05/06/2025

Célina Habtiche ; Zoé Abadie ; Carla Simon

ESME

Prédiction des retards dans le transport aérien.

Rapport Projet 10

Sommaire

[Choix et explication des sources de données choisies 2](#_Toc194947726)

[Web Scraping 3](#_Toc194947727)

[Liens et références 4](#_Toc194947728)

# Choix et explication des sources de données choisies

Sous-titre

Nous avons utilisé le site *Aviono.com* pour la collecte des données. Ce site nous a permis d’accéder à l’historique des vols depuis le 5 janvier jusqu’à aujourd’hui, incluant des informations telles que la destination, la compagnie aérienne, ainsi que les heures de départ estimées et réelles.

Afin de simplifier la centralisation de données, nous avons choisi de nous concentrer uniquement sur les vols au départ de l’aéroport de Orly situé à Paris.

Aujourd’hui, les données sont un atout stratégique précieux pour toutes les organisations pour piloter et développer des activités. Pour cela il est important d’utiliser les bonnes données afin de prendre les décisions les plus optimales ou encore d’établir des prédictions mathématiques. Cependant, il existe une quantité de données qui rend tout simplement impossible de trouver manuellement toutes les informations qui nous sont utiles, c’est à cette étape que nous introduisons le Web Scraping.

# Web Scraping

Qu’est-ce que c’est ?

Le Web Scraping est une méthode automatisée de data mining qui permet d’accéder de façon évolutive et organisée l’information récoltée sur des sites web pour l’analyser. Cette technologie est souvent utilisée par des Data Scientist, des Data Analyst ou encore des Business Analyst pour des objectifs de business intelligence ou de projets publicitaire, par exemple pour trouver des informations de contact, comparer des prix sur différents sites web… . Elle permet d’obtenir de grandes quantités de données à partir de sites web.

Comment ça fonctionne ?

Dans un premier temps, le protocole HTTP/HTTPS est utilisé pour transférer les données entre un serveur web et un navigateur. Le protocole http est un protocole de la couche application dans le modèle OSI de communication réseau qui permet d’échanger différents types de requêtes et de réponses dans notre cas pour consulter les données d’un site web.

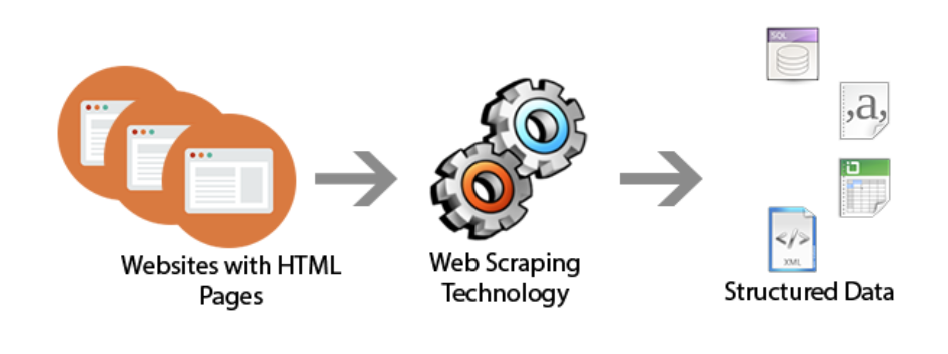


Figure : Architecture simplifiée du web scraping

Un site internet est composé de plusieurs pages web qui sont reliées entre elles par des hyperliens, chaque page web correspond à un fichier HTML renvoyé par le serveur, généralement accompagné de fichiers sources JavaScript et de feuilles de style CSS. L’ensemble des informations que nous recherchons se trouvent entièrement dans le fichier HTML transmis par le serveur.

# Liens et références

Sous-titre

<https://prevision-meteo.ch/climat/journalier/paris-orly/2025-02>

# Table des figures

[Figure 1: Architecture simplifiée du web scraping 3](#_Toc195005024)