

# Rapport d'Évaluation des Chatbots de Paris Sportifs

## 1. Matrices de Confusion

Matrice pour la classe « Recommandé » (One-vs-Rest)

	Prédit Recommandé	Prédit Non-Recommandé
Réellement Recommandé	40	10
Réellement Non-Recommandé	15	135

Matrice pour la classe « Éviter » (One-vs-Rest)

	Prédit Éviter	Prédit Non-Éviter
Réellement Éviter	50	20
Réellement Non-Éviter	10	110

## 2. Performances du Modèle

Classe	Precision	Recall	F1-score	Support
Éviter	0.83	0.71	0.77	70
Modéré	0.64	0.56	0.60	80
Recommandé	0.73	0.80	0.76	50

## 3. Observations sur les Erreurs Fréquentes et Biais

- La classe « Modéré » présente la plus faible F1-score (0.60), souvent confondue avec « Éviter » et « Recommandé ».
- Biais d'entraînement : classes extrêmes (« Éviter », « Recommandé ») surreprésentées par rapport à « Modéré ».
- Taux de faux négatifs pour « Éviter » (20/70 soit 29%) : certains cas d'évitement sont mal détectés.
- Quelques prédictions erronées dues à un déséquilibre des features numériques (cotes similaires).

## 4. Recommandations d'Amélioration

- Enrichir le jeu de données pour la classe « Modéré » (oversampling, collecte ciblée).
- Ajuster poids de classe dans la fonction de perte (CrossEntropy Loss avec weight).
- Tester des architectures plus profondes ou modèles ensemble (ensembles de MLP).
- Appliquer des techniques de régularisation (dropout, L2) et validation croisée.
- Mettre en place des outils d'explicabilité (SHAP, LIME) pour comprendre les décisions.