



Université Constantine 2
جامعة قسنطينة 2

Web Analytics and Natural Language Processing (WANLP)

Chapitre 04

Principes, Techniques et Outils du Web Analytics

Professeur BOURAMOUL Abdelkrim

Département IFA, Faculté NTIC

abdelkrim.bouramoul@univ-constantine2.dz

www.bouramoul.com

Etudiants concernés

Faculté/Institut	Département	Niveau	Spécialité
NTIC	IFA	Master 1	SDIA

Plan du Cours

Introduction au Web Analytics NLP

- Définition et importance du Web Analytics
- Objectifs et avantages du Web Analytics
- Outils et technologies utilisés en Web Analytics
- Exemple d'utilisation d'outil du Web Analytics

Collecte des données

- Méthodes de collecte des données
- Respect de la vie privée et conformité aux réglementations

Analyse des données

- Métriques et indicateurs de performance (KPIs)
- Analyse du trafic

Intégration du Web Analytics avec le NLP

- Utilisation avancée du NLP en Web Analytics
- Optimisation du contenu basée sur le NLP

Section 1

Introduction au Web Analytics NLP

Introduction au Web Analytics NLP

Définition du Web Analytics

Définition

- Le Web Analytics est l'ensemble des **méthodes** et **techniques** permettant de **collecter, mesurer, analyser** et **interpréter** les données générées par les **activités des utilisateurs** sur un **site web** ou une **application mobile**.
- L'objectif principal est de comprendre le **comportement des visiteurs** pour **optimiser la performance** et **l'expérience utilisateur** du site.



Introduction au Web Analytics NLP

Importance du Web Analytics

Prise de décision basée sur des données

- Fournir des perspectives pour guider les décisions stratégiques et opérationnelles.
- **Exemple** : Un site de **vente en ligne** constate **une baisse** de **conversions** sur sa **page de paiement**. Après analyse des données, il **simplifie le processus de paiement**, entraînant une **augmentation** notable des conversions.

Amélioration de l'expérience utilisateur

- Comprendre comment les utilisateurs interagissent avec le site pour optimiser leur parcours et de réduire les obstacles.
- **Exemple** : Un **blogueur** observe que les articles contenant des **images** et des **vidéos** engagent **plus les lecteurs**. En conséquence, il **enrichit ses futurs articles** de contenu multimédia, améliorant l'expérience utilisateur et l'engagement

Optimisation du retour sur investissement (ROI)

- Aider à évaluer l'efficacité des campagnes marketing et à allouer les ressources de manière plus efficace.
- **Exemple** : Une entreprise **analyse le ROI** de ses campagnes **publicitaires** sur divers canaux. Elle découvre que les **médias sociaux** sont les **plus rentables** et réaffecte une partie de son **budget publicitaire** pour maximiser son investissement

Introduction

Objectifs et avantages du Web Analytics

Mesurer la performance du site

- Suivre des indicateurs clés pour évaluer l'efficacité du site web.
- **Exemple** : Un site de e-commerce constate une baisse du **temps passé** sur le site. En réduisant le nombre de clics nécessaires pour effectuer un achat, le site augmente le temps de visite et les ventes.

Comprendre le comportement des utilisateurs

- Analyser comment les visiteurs interagissent avec le site pour identifier les points forts et les points à améliorer.
- **Exemple** : Un blog de voyage remarque que les articles sur les **destinations européennes** sont les plus lus. Il décide de se concentrer davantage sur ce thème pour attirer plus de lecteurs.

Optimiser les campagnes marketing

- Évaluer l'impact des différentes stratégies marketing pour maximiser le retour sur investissement.
- **Exemple** : Une entreprise de cosmétiques observe que ses **campagnes Instagram** ont un meilleur taux de conversion que celles sur Facebook. Elle redirige donc une partie de son budget vers Instagram pour maximiser son ROI

Introduction

Outils utilisés en Web Analytics

Google Analytics

- L'un des outils les plus populaires, offrant une gamme étendue de fonctionnalités pour suivre et analyser le trafic web.



Matomo (anciennement Piwik)

- Une alternative open-source à Google Analytics qui offre des fonctionnalités de suivi de la confidentialité.



Adobe Analytics

- Une solution plus avancée utilisée par les grandes entreprises pour des analyses approfondies.



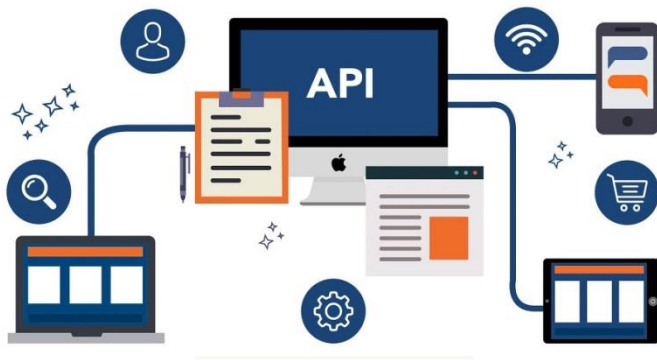
Introduction

Technologies utilisés en Web Analytics



Cookies & Tracking Pixel

- Utilisés pour collecter des données sur le comportement des utilisateurs.

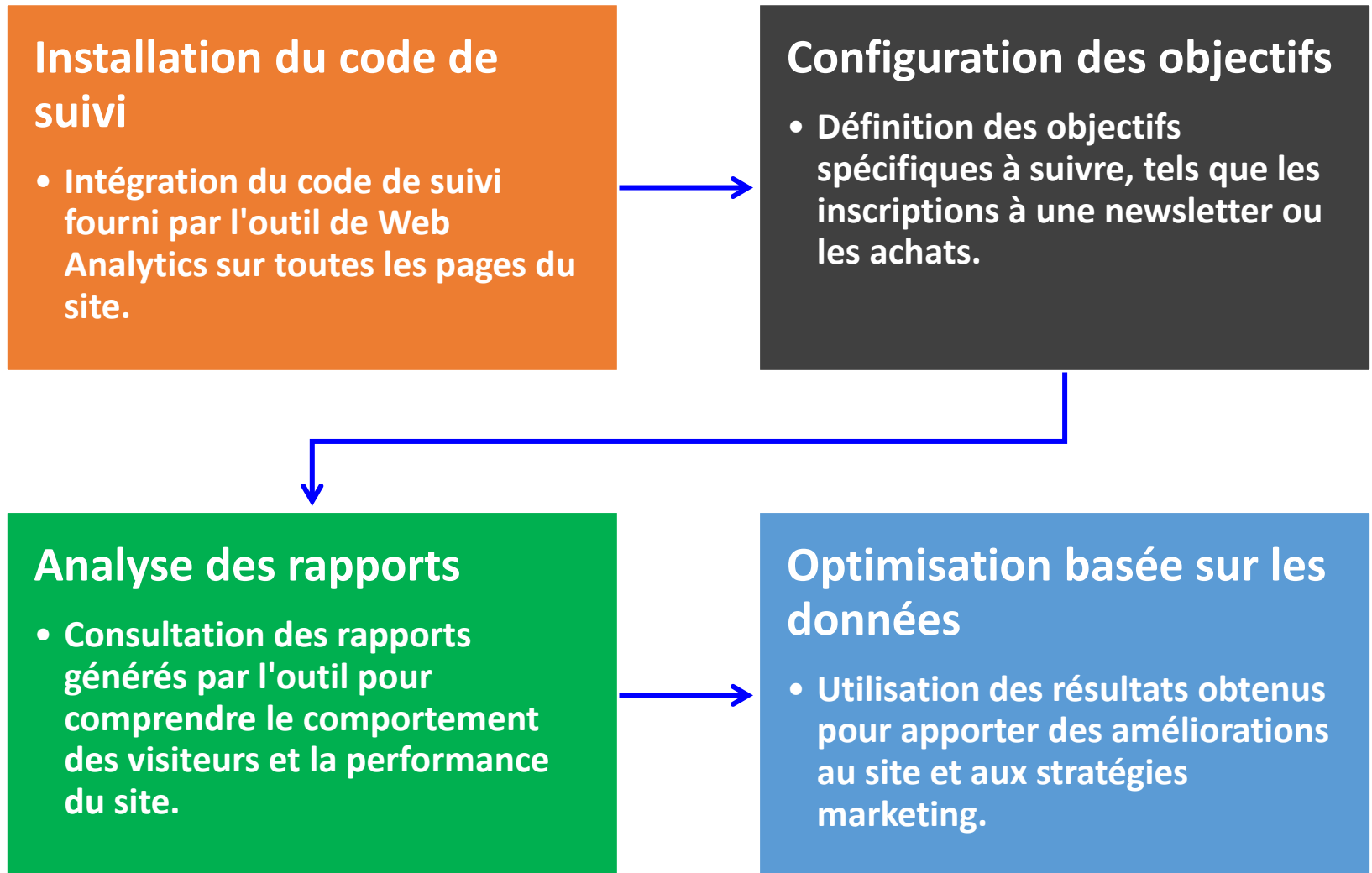


APIs

- Permettent d'intégrer les données de Web Analytics avec d'autres systèmes pour une analyse plus poussée.

Introduction

Exemple d'utilisation d'un outil de Web Analytics

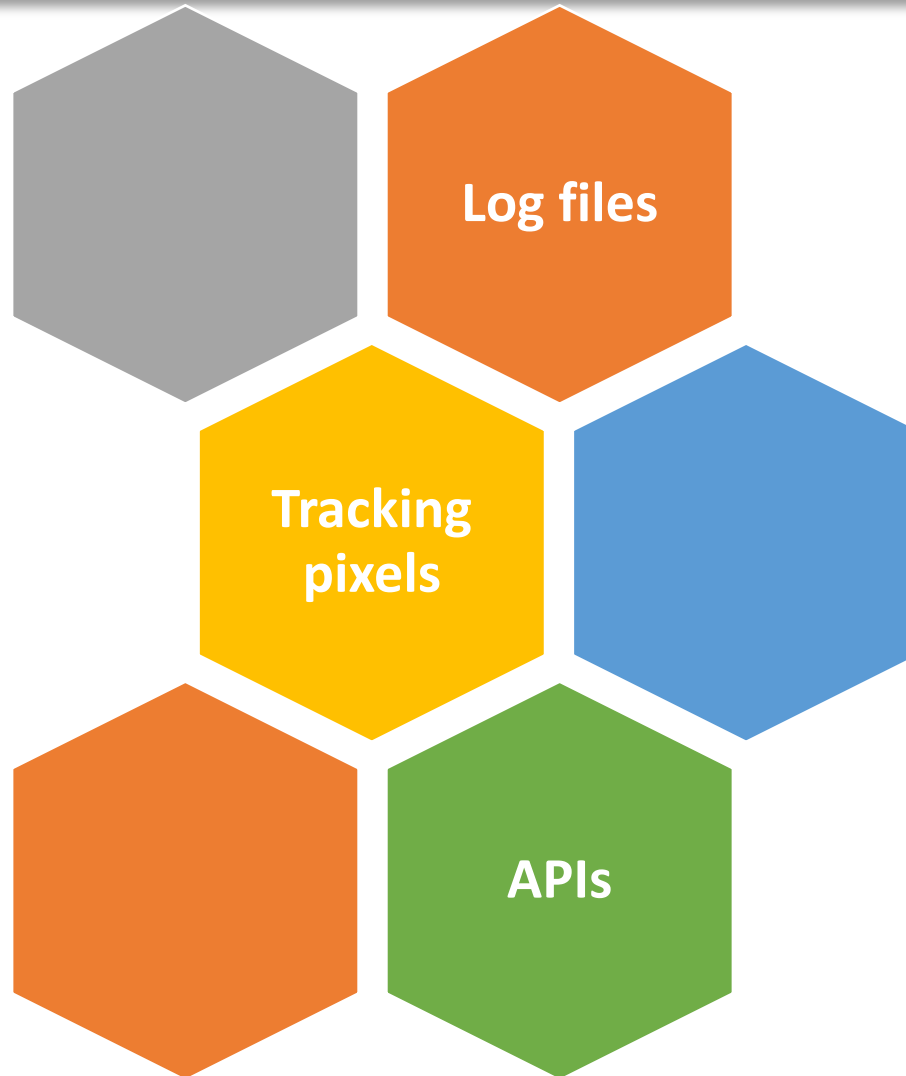


Section 2

Collecte des données

Collecte des données

Méthodes de collecte des données



Collecte des données


Méthodes de collecte des données

Fichiers journaux (Log files)

- Les serveurs web enregistrent chaque requête reçue dans des fichiers log. Ces fichiers contiennent des informations telles que l'adresse IP du visiteur, le type de navigateur, la date et l'heure de la visite, et la page demandée.
- Ils sont utiles pour analyser le trafic du site et identifier les tendances.

Exemple d'extrait de fichier journal

arduino

 Copy code

```
192.168.1.1 - - [01/Apr/2024:12:00:00 +0200]  
"GET /index.html HTTP/1.1" 200 512  
"- " "Mozilla/5.0 (Windows NT 10.0; Win64; x64)  
AppleWebKit/537.36 (KHTML, like Gecko)  
Chrome/58.0.3029.110 Safari/537.3"
```

Collecte des données

Méthodes de collecte des données

Sources des log files

Applications

Actions, des erreurs, ou d'autres événements

Bases de données

Transactions et modifications apportées aux données.

Réseaux et serveurs

Trafic réseau, accès aux ressources du serveur, performances système.

Pare-feu et points de terminaison

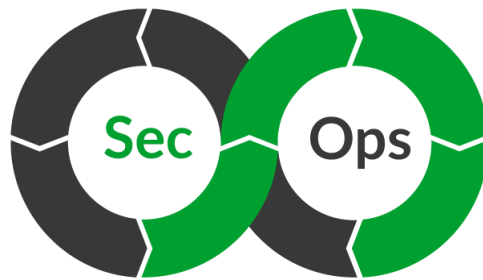
Tentatives d'accès et des blocages pour des raisons de sécurité.

Appareils IoT

Données sur l'usage et les performances des appareils connectés.

Collecte des données

Méthodes de collecte des données



Utilisation des fichiers journaux

- **ITOps** : Surveiller et maintenir les **infrastructures**, gérer les **charges de travail**, et assurer la continuité des activités.
- **DevOps** : Faciliter la **gestion du déploiement** continu, maintenir la **disponibilité** des applications, et identifier les **erreurs critiques**.
- **SecOps** : Améliorer la **sécurité** en découvrant les activités suspectes et en répondant rapidement aux incidents de sécurité.
- **Analystes informatiques** : Gérer la conformité, surveiller les coûts et les dépenses, et fournir des insights commerciaux.

Défis de la gestion des log file

- **Volume**: Augmentation massive des données à gérer.
- **Normalisation**: Différences de format entre les sources de journaux.

Collecte des données


Méthodes de collecte des données

Pixels de suivi (Tracking pixels)

- Un pixel de suivi est une petite image invisible (généralement de 1x1 pixel) insérée dans une page web ou un email.
- Lorsque la page ou l'email est chargé, le pixel est téléchargé depuis le serveur, enregistrant ainsi une visite ou une ouverture.

Exemple : Code HTML pour insérer un pixel de suivi :

html

 Copy code

```

```

Collecte des données

Méthodes de collecte des données

Utilisation des Tracking pixels sur Facebook

Exemple d'utilisation

- Facebook offre des pixels de suivi que l'on peut intégrer à un site pour surveiller l'efficacité des publicités diffusées sur Facebook, en traçant les visites et interactions des utilisateurs provenant de ces publicités.

Réglementations

- Les lois sur l'utilisation des outils de suivi en ligne varient d'un pays à l'autre.
- Par exemple, l'utilisation des pixels de suivi est réglementée dans l'union européenne, qui exige un consentement clair et informé des utilisateurs avant leur déploiement.

Collecte des données

Méthodes de collecte des données

Utilisation des Tracking pixels pour le Mailing

Exemple d'utilisation

- Une entreprise lance une **campagne d'emailing** et souhaite évaluer son efficacité, notamment en **vérifiant l'ouverture des emails**.

Mécanisme

- L'ouverture de l'email **active le code du pixel de suivi**, qui envoie alors un **signal aux serveurs** de l'entreprise, confirmant l'ouverture et **transmettant d'autres détails** comme le système d'exploitation, l'adresse IP, et la localisation de l'utilisateur.

Collecte des données


Méthodes de collecte des données

APIs

- Les interfaces de programmation d'applications (APIs) permettent de collecter des données à partir de différentes sources, telles que les réseaux sociaux, les bases de données ou d'autres services en ligne.

Exemple l'API pour collecter des tweets contenant un mot-clé spécifique.

python

 Copy code

```
import tweepy

# Configuration de l'API
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

# Collecte des tweets
for tweet in tweepy.Cursor(api.search, q="#exemple").items(10):
    print(tweet.text)
```

Collecte des données

Méthodes de collecte des données

Utilisation des API par Amazon

Intégration des API AWS

- Amazon intègre des API fournies par **Amazon Web Services** dans son infrastructure e-commerce.
- Ces API capture automatiquement des interactions spécifiques des utilisateurs telles que les clics, les recherches de produits, les pages visitées, et les transactions effectuées.

Traitement et Analyse des Données

- Les données collectées sont envoyées à des serveurs AWS où elles sont stockées et traitées.
- Des algorithmes d'analyse avancés examinent ces données pour identifier des tendances, évaluer l'efficacité des éléments de site, et comprendre les comportements d'achat des utilisateurs.
- Cela permet à Amazon de créer des profils de comportement qui facilitent la personnalisation des recommandations et l'amélioration continue de l'expérience utilisateur.

Collecte des données

Respect de la vie privée et conformité aux réglementations

GDPR (Règlement général sur la protection des données)

- Cette réglementation impose des règles strictes sur la collecte, le stockage et l'utilisation des données personnelles.

Demander le consentement explicite des utilisateurs avant de collecter leurs données.

CCPA (Loi sur la protection de la vie privée des consommateurs)

- Cette loi donne aux consommateurs le droit de savoir quelles données personnelles sont collectées à leur sujet et de refuser la vente de ces données.

Ajouter un lien "Ne pas vendre mes informations personnelles" sur votre site web.

Consentement des utilisateurs

- Il est important de demander et d'obtenir le consentement des utilisateurs avant de collecter leurs données, en leur fournissant des informations claires sur l'utilisation de leurs données.

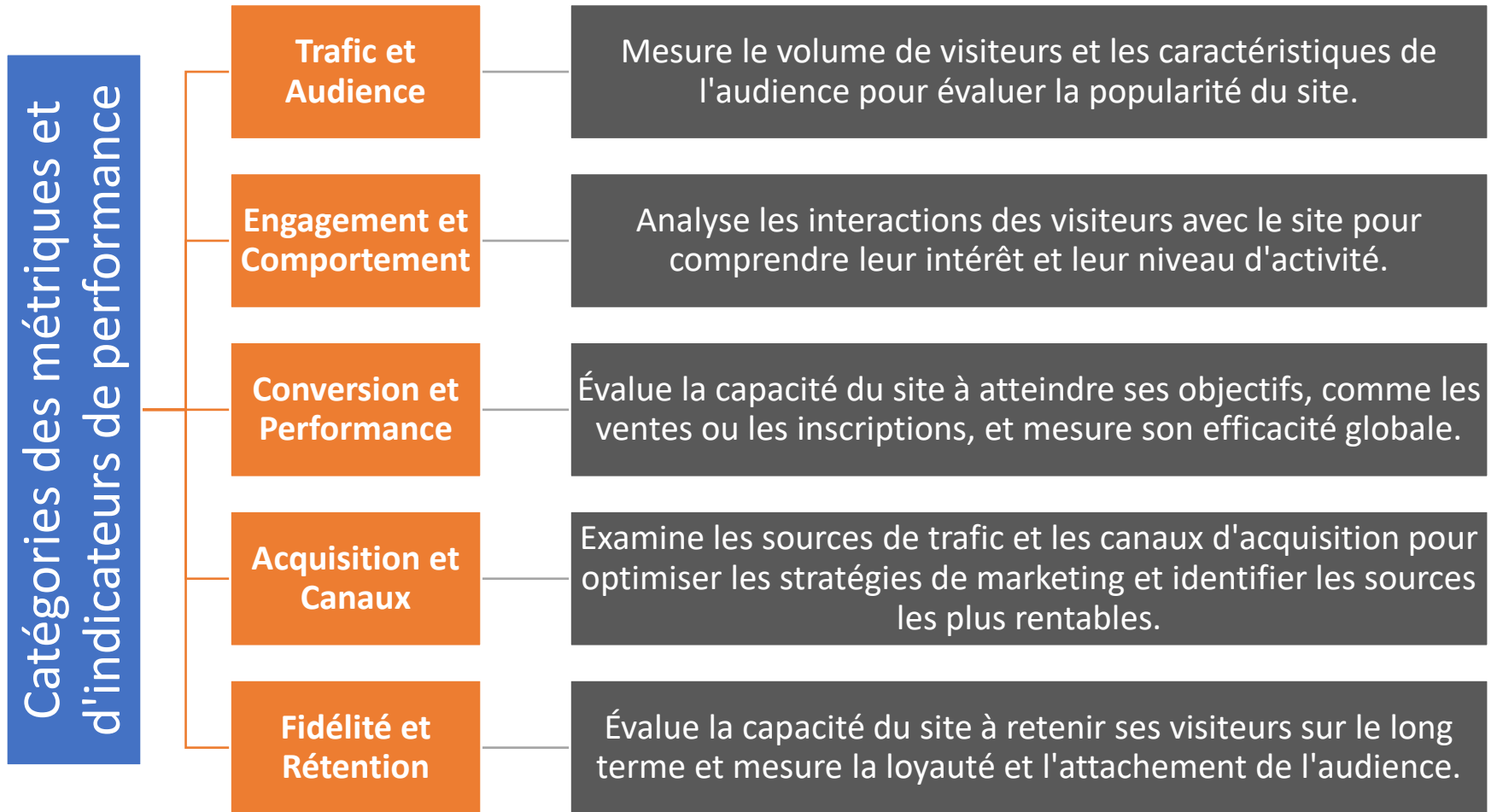
Utiliser des bannières de consentement aux cookies pour donner aux utilisateurs le choix d'accepter ou de refuser le suivi.

Section 3

Analyse des données

Analyse des données

Métriques et d'indicateurs de performance (KPIs)



Trafic et Audience

Visites

Nombre total de visites sur le site.

Visiteurs uniques

Nombre de personnes différentes ayant visité le site.

Taux de nouveaux visiteurs

Pourcentage de visiteurs n'ayant jamais visité le site auparavant.

Pages vues

Nombre total de pages consultées.

Pages par session

Nombre moyen de pages consultées par visite.

Durée moyenne de la session

Temps moyen passé sur le site par les visiteurs.

Engagement et Comportement

Taux de rebond

Pourcentage de visites d'une seule page sans interaction.

Profondeur de la visite

Nombre moyen de pages consultées avant de quitter le site.

Temps passé sur la page

Durée moyenne passée par les visiteurs sur une page spécifique.

Événements

Actions spécifiques effectuées par les utilisateurs (clics, téléchargements, interactions avec des éléments).

Conversion et Performance

Conversions

Nombre de fois où les objectifs du site ont été atteints.

Taux de conversion

Pourcentage de visiteurs ayant effectué une action souhaitée.

Valeur moyenne de la commande

Montant moyen dépensé par commande.

Revenu par visiteur

Revenu moyen généré par chaque visiteur du site.

Acquisition et Canaux

Sources de
trafic

Origine des
visiteurs (moteurs
de recherche,
réseaux sociaux,
sites référents).

Canaux
d'acquisition

Répartition du
trafic entre
différents canaux
(organique,
payant, direct,
etc.).

Coût par
acquisition

Coût moyen pour
acquérir un client
ou un lead.

Taux de clics

Pourcentage de
clics sur un lien
par rapport au
nombre total de
vues de ce lien.

Fidélité et Rétention

Taux de
retour

Fréquence
des visites

Durée entre
les visites

Pourcentage de
visiteurs revenant
sur le site.

Nombre moyen
de visites par
visiteur sur une
période donnée.

Temps moyen
écoulé entre les
visites d'un
même utilisateur.

Analyse des données

Analyse du trafic

Définition

- L'analyse du trafic permet de comprendre comment les visiteurs arrivent sur votre site, d'où ils viennent, et comment ils interagissent avec votre contenu.
- Cette analyse est importante pour optimiser votre stratégie de marketing en ligne et améliorer l'expérience utilisateur.



Sources de trafic

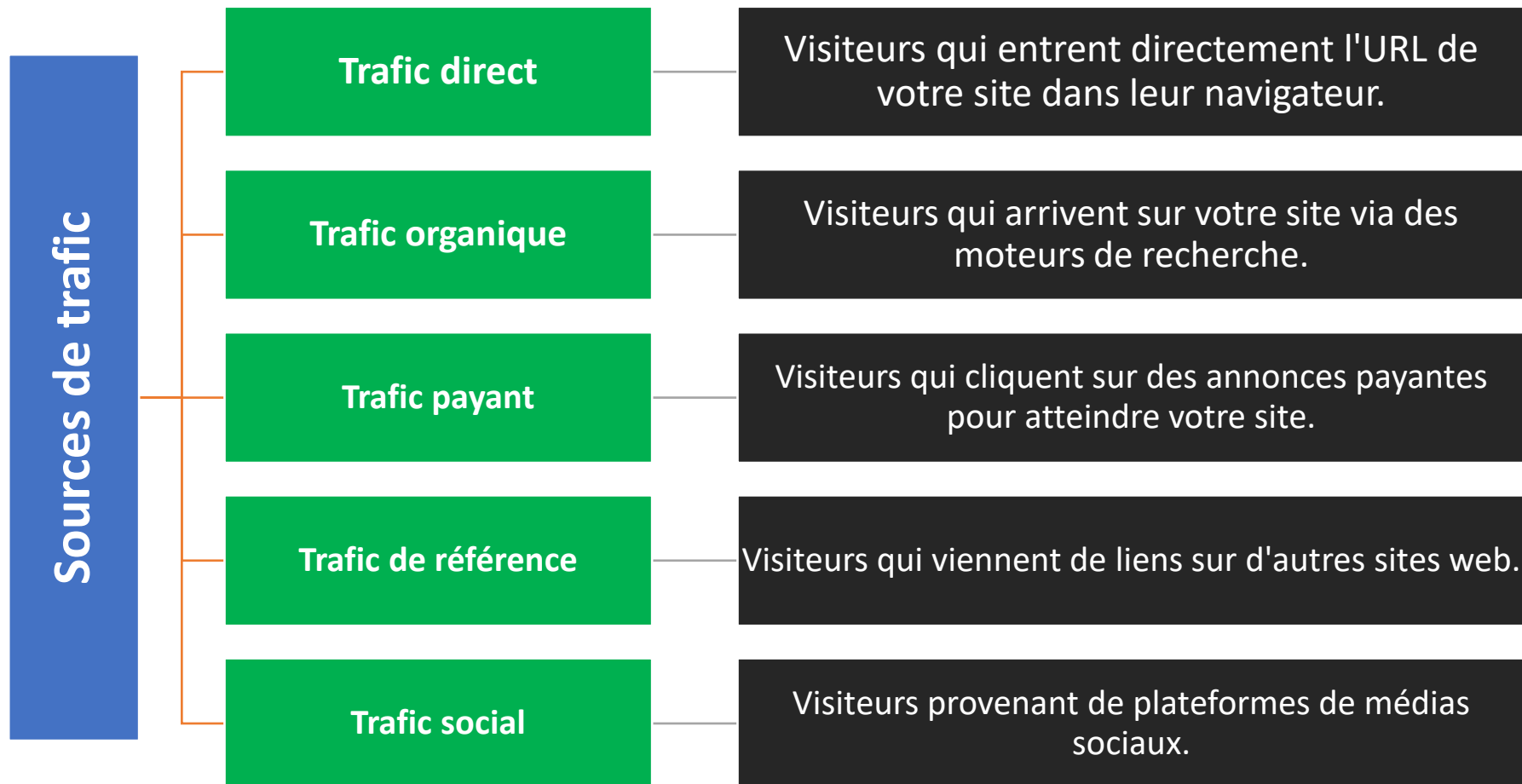
- Les sources de trafic indiquent d'où proviennent vos visiteurs.

Exemple de scénario

- Vous constatez une augmentation du trafic organique après avoir optimisé votre contenu pour le référencement (SEO).
- Cela indique que votre site apparaît plus haut dans les résultats de recherche, attirant ainsi plus de visiteurs.

Analyse des données

Analyse du trafic



Géolocalisation

- La géolocalisation permet de savoir d'où viennent vos visiteurs, à l'échelle des pays, régions ou villes.
- Cette information est utile pour cibler vos campagnes marketing et adapter votre contenu à des audiences spécifiques.

Exemple de scénario

- Vous remarquez que la majorité de vos visiteurs viennent des Constantine et du Skikda.
- Vous décidez de lancer une campagne promotionnelle ciblant ces régions pendant les vacances scolaire.

Section 4

Intégration du Web Analytics avec le NLP

Intégration du Web Analytics avec le NLP

Utilisation avancée du NLP en Web Analytics

Objectif

- L'intégration du NLP dans le web analytics permet d'approfondir l'analyse des données textuelles et d'extraire des éléments plus parlant sur le comportement et les préférences des utilisateurs.

Utilisation

Analyse sémantique

Comprendre le sens et le contexte des mots et des phrases dans les contenus textuels. Elle permet de capturer les subtilités de la communication humaine, allant au-delà de la simple identification des mots-clés.

Modélisation de sujets

Identifier automatiquement les thèmes principaux dans un ensemble de textes. Elle est particulièrement utile pour analyser de grands volumes de données textuelles et découvrir les sujets de discussion prédominants

Intégration du Web Analytics avec le NLP

Utilisation avancée du NLP en Web Analytics

Exemple 01 : Analyse sémantique du site d'Amazon

Contexte

- Analyse des avis clients sur les produits d'Amazon.

Application

- Utilisation de l'analyse sémantique pour identifier les aspects spécifiques des produits qui génèrent des réactions positives ou négatives.

Résultat

Aspect du produit	Sentiment moyen	Exemple d'avis
Qualité du matériel	Positif	"Le matériau est robuste et durable."
Design	Négatif	"Le design est démodé et peu attrayant."
Fonctionnalité	Positif	"Les fonctionnalités sont intuitives et utiles."

Intégration du Web Analytics avec le NLP

Utilisation avancée du NLP en Web Analytics

Exemple 02 : Analyse des avis clients sur TripAdvisor

Contexte

- TripAdvisor, un site populaire d'avis de voyageurs, souhaite analyser les commentaires laissés par les utilisateurs sur différents hôtels et destinations.

Application

- L'analyse sémantique est utilisée pour extraire des informations précieuses sur les aspects les plus appréciés ou critiqués par les clients (par exemple, la propreté des chambres, la qualité du service, l'emplacement de l'hôtel).

- Les hôtels peuvent utiliser ces informations pour identifier les domaines à améliorer et mettre en avant leurs points forts dans leur communication marketing.

Résultat



Intégration du Web Analytics avec le NLP

Utilisation avancée du NLP en Web Analytics

Exemple 03 : Modélisation de sujets pour un Forum de discussion sur la technologie

Contexte

- Analyse des discussions sur un forum populaire de technologie.

Application

- Utilisation de la modélisation de sujets pour filtrer et organiser ces discussions en thèmes principaux et identifier les tendances émergentes.

Principaux sujets

- Les algorithmes de NLP permettent de distinguer des sujets récurrents tels que "smartphones", "réalité virtuelle", "sécurité informatique" et "développement logiciel".

Visualisation

- Un nuage de mots est créé pour chaque sujet identifié, mettant en évidence les termes les plus fréquents et les plus significatifs associés à ces thèmes.

Intégration du Web Analytics avec le NLP

Utilisation avancée du NLP en Web Analytics

Exemple 04 : Analyse des tendances sur Twitter

Contexte

- Une entreprise souhaite suivre les tendances émergentes et les sujets de discussion sur Twitter relatifs à son secteur d'activité.

Application

- La modélisation de sujets est appliquée aux tweets pour identifier les thèmes principaux et leur évolution dans le temps.

- L'entreprise peut détecter rapidement les tendances émergentes et ajuster sa stratégie de contenu et de communication pour rester pertinente et engager son audience.

Résultat



Intégration du Web Analytics avec le NLP

Optimisation du contenu basée sur le NLP

Principe

- Le NLP analyse les interactions des utilisateurs sur un site, révélant leurs préférences pour créer des profils d'intérêts.
- Ces profils permettent de personnaliser le contenu du site, proposant des articles et produits adaptés aux goûts de chaque utilisateur.

Exemple 01 : Amazon

- Leader mondial du e-commerce, est réputé pour son système de recommandation personnalisé.
- L'entreprise utilise des techniques avancées de NLP pour analyser les avis des clients et les comportements d'achat.

Exemple 02 : Netflix

- Géant du streaming vidéo, se distingue par la personnalisation des recommandations de contenu pour ses utilisateurs.
- Cette personnalisation est possible grâce à une combinaison sophistiquée d'algorithmes de recommandation et du NLP.

Intégration du Web Analytics avec le NLP

Optimisation du contenu basée sur le NLP

Analyse des avis

- Extraction des informations clés des avis laissés par les clients. Par exemple, si de nombreux avis mentionnent la "longue durée de vie de la batterie" d'un smartphone, ce trait sera considéré comme un point positif pour ce produit.

Création de profils d'intérêts

- Combinons les données des avis avec les historiques d'achat et de navigation, Amazon crée des profils d'intérêts pour chaque utilisateur. Ces profils incluent des préférences pour certaines caractéristiques de produits, marques ou catégories.

Personnalisation du contenu

- Sur la base de ces profils, Amazon personnalise la page d'accueil de chaque utilisateur, mettant en avant les produits susceptibles de l'intéresser. Par exemple, un utilisateur passionné de photographie se verra proposer des appareils photo haut de gamme et des accessoires connexes.



Résultats

La personnalisation a permis à Amazon d'accroître significativement l'engagement des utilisateurs et les taux de conversion

Les clients bénéficient d'une expérience d'achat plus pertinente et adaptée à leurs besoins, tandis qu'Amazon optimise ses ventes et sa fidélisation client.

Intégration du Web Analytics avec le NLP

Optimisation du contenu basée sur le NLP

Analyse des interactions et des commentaires

- Netflix analyse les données des utilisateurs, y compris les visionnages, évaluations et recherches, en utilisant le NLP pour extraire des insights des critiques et commentaires.

Création de profils d'intérêts

- Le NLP permet à Netflix de déterminer les préférences des utilisateurs, comme les genres favoris et les acteurs préférés, pour créer des profils d'intérêts détaillés.

Personnalisation des recommandations

- Netflix utilise ces profils pour personnaliser les recommandations, en alignant les caractéristiques des contenus avec les préférences des utilisateurs, offrant ainsi une expérience hautement personnalisée.



Résultats

Cette stratégie de personnalisation a prouvé son efficacité en fidélisant les abonnés et en optimisant les taux de visionnage, améliorant la satisfaction et l'engagement des utilisateurs.



Université Constantine 2
جامعة قسنطينة 2

Fin de Chapitre 04