

## LEARNING ANALYTICS : QUE DIT LA RECHERCHE SCIENTIFIQUE ?

### **Analytics, Data Sciences & Data Mining**

Le terme Analytics désigne de façon usuelle des techniques informatiques, mathématiques et statistiques pour révéler une information pertinente, à partir de très larges ensembles de données. Par extension, les Analytics permettent de décrire, analyser, prédire et prescrire des actions, passées et à venir, dans une quête de performances et d'efficacité.

Big Data : désigne un ensemble de données caractérisé par cinq dimensions (modèle des 5 V) : un volume massif, une collecte vélocité, une information variée et des modèles variables, une véracité à contrôler. Par contraste, on désigne par l'expression *Thick Data* des données qui permettent de comprendre des contextes.

Data Mining : consiste à appliquer des algorithmes, issus de la statistique ou de l'intelligence artificielle, pour extraire de l'information (utile et inconnue) à partir de gros volumes de données non structurées.

#### DE QUELLES DONNÉES PARLE-T-ON ?

Il est question de données qualitatives et de traces : les données qualitatives correspondent par exemple aux réponses des usagers à des formulaires, alors que les traces regroupent l'ensemble des interactions d'un usager avec son environnement numérique d'apprentissage.

### **Données d'apprentissage : deux jeunes communautés scientifiques en cours d'hybridation**

Big Data, intelligence artificielle, Data Mining, Analytics sont devenus des termes usuels pour penser et animer notre monde. Le domaine éducatif est longtemps resté étranger à cette mutation.

Cependant, de nombreuses communautés œuvrent dès les années 1990 à la fouille de données éducatives, la recherche sur les environnements informatiques pour l'apprentissage humain, la collaboration assistée par ordinateur.

La communauté scientifique AIED (*Artificial Intelligence in Education*) est la plus ancienne (1993). En 2007, la communauté scientifique IEDMS<sup>1</sup> (pour *International Educational Data Mining Society*) émerge, puis une autre en 2011, avec la création de SoLAR (*Society for Learning Analytics Research*).

Dans ces deux dernières communautés (IEDMS et SoLAR), les chercheurs partagent un même enjeu : à partir des traces numériques d'apprentissage, améliorer les expériences d'apprentissage et leur environnement. Pourtant, ils définissent différemment leur champ d'expertise.

#### QUELLES SONT CES DIFFÉRENCES ?

Spécialisée en fouille de données, la communauté EDM priorise le développement de nouveaux outils et algorithmes pour révéler les modèles enfouis dans les données d'apprentissage. Ces modèles touchent à des connaissances pointues, par exemple, dans l'apprentissage du calcul. Dans cette communauté, la conception d'algorithmes vise ainsi à donner au logiciel la capacité de prédire les résultats d'un apprenant et de personnaliser sa stratégie d'apprentissage. Dans le sillon de SoLAR et de leur domaine de recherche Learning Analytics, en revanche, modélisation et visualisation des données sont transmises aux acteurs de l'apprentissage (apprenants, personnels d'éducation, enseignants, etc.). C'est là la différence fondamentale entre EDM, dont le produit de la recherche alimente une machine, et Learning Analytics, qui visent à amplifier le rôle décisionnel des acteurs de l'apprentissage. Les Learning Analytics deviennent un moyen, pour la communauté éducative, d'améliorer l'expérience d'apprentissage. Le véritable enjeu est alors de ne pas noyer les acteurs dans les données mais de leur donner des moyens pertinents de les explorer<sup>2</sup>.

1. Aussi appelée communauté EDM.

2. Charleer S., Klerkx J., Duval E. (2014), « Learning Dashboards », *JLA*, vol. 1, n° 3, p. 199-202.

# Questions de recherche en science des données de l'apprentissage

Les tenants des Learning Analytics partent du principe que la numérisation des activités d'apprentissage va offrir aux acteurs de la chaîne éducative des données robustes, issues du terrain. Qu'ils mobilisent ces indicateurs en présentiel comme à distance, en temps réel comme en différé, à un niveau personnel ou agrégé, ces acteurs seraient constamment en position de faire des choix éclairés pour améliorer leur expérience.

Nous présentons dans la suite quatre questions de recherche. Cette présentation ne saurait rendre compte de l'ensemble des travaux entrepris dans le domaine des sciences des données de l'apprentissage, mais il prétend éclairer les questionnements les plus actifs et les recherches les plus abouties.

Pour chacun des thèmes retenus, ces recherches peuvent être motivées par des logiques diverses : d'une part, des protocoles de recherche fondamentale ambitionnent de capter la moindre interaction de l'apprenant pour agir sur son apprentissage et l'améliorer (Learning Analytics) ; d'autre part, une recherche davantage appliquée se soucie de donner aux acteurs institutionnels les moyens de mesurer les flux d'apprenants et d'agir sur les performances du système (Academic Analytics).

## PRÉDIRE LA PROGRESSION DE L'APPRENANT

Avec l'acquisition en temps réel de données d'apprentissage toujours plus fines et massives, les chercheurs ont développé des techniques de fouille de données pour classer des profils, les orienter en fonction des capacités estimées, adapter les contenus, déployer des stratégies d'engagement et lutter contre le décrochage.

## MESURER LES INTERACTIONS SOCIALES

En s'appuyant sur les apports fondamentaux de Lev Vitgovsky et du socio-constructivisme, une communauté scientifique s'attache à évaluer le rôle des interactions sociales au sein des « environnements informatiques pour l'apprentissage humain » (EIAH) : la communauté CSCL (*Computer-Supported Collaborative Learning*), née en 1995, porte cette conviction que la cognition et l'action humaines sont socialement et culturellement médiées. L'objectif est donc de comprendre et valoriser ces processus d'apprentissage collectif à partir des interactions sociales médiées par des dispositifs numériques.

## ANALYSER LE DISCOURS

Le langage est l'un des premiers vecteurs par lequel les apprenants construisent du sens : son usage est influencé par les buts, les sentiments et les relations, très variables selon les contextes.

## DONNER À VOIR L'APPRENTISSAGE

À l'origine des Learning Analytics, les Visual Data Analytics (analyses visuelles des données) reposent sur deux principes essentiels : donner aux acteurs de la communauté éducative un pouvoir de décision par l'exploration du modèle qui sous-tend les données d'apprentissage elles-mêmes. Ces tableaux visent à seconder l'enseignant dans le développement de l'attention, la compréhension et la métacognition des apprenants.

Ces quatre questions de recherche ont solidement établi la plus-value du champ scientifique des Learning Analytics et fixé leur visée ultime : élever chaque acteur en décideur, adaptant ainsi la notion d'*empowerment* (ou d'autonomisation) au monde de l'éducation. Les dispositifs expérimentaux se diversifient, qui soutiennent la mutation de l'apprentissage : l'instrumentation de la relation entre enseignant et enseignés peut permettre une proximité croissante, un suivi plus personnalisé et l'émergence d'une communauté d'apprentissage. Pourtant, un défi majeur se pose aujourd'hui, pour passer de protocoles expérimentaux à la diffusion de solutions innovantes.