BIG DATA LIVE

Session #1

Inria INVENTEURS DU MONDE NUMÉRIQUE

tinyclues

Compte-rendu de l'atelier du 29 octobre

à Cap Digital

Compte-rendu

Contact :
ecosysteme@capdigital.com
Site Web :
www.capdigital.com/strategies/

cap-digital Paris Region

Sommaire

Speakers

Inria INVENTEURS DU MONDE NUMÉRIQUE — scikits learn machine learning in Python

Gaël Varoquaux**, chercheur à l'**INRIA,** project leader **Scikit-learn

tinyclues

David Bessis**, président de **tinyclues

I. Apprentissage statistique : créer des machines intelligentes

*par **Gaël Varoquaux**, chercheur à l'**INRIA,** project leader **Scikit-learn***

*Chercheur à l'**INRIA** sur l'analyse du cerveau à partir d'imagerie médicale dans l'équipe PARIETAL, **Gaël Varoquaux** est également l'un des porteurs du projet **Scikit-learn**, qui vise à démocratiser les techniques de machine-learning aux industriels et aux chercheurs issus d'autres domaines que les mathématiques.*

L'apprentissage statistique en deux mots

Le terme *d'intelligence artificielle* apparaît dans les années 80, pour désigner la conception manuelle de règles de décision, implémentées ensuite dans les machines. L'IA posait ainsi les bases de l'apprentissage : la décennie suivante, le « machine learning » fait son apparition, et consiste à créer ces règles de décisions à partir d'un échantillon d'observations. Les statistiques prennent de l'importance dans les années 2000 pour modéliser le bruit dans les observations. Aujourd'hui, le big data propose une forme d'apprentissage où des règles simples sont déduites d'un grand nombre d'observations.

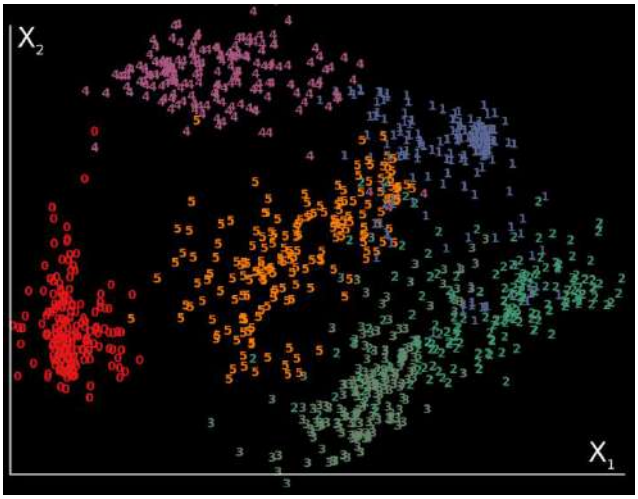
Clin d'œil à l'assistance, Gaël Varoquaux cite Steve Jurvetson, VC de référence dans le domaine en Silicon Valley : « Big data isn't actually interesting without machine learning ».

Pour illustrer son propos, il prend l'exemple de l'apprentissage statistique sur la reconnaissance de visage. Une méthode consiste à stocker toutes les images connues dans une base de données (bruitées). Si une nouvelle image (bruitée elle aussi) est présentée, on recherche alors l'image partageant le plus de similitudes. C'est la méthode des « plus proches voisins ».

Introduisant la notion de **données d'apprentissage** (celles sur lesquelles l'algorithme est affiné) et les **données de tests** (celles sur lesquelles on mesurera le taux d'erreurs de l'algorithme), il présente plusieurs problèmes entrant en ligne de compte concernant l'apprentissage statistique. Premièrement, le niveau de **bruit**, qui fait chuter naturellement le taux de prédiction. Ensuite, la fraction utile de l'image proposée à l'algorithme a également un impact sur le taux de prédiction. Le concept important est celui des descripteurs, c'est à dire une des caractéristiques dérivées de l'image qu'on utilise. Le niveau de bruit et le nombre de descripteurs représentent les difficultés clés pour l'apprentissage. Si il y a une valeur à prédire connue à l'avance, comme le nom des personnes sur les photos, on parlera alors de tâche supervisée, et donc de tâche non supervisée dans le cas contraire, ce qui permet de détecter des structures sous-jacentes à l'échantillon de données.

Sur un jeu de données cherchant à prédire une grandeur continue, la comme taille et le poids d'un individu, on effectuera une régression, cas particulier de l'apprentissage supervisé. Si une régression linéaire n'explique pas bien toutes les données, on pourrait chercher un modèle plus précis, permettant ainsi de minimiser l'erreur sur l'ensemble des données de l'échantillon. Mais cela entraîne le risque de sur-apprentissage : minimiser l'erreur n'est pas toujours favorable car cela présente le risque d'apprendre le bruit. Des modèles simples sont à privilégier, en équilibrant le nombre de paramètres à apprendre avec la quantité de données (concept de « régularisation »). Les mêmes concepts s'étendent sur des problèmes à plusieurs descripteurs, par exemple intégrant des facteurs génétiques, mais plus le nombre des descripteurs est grand, plus il y aura besoin de données. C'est la « malédiction de la dimensionnalité ».

Gaël Varoquaux présente ensuite le problème dit de « classification », par opposition à régression, qui cherche à prédire des variables par catégories, avec comme exemple à l'appui la classification des chiffres par reconnaissance d'écriture. Il présente un graphique montrant une classification à partir de deux descripteurs, où l'on remarque par exemple la difficulté à différencier les chiffres « 2 » et « 3 ».



On peut également chercher des catégories dans les observations sur des données non labélisées, plus courantes que les données labélisées. C'est alors un problème non supervisé, dit de « clustering ». Finalement, un autre problème d'apprentissage est celui des systèmes de recommandation. Par exemple, recommander des films à des utilisateurs en fonction des notes qu'ils ont données à des films qu'ils ont déjà vus. La difficulté est alors le peu de recoupement entre les utilisateurs et les films : deux utilisateurs ont rarement vu les mêmes films...

L'apprentissage statistique présente donc des défis aussi bien statistiques que computationnels. Il faut garder à l'esprit que l'accès aux données est souvent un plus grand frein que la puissance de calcul. Les stratégies adoptées dans des contextes de big data sont :

- Une réduction de données à la volée, rapide, mais respectant les propriétés statistiques du dataset, et une limite de la charge mémoire et disque ;
- Des algorithmes en ligne, c'est-à-dire qui consomment les données en flux et convergent vers des grandeurs moyennes ;
- Le parallélisme par découpage des données, cohérent par rapport à la structure statistique du dataset et adapté aux unités de calcul ;
- Une minimisation des temps de latence pour l'accès aux données (caching), sans recalculer la même chose, et en faisant appel à de la compression pour limiter la bande consommée.

De l'activité des neurones à la pensée

Le défi de mise en œuvre est donc réel, ce qui amène Gaël Varoquaux à présenter la thématique de ses recherches où il exploite des jeux de données conséquents (de l'ordre du téraoctet) grâce à de l'apprentissage statistique.

Au sein de l'équipe PARIETAL [1], l'exploitation d'imagerie fonctionnelle (enregistrement en continu de l'activité neuronale) permet de chercher à comprendre le lien bilatéral entre activité cérébrale et fonction cognitive. Pour cela, il faut explorer et mieux appréhender comment le substrat neural, les stimuli et les mécanismes de décisions sont reliés.

Connaissant les zones du cerveau activées, on peut prédire la fonction cognitive que le cerveau est en train de réaliser… et réciproquement. La force de l'apprentissage statistique permet de réaliser des « méta-analyses », recoupant beaucoup de jeux de données, utiles car chaque expérience ne touche qu'à peu de domaines cognitifs. L'équipe de recherche construit donc un atlas cognitif, fruit des méthodes d'apprentissage et de big data dans une discipline de sciences fondamentales.

Scikit-learn : une boîte à outils de l'apprentissage

Issue de travaux d'une thèse (David Cournapeau), cette librairie logicielle a été reprise pour besoins de travaux de recherche à l'INRIA. En effet, en 2010, le laboratoire décide d'investir dans ce projet d'apprentissage statistique, car ce dont avaient besoin les chercheurs pour travailler sur ces outils n'existait pas ou était réservé aux spécialistes en statistiques.

Nommée « scikit-learn », cette boîte à outils de l'apprentissage statistique vise :

- à mettre à disposition les algorithmes d'apprentissage statistique pour tous, sans domaine d'application spécifique, et ne requérant pas de compétences en machine learning ;
- à être une librairie logicielle de qualité, aux interfaces pensées pour l'utilisateur ;
- à être développée de manière communautaire, sous licence BSD, par des contributeurs très variés, répartis en Europe (historiquement, en France), aux Usa et en Asie du Sud-Est.

Scikit-learn est non pas un programme mais une bibliothèque Python (un langage haut niveau, ie plus proche du langage humain que du code binaire), c'est-à-dire une compilation de fonctions, prêtes à être utilisées par des programmes. Il apparaît ainsi plus expressif, polyvalent et facile à intégrer dans des contextes métiers variés. Selon Gaël Varoquaux, le choix de Python s'est imposé par son caractère interactif, facile à débuguer et d'application

[1] http://www.inria.fr/equipes/parietal

générale, par ailleurs soutenu par un écosystème très dynamique.

Comparé à d'autres librairies de machine learning, les performances opérationnelles de Scikit-learn sont parmi les meilleurs pour une série d'algorithmes courants, grâce à des optimisations au niveau des algorithmes eux-mêmes, et non à un plus bas niveau. Par ailleurs, la minimisation des copies des données contribuent à ces performances.

Scikit-learn est un cas de développement communautaire [2] remarquable par le nombre de contributeurs : environ 200), dont une trentaine de développeurs cœur, appuyés par un ingénieur INRIA à plein temps. Sur la base du modèle COCOMO [3], le coût de développement cumulé du projet est évalué à 6 millions de dollars.

Gaël Varoquaux pose enfin les facteurs clés de succès que l'initiative Scikit-Learn met en lumière :

1. Un « marché » dynamique, en l'occurrence une communauté Python active et des besoins en machine learning croissants
2. Des développements transparents
3. Une identité du projet non monopolisée en tant qu'INIRIA, malgré ses investissements conséquents La grande qualité technique du projet, contribuant à recruter des développeurs de haut niveau
4. La valorisation de ses contributeurs
5. Un recrutement actif de nouveaux développeurs
6. Des efforts de communication et de marketing autour de la solution et de ses bénéfices, passant entre autres par les initiatives « sprint », formes de hackathons autour des core developpers

Cela dit, la bonne conduite du projet a tendance à être fragilisée par le constat suivant : une fuite des cerveaux jugée sérieuse Les talents des développeurs étant naturellement repérés par le secteur privé, leur contribution active à la solution cesse.

En conclusion, Gaël Varoquaux rappelle les points clés de sa présentation :

- L'apprentissage statistique confère de la valeur au big data ;
- Il s'appuie sur les statistiques, et l'informatique théorique et appliquée ;
- Toutes les disciplines scientifiques, pas uniquement les neurosciences, connaissent un essor du big data
- Le logiciel est critique, et le développement libre bien adapté à son émergence et son développement.

Sur le continuum entre imagerie médicale et l'algorithmique de pointe, Gaël Varoquaux insiste sur le fait que l'apprentissage statistique doit être instancié avec la connaissance métier : c'est un outil extrêmement puissant qu'il ne faut pas hésiter à s'approprier au vu de ce qu'il permet pour intensifier le travail de recherche. Nous sommes trop peu de nos jours à maîtriser les données ET le code. C'est encore trop souvent soit l'un soit l'autre.

Sur l'opportunité de l'apprentissage statistique sur différents pans des sciences fondamentales comme la physique, il souligne que la connaissance métier reste indispensable, la valeur de ces outils considérée à sa juste valeur (sans surestimation), et ne saurait constituer seul le fondement de l'inférence scientifique.

Sur les compétences nouvelles pouvant s'imposer à des métiers comme le marketing, il affirme que ces capacités de compréhension et de maîtrise des outils informatiques vont devenir de plus en plus incontournables. Il se félicite que l'enseignement de l'informatique évolue, en classes préparatoires et même au lycée, et souligne que de nombreux dirigeants de ces entreprises sont issus de Grandes Ecoles où ces principes sont enseignés. Charles Huot rappelle l'initiative récente d'une formation dédiée portée par l'Institut Mines-Télécom[4], et une autre à venir par le LiP6 [5] (Laboratoire d'Informatique de l'Université Paris 6).Sur l'évolution des solutions, de plus en plus packagées (« boîtes noires »), il confirme la tendance tout en reprécisant que Scikit-Learn doit être adapté dans un contexte métier en tant que librairie, comme cela a été le cas pour la neuro-imagerie.

Sur l'enjeu de la personnalisation, qu'il s'agisse de marketing ou de pharmacologie, et la clusterisation des populations associées, Gaël Varoquaux confirme que la démarche est cohérente, comme en témoigne le challenge de Netflix, les traces existantes sur de telles bases de données, où l'on a depuis longtemps dépassé l'approche de la moyenne.

Sur la difficulté à gérer le bruit à mesure que les bases de données s'étendent, il souligne le fait que le frein est avant tout l'accès à du stockage performant, combinant à la fois volume de stockage et puissance de calcul. Ceci étant, les

[2] http://www.ohloh.net/p/scikit-learn
[3] http://fr.wikipedia.org/wiki/Constructive_Cost_Model

[4] http://www.telecom-paristech.fr/formation-continue/masteres-specialises/big-data.html
[5] http://www.lip6.fr/

modèles évoluent : les algorithmes sont poussés pour tenir compte de plus de paramètres…

II. Apprentissage dans de grandes bases de données e-commerce

*par **David Bessis**, président de **tinyclues***

***David Bessis** a fondé **tinyclues** après 10 ans de recherche en algèbre et géométrie dans des instituts tels que l'Université de Yale ou le CNRS. Docteur en mathématiques pures, il se consacre désormais au machine learning dans le contexte des grandes bases de données clients. Avec Jakob Haesler (précédemment chez McKinsey), il a co-fondé en 2010 la société tinyclues. Aujourd'hui composée d'une équipe de douze personnes, tinyclues opère une plateforme cloud de CRM prédictif qui permet aux annonceurs, et notamment aux acteurs du e-commerce, d'optimiser la qualité de leurs ciblages. Illustré par le cas PriceMinister, David nous présentera comment l'équipe marketing de ce grand e-Commerçant français se sert quotidiennement de la plateforme de tinyclues pour dépasser ses newsletters génériques et bâtir une stratégie de communication personnalisée.*

Big data ?

David Bessis commence sa présentation en prenant ses distances avec la définition classique des « 3V » de Gartner (volume, vélocité et variété), et annonce que son exposé suivra une approche orientée produits, plutôt que sur les algorithmes ou même les architectures technologiques pressenties.

Rappelant la décroissance fulgurante du coût du stockage de la donnée, il l'illustre par la réflexion de Jay Parikh, VP de l'ingénierie infrastructure de Facebook en Juin 2012, qui annonçait que 100 Po allaient devenir ennuyeux pour lui d'ici quelques années, sachant qu'aujourd'hui la société en stocke 180 par an. Ces 100 Po représentent par exemple la taille des génomes de tous les êtres humains[6]…

Machine learning, vers les structures implicites

Entre la révolution d'Altavista, moteur de recherche pionnier révélant la donnée non structurée du web, et l'approche « portail » de Yahoo à la même époque structurant la donnée par indexation manuelle, le paradigme Google permettra alors d'accéder à la structure implicite du web, constituée de connexions allant au-delà des simples liens hypertextes entre page. Le jeune chercheur qu'il était alors, ayant compris que la révolution était déjà actée, poursuivit sa carrière dans des institutions de renommée internationale (Yale, CNRS…).

Il prit toutefois conscience du potentiel du machine learning, paradigme permettant d'apprendre des données. Il partage son enthousiasme en démontrant comment cette approche permet, en partant d'une photographie d'un bâtiment d'Oxford partiellement masquée par des personnages, d'extraire la connaissance de l'image en apprenant, grâce à des clichés du même bâtiment pris sous un autre angle[7], la scène entière.

Les datasets d'e-commerce

Ayant démontré la force de la méthode, il explique alors la démarche qu'utilise la société qu'il a cofondée pour traiter les bases de données de type e-commerce. Ces bases sont fortement complexes tous les clients sont reliés à des produits, mais au travers de campagnes, provoquant emails et clics, ou encore de pages vues ou de recherches. Ces

[6] 1Go par personne, soit 10 000Po pour l'Humanité, pouvant être compressée raisonnablement à un ratio de 1%

[7] http://www.di.ens.fr/willow/research/inpainting/

bases complexes sont le use case principal qu'il s'est proposé d'expliciter.

Les travaux menés par tinyclues pour PriceMinister entrent parfaitement dans cette typologie de bases : 18 millions de clients, 200 millions de produits référencés, historique des recherches, clics web et des données de campagnes d'emailing. La solution de targeting proposée permet par exemple de générer une courbe de gain obtenue par machine learning, prédisant la probabilité d'achat en fonction du volume de la cible. Cette courbe étant obtenue par apprentissage statistique, permet ainsi d'envoyer moins de sollicitations à des personnes a priori plus sensibles au message planifié. Convaincu des résultats obtenus sur ce cas client, il suggère de dépasser la sociologie traditionnelle définissant a priori des classes d'agents, et de tirer profit de l'apprentissage pour s'appuyer sur des modèles prédictifs éprouvés. La sociologie s'apprend sur les données et ne se résume plus uniquement à un input du marketing.

Les modèles prédictifs

David Bessis précise alors les principes fondateurs des modèles prédictifs, en soulignant leurs forces mais aussi les précautions à prendre avec de telles approches. Par exemple, une régression linéaire (fournir une relation de proportionnalité entre deux variables à une constante près), si elle est la plus simple approche, peut vite montrer ses limites. La note de bas de page courante dans le milieu banquier *"Past performance is not a guarantee of future performance. »* est alors plus claire. Tout ceci implique donc de disposer de deux jeux de données : des données d'apprentissage, sur lesquelles les coefficients seront estimés, et des données de tests (distinctes !), sur lesquels des mesures de performance pourront être effectuées.

La robustesse de ces approches est un compromis à réaliser sur la complexité du modèle. Plus il est complexe (ordre de grandeur du nombre de coefficient s'approchant du nombre de données dans l'échantillon), plus il sera robuste sur les données de tests, mais plus les biais de sélection sur ces données génèreront des écarts sur un autre jeu de données lors de tests. C'est le risque de sur-apprentissage statistique, qui doit être évité en suivant des protocoles agissant comme juge de paix sur la robustesse du modèle.

Le graal du *machine learning* est d'obtenir des abstractions supérieures au jeu de données. David Bessis illustre ce principe en s'appuyant sur une vue en coupe de cellules rétiniennes, présentant des schémas distincts en fonction de l'échelle à laquelle on observe cette vue. Le *machine learning* permet de construire des cartes de descripteurs, ces derniers liant par exemple un jeu de données d'utilisateurs et un autre jeu de données qui serait un catalogue de films. C'est le problème qu'avait posé Netflix à la communauté du *machine learning* en 2006, proposant à ce titre 1 million de dollars au meilleur algorithme.

La sociologie… apprise

Un film peut en effet être décrit selon plusieurs critères : la violence, le sexe, le temps d'apparition d'une certaine actrice, la luminosité de la pellicule, l'adaptation d'un roman particulier, la présence de sabres de samouraï… Le nombre de critères potentiels est infini, or l'apprentissage statistique permet de minimiser l'erreur d'une fonction liant les bases de données des utilisateurs à celui des films présents en catalogue. C'est alors qu'apparaîtront les réels critères influant sur les préférences de tel utilisateur pour tel film.

Cette approche a été menée par tinyclues sur plusieurs jeux de données (versus un catalogue de produit) pour en définir les « méta-caractéristiques » les liant de manière identique à un certain type de produits : des prénoms, des surnoms, des codes postaux,… De la même manière, des familles de produits peuvent aussi être régénérées, ouvrant des possibilités robustes pour de la recommandation.

De l'algorithme au produit

Si Hal Varian, Chief Economist chez Google, affirmait que le métier le plus « sexy » dans les 10 années à venir serait celui de statisticien, ce qu'étayait le rapport de McKinsey sur le Big Data (évoquant un besoin aux US de 140 à 190 000 profils de datascientists), David Bessis pondère fortement cette intuition. Tout d'abord, il est peu concevable que le système de formation puisse répondre à ces besoins. Ensuite, un algorithme a priori aussi puissant que celui qui avait été couronné dans le challenge Netflix n'est pas forcément destiné à être implémenté (ce qui s'est effectivement produit pour Netflix, les coûts d'implémentation ayant été jugés trop élevés par rapport au gain attendu).

Ceci mène David Bessis à formuler deux idées essentielles. La première est qu'il faut distinguer un algorithme des produits qui sont attendus par les firmes pour effectivement extraire la valeur de leurs données. L'approche de tinyclues est en ce sens très claire, leur plateforme étant adaptée aux besoins métier des fonctions marketing de chaque client. Ensuite, outre les principes qu'il a explicités sur le *machine learning* et les attendus possibles, il insiste sur une culture du cloud qui est à encourager et à privilégier (des applications en SaaS et clé en main, pour dépasser la « simple » algorithmique).

Sur son parcours et son passage de la recherche à l'industrie, David Bessis précise qu'il a passé près de 10 ans en recherche fondamentale, sur un sujet très aride (algèbre et géométrie). Il a commencé par des courtes missions de consulting qui l'ont rapidement mené au poste de directeur R&D d'une agence de marketing, avant de se lancer dans l'aventure entrepreneuriale avec tinyclues.

Sur les limites du modèle linéaire, en particulier dans le domaine sensoriel (l'attrait pour un aliment peut dépendre du goût sucré, mais pas indéfiniment), David Bessis précise que les modèles linéaires sont effectivement les plus simples à décrire, et il suffit d'ajuster les coefficients pour minimiser une fonction d'erreur. Précisant sa déformation professionnelle en tant qu'ancien chercheur en algèbre, il rappelle qu'outre des modèles neuronaux ou logistiques, il est aussi possible d'emboîter ces modèles pour atteindre la performance souhaitée.

Sur le récent article du Times Magazine sur *Google VS death*, il affirme qu'en pharmacologie, les datasets contiennent des datasets relationnels relativement similaires en termes de complexité aux datasets du e-commerce. Il ne cache pas l'intérêt intellectuel qu'une telle action représenterait pour lui ! Gaël Varoquaux précise qu'un frein majeur est l'accès aux sources de ces données, piégées chez les différents laboratoires pharmaceutiques et par ailleurs difficilement harmonisables auprès de l'ensemble du corps médical.

Sur le sujet des infrastructures, David Bessis précise que tinyclues ne communique pas sur leur choix. Un débat naît sur la valeur des données, qui se révèlerait d'après un participant dès le stockage, au moment où de nombreuses organisations confient leurs données à des acteurs américains. Il souligne que les problématiques techniques sont en passe d'être résolues, et que selon les contextes métiers, il convient de trouver les solutions sur étagères adaptées au dit contexte.

David Bessis est interrogé sur le vocable de « prédiction », et précise alors que la promesse des outils de tinyclues est d'aider à prendre les bonnes décisions et à les rendre exécutables. Par ailleurs, et certes plus facilement chez les pure-players, l'organisation des structures que propose tinyclues évolue, et réallouent leurs ressources sur des postes plus orientés sur la prise de décision.

Sur le modèle économique de la société, il confirme que l'approche basée sur l'apprentissage statistique est à rendement marginal croissant, ce qui signifie que la performance prédictive de ces outils croît d'autant plus vite qu'ils sont confrontés à de nouveaux jeux de données. Toutefois, il précise que cette connaissance acquise par leur solution est strictement encadrée, évidemment par la CNIL concernant les données personnelles, mais plus largement par tous les descripteurs obtenus par apprentissage qui pourraient révéler une identité. Cette approche est explicitée dans les conditions générales de ventes de la société.

Sur le lien entre big data et le plus classique décisionnel, David Bessis explique que les solutions actuelles sont conçues pour être clés-en-mains, et à des publics qui jusqu'alors ne pouvaient opérer des solutions décisionnelles sans l'aide de profil statisticien.

Enfin, interrogé sur les risques de recommander un produit prohibé, par sa nature ou à internaute d'un certain âge, il précise que ces outils peuvent être complétés de couches de règles métiers, pouvant empêcher une prédiction (qui n'a pas pour autant moins de chances de se produire !) d'être actionnée.

Nous tenons à remercier Charles Huot, fondateur et corporate development de la société TEMIS, pour avoir accepté d'animer cette session. Président de la commission connaissance au sein de Cap Digital, vice-président Aproged, trésorier du GFII, Charles Huot est également président du comité éditorial de l'alliance Big Data, dont il a invité les participants à rejoindre le réseau social pour poursuivre les échanges.

A propos du groupe de réflexion

« Big Data Live » est un groupe de réflexion né à l'initiative de Havas Média et Cap Digital, sur le constat que les tissus académique et industriel français présentaient des talents compétitifs à l'échelle internationale dans le domaine. Havas Media et Cap Digital se joignent donc pour organiser un évènement permettant de les promouvoir et de les questionner sur les ingrédients de cette économie de la donnée.

Cap Digital

Pôle de compétitivité de la filière des contenus et services numériques, Cap Digital regroupe plus de 700 entreprises dont environ 650 TPE/PME et 25 grands groupes.

Plus de 50 organismes de Recherche et de Formation sont en relation directe avec les entreprises au travers de plusieurs centaines de projets collaboratifs en cours. Une communauté d'investisseurs est également très présente pour soutenir les actions de développement des entreprises.

Cap Digital œuvre à faire de la Région Île-de-France l'une des références mondiales du numérique, qui représente à lui seul un marché mondial de 3 000 milliards d'euros, tant d'un point de vue industriel que stratégique. Cap Digital organise le festival Futur en Seine, rendez-vous mondial annuel des forces vives de la création, de l'innovation et de l'économie numérique désireuses d'exposer, rencontrer, débattre, d'exprimer et de partager une vision du futur avec le grand public.

www.capdigital.com/strategies/bigdata

Contact Presse : Gaëlle Couraud / 06 33 54 93 90 / gaelle.couraud@capdigital.com

L'action de Cap Digital est soutenue par :

Havas Media

Havas Media est le réseau media du Groupe Havas Media, réseau présent dans 126 pays.

Notre mission est d'unir des marques et des personnes par des connexions significatives et de les accompagner jusqu'au succès. Nous aidons nos clients grâce à nos équipes de spécialistes dédiés à l'expertise média, à la stratégie, à la gestion internationale, au digital, au mobile, aux réseaux sociaux, au divertissement et au sport. Notre structure simplifiée et intégrée nous a permis de construire une des équipes globales les plus intégrées, vives et réactives du marché.

Pour plus d'informations, allez sur notre site www.havasmedia.com ou suivez-nous sur Twitter @HavasMedia