LIFPROJET – RC1 : Data Mining

Avalanches

Aymeric TOUCHE & Gérome FERRAND

Enseignant référent Mr Cazabet

2020

Université Lyon I

Table des matières

[I. Le sujet 2](#_Toc37676098)

[A. Le choix 2](#_Toc37676099)

[B. Nos objectifs 2](#_Toc37676100)

[II. Le travail réalisé 2](#_Toc37676101)

[A. La recherche de données 2](#_Toc37676102)

[B. La création d’un site web 3](#_Toc37676103)

[C. La carte interactive 3](#_Toc37676104)

[D. Le scrapper 3](#_Toc37676105)

[III. Organisation du travail 3](#_Toc37676106)

[A. Répartition 3](#_Toc37676107)

[B. Synchronisation 4](#_Toc37676108)

[IV. Les défis 4](#_Toc37676109)

# Le sujet

## Le choix

Nous avons choisi le sujet RC1 – Data Mining.

La recherche et l’exploitation de données est un sujet qui nous a tout de suite intéressé.

Nous avons choisi un support web pour notre application car nous voulions progresser dans ce domaine.

L’un d’entre nous étant passionné de montagne, nous nous sommes très vite orientés vers la récupération de données concernant les avalanches, principalement dans les Alpes. Ce sujet n’ayant jamais été traité lors des sessions précédentes, cela nous a conforté dans notre choix.

L’idée est ici de donner un maximum d’informations sur les avalanches recensées de la manière la plus simple possible et d’en ressortir leurs principales caractéristique grâce un une base de données fournie.

## Nos objectifs

Nos objectifs de départs étaient les suivants :

* Collecter des données sur le net
* Analyser les caractéristiques principales d’une avalanche
* Cartographier les avalanches en fonction de ces caractéristiques
* Etablir une analyse graphique via des statistiques
* Faire ressortir des indicateurs et en dégager un sens
* Rester ouverts aux possibilités offertes par les jeux de données collectés

# Le travail réalisé

## La recherche de données

Dans un premier temps, nous avons recherché des jeux de données sur internet et notamment sur google DataSearch.

* 1er jeu de données : un fichier geojson rapportant une photo interprétation des avalanches dans les Alpes. Des personnes ont étudié la forme des terrains pour en déduire si une avalanche y avait eu lieu.
* 2e jeu de données : Fichier csv par année recensant les accidents mortels ou non dus aux   
  avalanches, toujours dans les Alpes.
* 3e jeu de données : Un site internet qu’il a fallu parcourir via un programme afin de récupérer les données intéressantes et similaires au jeu de données précédents et de créer un fichier geojson.

Nous avons pris contact avec le propriétaire du premier jeu de données (l’institut national de recherche pour l’agriculture, l’alimentation et l’environnement) afin d’obtenir des informations complémentaires, malheureusement les délais pour obtenir ces données et signer une licence de mises à disposition étaient trop long.

## La création d’un site web

Le site web a été développé via le Framework Slim PHP qui permet une architecture MVC et une gestion simple et efficace des requêtes http.

Son utilisation n’est pas essentielle à un tel projet, néanmoins nous voulions aller plus loin dans la programmation web.

Le choix du moteur de templates twig nous a permis de factoriser notre code afin d’éviter toutes formes de redondances, c’est un outil qui s’utilise naturellement avec Slim PHP.

Le site web a par la suite été déployé sur un serveur mis à disposition par Mr Cazabet.

## La carte interactive

Le site MapBox.com permet de créer un fond de carte que nous avons adapté au sujet, avec les reliefs. La carte interactive a été créée grâce à la bibliothèque MAPBOX GL JS qui est une librairie JavaScript.

Cet outil permet de créer différents layers traduisant les fichiers geojson préalablement créés ou récupérés. Ces layers sont ajouté à la carte, les avalanches sont alors représentées sur la carte.

Des scripts en JavaScript ou JQuery sont implémentés afin d’afficher les différents layers selon ce que souhaite l’utilisateur.

## Le scrapper

Ijqfhqskjd

# Organisation du travail

## Répartition

Trois tâches ont rapidement été identifiées.

La première étant de trouver des donnée, la seconde de créer une application web.

La recherche de données a été réalisée en binôme, nous avons passé du temps à deux là-dessus.

Par la suite l’un de nous s’est occupé principalement de l’application web en elle-même et du traitement des données récupérées afin de les afficher correctement sur la carte et de rendre l’application interactive.

L’autre s’est chargé de créer un scrapper en python afin de parcourir le site data-avalanche.org, de récupérer les données concernant les avalanches cataloguées et de les traduire dans des fichiers geojson.

Ces fichiers ont ensuite servi de base de données pour l’application web, en plus de celles déjà mise à disposition directement sur d’autre sites.

## Synchronisation

Un dépôt GitHub a été créé, chacun pouvait donc avancer sur l’implémentation des fonctionnalités qu’il avait à réaliser.

Au début du projet nous travaillions souvent ensemble à l’université ou chez nous. Tout le travail était réalisé en lors de ces séances fréquentes afin que chacun puisse participer à toutes les fonctionnalités.  
En revanche depuis le début du confinement, chacun les deux aspects du projet est devenu assez hermétiques et nous travaillions sur nos parties respectives, la communication étant devenue plus difficile.

# Les défis

Nous avons dû nous former sur de nouvelles technologies connues ou non pour mener à bien ce projet.

Nous ne voulions pas faire un site web classique en tout html bien que cela aurait été possible et n’aurait rien changé au rendu, mais nous n’aurions pas progressé. Pour cela il a fallu se renseigner sur les Frameworks PHP existant et en sélectionner un adapté à la structure et l’envergure du projet.

Plusieurs jours ont été nécessaires afin de nous familiariser avec Slim PHP et ses techniques.

Au dernier moment nous avons découvert un conflit entre le serveur partagé hébergeant notre site et Slim PHP concernant les requêtes http. Une solution de secours tout html a été conçue en quelques dizaines de minutes. Finalement nous avons pu gérer ce problème et conserver une architecture plus avancée.

La méthode d’affichage des données sur une carte interactive nous était totalement étrangères, et nos compétences en JavaScript et JQuery limitées. Nous nous sommes donc formés pendant des jours sur ces technologies. L’outil leaflet proposé à été remplacé par mapbox qui est plus complet et possède une documentation très fournie ainsi qu’un système d’hébergement.   
Cette solution d’hébergement nous a été très utile concernant la partie photo-interprétation, les données étant trop volumineuses nous avons pu ainsi en déléguer le traitement ce qui a permis un gain de temps considérable au chargement de la carte.

Enfin, comme cité précédemment, la crise sanitaire et le confinement que nous vivons a été un réel frein à l’avancement de notre projet. Collaborer à distance aura constitué un défi majeur que nous avons relevé en modifiant notre stratégie de fonctionnement.