



## Une plateforme pour partager le travail des data scientists

Entre les données et les connaissances qu'on en tire, le travail des data scientists est souvent un secret inaccessible. Pour ouvrir cette boîte noire, le Swiss Data Science Center a développé la plateforme Renku. Notre rédaction a rencontré ses concepteurs. Rodolphe Koller

«Les données, ce n'est pas la connaissance». Olivier Verscheure souhaite mettre les choses au clair dès le début de notre échange dans son bureau de l'EPFL: on ne va pas parler de data mais bien du savoir qui en est extrait par les data scientists et les experts. Le responsable du Swiss Data Science Center (SDSC) - une initiative commune des deux EPF lancée en 2017 - nous reçoit avec Eric Bouillet, son collègue responsable de la plateforme Renku. Nous sommes justement là pour parler de cet outil qui matérialise une promesse faite dès l'inauguration du SDSC de «comblé le fossé qui existe entre les data scientists et les experts de domaines scientifiques spécifiques».

### Ouvrir la boîte noire du travail sur la donnée

Faisant office de middleware entre les données et les outils d'analyse, la plateforme Renku enregistre de manière organisée le travail des data scientists - les jeux de données choisis, les sélections et nettoyages opérés, les modèles et algorithmes appliqués, etc. - de façon à dresser une filiation entre les données et la connaissance qui en est extraite. L'outil donne ainsi de la transparence à la boîte noire du travail sur la donnée. Pour la science, l'enjeu est énorme: «Les expérimentations deviennent répliquables instaurant de la confiance quant à leurs résultats», explique Olivier Verscheure. De plus, Renku fonctionnant comme un outil de partage, les data scientists peuvent profiter des recettes développées par leurs collègues.

### Coder les expérimentations

Au fur et à mesure de son emploi, Renku développe ainsi une gigantesque arborescence des voies tracées entre les données et les connaissances. Pour faciliter cette capture

d'informations, la plateforme qui s'appuie sur Git, enregistre un grand nombre d'informations de manière automatisée via Docker (données, bibliothèques, versions, etc.). «Nous avons renoncé à tout enregistrer sans quoi il y aurait trop de bruit, explique Eric Bouillet. Ce sont les utilisateurs qui décident de ce qui est conservé dans la plateforme - l'automatisation leur facilite la tâche».

A terme, Renku pourrait intégrer des ontologies spécifiques à certains domaines - sciences de la vie, environnement, etc. - pour coder les connaissances métier résultant de l'analyse de données. La plateforme pourrait également renseigner les experts sur le fait que tel algorithme fonctionne particulièrement bien avec tel jeu de données.

### Du monde académique à l'industrie

La plateforme Renku est aujourd'hui exploitée et employée par les chercheurs de diverses facultés en collaboration avec les data scientists du SDSC. En raison de la sensibilité de leurs données, les sciences de la vie de l'EPFL disposent de leurs propres instances. Hors milieu universitaire, Renku est notamment utilisé par la société suisse Bühler, spécialisée dans les machines pour l'agroalimentaire et pionnière en matière d'intelligence artificielle. Outre l'apport dans le travail quotidien des métiers travaillant autour de la donnée, la plateforme a en effet de quoi séduire l'industrie. Documenter la chaîne de création de valeur entre la data et la connaissance, c'est aussi s'assurer que l'entreprise conserve ces savoirs et savoir-faire, notamment lorsque les équipes changent. Les données sont déjà considérées comme un capital important pour l'entreprise, alors que dire des opérations qui les transforment en connaissances...

«Les expérimentations deviennent répliquables instaurant de la confiance quant à leurs résultats.»

Olivier Verscheure, Swiss Data Science Center