

Les fondamentaux du Big Data

Table des matières

I. Big Data : qu'est-ce que c'est ?	3
II. Exercice : Quiz	9
III. L'application du Big Data dans le domaine du marketing	9
IV. Exercice : Quiz	12
V. Essentiel	12
VI. Auto-évaluation	13
A. Exercice	13
B. Test	13
Solutions des exercices	14

I. Big Data : qu'est-ce que c'est ?

Contexte

L'explosion des technologies de l'information et de la communication a coïncidé avec la production de données numériques en masse. Ces données se sont vite révélées être un véritable filon, dont l'exploitation est devenue stratégique. Cette situation a progressivement conduit à l'émergence du concept de « *Big Data* » qui a permis de cerner le potentiel de la donnée et d'en faire un élément incontournable dans les interactions interhumaines comme entre l'homme et la machine. Comprendre le Big Data, ses enjeux et ses différentes applications est nécessaire pour en saisir la portée.

Définition

Les activités humaines sur Internet, les informations transmises et captées par les objets connectés, tout cela forme un tout : le Big Data. C'est un ensemble constitué d'une masse de données numériques provenant de plusieurs sources : les discussions sur différents médias en ligne (réseaux sociaux, messageries instantanées), les applications mobiles, les opérations e-commerces, les bases de données des industries et des entreprises. L'institut Gartner en donne une définition explicite en ces termes : « *Le Big Data est une forte volumétrie, haute vélocité et grande variété de données qui exigent des techniques innovantes et rentables de traitement d'information pour une meilleure prise de décision.* »

Bien qu'il soit difficile d'attribuer la paternité du terme Big Data à un individu en particulier, on reconnaît en John Mashey son principal vulgarisateur. Informaticien et chercheur à Silicon Graphics, il avait déjà esquissé en 1987 les contours du Big Data. Aujourd'hui, le concept a évolué et s'est enrichi de nouveaux principes.

Si l'on évoque souvent la profusion des données pour parler du Big Data, ce seul élément ne suffit pas à le caractériser. Le Big Data existe surtout au travers de la règle des 3 V édictée par Gartner. Les 3 V renvoient au volume, à la vélocité et à la variété des données constitutives de Big Data :

- Le **Volume**, c'est la masse ou la quantité des informations générées et dont la mesure s'exprime en différents ordres de mesures : pétaoctets, exaoctets, zettaoctets, yottaoctets ou en bronto-octets.
- La **Vélocité** fait référence à la vitesse à laquelle les données sont créées, recueillies, diffusées et analysées.
- La **Variété**, ce sont les propriétés que présentent les données, c'est-à-dire format, nature et structure. Elles peuvent être structurées (dates, numéros de téléphone, numéros de cartes de crédit, adresses), semi-structurées (données produites en continu, données XML, journaux ou logs des serveurs) ou non structurées (images, fichiers audio et vidéo, emails, textes, commentaires sur les réseaux sociaux).

Désormais, deux autres V viennent s'ajouter aux 3 V principaux, ce qui étend la liste à 5 V. Il s'agit notamment de la valeur et de la véracité des données. Par **valeur**, on entend ce que l'on peut tirer de la donnée en termes de connaissance et de profit. Quant à la **véracité**, elle détermine le crédit qu'on peut accorder aux sources à l'origine des données.

Vendez+ DATA

Big Data : Les 5V

- ▷ **Volume**
Enormément de données
- ▷ **Véracité**
La problématique de la fiabilité
- ▷ **Vélocité**
Toujours plus de données
- ▷ **Valeur**
La problématique du tri des données
- ▷ **Variété**
Des sources et formats

VENDEZ.PLUS

Les 5 V du Big data

Twitter exprime assez bien le principe de vélocité introduit un peu plus haut. On peut l'apprécier à la lumière de ces statistiques fournies par le site Planétoscope¹ :

« Chaque seconde, environ 5 900 tweets sont expédiés sur le site de microblogging Twitter. Cela représente 504 millions de tweets par jour ou 184 milliards de tweets par an. Cette masse d'information vient alimenter le flot d'informations (Big Data) publié par l'humanité chaque jour sur Internet. »

Pour se faire une idée du volume des mégadonnées, il suffit de s'intéresser à d'autres statistiques (findstack²) : « D'ici 2025, les experts indiquent que plus de 463 exaoctets de données seront créés chaque jour. ». À titre de rappel, selon l'ordre de mesure de stockage des données, 1 Exaoctet (Eo) = 1 024 Pétaoctets (Po), dont l'unité (1 Po) équivaut à 1 024 Téraoctets (To). Ce comparatif permet de se représenter la gigantesque masse de données qui pourra être produite sur une base quotidienne.



1 <http://www.planetoscope.com/>

2 <https://findstack.com/fr/big-data-statistics/>

Aperçu des unités de mesure en informatique**Les enjeux du Big Data**

Aujourd'hui, on ne fait plus grand cas de l'accès à la donnée. On s'inquiète davantage de qui l'exploite, comment et à quelles fins. Pour les entreprises, il importe de développer une stratégie Big Data à même de répondre aux problématiques de la satisfaction client, de la productivité et de l'innovation. On tend ainsi vers des enjeux sur trois points principaux : technologie, organisation et économie. Le défi, c'est de réussir à concilier la bonne technologie et la bonne organisation pour des performances économiques optimales.

Big Data, quels enjeux technologiques pour les entreprises ?

Le Big Data reste un défi de taille à relever pour l'entreprise et sa DSI (Direction des Services Informatiques). On entre ici dans le cadre de la *Business Intelligence* ou informatique décisionnelle. Se doter d'outils capables de traiter des données massives en temps réel et de fournir des analyses prédictives devient impératif. Le premier problème qui se pose est celui du stockage. Au regard de la quantité de données à gérer, les *data centers* apparaissent clairement comme les supports de stockage privilégiés. Diverses technologies permettent de répondre à la problématique de stockage de données volumineuses. Dans le domaine du traitement du Big Data, Hadoop est sans nul doute la référence. Plusieurs plateformes se sont basées sur son architecture pour tirer profit de sa capacité à prendre en charge de grandes quantités de données (de l'ordre de plusieurs pétaoctets).

Hadoop, c'est quoi ? « *C'est un framework libre et Open Source écrit en langage Java dont le but est de faciliter la création d'applications distribuées et scalables. Hadoop est donc conçu pour réaliser des traitements sur des volumes de données importants, de l'ordre de plusieurs pétaoctets.* » (Source : Saagie¹).

Il comporte plusieurs modules :

- Un système en charge des fichiers distribués, appelé HDFS (Hadoop Distributed File System),
- Map Reduce pour le traitement distribué des données,
- Un système dédié à la gestion des ressources du cluster (serveurs en réseau), Yarn (acronyme de Yet Another Resource Negotiator).

Dans un environnement distribué, le cluster Hadoop permet de conserver les données même quand les autres nœuds présentent des défaillances. L'architecture de Hadoop permet d'étendre le traitement des données structurées et non structurées sur plusieurs nœuds, ce qui augmente considérablement leur disponibilité, y compris en cas de panne. En dehors des solutions basées sur Hadoop, les entreprises peuvent aussi opter pour le cloud computing et sur les bases de données relationnelles SQL et non relationnelles NoSQL. Chacune de ces solutions a ses avantages, mais aussi ses inconvénients.

Complément

- **Hadoop**
 - Points forts :
 - Prise en charge des calculs simultanés ou parallèles,
 - Libre d'accès (open source),
 - Haute disponibilité des données, même en cas de défaillances techniques,
 - Manipulation possible de grandes quantités de données,
 - Rapidité de traitement des données.

1 <https://www.saagie.com/fr/blog/hadoop-et-big-data/>

- Points faibles :
 - Limité dans les calculs en temps réel,
 - Difficile à prendre en main.
- **Cloud computing**
 - Points forts :
 - Optimisation de l'utilisation des ressources,
 - Maintenance informatique simplifiée,
 - Dématérialisation du parc informatique.
 - Point faible :
 - Vulnérabilité des données aux cybermenaces.

L'adoption des systèmes de gestion de bases de données pose aussi le problème de leur compatibilité avec les solutions déjà présentes. Dans ce cas précis, le service informatique est tenu de veiller à l'interopérabilité entre les différents systèmes. La fiabilité des données collectées doit aussi être questionnée afin de réduire le taux d'erreurs au maximum. Prendre cette précaution est nécessaire pour ne pas fausser le processus décisionnel. Côté sécurité, la gestion des autorisations (accès aux outils Big Data) est essentielle.

Les enjeux organisationnels du Big Data

Le Big Data semble offrir les mêmes chances aux entreprises étant donné la facilité d'accès aux données. Mais la surabondance de l'information s'est peu à peu muée en une problématique que peu d'entités arrivent finalement à résoudre : l'identification de la donnée utile pour en faire un argument différenciant. Même si la quantité des données disponibles reste phénoménale, celles ayant de la valeur sont plutôt minoritaires et sont difficilement décelables. Pour isoler le bon substrat, il faut déployer des ressources matérielles et humaines suffisamment qualifiées. La mise en place d'un système de classification est donc nécessaire pour hiérarchiser les données.

Cela induit une stratégie de gouvernance des données dont la direction doit être confiée à un spécialiste de la donnée, idéalement un data scientist. Son profil est celui d'un professionnel avec des connaissances en statistiques, en data management et en informatique, qui sait évaluer les données dont il dispose. Il a un rôle de conseil auprès de la direction des systèmes d'information pour laquelle il produit des rapports qui peuvent influencer les décisions managériales.

Le data scientist occupe une place centrale dans l'entreprise. Il interagit avec les pôles commerciaux, financiers et marketing. En plus de la collecte et de l'analyse des données, il a la capacité d'identifier en amont les solutions de stockage et les outils d'analyse les plus pertinents. Il a aussi pour mission de développer des modèles prédictifs et des tableaux de bord qui facilitent la lecture de ses évaluations. Pour l'acquisition des compétences en sciences des données, l'entreprise peut choisir de former des ressources en interne ou opter pour le recrutement d'un nouveau profil.

Complément

Au-delà de la nécessité de confier la stratégie de gouvernance des données à un spécialiste, le principal enjeu reste celui de la conduite du changement, étant donné que le Big Data contribue à la transformation de l'entreprise. Une synthèse du Cigref parue en 2013 le précise assez bien : « *Le Big Data est une composante de la transformation de l'entreprise, il est transversal et touche tous les métiers. En ce sens, le Big Data n'est pas un projet SI : c'est une manière nouvelle de penser et d'appréhender l'information. Il s'agit donc davantage d'une (r) évolution culturelle et technologique que d'un nouveau sujet SI.* »

Pour implémenter durablement une « *transformation data* », l'analyse de la donnée ne doit pas peser uniquement sur la DSI. Elle doit concerner tous les métiers de l'organisation. Cette approche permet aux différents services de mettre à profit leurs connaissances du terrain et des clients pour identifier plus rapidement les données ayant de la valeur. Dans cette optique, une formation des équipes à l'extraction et à la manipulation des données s'impose.

Big Data, les enjeux économiques pour l'entreprise

Porteur d'innovation, le Big Data sert les ambitions économiques de l'entreprise, à condition que des ressources humaines et technologiques soient disponibles. L'amélioration des résultats financiers passe par de meilleurs produits et des services de qualité, ce qui nécessite une juste exploitation de la donnée. Précisons que la conception et le déploiement de l'architecture du Big Data en entreprise ont un coût. Il y a par conséquent une notion de retour sur investissement à prendre en compte. Les bénéfices du Big Data peuvent s'apprécier sur plusieurs fonctions, à savoir le marketing, la logistique, les finances et les ressources humaines.

Comment ça se passe ? Pour un service marketing, l'accès à des données pertinentes est une véritable aubaine, puisqu'il ouvre la voie à une amélioration des services de l'entreprise. Avec le Big Data, les stratégies de marketing gagnent en efficacité grâce à un ciblage plus précis. On passe ainsi du marketing de masse à un marketing plus humain et personnalisé. Les comportements et les habitudes du client étant désormais connus, notamment par le truchement des données issues des réseaux sociaux, il est plus aisé de lui proposer des produits ou des services dont il a envie. Pour aboutir à des résultats probants et durables, le recours à l'intelligence artificielle est indispensable.

L'apport de cette technologie permet de produire une procédure d'analyse complète, c'est-à-dire descriptive, prédictive et enfin prescriptive. Sur un plan purement stratégique, en plus d'une meilleure connaissance du client, le Big Data donne les moyens d'affiner l'analyse concurrentielle. Cela aide à la préparation de campagnes marketing plus performantes. En plus, la donnée offre l'occasion d'analyser plus finement la satisfaction des clients dans le but de déceler les indicateurs pouvant servir à la fidélisation et à l'amélioration des services à la clientèle.

Associer le Big Data à la logistique permet d'optimiser plusieurs composantes de la chaîne logistique, comme la gestion des stocks, le réapprovisionnement et la gestion de la livraison. En équipant, par exemple, ses véhicules de capteurs, une entreprise de transport peut, à partir des données récoltées, identifier les facteurs qui influencent le plus la consommation de carburant. En recoupant les informations analysées, elle a la possibilité de créer des itinéraires pensés pour réduire le temps de trajet et écourter les délais de livraison.

Quand il est appliqué aux ressources humaines, le Big Data se révèle être un formidable outil qui peut notamment être mis au service du recrutement. Pour un DRH en quête d'un profil particulier, l'intuition comme seule arme ne suffit pas. La donnée suggère une approche plus scientifique, basée sur des statistiques et des faits. Ainsi, avec un algorithme nourri au Big Data, le processus de recrutement peut être non seulement accéléré, mais aussi amélioré. La data contribue par ailleurs au renforcement de la marque employeur, un critère déterminant dans la gestion des talents, qu'il s'agisse de les attirer ou de les fidéliser.

Les innovations liées au Big Data

Avec le Big Data sont apparues des problématiques dont la résolution dépend de la capacité des organisations à innover. Considérée littéralement comme une matière première, la data se révèle être un véritable accélérateur de progrès dans plusieurs domaines, notamment celui de la data science (science des données) et de l'Intelligence Artificielle (IA).

Définition La data science, qu'est-ce que c'est ?

La data science, ou science des données, est la combinaison de « *plusieurs domaines, dont les statistiques, les méthodes scientifiques, l'Intelligence Artificielle (IA) et l'analyse des données, pour extraire de la valeur des données* » (Oracle¹). La data science fait référence à l'ensemble de la procédure menant à la valorisation de la donnée : analyse, manipulation, agrégation et nettoyage. Pratiquée par les data scientists, elle est indispensable à la mise en place de toute stratégie Big Data.

1 <https://www.oracle.com/fr/data-science/what-is-data-science/>

La science des données est étroitement liée au Big Data, même si la data science peut s'appliquer à de faibles volumes de données. Toutefois, la discipline a connu un essor considérable du fait de la croissance exponentielle des données produites par l'homme. L'irruption des nouvelles technologies englobant les objets connectés et l'utilisation massive des réseaux sociaux, des moteurs de recherche et des applications mobiles suggèrent une augmentation de la data.

Mais la donnée n'a de sens que si elle est exploitable, et c'est là tout l'intérêt de la data science. En sondant différents « *lacs de données* » (*data lakes*), les data scientists cherchent à identifier des tendances, des points de connexion et d'intérêt afin de générer des modèles à l'intention de ceux qui ont le pouvoir de décision dans l'entreprise.

L'exploitation de la donnée se fait en plusieurs étapes :

- **La collecte**
- **L'entreposage dans des centres de données**
- **La classification**
- **L'analyse**
- **Le reporting**

En tant que technologie, la science des données se présente comme une composante majeure des outils d'aide à la décision, de prédiction et d'automatisation de tâches. On la retrouve également dans les technologies de reconnaissance textuelle, vocale ou faciale. Les domaines d'usage de cette science sont légion et ne cessent de s'élargir.

Complément

Pour des prédictions météorologiques, la data science peut être mise à contribution pour analyser des données fournies par plusieurs sources : satellites, bateaux, radars, avions. À partir de cette analyse, on établit des modèles pour prévoir le temps sur un intervalle donné, mais aussi anticiper la survenue de certains aléas climatiques.

Autre exemple parlant : celui de la data science appliquée au commerce du détail. Pour améliorer leur moteur de recommandation, les plateformes e-commerce scrutent l'historique des achats des utilisateurs, la liste des articles figurant dans leur *wishlist* ou dans leur panier. Sont aussi pris en compte les produits avec lesquels il a eu une quelconque interaction (commentaire, notation). En croisant ces données avec celles d'autres clients, le moteur de recommandation suggère différents produits au client.

Définition L'intelligence artificielle, qu'est-ce que c'est ?

On définit l'Intelligence artificielle comme « *des systèmes ou des machines qui imitent l'intelligence humaine pour effectuer des tâches et qui peuvent s'améliorer en fonction des informations collectées grâce à l'itération* » (Oracle¹).

L'Intelligence artificielle peut évoluer par le truchement d'un processus d'apprentissage connu sous le terme de « *Machine Learning* » ou « *Deep Learning* ». Elle apprend en se « *nourrissant* » de données. Plus de données implique l'assimilation d'une grande quantité d'informations et une IA plus performante et autonome. L'IA a comme matière première le Big Data, d'où elle tire ses ressources. Si les débuts de l'IA ont été quelque peu balbutiants, en partie à cause de la méfiance suscitée par cette nouvelle technologie, aujourd'hui elle occupe une place de choix sur la carte technologique.

Même si une certaine retenue demeure du fait de sa capacité à prendre des décisions de façon autonome, son potentiel ne fait aucun doute. Big Data et IA sont complémentaires l'un de l'autre. Une IA qui a accès à des données massives gagne en intelligence et peut exploiter plus efficacement le Big Data pour une gestion des données optimisée. Il s'agit par conséquent de deux systèmes qui sont appelés à converger et à s'entremêler.

1 <https://www.oracle.com/be-fr/artificial-intelligence/what-is-ai/>

Une IA plus intelligente est utile aux outils de prise de décision, à l'automatisation des tâches, au diagnostic et à la résolution des problèmes. Portée par le Big Data, l'intelligence artificielle permet d'améliorer les connaissances déjà acquises et d'explorer de nouvelles pistes, et cela dans divers domaines. L'amélioration de l'IA se remarque notamment dans la conception de programmes dédiés à l'automatisation de la prise de décision ou dans le développement de logiciels d'analyse de données plus précis : des machines trouvent du sens à la donnée et accélèrent la prise de décision, des promesses peuvent être tenues avec une IA améliorée.

Exercice : Quiz

[solution n°1 p.15]

Question 1

Un grand volume de données suffit à définir le Big Data.

- Vrai
- Faux

Question 2

Le Big Data peut se définir au-delà des 5 V.

- Vrai
- Faux

Question 3

Hadoop est une solution Big Data.

- Vrai
- Faux

Question 4

Les PME et les PMI peuvent aussi profiter du Big Data.

- Vrai
- Faux

Question 5

L'exploitation du Big Data présente des inconvénients.

- Vrai
- Faux

III. L'application du Big Data dans le domaine du marketing

Le Big Data a profondément bouleversé l'univers du marketing en y apportant une nouvelle façon de créer les produits et les services. Si le marketing traditionnel ne considérait le consommateur que dans un cadre précis, le marketing à l'ère du digital est plus enveloppant et prend le consommateur dans son ensemble : il est constamment scruté, dans toutes ses interactions, même quand il n'est pas dans une logique de consommation. Il est sondé au travers de ses échanges sur les réseaux sociaux, des photos qu'il publie ou des publications qu'il commente : toutes les traces laissées sur la toile et dans les points de vente physiques ou phygitals.

Un marketing transformé

Nourri au Big Data, le marketing est moins attentiste. Les données récoltées sont analysées en flux continu pour déterminer la prochaine action à déclencher en fonction du comportement de l'internaute. Le Big Data permet aux entreprises de mieux penser leur positionnement concurrentiel, leur stratégie de prix ou leur offre de services ou de produits grâce à une meilleure étude de leur environnement. On s'éloigne progressivement du marketing de masse pour se rapprocher d'un marketing personnalisé et plus prédictif. Personnaliser, c'est aussi tenir compte du contexte dans lequel se situe le client pour mieux affiner son offre ou sa communication.

Le Big Data appliqué au marketing, c'est :

- Une communication et des publicités plus ciblées,
- Une individualisation du consommateur qui bénéficie d'une expérience utilisateur optimisée et accède à des contenus correspondant à ses centres d'intérêt,
- Une prédictibilité améliorée qui permet de mieux anticiper les attentes du consommateur.

Méthode

Pour augmenter le taux de conversion de son site, une marque peut recourir au reciblage publicitaire ou au retargeting. Comment ça marche ?

Elle commence par déposer des cookies de retargeting sur son site internet. Grâce à ces cookies, elle peut identifier les internautes qui ont visité son site, même quand ils se rendent sur un autre site. Une fois l'identité de la cible confirmée, cette dernière se voit proposer du contenu publicitaire qui la redirige sur le site de la marque.

Les formats publicitaires non intrusifs sont privilégiés afin d'éviter toute altération de l'expérience utilisateur. C'est notamment le cas des publicités natives qui s'adaptent parfaitement aux codes et au design des médias supports.



Schéma illustratif du retargeting

Le Big Data au service d'une expérience client plus évoluée

Le Big Data donne aux entreprises les moyens de leurs ambitions : mieux connaître le client pour mieux appréhender ses besoins.

L'expérience client est plus que jamais au centre des préoccupations des entreprises qui en ont fait un cheval de bataille pour se démarquer de leurs concurrents et s'affirmer auprès de leur public. Comme l'indique le site Définition marketing¹, l'expérience client fait référence à « *l'ensemble des émotions et sentiments ressentis par un client avant, pendant et après l'achat d'un produit ou service. C'est le résultat de l'ensemble des interactions qu'un client peut avoir avec la marque ou l'entreprise* ».

Une expérience client positive reste un levier d'acquisition et de fidélisation pour les entreprises. Mais la construction d'une telle relation est difficile sans données, ce qui rend l'adoption du Big Data plus que légitime. Bon nombre d'entreprises ont fini par intégrer cette réalité, ce qu'une **étude** menée par Gartner en 2014 corrobore. En effet, dans cette enquête, 73 % des sondés avouent avoir déjà investi dans le Big Data ou pensent le faire dans les deux années à venir.

Les innovations apportées par le Big Data à l'expérience client sont les suivantes :

- Une meilleure anticipation des demandes des clients grâce à des outils prédictifs améliorés,
- La création de produits innovants mieux adaptés aux besoins des clients,
- La création d'interactions plus dynamiques entre l'entreprise et les clients,
- Un engagement client plus élevé,
- Une optimisation de l'identification et de la résolution des points bloquants.

Pour attirer l'attention du client et l'amener jusqu'à l'acte d'achat, la personnalisation est la clé. La création d'un parcours client personnalisé est la conséquence d'un processus basé sur la compréhension du client. Comprendre un client, c'est savoir ce qu'il recherche à tel moment en se basant sur une analyse de son historique d'achat. Il devient ainsi plus aisé de segmenter sa base de clients en tenant compte des comportements d'achat.

Exemple Netflix et Amazon

Rien ne saurait mieux évoquer la personnalisation de l'expérience client que Netflix. La plateforme de streaming s'appuie sur une intelligence artificielle pour recommander du contenu à ses abonnés. Tout part d'une étude comportementale du spectateur par un algorithme. Ce dernier détecte les contenus les plus appréciés et ceux qui le sont moins. Ces informations servent à créer des segments regroupant des spectateurs ayant des comportements en commun. Les membres du groupe se voient ainsi proposer des contenus que les autres ont aimés.

Netflix tient également compte des informations du profil de ses clients pour mettre en avant des films ou séries correspondant à leurs thématiques favorites. La plateforme use également d'un modèle qui considère le contexte dans lequel évolue le spectateur pour lui recommander du contenu approprié. Selon qu'il soit en France, en Espagne ou au Canada, les recommandations varient. L'association de ces différents modèles permet d'améliorer grandement l'expérience client.

Amazon a aussi recours à l'étude comportementale pour affiner les recommandations faites à ses clients. Le site regroupe les clients présentant des similitudes dans leurs profils en différents segments. Les clients d'un même groupe auront des suggestions sur la base des produits achetés par d'autres membres. Amazon arrive à réaliser ce filtrage collaboratif par la collecte de diverses données sur ceux qui visitent son site :

- Les données transactionnelles (historique des achats, produits consultés, contenu du panier d'achats, détails de paiement et de livraison),
- Les données comportementales (pages les plus visitées, temps de visite moyen, paniers abandonnés, produits aimés),
- Les données déclaratives (informations communiquées volontairement par le client : nom, prénoms, âge, sexe, profession, etc.).

1 <https://www.definitions-marketing.com/definition/experience-client/>

Les données comportementales constituent la plus grande part des données recueillies (plus de 90 %). Du fait de leur nature, elles permettent de spécifier le profil des clients et favorisent la personnalisation de l'expérience client de façon instantanée.

Exercice : Quiz

[solution n°2 p.15]

Question 1

Le reciblage publicitaire contribue à l'augmentation du taux de conversion.

- Vrai
- Faux

Question 2

Le retargeting peut être considéré comme un levier de fidélisation.

- Vrai
- Faux

Question 3

Le retargeting ne concerne que les prospects.

- Vrai
- Faux

Question 4

Les cookies sont utilisés uniquement à des fins publicitaires.

- Vrai
- Faux

Question 5

L'exploitation du Big Data permet aux entreprises de booster leur visibilité.

- Vrai
- Faux

V. Essentiel

Phénomène dont le potentiel demeure largement inexploité, le Big Data sert la stratégie d'entreprises évoluant dans divers secteurs d'activité : éducation, communication, commerce, banque, transport, industrie, etc. Il permet de tirer profit du flux de données généré en continu par les activités de l'humanité et concourt au développement de solutions technologiques avancées.

On peut sans nul doute considérer le Big Data comme un acteur incontournable du rapprochement de la technologie avec l'humain. Il a notamment participé au changement de paradigme du discours marketing des marques, qui conçoivent désormais leurs services et leurs produits en se basant sur les envies et les besoins de leurs cibles.

Composante essentielle de la data science et de l'intelligence artificielle, le Big Data est plus que jamais au service de l'innovation, que ce soit en termes d'automatisation de tâches, de prédictivité ou d'aide à la prise de décision. Bien exploité, c'est un formidable levier de croissance économique et humaine dont on peut difficilement se passer.

L'exploitation du Big Data requiert des compétences et des outils adaptés, mais pas seulement. À l'échelle d'une organisation, cela engendre des changements organisationnels qu'il faut savoir appréhender et gérer tout en conservant la qualité des services.

VI. Auto-évaluation

A. Exercice

1. Vous êtes le responsable des ventes d'un site de vente de voitures et vous souhaitez améliorer votre moteur de recommandations afin d'augmenter les ventes du site.
2. Vous dirigez une PME qui souhaite bénéficier des atouts du Big Data, mais vous ne savez pas par où commencer.

Question 1

[solution n°3 p.16]

Présentez la démarche que vous allez suivre pour parvenir au résultat voulu.

Question 2

[solution n°4 p.17]

Décrivez le processus à suivre pour un déploiement progressif du Big Data à l'échelle d'une PME.

B. Test

Exercice 1 : Quiz

[solution n°5 p.17]

Question 1

Le Big Data est uniquement stocké dans les data centers.

- Vrai
- Faux

Question 2

L'exploitation du Big Data est une solution pour le développement économique des entreprises.

- Vrai
- Faux

Question 3

Les seules tâches du data scientist consistent à collecter et à analyser les données.

- Vrai
- Faux

Question 4

Les mégadonnées sont essentielles au développement de l'intelligence artificielle.

- Vrai
- Faux

Question 5

Le véritable enjeu du Big Data, c'est l'analyse des données.

- Vrai
- Faux


Solutions des exercices

Exercice p. 9 Solution n°1**Question 1**

Un grand volume de données suffit à définir le Big Data.

Vrai

Faux


 Pour parler de Big Data, il faut un certain volume, mais la variété et la vitesse des données compte également. Le Big Data, c'est un ensemble de petites données qui, mises ensemble, forment des mégadonnées.

Question 2

Le Big Data peut se définir au-delà des 5 V.

Vrai

Faux


 En plus des 3 V ou 5 V que l'on a l'habitude de citer, d'autres V peuvent caractériser le Big Data, notamment la *vulnérabilité*, la *volatilité*, la *visualisation* et la *validité*. En fonction des sources, on compte ainsi entre 3 V et 10 V.

Question 3

Hadoop est une solution Big Data.

Vrai

Faux


 Hadoop est un framework qui permet de créer des solutions de traitement du Big Data.

Question 4

Les PME et les PMI peuvent aussi profiter du Big Data.

Vrai

Faux


 Le Big Data est un réel atout pour les PME ou PMI qui disposent d'un volume de données relativement réduit, ce qui en facilite le contrôle. En outre, ces entreprises peuvent recourir à des solutions cloud pour bénéficier de solutions Big Data à des coûts maîtrisés.

Question 5

L'exploitation du Big Data présente des inconvénients.

Vrai

Faux

 En dépit de ses avantages, le Big Data n'est pas sans inconvénient. Les données étant vulnérables, il convient d'assurer leur sécurité, ce qui exige des compétences et des ressources techniques. La manipulation des données doit aussi respecter des conditions de confidentialité en conformité avec les normes en vigueur.


Exercice p. 12 Solution n°2

Question 1

Le reciblage publicitaire contribue à l'augmentation du taux de conversion.

Vrai

Faux


 Grâce au retargeting, on peut amener un prospect ayant montré de l'intérêt pour une marque à passer à l'acte d'achat.

Question 2

Le retargeting peut être considéré comme un levier de fidélisation.

Vrai

Faux


 Rester en contact avec une marque à laquelle on s'est intéressé peut être rassurant pour un prospect ou un client.

Question 3

Le retargeting ne concerne que les prospects.

Vrai

Faux


 Même un client qui est déjà passé à l'action peut être reciblé dans l'optique de la réalisation de ventes additionnelles.

Question 4

Les cookies sont utilisés uniquement à des fins publicitaires.

Vrai

Faux


 Les cookies servent aussi à la collecte de données utilisateurs et contribuent à l'amélioration de l'expérience client.

Question 5

L'exploitation du Big Data permet aux entreprises de booster leur visibilité.

Vrai

Faux

 Grâce aux données clients, une entreprise a plus de facilité à faire correspondre son image de marque aux attentes de son public et à construire une expérience client optimale.

Nous commencerons par analyser le parcours du prospect sur le site afin d'identifier ses centres d'intérêt. En fonction des interactions ou des recherches effectuées (modèles ou marques de voitures), nous pouvons dégager un modèle comportemental que nous intégrerons dans un groupe d'utilisateurs ayant les mêmes critères de recherche. Pour inciter le visiteur à revenir sur le site, nous lui demandons son email en contrepartie d'informations sur des modalités de paiement avantageuses ou des remises exceptionnelles. Cette technique peut être couplée à du retargeting. Dans les communications email, nous lui suggérons des voitures qui correspondent aux critères du segment de clients auquel il appartient, avec à la clé une offre avantageuse. Pour affiner les recommandations, une fois de retour sur le site, on peut lui proposer de configurer sa voiture favorite en échange d'une estimation de prix. On obtient ainsi des détails techniques supplémentaires par rapport à ses préférences.

p. 13 Solution n°4

Il est important de débiter par la définition d'un cas d'usage justifiant le recours au Big Data (personnalisation des prix, meilleure connaissance client, etc.).


Une fois l'objectif identifié, on fait le tour des données internes et externes dont nous disposons. Si nous visons une meilleure connaissance client, l'ERP, le CRM et les réseaux sociaux constituent des sources de premier choix. Pour l'interconnexion des informations provenant de sources internes et externes, nous utilisons une API.

Enfin, il nous faut un outil capable d'exploiter les mégadonnées à notre disposition. Puisque notre budget de départ est restreint et que nous débutons avec un volume de données relativement faible, nous opterons pour une solution NoSQL cloud. Nous faisons ainsi l'économie d'infrastructures Big Data physiques, tout en bénéficiant de services d'analyse de données à un coût abordable.

Exercice p. 13 Solution n°5


Question 1

Le Big Data est uniquement stocké dans les data centers.

- Vrai
- Faux
-  Bien que les data centers fassent partie des supports de stockage préférés du Big Data, d'autres solutions existent, notamment le cloud.


Question 2

L'exploitation du Big Data est une solution pour le développement économique des entreprises.

- Vrai
- Faux
-  En réussissant à extraire les informations pertinentes des mégadonnées, les entreprises ont une plus grande facilité à élaborer des offres proportionnées aux besoins de leurs cibles.

Question 3

Les seules tâches du data scientist consistent à collecter et à analyser les données.


- Vrai
- Faux
-  En plus de la collecte et de l'analyse des données, le data scientist a aussi la responsabilité du choix des solutions de stockage et de la veille technologique, entre autres missions.

Question 4

Les mégadonnées sont essentielles au développement de l'intelligence artificielle.

Vrai

Faux


 Sans mégadonnées, l'intelligence artificielle n'a plus accès à la matière qui lui permet d'apprendre. Son évolution est donc limitée.

Question 5

Le véritable enjeu du Big Data, c'est l'analyse des données.

Vrai

Faux

 Pouvoir accéder aux informations utiles contenues dans les mégadonnées, tel est le principal défi des entreprises. Surmonter cet obstacle est essentiel pour élaborer des modèles prédictifs justes et optimiser les processus de prise de décision.