



IFAC

CHU de Nantes

Institut
Fédératif des
**Addictions
Comportementales**



ARTICLE DE FEVRIER 2025 - Identification précoce du jeu problématique dans un casino en ligne : une étude de machine learning sur les données de compte des joueurs

Retrouvez ci-dessous la synthèse de l'article du mois.

Pourquoi avoir fait cette recherche ?

Le jeu problématique est un enjeu majeur de santé publique avec des conséquences économiques et sociales importantes. L'essor des jeux d'argent et de hasard en ligne permet de collecter des données détaillées sur le comportement des joueurs, offrant une opportunité unique pour une identification précoce des comportements de jeu à risque.

Cependant, les méthodes habituelles de détection des comportements de jeu problématique présentent plusieurs limites. Les questionnaires autorapportés sont sujets aux biais de sous-déclaration et d'autoévaluation. Le suivi du comportement de jeu (tracking) comme l'analyse des dépôts, retraits et pertes, peut aider à identifier des tendances préoccupantes, mais son interprétation est complexe et nécessite des méthodes d'analyses de données avancées ce qui en fait un outil rarement utilisé pour des détections en temps réel. Enfin, les approches basées sur des seuils fixes, comme définir un montant maximal de dépôt autorisé, manquent de flexibilité et ne tiennent pas compte des profils individuels des joueurs.

Or, les récentes avancées de l'intelligence artificielle en matière de machine learning permettraient aujourd'hui de contourner ces limites.

Quel est le but de cette recherche ?

L'objectif de cette recherche était donc d'exploiter les techniques de machine learning pour prédire les comportements de jeu problématique à partir des données de compte des joueurs, ainsi que d'évaluer la stabilité temporelle de ces prédictions en

réduisant progressivement la quantité de données utilisées pour entraîner le modèle.

Plus précisément, l'étude visait à développer un modèle prédictif basé sur les comportements de jeu (fréquence des paris, pertes, durée des sessions, etc.), à tester si ces prédictions restent robustes lorsque la quantité de données d'entraînement est progressivement réduite (30, 60, 90 jours), et à évaluer la possibilité d'intégrer ces modèles dans des stratégies d'intervention précoce mises en place par les opérateurs de jeux.

Comment les chercheurs ont-ils fait pour répondre à cet objectif ?

Les participants étaient des joueurs basés en Suède (n=35 048) enregistrés sur la plateforme de jeux en ligne LeoVegas, sur une période de 4,5 ans (2019-2023).

Les données collectées sur les joueurs portaient sur des :

Données financières (montants des dépôts et retraits, échecs de paiement, solde net, utilisation des bonus et promotions)

Comportements de jeu (fréquence des paris, taille des mises, durée des sessions de jeu, mises croissantes après des pertes)

Outils de jeu responsable (utilisation des outils d'auto-exclusion, paramétrage de limites de dépôt ou de temps de jeu, demandes d'assistance au support client pour des problèmes liés au jeu)

Données démographiques et administratives (âge, sexe, localisation des joueurs, nombre de comptes par joueur)

Ces données ont d'abord été prétraitées pour extraire des tendances selon les différentes variables. Ensuite, deux techniques de machine learning ont été utilisées permettant d'identifier les variables les plus pertinentes (SHAP SHapley Additive exPlanations) et de sélectionner les meilleures variables pour maximiser la précision des prédictions de jeu problématique (GMLVQ Generalized Matrix Learning Vector Quantization). Puis une étape de modélisation (XGBoost Extreme Gradient Boosting) a permis d'identifier les joueurs à faible risque et ceux à risque élevé de jeu problématique. Enfin, une évaluation de la stabilité temporelle du modèle a été réalisée sur des périodes réduites (30, 60 et 90 jours) pour voir si les prédictions restaient fiables avec moins de données historiques.

Quels sont les principaux résultats à retenir ?

Le modèle utilisé a montré une bonne capacité à identifier les joueurs à risque et une précision élevée des prédictions. Les variables les plus influentes d'un comportement de jeu problématique retrouvées étaient : le comportement de poursuite des pertes (mise croissante après des pertes), la tendance du solde net (évolution des gains/pertes sur le temps) et les montants maximaux déposés (gros dépôts répétés

en peu de temps). De plus, la performance du modèle est restée stable et robuste même en réduisant la quantité de données d'entraînement. Ainsi, les modèles entraînés sur 30, 60 et 90 jours de données ont produit des prédictions comparables à celles basées sur l'ensemble des 4,5 ans de données. Ces résultats indiquent que les opérateurs de jeux pourraient détecter les comportements de jeu problématique précocement, en seulement 30 à 60 jours de suivi.

Les points clés à retenir

Le machine learning appliqué aux jeux d'argent et de hasard en ligne peut identifier précocement les joueurs à risque avec une bonne précision. Les modèles restent fiables même avec une quantité réduite de données, facilitant leur utilisation en temps réel. Cette approche permettrait d'améliorer les stratégies de jeu responsable en détectant rapidement les comportements de jeu problématique. Ces résultats nécessiteraient d'être répliqués sur d'autres plateformes de jeu et chez d'autres opérateurs.

Plus d'informations sur cette recherche :

Andersson, S., Carlbring, P., Lyon, K., Bermell, M., & Lindner, P.

Insights into the temporal dynamics of identifying problem gambling on an online casino: A machine learning study on routinely collected individual account data. Journal of Behavioral Addictions (2025)

[» Lien](#)