

Extraction de signaux comportementaux d'utilisateur et prédiction de conclusion de vente

Diana Nurbakova*, Sylvie Calabretto*
Timothée Saumet **, Jialiang Wei***

*LIRIS UMR 5205 CNRS, INSA Lyon - Université de Lyon
diana.nurbakova@insa-lyon.fr, sylvie.calabretto@insa-lyon.fr

**Tilkee SAS

tim@tilkee.com

***AKKA Technologies

jialiang.wei@outlook.com

*Le problème de prédiction de conclusion de vente, PCV (deal closure prediction) consiste en l'estimation de la probabilité de convertir une opportunité (un client potentiel) en une vente. Ainsi, il est souvent formulé et abordé en tant que problème de classification binaire auquel sont appliqués des algorithmes d'apprentissage automatique (e.g. (Eitle et Buxmann, 2019)) et dont l'enjeu principal est l'extraction de features de différentes sources (e.g. (Aggour et Hoogs, 2013)). Dans le contexte industriel, ces features sont dans la majorité des cas directement issues de données du système de gestion de la relation client (GRC). Cependant, le problème de la qualité des données, en particulier les données manquantes et/ou erronées, persiste. Il est peut être lié au renseignement manuel par l'équipe commerciale mais aussi à l'absence de traces automatiques d'interactions des commerciaux avec les opportunités. Plusieurs travaux existants s'intéressent aux modèles de pertinence de document / résultats de recherche basés sur *les taux de clics* (CTR) (e.g. (Yu et al., 2019; Zheng et al., 2019)). Ainsi, ils considèrent peu l'engagement post-clic de l'utilisateur (Grinberg, 2018), en s'appuyant souvent sur le temps passé sur la page (dwell time) (Kim et al., 2014; Nurbakova et Saumet, 2020). Ceci est principalement lié à l'absence de données de log de traçage d'interactions d'utilisateur. Cependant, les interactions d'utilisateur avec le contenu des pages affichées (Huang et al., 2012; Grinberg, 2018) ont été utilisées afin d'étudier l'intention utilisateur ainsi que son attention lors de la lecture. Contrairement à ces travaux existant, nous cherchons à relier les interactions d'utilisateur lors de sa lecture de documents commerciaux avec son intention de conclure la vente. Dans (Nurbakova et Saumet, 2020), nous avons montré que la précision de la PCV basée uniquement sur le comportement de lecture de documents est relativement haute. Or, les features créées et utilisées par cette approche sont principalement de multiples agrégations de temps passé sur les pages de documents et des fréquences de visites. Ainsi, elles ne sont pas suffisamment discriminantes et n'exploitent pas le potentiel de tracking.*

L'objectif principal de ce projet consiste en l'élaboration d'un algorithme d'extraction de signaux comportementaux d'utilisateur caractéristiques de son intérêt commercial et issus de sa lecture de document numérique. Ceci va servir de base à l'approche de PCV. Pour atteindre l'objectif visé, nous proposons une étude en plusieurs **étapes** : (1) Analyse de l'impact