

# Prédiction des événements rares : application à la prédiction du clic.

Slimane Makhlouf\*, Avner Bar-Hen\*\*, François-Xavier Jollois\*\*\*

\*Velvet Consulting

slimane.makhlouf@hotmail.fr,

\*\*CNAM

avner.bar-hen@lecnam.net

\*\*\*LIPADE, Université Paris Cité

francois-xavier.jollois@u-paris.fr

**Résumé.** Lors de l’affichage de publicités en ligne, un système d’enchères en temps réel est mis en place pour choisir l’annonceur. Du côté de ce dernier, à partir d’informations sommaires (majoritairement catégorielles), le but est de prévoir si l’utilisateur va cliquer sur l’affichage ou non (et donc de choisir d’enchérir ou non). Les clics étant rares, de l’ordre de 1 pour 1000 en moyenne, il y a un fort déséquilibre entre les deux classes (clic ou non-clic). Dans ce cadre, les modèles de prédiction habituels sont biaisés, puisqu’ils ne prennent pas en compte ce déséquilibre. Les mesures d’évaluation, telles que  $AUC_{ROC}$ , sont également défectueuses lorsqu’elles sont appliquées à ces événements rares. Dans cet article, nous étudions ces biais dans le cadre des enchères temps réel et proposons une nouvelle mesure d’évaluation. Nous concentrons notre analyse sur la prédiction du clic ( $pCTR$ ), en considérant des algorithmes de prédiction linéaires (régression logistique) et non linéaires (Deep Factorization Machine). Nous évaluons la performance des modèles à l’aide d’une fonction d’évaluation probabiliste spécifique incluant les coûts liés à l’enchère, et nous la comparons à plusieurs mesures d’évaluation classiques. Cette mesure met en évidence les limites des métriques classiques dans les problèmes de prédiction fortement déséquilibrés. Elle permet ainsi une meilleure évaluation des modèles spécifiques à la prédiction du clic et fournit également des indications quant à la rentabilité de la campagne d’affichage.

## 1 Introduction

Ces dernières années, le marché de la publicité en ligne a connu une croissance exponentielle. Dans le même temps, les méthodes d’attribution de ces publicités ont beaucoup évolué. Le real-time bidding (RTB) est le principal moyen pour les annonceurs d’acheter et d’afficher leurs publicités en ligne. Les algorithmes RTB sont conçus pour optimiser les transactions afin que les éditeurs reçoivent le meilleur prix par affichage remporté et que les annonceurs puissent diffuser leurs annonces auprès du public le plus pertinent au moindre coût. Chaque emplacement publicitaire disponible est mis aux enchères sur une plateforme AdExchange où chaque