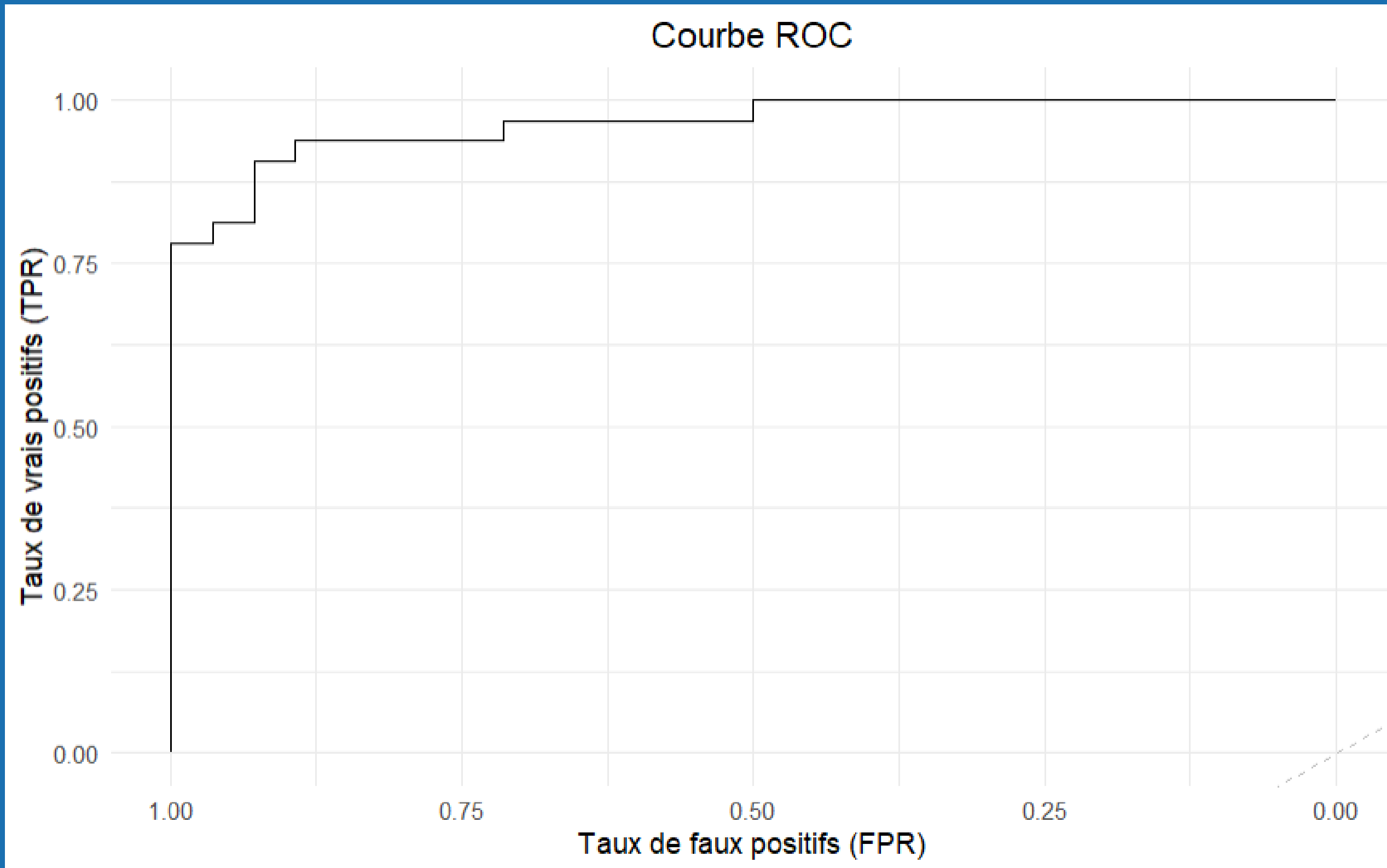
# MODELE ARBRE DE DECISION

AUC : 0.77



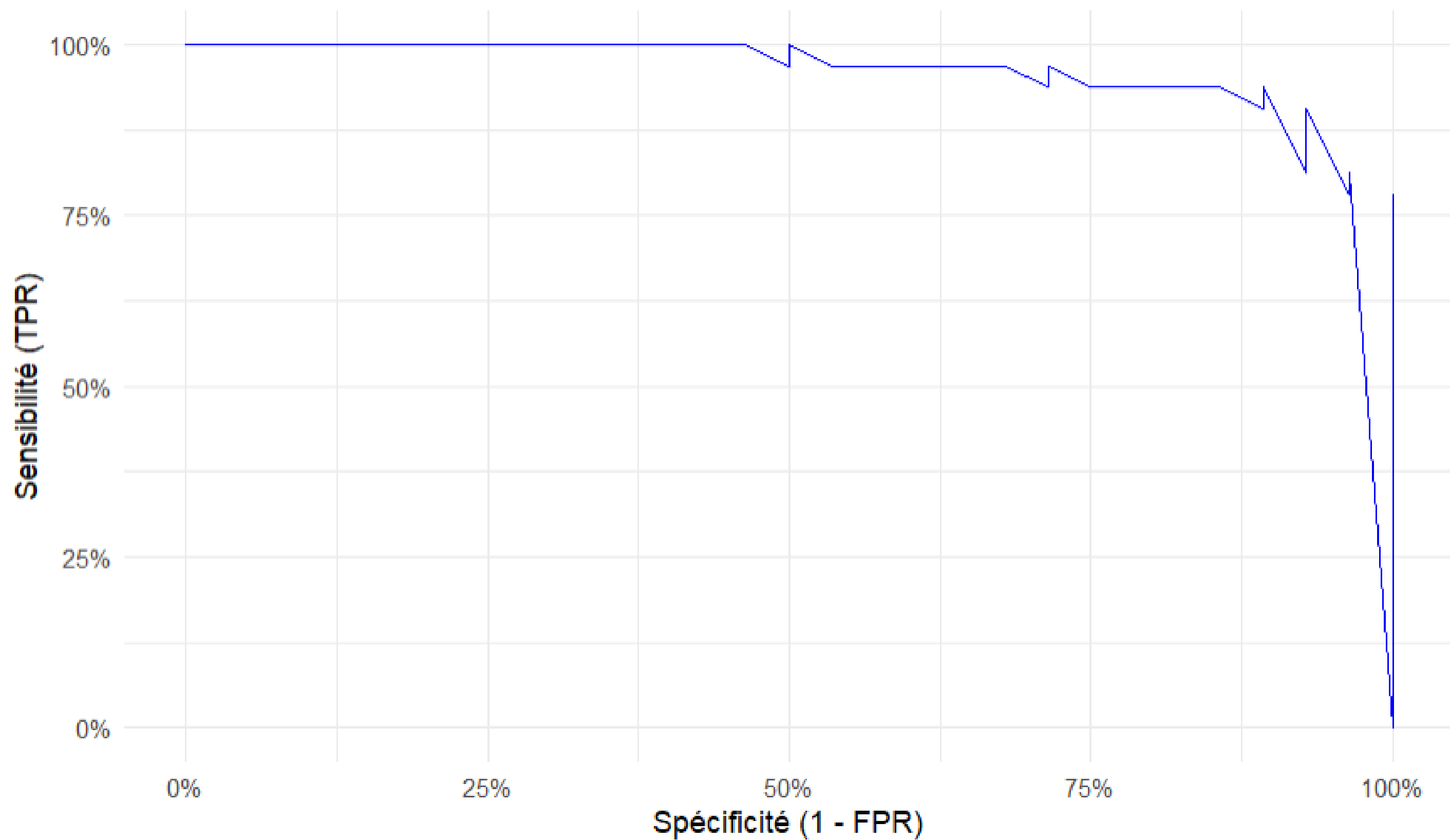


# MODELE REGRESSION LOGISTIQUE



AUC : 0.964

Courbe Sensibilité vs. Spécificité



# SCORE AVEC REGRESSION LOGISTIQUE

