

DATA & ENTREPRISES :

des opportunités à saisir et des enjeux à relever

Pierre-Henri Cumenge (X02), cofondateur et CTO de Sicara, nous présente cette société et son positionnement dans le monde de la data et de l'intelligence artificielle. Il revient notamment sur le périmètre d'action de Sicara et dresse un état des lieux de la place de la data dans le monde de l'entreprise. Rencontre.



Pierre-Henri Cumenge (X02)

Que propose Sicara ?

Sicara est une société de services spécialisée dans la data. Nous travaillons avec nos clients pour construire des solutions qui couvrent l'ensemble de leur chaîne data. Historiquement, nous nous sommes d'abord positionnés sur des produits de reconnaissance d'images. Nous nous sommes très vite rendu compte que la compétence purement algorithmique ne suffisait pas. Il nous a fallu élargir notre périmètre de compétences. D'une part, nous avons développé une compétence sur le data engineering qui est la capacité à intégrer, agréger et mettre à disposition des flux de données issus de diverses sources dans un entrepôt de données. D'autre part, nous intervenons sur des phases vision produits afin de concevoir des solutions utilisables et utilisées par les utilisateurs finaux. En effet, cette typologie de projets requiert une compréhension fine d'enjeux spécifiques à la donnée et des expertises dont nos clients ne disposent pas forcément en interne : par exemple la manière de mesurer la

pertinence d'un algorithme au regard des enjeux métier, ou le contrôle de la qualité de la donnée. Dès le départ, nous avons aussi accordé une attention particulière à la qualité du développement logiciel pour accélérer la mise en production des projets et produits. Une de nos missions est de garantir que les produits soient intégrés aux systèmes d'information de nos clients et que les applications data qui en résultent soient robustes et pérennes. Il est donc critique que nos data scientists aient aussi la compétence nécessaire pour non seulement développer des algorithmes performants, mais aussi pour les exposer au sein d'une base de code facile à déployer, maintenir et monitorer.

Au-delà, nous avons la particularité et l'avantage de faire partie du groupe Theodo qui regroupe 10 entités aux expertises complémentaires. Cela nous permet de couvrir l'ensemble des expertises nécessaires au déploiement de solutions data, par exemple pour mettre en place l'infrastructure cloud sous-jacente ou développer les applications web ou mobiles exploitant les données et algorithmes sur lesquels nous travaillons. Cela permet à nos clients de simplifier la gestion de leurs projets avec une approche de guichet unique. Au-delà de cette capacité à développer un produit de bout en bout pour nos clients, nous nous appuyons aussi sur deux autres sociétés du groupe respectivement spécialisées dans le domaine de la santé et de la finance. Pour nos clients positionnés sur ces secteurs, c'est la possibilité d'avoir accès à des experts métier.

Pouvez-vous nous donner des exemples pour illustrer votre périmètre d'action ?

Nous avons construit une application de facturation automatique des plateaux repas pour un acteur spécialisé dans la fabrication de matériel pour la restauration. Cela permet de réduire les files d'attente aux caisses et aux clients de manger chaud !

Concrètement, quand ces derniers posent leur plateau sur une borne, des photos sont prises et analysées grâce aux techniques de reconnaissance d'images, ce qui permet de reconnaître les plats et de calculer instantanément le montant à facturer. Pour ce projet, nous sommes partis d'une feuille blanche. Le développement de la solution technique a nécessité une importante phase de R&D initiale. Aujourd'hui, le projet est en plein déploiement avec plusieurs dizaines de restaurants déjà équipés. L'objectif est maintenant d'accélérer ce déploiement à moindre coût afin que notre client puisse proposer cette solution à l'ensemble de sa clientèle.

Dans l'univers médical, pour un grand laboratoire pharmaceutique, nous avons travaillé sur une solution visant à digitaliser les études cliniques menées dans le domaine des maladies neurodégénératives. Cela a permis un gain moyen de 18 mois sur l'ensemble du processus qui contribue à réduire les coûts et accélérer la recherche sur les maladies neurodégénératives. Les études cliniques sont une étape indispensable à la mise sur le marché de nouveaux médicaments, durant lesquelles le laboratoire sollicite des patients qui se sont portés volontaires. Ce sont des phases longues et coûteuses. Les patients participant doivent se rendre régulièrement à l'hôpital pour participer aux tests visant à collecter les données. Nous avons travaillé sur une application mobile permettant de faire ces tests depuis chez soi sur un smartphone, avec deux impacts : limiter le temps pris aux médecins pour réaliser les tests sur place, donc le coût d'une étude, et élargir le pool de patients potentiels à l'ensemble des zones géographiques, ce qui augmente les chances de mener à bien les études sur des maladies rares. Sur ce projet, nous avons collaboré avec d'autres start-up du groupe : BAM pour le développement de l'application mobile, Padok pour l'ingestion, le traitement et la mise à disposition de

la donnée, et Hokla qui apporte son expertise en développement de dispositifs médicaux. Nous sommes mobilisés sur ce projet complexe depuis deux ans. Il implique un contrôle optimal de la donnée, de son utilisation et de la documentation des flux parce que nous intervenons sur un secteur soumis à de très fortes contraintes réglementaires. Enfin, nous avons aussi récemment accompagné une entreprise à deux niveaux : la mise en place d'une solution technique et celle de son équipe data. Cette société fournit des informations et de la donnée très fine sur la fréquentation des lieux. Elle aide, par exemple, à optimiser la prise de décision d'une enseigne sur un futur emplacement. Nous les avons aidés à concevoir et mettre en place l'architecture technique de leur data lake et des flux d'ingestion et d'agrégation de données, et à poser les bases de leur organisation data. Aujourd'hui, ils disposent de leur propre équipe et poursuivent de manière autonome le développement de leur solution.



Sur l'ensemble de ces enjeux et problématiques, quel est le niveau des entreprises ? Qu'observez-vous ? Quels sont les freins qui persistent ?

C'est très hétérogène ! À notre niveau, nous observons un écart qui se creuse entre les entreprises qui ont une démarche proactive vis-à-vis de la data et ceux qui n'en ont pas encore mesuré toute l'importance. La différence vient en général d'un sponsorship fort au niveau comex. En parallèle, d'autres secteurs ont été contraints de se lancer par la pression du marché ou la réglementation. C'est notamment le cas du secteur bancaire. Sur un sujet comme le KYC (Know Your Customer), le contrôle de la qualité de la donnée et sa maîtrise sont une obligation réglementaire assortie de fortes pénalités, ce qui a poussé les acteurs concernés à s'intéresser fortement à la valeur ajoutée qu'une bonne maîtrise de la donnée peut apporter. Encore aujourd'hui, il y a une forme de sous-investissement dans la donnée. Je suis très souvent en contact avec des entreprises qui ont recruté des data scientists en interne pour développer des algorithmes, mais qui, en amont, n'ont pas réalisé le travail nécessaire de récupération, de nettoyage et d'homogénéisation de la donnée, qui est la matière première que les data scientists vont exploiter. Sur un plan organisationnel, la bonne connexion et coopération des équipes data avec le reste des équipes pour fournir de la data aux métiers et concevoir des produits data avec une véritable utilité est une nécessité absolue. Les modèles organisationnels autour de la data constituent un sujet de discussion et d'expérimentation très fort en ce moment. Nous observons notamment un mouvement vers la décentralisation de la responsabilité de la donnée afin d'impliquer plus fortement les pro-

ducteurs de données. En parallèle, nous nous dirigeons vers un rééquilibrage entre les parties data science et data engineering. Comme évoqué plus haut, le data engineering permet une mise à disposition d'une donnée de bonne qualité. Cette prise de conscience qui s'opère depuis environ 5 ans crée, et va continuer à créer, une forte tension sur les profils de data engineers sur le marché du travail. Si, au cours des dernières années, nous avons vu apparaître de nombreuses formations et masters spécialisés en data science, l'offre reste assez limitée sur la partie data ingénierie. D'ailleurs, en interne, actuellement, nous recrutons essentiellement sur cette partie data engineering et formons d'anciens data scientists ou développeurs au data engineering. En parallèle, nous recherchons aussi des profils avec une spécialisation dans la partie analytique et business intelligence, c'est-à-dire des experts capables de mettre en place des tableaux de bord pour rendre la donnée accessible visuellement aux décideurs.

Et pour conclure, quelles sont les tendances qui vous intéressent ? Comment vous projetez-vous ?

Nous évoluons dans un secteur en pleine effervescence. De plus en plus d'entreprises se dotent de Chief Data Officers (CDO) et ont pris conscience que le sujet de la data doit être traité dans les sphères décisionnelles de leur organisation. Pourtant, actuellement, peu de CDO sont membres des comités de direction des entreprises. Je pense que cette situation va évoluer et que dans quelques années, on trouvera anormal de ne pas avoir un CDO membre d'un comité de direction. En parallèle, sur le plan technique et technologique, le rythme des évolutions

accélère. Il y a un intérêt croissant pour le concept de data centric AI ; la gestion du cycle de vie d'un algorithme ou encore le SSL (Self-Supervised Learning) afin de réduire les coûts relatifs à l'entraînement des algorithmes sur d'importantes volumétries de data. Côté traitement de données, nous observons un usage croissant des data plateformes, comme Databricks ou Snowflake avec qui nous avons mis en place des partenariats afin d'accélérer les développements de nos projets client. Ce type de plateformes leur permet d'avoir une solution intégrée pour l'ensemble des traitements et exposition de données, donc de limiter la complexité technique de leurs solutions data.

Au delà de la communauté data, l'explosion des performances des LLM (Large Language Models) et leur accessibilité au plus grand nombre a le potentiel de modifier en profondeur la manière de travailler pour un grand nombre de métiers. Dans les métiers du numérique par exemple, l'assistance à la génération de code ou à la recherche d'informations pour déboguer son code devraient transformer le quotidien des développeurs. Tous les métiers touchant à la création de contenu ou à la recherche d'information sont concernés, et le changement pourrait être très rapide. Ce qui est fascinant ici pour moi c'est que nous assistons à des avancées qui arrivent beaucoup plus tôt que ce que j'aurais imaginé. Il y a 5 ans, si on m'avait demandé combien d'années de recherche étaient nécessaires pour obtenir un système d'intelligence artificielle capable de produire ce que les outils d'openAI par exemple font aujourd'hui, j'aurais probablement répondu plus de 10 ans. L'engouement récent pour chatGPT parle de lui-même : il a été testé par 1 million d'utilisateurs en moins de 3 jours. ×