

**Arnaques en ligne**

**Projet tableau de bord**

Janvier à mars 2021

Université Paul Sabatier Toulouse

**Fournisseurs :**

**AZOULAY Nathan – KIRED Nour MANSON Marianne – ROSSI Valentine**

**Clients :**

**BAHSOUN Wahiba MOKADEM Riad**

**VEDIS Théo**

**Historique du document**

| **Date N° de version Chapitre concerné Objet de la modification** |
| --- |
| 28 janvier 2021 v0 Création du document |
| 8 février 2021 v0.1 Chapitres I, II, III et VI Début de la rédaction |
| 9 février 2021 v0.2 Chapitre III  Création du Gantt prévisionnel  Chapitre IV  Début de la rédaction |
| 16 février 2021 v0.3 Annexes Ajout de la charte qualité Ajout du cahier des charges |
| Chapitre IV  Modification des SADT  27 février 2021 v0.4  Table des figures  Remplissage automatique  Sommaire  Remplissage automatique |
| Chapitre II  Mise à jour du contenu  Mise à jour du contenu  Chapitre III  Création du Gantt réel  3 mars 2021 v1.0  Chapitre IV  Mise à jour du contenu  Chapitre VI  Mise à jour des revues  Chapitre VII  Début de la rédaction |
| 8 mars 2021 v1.1 Chapitre IV Mise à jour du contenu |
| Modification des SADT  13 mars 2021 v1.2 Chapitre IV  Mise à jour du contenu  Chapitre VI  Mise à jour des revues |
| Modifications suite à la  15 mars 2021 v2 Tout le document  relecture des membres  de l’équipe |
| 19 mars 2021 v2.1 Chapitre IV Mise à jour du contenu |
| Modifications suite à la  22 mars 2021 v2.2 Tout le document  relecture des membres  de l’équipe |

Arnaques en ligne 2

**Sommaire**

Table des figures......................................................................................................................... 4 Chapitre I : Objet, but et domaine d’application ....................................................................... 5 I. Objet du document.............................................................................................................. 5 II. But du document................................................................................................................ 5 III. Domaine d’application....................................................................................................... 5 Chapitre II : Documents de référence et terminologie .............................................................. 6 I. Documents applicables........................................................................................................ 6 II. Documents de référence .................................................................................................... 6 III. Terminologie...................................................................................................................... 8 Chapitre III : Organisation de développement........................................................................... 9 I. Ressources humaines .......................................................................................................... 9 II. Gestion de projet.............................................................................................................. 10 Chapitre IV : Démarche de développement............................................................................. 12 I. Recherche des données..................................................................................................... 13 II. Préparation des données.................................................................................................. 14 III. Valorisation des données ................................................................................................ 14 IV. Visualisation des données............................................................................................... 17 Chapitre V : Gestion de configuration...................................................................................... 24 I. Rechercher les données..................................................................................................... 24 II. Préparer les données........................................................................................................ 24 III. Valoriser les données....................................................................................................... 25 IV. Visualiser les données..................................................................................................... 25 Chapitre VI : Assurance qualité ................................................................................................ 27 Chapitre VII : Bilan du projet.................................................................................................... 32 I. Points négatifs ................................................................................................................... 32 II. Points positifs.................................................................................................................... 32 Annexes.................................................................................................................................... 34 1. Cahier des charges............................................................................................................ 34 2. Charte qualité ................................................................................................................... 41

Arnaques en ligne 3

**Table des figures**

Figure 1 : Organigramme............................................................................................................ 9 Figure 2 : Diagramme de Gantt prévisionnel ........................................................................... 11 Figure 3 : Diagramme de Gantt réel......................................................................................... 11 Figure 4 : SADT - Etape A-0....................................................................................................... 12 Figure 5 : SADT - Etape A0........................................................................................................ 12 Figure 6 : SADT - Etape A1........................................................................................................ 13 Figure 7 : SADT - Etape A2........................................................................................................ 14 Figure 8 : SADT - Etape A3........................................................................................................ 14 Figure 9 : Modèle Conceptuel de Données.............................................................................. 16 Figure 10 : SADT - Etape A4...................................................................................................... 17 Figure 11 : Evolution de l’intérêt pour les différents types d'arnaque dans le temps ............ 18 Figure 12 : Proportion des types d'arnaque de 1999 à 2005................................................... 19 Figure 13 : Proportion des types d'arnaque de 2006 à 2010................................................... 19 Figure 14 : Proportion des types d'arnaque de 2011 à 2016................................................... 19 Figure 15 : Proportion des types d'arnaque de 2016 à nos jours............................................ 19 Figure 16 : Evolution du thème des articles dans le temps ..................................................... 20 Figure 17 : Mots clefs qui ressortent le plus dans les articles par rapport au thème description................................................................................................................................ 20 Figure 18 : Mots clefs qui ressortent le plus dans les articles par rapport au thème détection .................................................................................................................................................. 20 Figure 19 : Nombre d'articles écrits de 1999 à 2005 ............................................................... 21 Figure 20 : Nombre d'articles écrits de 2006 à 2010 ............................................................... 21 Figure 21 : Nombre d'articles écrits de 2011 à 2015 ............................................................... 21 Figure 22 : Nombre d'articles écrits de 2016 à 2022 ............................................................... 21 Figure 23 : Evolution de la nationalité des auteurs par rapport au thème au cours du temps .................................................................................................................................................. 22 Figure 24 : Nuage de mots des mots clefs de 1999 à 2005...................................................... 22 Figure 25 : Nuage de mots des mots clefs de 2006 à 2010...................................................... 22 Figure 26 : Nuage de mots des mots clefs de 2011 à 2015...................................................... 22 Figure 27 : Nuage de mots des mots clefs de 2016 à 2020...................................................... 22

Arnaques en ligne 4

**Chapitre I : Objet, but et domaine d’application**

**I. Objet du document**

Ce document présente notre travail pour le projet tableau de bord.

Ce projet tableau de bord est réalisé dans le cadre de la formation SID (Statistique et Informatique Décisionnelle). Celui-ci nous permet d’apprendre à restituer nos compétences et connaissances acquises, grâce à nos professeurs ainsi qu’à nos expériences personnelles.

Le projet a duré deux mois, de fin janvier à fin mars, au cours desquels nous avons parfois travaillé ensemble, parfois en sous-groupe sur différentes tâches réparties équitablement et en fonction des compétences de chacun des cinq membres que compte notre groupe.

Il s’agit de développer un système d’aide à la décision à partir de la visualisation de données textuelles issues des bases de données en ligne.

**II. But du document**

Le projet est une mise en situation sous le format Client-Fournisseur où Mme Bahsoun et M. Mokadem sont les clients et nous, les étudiants, sommes leurs fournisseurs. Le document est donc destiné à présenter notre travail à nos professeurs.

**III. Domaine d’application**

Nous nous sommes penchés sur différents sujets, tels que l’IA et la santé, l’IA et la sécurité ou encore le télétravail, pour finir par nous intéresser à l’évolution des arnaques en ligne. En effet, après étude de la quantité et la qualité des données correspondantes pour chacun des sujets, c’est celui qui nous semblait le plus adéquat.

Arnaques en ligne 5

**Chapitre II : Documents de référence et terminologie**

**I. Documents applicables**

Dans le cadre de la mise en situation Client-Fournisseur, nous avons suivi pas-à-pas le cahier des charges fourni par le client au début du projet. Celui est répertorié en annexe.

Notre groupe disposant d’un responsable d’assurance et contrôle qualité, nous avons choisi de créer une charte qualité afin que chaque membre du groupe puisse se tenir à une organisation de code similaire, ce qui rend un livrable plus propre et structuré.

**II. Documents de référence**

Concernant les ressources qui nous ont permis de réaliser ce projet, elles sont multiples. D’une part, les connaissances acquises lors des différents cours suivi tout au long de nos formations. D’autre part, des ressources (principalement en ligne à la vue des conditions sanitaires actuelles) extérieures qui nous ont permis de compléter nos connaissances et compétences vis-à-vis de notre sujet.

a. Formation SID

Nous nous sommes d’abord basés sur les cours dispensés dans notre cursus : - l’extraction d’informations ;

- les entrepôts de données (data warehouse) ;

- les bases de données ;

- l’analyse statistique ;

- la visualisation de données ;

- le génie logiciel ;

- la gestion de projet ;

- la gestion de configuration ;

- la programmation.

b. Documentation pour la recherche d’articles

Pour la recherche d’articles :

- support.google.com/websearch/answer/2466433?hl=fr : pour comprendre comment affiner nos équations de recherche ;

- scholar.google.com : outil de recherche doté d’une bibliothèque colossale, nous ayant aiguillés dans le choix de notre sujet ainsi que dans la recherche et le tri d’articles ;

- link.springer.com : notre source de données.

Arnaques en ligne 6

c. Librairies utilisées

Pour la documentation Python, nous avons utilisé les librairies suivantes.

**Librairies pour les fichiers :**

- json : écriture/lecture des données dans des fichiers json ; - PIL : lecture/écriture d’images.

**Librairies pour la base de données :**

- cx\_oracle : type base de données ;

- sqlalchemy : liaison avec la base de données.

**Librairie pour les mathématiques :**

- numpy : type vecteur.

**Librairies pour la visualisation :**

- matplotlib : visualisation graphique ;

- seaborn : visualisation graphique ;

- WordCloud : nuage de mots.

**Librairies pour le scraping :**

- bs4 : parseur html ;

- requests : requête http.

**Librairies pour le système :**

- os : gestion system ;

- sys : gestion system.

**Libraires pour le traitement de texte et le nettoyage des données :** - collections : type de données ;

- gensim : topic modelling ;

- nltk : traitement de texte ;

- operator : ensemble de fonctions ;

- pandas : data frame ;

- re : expression régulière ;

- sklearn : machine learning ;

- spacy : traitement automatique des langues.

**Librairies pour le format :**

- datetime : type date ;

- pandas : data frame ;

- time : récupération de l’heure.

**Librairie autre :**

- TQDM : estimation du temps des boucles.

Arnaques en ligne 7

**III. Terminologie**

a. Base de données

**Entité :** Objet concret ou abstrait qui peut être reconnu distinctement et qui est unique. **Association (ou relation) :** Lien entre plusieurs entités.

**Attribut (ou une propriété) :** Caractéristique associée à une entité.

**Dictionnaire de données :** Liste de données référence nécessaire à la conception d’une base de données.

**Schéma entité/association :** Modèle de représentation des données sous forme d’entités (avec des attributs) et d’associations.

**MCD (Modèle Conceptuel de données) :** Modèle qui permet de présenter les différentes classes d’objets ainsi que leurs associations et leurs règles de comportement. **Schéma relationnel :** Ensemble des noms des relations et description des attributs pour chaque relation. Ce modèle représente la structure logique associée au modèle relationnel. **Clef primaire :** Attribut ou ensemble d’attributs qui assurent l’unicité d’une observation dans une table.

**Clef étrangère :** mécanisme permettant de garantir l’intégrité référentielle entre les données de deux tables.

b. Programmation

**iDF (inverse Document Frequency) :** Fréquence du mot dans la collection. **TF (Term Frequency) :** Fréquence du mot dans un document.

**JSON (JavaScript Object Notation – Notation Objet issue de JavaScript) :** Format de données textuelles dérivé de la notation des objets du langage JavaScript permettant de représenter de l’information structurée.

**Scraping :** Le scraping définit de façon générale une technique permettant d'extraire du contenu (des informations) d'un ou de plusieurs sites web de manière totalement automatique.

**Parsing :** Processus de transformation des données de type primitif, tel qu’une chaîne de caractères, en un autre type plus exploitable.

c. Planification

**Gantt :** Diagramme utilisé en gestion de projet. Il permet de représenter visuellement l’état d’avancement de chaque tâche du projet concerné.

**SADT (Specified Analysis and Design Technic) :** Démarche de modélisation d’un système complexe ou d’un processus opératoire.

Arnaques en ligne 8

**Chapitre III : Organisation de développement**

**I. Ressources humaines**

Chef de projet 

VEDIS Théo

Adjoint 

AZOULAY Nathan

Responsable de gestion de configuration 

MANSON Marianne

**Le chef de projet :**

Responsable d'assurance et contrôle qualité 

KIRED Nour

*Figure 1 : Organigramme*

Responsable de rédaction 

ROSSI Valentine

Le chef de projet est chargé de manager une équipe et gérer le bon déroulement du projet. Il a pour missions principales :

- le suivi de la conception du projet (besoin client, livrables, critères de réception…) ; - le pilotage du projet en organisant et coordonnant l’équipe ;

- le respect du cahier des charges fourni par le client (délais, objectifs, budget…).

**L’adjoint :**

L’adjoint du chef de projet aide le chef de projet sur ses missions en cas de besoin.

**Le responsable de gestion de configuration :**

Le rôle du responsable de gestion de configuration est de gérer la description technique d’un système (et de ses composants) en suivant toutes les étapes d’évolution de celui-ci. Il s’agit donc de contrôler l’ensemble des processus permettant d’assurer la conformité d’un produit aux exigences du cahier des charges, tout au long du cycle de vie.

**Le responsable assurance et contrôle qualité :**

La mission du responsable assurance et contrôle qualité est de s’assurer de la qualité informatique des livrables fournis au client ainsi que de leur lisibilité. Il s’appuie sur une charte qualité dont il a la charge. Il est le garant du bon niveau de qualité et de l'adaptation parfaite aux besoins du client.

**Le responsable de rédaction :**

Le responsable de rédaction est chargé de l’ensemble des aspects rédactionnels de tout document. Il en assure la responsabilité des contenus et la cohérence, et pilote l’évolution des documents (notamment leurs historiques et versions). Il s’assure également de la relecture des documents et veille à leur validation.

Arnaques en ligne 9

**II. Gestion de projet**

a. Planification

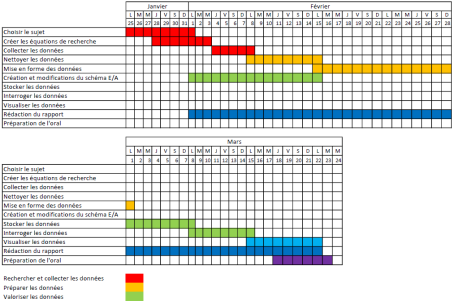
Nous avons réparti les tâches et les rôles à travers un tableau de répartition et un diagramme de Gantt prévisionnel. Cependant, les délais prévus se sont révélés différents de la réalité. Nous avons donc ajusté le diagramme en fonction de notre avancement, ce qui nous a donné un diagramme de Gantt réel.

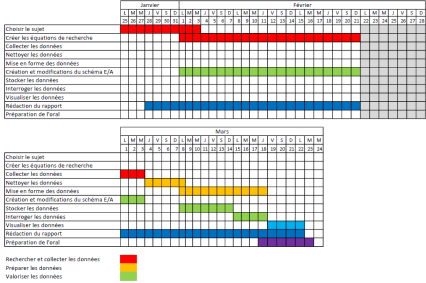
b. Répartition des tâches

| **Nom de la tâche Personnes concernées** | |
| --- | --- |
| **Rechercher et collecter les données** | |
| Choisir le sujet - Tout le groupe | |
| Créer les équations de recherche - Tout le groupe | |
| - KIRED Nour  Récupérer les données  - MANSON Marianne  - VEDIS Théo | |
| **Préparer les données** | |
| Rechercher les anomalies des données - MANSON Marianne | |
| - KIRED Nour  Nettoyer les données  - MANSON Marianne  - VEDIS Théo | |
| - KIRED Nour  Mettre en forme des données  - MANSON Marianne  - VEDIS Théo | |
| **Valoriser les données** | |
| Créer le schéma E/A - KIRED Nour  - ROSSI Valentine | |
| Créer la structure de la BD - ROSSI Valentine  - VEDIS Théo | |
| Stocker les données - ROSSI Valentine  - VEDIS Théo | |
| - KIRED Nour  Interroger les données  - ROSSI Valentine  - VEDIS Théo | |
| **Visualiser les données** | |
| - KIRED Nour  - MANSON Marianne  Analyser les données  - ROSSI Valentine  - VEDIS Théo | |
| - KIRED Nour  Sélectionner les résultats les plus  - MANSON Marianne  pertinents  - ROSSI Valentine  - VEDIS Théo | |
| - KIRED Nour  - MANSON Marianne  Représenter les résultats  - ROSSI Valentine  - VEDIS Théo | |
| **Qualité des notebooks Python** | - KIRED Nour |
| **Rédaction du rapport** | - ROSSI Valentine |
| **Préparation de l’oral** | - Tout le groupe |

Arnaques en ligne 10

c. Diagrammes de GANTT

*Figure 2 : Diagramme de Gantt prévisionnel*

*Figure 3 : Diagramme de Gantt réel*

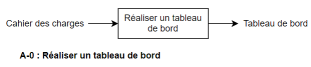
Arnaques en ligne 11

**Chapitre IV : Démarche de développement**

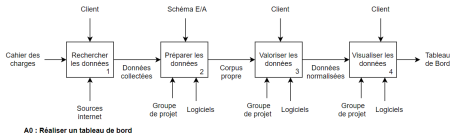
Avant toute chose, nous avons dû choisir un sujet sous divers critères : - ayant suffisamment de données à collecter pour être pertinent ;

- dont la problématique soit jugée adéquat et intéressante par le client ; - que nous le trouvions attrayant et plaisant à étudier.

Tous les schémas SADT présentés dans ce chapitre ont été réalisés grâce au logiciel *drawio* en ligne, qui permet de réaliser toutes sortes de graphiques ou diagrammes. Voici, ci-dessous, le développement global du projet :



*Figure 4 : SADT - Etape A-0*

*Figure 5 : SADT - Etape A0*

Le système développé dans le cadre de ce projet se décline en plusieurs étapes du processus décisionnel, qui peuvent être regroupées en quatre grandes parties citées ci-après : - recherche et collecte des données ;

- préparation des données ;

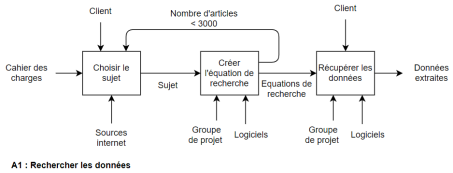
- valorisation des données ;

- visualisation des données.

Afin de schématiser notre processus de développement suivi dans le cadre du projet, la méthode SADT nous semble la plus pertinente.

Arnaques en ligne 12

**I. Recherche des données**

*****Figure 6 : SADT - Etape A1*

Notre sujet d’étude étant les arnaques en ligne, il était nécessaire de lister les arnaques les plus communes basées sur notre expérience, celle de notre entourage et de compléter ce listage par des recherches approfondies. Nous nous sommes donc appuyés sur les sites internet suivants :

- https://arnaqueinternet.com/ ;

- https://domaine-legal.com/ ;

- https://www.kaspersky.fr/ ;

- https://www.abcargent.com/.

Tout d’abord, pour la collecte des données, nous avons décidé de nous concentrer sur les abstracts d’articles de différentes sources. Cependant, via nos équations de recherche, beaucoup n’étant pas pertinentes quant à la quantité (< 3000 articles) et/ou la qualité des données, nous avons choisi de n’en garder qu’une seule : https://link.springer.com/. La question de la langue ne s’est donc pas posée pour nous puisque notre meilleure source, qui est donc notre seule source au final, est en anglais. De plus, ce site nous donne accès à des documents scientifiques de tous genres, ce qui nous permet de varier nos contenus entre revues, livres, ouvrages de références et autres, tout en étant sûrs d’avoir un contenu sérieux : il était hors de question pour nous d’avoir des articles people par exemple.

Concernant nos équations de recherche (en anglais), nous en avons finalement retenu qu’une : (internet OR online) AND (scam OR fraud).

Pour effectuer la collecte des données, nous avons choisi de faire un programme Python grâce à ces deux librairies :

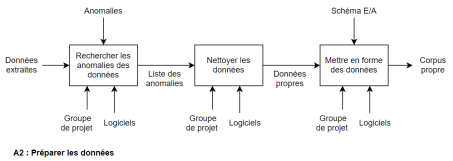
- *requests* : utile pour la récupération du code html d'une page ;

- *bs4* (*BeautifulSoup*) : pratique et efficace pour le parsage du code html ; afin d’exécuter nos équations de recherche (citées au-dessus) et de scrapper les articles qui en ressortent. Une fois les données collectées, nous avons choisi de les stocker dans un fichier JSON, ce qui permet d’avoir des listes de listes propres.

Arnaques en ligne 13

Plusieurs réunions client/fournisseur ont été nécessaires afin de valider notre choix de sujet ainsi que la pertinence de nos données extraites. Cela nous a permis de correctement saisir les demandes et attentes du client afin que notre projet y adhère au maximum.

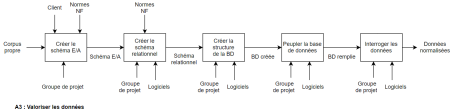
**II. Préparation des données**

*****Figure 7 : SADT - Etape A2*

La préparation des données est une étape importante dans le déroulement du projet : les données extraites ne sont pas exploitables et utilisables en l’état. Il s’agit donc tout d’abord de rechercher des anomalies possibles. Une fois cela fait, nous avons nettoyé et mis en forme les données (grâce aux librairies json et datetime) afin de pouvoir les insérer correctement dans la future base de données, ce qui nous donne ainsi un corpus propre.

Simultanément, nous avons élaboré le schéma entité/association de la future base de données ainsi que son schéma relationnel associé (cf II. Valorisations des données). Celui-ci nous est utile ici, notamment dans la mise en forme des données afin que ces dernières correspondent parfaitement à notre future base de données et qu’aucune modification ne soit nécessaire de ce côté-là.

**III. Valorisation des données**

*****Figure 8 : SADT - Etape A3*

Suite à la situation sanitaire actuelle, pour une question de facilité et pour adhérer au cahier des charges, nous avons choisi Oracle comme système de gestion de bases de données. De ce fait, la base de données a été implémentée via le logiciel SQL Developper, que nous connaissons bien car fréquemment utilisé en cours.

Arnaques en ligne 14

a. Stockage des données

Lors de l’élaboration du schéma entité/association (créé grâce au logiciel *mocodo* en ligne), il nous a été nécessaire de faire valider notre démarche et nos choix auprès du client. Une fois fait, nous avons pu commencer à créer la structure de notre base de données : création des tables (tout en faisant attention aux clefs primaires et étrangères), des procédures et des vues associées.

| **Dictionnaire des données** |
| --- |
| **Exemple de**  **Nom**  **Nom de l’attribut Description de l’attribut Type de**  **valeur que peut**  **de la**  **l’attribut**  **prendre**  **table**  **l’attribut** |
| id\_arnaque Fait référence à id\_arnaque de  la table ARNAQUE. NUMBER 392  **R**  **I**  id\_acteur Fait référence à id\_acteur de la  **B**  table ACTEUR. NUMBER 658  **U**  **S** |
| id\_auteur Identifiant de l’auteur. NUMBER 127 nom Nom et prénom de l’auteur. VARCHAR2 Mohamed Chawki  International  **R**  affiliation Organisme auquel l’auteur est  Association of  **U**  **E**  rattaché pour la publication. VARCHAR2  **T**  Cybercrime  **U**  **A**  Prevention  paysPays où se trouve l’auteur au  moment de l’arnaque. VARCHAR2 France |
| id\_article Fait référence à id\_article de la  table ARTICLE. NUMBER 9480  **E**  **R**  **I**  id\_auteur Fait référence à id\_auteur de la  **R**  table AUTEUR. NUMBER 127  **C**  **E** |
| Identifiant de l’acteur de  id\_acteur  l’arnaque, qu’il soit victime ou  NUMBER 658  arnaqueur.  **R**  **U**  nom Nom et prénom de l’acteur. VARCHAR2 Chris Barton  **E**  **T**  **C**  demographie Caractéristique de l’acteur. VARCHAR2 - **A**  nationalité Nationalité de l’acteur. VARCHAR2 France |
| id\_arnaque Fait référence à id\_arnaque la  table ARNAQUE. NUMBER 392  **R**  **E**  id\_acteur Fait référence à id\_acteur de la  **E**  table ACTEUR. NUMBER 658  **R**  **C** |
| id\_arnaque Identifiant de l’arnaque. NUMBER 392 type\_arnaque Le type de l’arnaque. VARCHAR2 phishing date\_arnaque Date à laquelle s’est produite  l’arnaque. DATE 2015-01-01  **E**  (YYYY-MM-DD)  **U**  **Q**  plateforme\_utilisee Plateforme utilisée pour  **A**  l’arnaque. VARCHAR2 email  **N**  **R**  **A**  nb\_victimes Nombre de personnes victimes  de l’arnaque. NUMBER 11 |

Arnaques en ligne 15

| **E**nb\_arnaqueurs Nombre de personnes ayant  créé l’arnaque. NUMBER 3  **U**  **Q**  id\_article Fait référence à id\_article de la  **A**  table ARTICLE. NUMBER 9480  **N**  **R**  **A** |
| --- |
| **F**id\_article Fait référence à id\_article de la  table ARTICLE. NUMBER 9480  **E**  **L**  **C**  motclef Mot clef correspondant à  **T**  l’article en question. VARCHAR2 scan  **O**  **M** |
| **R**id\_arnaque Fait référence à id\_arnaque de  la table ARNAQUE. NUMBER 392  **E**  **R**  id\_acteur Fait référence à id\_acteur de la  **T**  table ACTEUR. NUMBER 658  **N**  **O**  **C** |
| id\_article Identifiant de l’article. NUMBER 9480 date\_article Date de la création de l’article. DATE 2013-06-07  **E**  (YYYY-MM-DD)  **L**  **C**  **I**  **T**  titre Titre de l’article. VARCHAR2 Romance Fraud  **R**  **A**  theme Thème de l’article. VARCHAR2 - |



*Figure 9 : Modèle Conceptuel de Données*

**Schéma relationnel correspondant :**

SUBIR (#id\_arnaque, #id\_acteur)

AUTEUR (id\_auteur, nom, affiliation, pays)

ECRIRE (#id\_article, #id\_auteur)

ACTEUR (id\_acteur, nom, demographie, nationalite)

CREER (#id\_arnaque, #id\_acteur)

ARNAQUE (id\_arnaque, type\_arnaque, date\_arnaque, plateforme\_utilisee, nb\_victimes, nb\_ arnaqueurs, #id\_article)

MOTCLEF (#id\_article, mot\_clef)

CONTRER (#id\_arnaque, #id\_acteur)

ARTICLE (id\_article, date\_article, titre, theme)

Arnaques en ligne 16

Une fois la base de données créée selon les schémas précédents, il nous a fallu y stocker les données. Pour cela, nous avons fait une fonction sous Python qui permet d’utiliser les procédures (préalablement codées) pour faire les insertions de données.

b. Interrogation des données

Une partie du nettoyage des données s'est faite directement dans la base de données lors de l’interrogation des données. Lire les données dans l'affichage Oracle a permis de détecter des anomalies que nous avions manquées. Nous avons donc modifié les pays des auteurs, par exemple, « United Kingdom » a été transformé en « UK » ou encore « Republic of China » a été modifié en « China ».

Nous avons utilisé différentes vues pour réaliser la visualisation de données : - Evolution du type d’arnaque dans le temps

- Evolution du thème de l'arnaque dans le temps

- Mots clefs en fonction de la détection des fraudes

- Mots clefs en fonction de la description des arnaques

- Nombre d’article par pays et par année

- Nombre d’article par thème et par pays

**IV. Visualisation des données**

*****Figure 10 : SADT - Etape A4*

Nous avons fait notre visualisation des données grâce à la récupération des vues créées en SQL, sur la base de données. Ainsi, tous les résultats ont été créés sur un notebook, en Python, avec les librairies suivantes : create\_engine, matplotlib.pyplot, pandas et numpy.

a. L’intérêt pour les types d’arnaque au cours du temps

Notre classification des différents types d'arnaques n'est pas très efficace. En effet, sur plus de 5000 articles, 4000 sont simplement placés en 'others'. Le problème vient de difficultés techniques mais aussi du fait que beaucoup d'articles parlent des arnaques de manière générale.

Arnaques en ligne 17

| **Arnaque Description** |
| --- |
| Spam emails Tout type d'arnaques reposant sur des courriels. |
| La fraude 419 (aussi appelée arnaque nigériane) est une arnaque  qui se présente généralement sous la forme d'un courriel. Dans ce  dernier une personne affirme posséder une importante somme  d'argent (plusieurs millions de dollars en héritage, pots-de-vin,  419 scam  comptes tombés en déshérence, fonds à placer à l'étranger à la  suite d'un changement de contexte politique, etc.) et fait part de  son besoin d'utiliser un compte existant pour transférer  rapidement cet argent. C'est un type particulier de "spam emails". |
| Toutes les arnaques où la victime croit investir (dans une  Investment scam  entreprise, une organisation). Ce type d'arnaque peut regrouper  les fraudes 419. |
| Recruitment scam Arnaque à l'emploi. Une personne se fait passer pour un employeur pour récupérer des informations personnelles. |
| Bank scam Toutes arnaques impliquant des banques (banques en ligne frauduleuses). |
| Arnaques aux sentiments. Des personnes malintentionnées  gagnent la confiance de leur victime (en particulier sur des sites  Romance scam  de rencontres) pour leur soutirer de l'argent, des informations  personnelles, des coordonnées bancaires. |
| Virtual currency scam Toutes les arnaques impliquant des monnaies virtuelles (en particulier bitcoin). |
| Consumer scam Toutes les arnaques où la victime est trompée lors d'un achat en ligne (marketing trompeur, faux sites, etc) |
| Toutes les arnaques liées aux enchères en ligne (vendeur  Auction scam  frauduleux, arnaqueurs tiers qui utilise les données de l'enchère  pour monter son arnaque). |



*Figure 11 : Evolution de l’intérêt pour les différents types d'arnaque dans le temps*

L’intérêt pour l’étude des arnaques à augmenter au cours du temps, particulièrement depuis 2015. En 2021, une chute de l’intérêt est visible car l’année n’est pas encore terminée, les données de cette année-ci ne sont donc pas pertinentes.

Arnaques en ligne 18

*Figure 12 : Proportion des types d'arnaque de 1999 à 2005 Figure 13 : Proportion des types d'arnaque de 2006 à 2010*

*Figure 14 : Proportion des types d'arnaque de 2011 à 2016 Figure 15 : Proportion des types d'arnaque de 2016 à nos jours*

Quelque soit les tranches d’années, l’intérêt pour les arnaques de consommations en ligne (consumer scam) reste le plus important. Plusieurs facteurs peuvent expliquer cette importance :

- les arnaques de consommation en ligne peuvent prendre plusieurs formes et donc regrouper beaucoup d’arnaques différentes ;

- une raison technique : c'est aussi le type d'arnaque pour lequel nous avons trouvé le plus de mots-clefs.

Les arnaques impliquant de la monnaie virtuelle (virtual currency scam), deviennent un sujet pour les auteurs réellement à partir des cinq dernières années. En effet, même si le concept de cryptomonnaie existe depuis les années 1990, il a été popularisé en 2009 avec l'invention du Bitcoin. Il attire petit à petit l'attention les années suivantes, en particulier en 2017 après une première montée en valeur (jusqu'à plus de 11 000€). Cette attention et forte valeur a pu faire émerger un nouveau type d'arnaque (en plus de potentiellement mettre en lumière ce qui existait déjà).

En revanche, les arnaques liées aux enchères en ligne (auction scam) sont de moins en moins un sujet en vogue. Les arnaques aux sentiments (romance scam), les arnaques à l’emploi (recruitment scam) et les arnaques à l’investissement (investment scam) restent une minorité de sujets d’articles au sein de notre source.

b. Les thèmes des articles

A partir des articles, nous avons dégagé deux thèmes: détection et description. Détection concerne les documents qui parlent de la détection et de la lutte contre les arnaques. Description signifie que le papier dépeint un type d’arnaque.

Arnaques en ligne 19



*Figure 16 : Evolution du thème des articles dans le temps*

Notre source de données nous propose globalement plus d’articles parlant de description d’arnaque que de détection. Cela semble cohérent vis-à-vis de notre équation de recherche et du sujet abordé. Il est à noter que l’écart entre les deux thèmes s’intensifie à partir de l’année 2015.

c. Les mots clefs des articles



*Figure 17 : Mots clefs qui ressortent le plus dans les articles par rapport au thème description*

*Figure 18 : Mots clefs qui ressortent le plus dans les articles par rapport au thème détection*

Concernant le terme de description des arnaques, les mots clefs les plus courants sont internet, cybercrime, fraud, ce qui semble cohérent vis-à-vis de notre sujet. Les cartes de crédit et l’identité reviennent souvent car les articles sur les fraudes en ligne y font également souvent référence.

Du côté des détections de fraude, il n’y a pas de mot-clefs flagrants, qui se détacheraient plus particulièrement des autres, contrairement aux descriptions. Cela vient sûrement du fait que notre source comporte beaucoup moins d’articles parlant de ce thème. On peut toutefois noter un nombre important de mot-clefs liés aux nouvelles technologies (machine learning, deep learning, internet of things/iot, big data). On peut en déduire qu'il y a un intérêt pour appliquer ces dernières à la détection des fraudes.

Arnaques en ligne 20

d. Nationalité des auteurs d’articles

*Figure 19 : Nombre d'articles écrits de 1999 à 2005 Figure 20 : Nombre d'articles écrits de 2006 à 2010*

*Figure 21 : Nombre d'articles écrits de 2011 à 2015 Figure 22 : Nombre d'articles écrits de 2016 à 2022*

De plus en plus de pays apparaissent au fil des années, cela peut venir de deux raisons : soit les auteurs n’étaient pas intéressés par les arnaques dans les pays non illustrés, soit le fait que notre source soit exclusivement en anglais biaise les résultats. Par exemple l'utilisation du site par des scientifiques non-anglophones peut s'être démocratisée seulement durant ces dernières années.

Depuis les années 2000, le pays le mieux représenté à travers les articles est donc les USA, sûrement pour les raisons citées ci-dessus. Une majorité des pays sont très peu représentés par les auteurs. En effet, seuls la Chine, l’Inde, l’Allemagne, l’Australie, le Canada, le Royaume Uni, l’Italie et les Pays Bas ont une quantité d’auteurs d’articles conséquente.

Arnaques en ligne 21

*Figure 23 : Evolution de la nationalité des auteurs par rapport au thème au cours du temps*

L’Inde et la Chine, en plus des USA, ont des auteurs qui écrivent beaucoup sur les détections d’arnaque comparé aux autres pays ainsi qu’au thème de description d’arnaque. On peut par ailleurs noter que le Nigéria (qui a pourtant un type d'arnaque portant son nom : l'arnaque nigériane ou fraude 419) ne semble pas avoir beaucoup participé dans ce domaine. On trouve d'ailleurs dans notre source un article déplorant l'absence de points de vue de scientifiques nigériens sur cette problématique.

e. Mots clefs dominants par tranche d’années

*Figure 24 : Nuage de mots des mots clefs de 1999 à 2005 Figure 25 : Nuage de mots des mots clefs de 2006 à 2010*

*Figure 26 : Nuage de mots des mots clefs de 2011 à 2015 Figure 27 : Nuage de mots des mots clefs de 2016 à 2020*

Arnaques en ligne 22

Entre les années 2000 et 2010, l'intérêt des arnaqueurs pour les données personnelles (identity theft) s'ajoute à celle des données financières (credit card). Concernant les années 2015 à 2020, l’impact des nouvelles technologies est plus fort : en solution aux arnaques le « machine learning » apparaît, quant aux menaces vis-à-vis, il y a la « cryptomonnaie » qui amène de nouvelles formes d’arnaque.

f. Discussion

En conclusion, nous avons pu constater une prédominance des pays anglophones dans nos résultats. Cependant, nous étant limités à la langue anglaise, il nous semble cohérent que la majorité des articles soient de nationalité des USA notamment.

De plus, au cours du temps, il est à constater que les nouvelles technologies sont sources de nouvelles menaces, via la cryptomonnaie par exemple. De nouvelles solutions ont également tendance à apparaître simultanément tels que le machine learning ou encore le big data par exemple.

Arnaques en ligne 23

**Chapitre V : Gestion de configuration**

Tout au long de ce projet, nous avons travaillé sur nos ordinateurs personnels sous Windows pour certains, sur les ordinateurs prêtés par la formation sous Linux pour d’autres. En effet, compte tenu de la situation sanitaire actuelle, cela représente une facilité d’organisation et évite des déplacements inutiles sur l’université.

La totalité de nos documents (codes, rapports, PowerPoint, données…) est organisée et gérée via Google Drive qui permet à tous les membres du groupe d’avoir constamment accès à tout. Ce type d’organisation est très pratique pour le travail en équipe et notamment à distance. Nous avions pensé à mettre notre code en commun sur Github mais personne de notre groupe ne connaissait suffisamment bien l’outil pour pouvoir l’adopter et expliquer aux autres membres comment l’utiliser. C’est pourquoi Google Drive nous a semblé le plus adéquat, étant connu et couramment utilisé par tous.

Concernant le langage de programmation, nous avons opté pour le langage Python, utilisé soit sur Jupiter Notebook, soit sur Google Colab (les données récupérées ne présentant pas de devoir de confidentialité, l'utilisation de cet outil ne pose aucun problème). Afin de respecter la gestion de configuration, nous avons choisi de faire des tableaux récapitulatifs des différentes versions de chaque notebook.

**I. Rechercher les données**

**g3\_recuperation\_articles :** Récupération des articles à partir de l’équation de recherche

| **Date N° de version Objet de la modification Résultats** |
| --- |
| 1er mars 2021 v0 Création du fichier 5940 articles |
| 10 mars 2021 v1 Correction et documentation 6000 articles |

**g3\_collecte :** Récupération des informations dans les articles

| **Date N° de version Objet de la modification Résultats** |
| --- |
| Création du fichier  Compilation des différentes  9 mars 2021 v0  Données brutes  fonctions de scraping pour  chaque type d’article |

**II. Préparer les données**

**g3\_nettoyage\_donnees :** Nettoyage des données

| **Date N° de version Objet de la modification Résultats** |
| --- |
| 10 mars 2021 v0 Création du fichier Pas de résultat |
| 12 mars 2021 v1 Création de résultats clean\_articles\_5629.json |
| Nouveau format pour les  15 mars 2021 v2  auteurs : {"NomPrenom" :  clean\_articles\_final.json  ["affiliation","Pays"]} |

Arnaques en ligne 24

**g3\_preparation\_de\_donnees\_theme :** Création des thèmes des articles

| **Date N° de version Objet de la modification Résultats** |
| --- |
| Création des thèmes avec  15 mars 2021 v0  l’abstract (utilisation  Pas de résultat  kmeans) |
| 17 mars 2021 v1 Création des thèmes avec  l’abstract (utilisation lda) Pas de résultat |
| 18 mars 2021 v2 Création des thèmes avec le  titre (utilisation lda) themes.json |

**g3\_analyse\_texte\_type\_arnaque :** Création des types d’arnaques

| **Date N° de version Objet de la modification Résultats** |
| --- |
| 17 mars 2021 v0 Création des types  d’arnaques type\_scams.json |

**g3\_creation\_json\_insertion :** Mise en forme des données pour insertion avec l’ajout des thèmes des articles

| **Date N° de version Objet de la modification Résultats** |
| --- |
| 17 mars 2021 v0 Création du fichier Données prêtes à l’insertion |

**III. Valoriser les données**

**g3\_alimentation\_bd :** Alimentation de la BD

| **Date N° de version Objet de la modification Résultats** |
| --- |
| Création du fichier  9 mars 2021 v0  Fonctions d’alimentation de  Pas de résultat  la BD. |

**g3\_insertion\_bd :** Insertions dans la base de données

| **Date N° de version Objet de la modification Résultats** |
| --- |
| Création du fichier  17 mars 2021 v0  Insertion dans la base de  Pas de résultat  données |

**IV. Visualiser les données**

Nous avons créé un fichier final contenant notre tableau de bord :

**g3\_presentation\_wordCloud :** Création du nuage de mots de la page de garde

| **Date N° de version Objet de la modification Résultats** |
| --- |
| 18 mars 2021 v0 Création du fichier  Création du nuage de mots. Nuage de mots |

Arnaques en ligne 25

**g3\_dashboard :** Tableau de bord final

| **Date N° de version Objet de la modification Résultats** |
| --- |
| Création du fichier  Sorties graphiques  19 mars 2021 v0  Génération des visualisations  au format png  des données |

Arnaques en ligne 26

**Chapitre VI : Assurance qualité**

Compte tenu de la situation sanitaire actuelle, chaque revue citée ci-dessous s’est déroulée à distance, via les logiciels zoom ou discord.

Les personnes indiquées entre parenthèses étaient présentes lors de la revue mais passives vis-à-vis du sujet discuté.

**Revue n°1 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personnes présentes (client) :*

- BAHSOUN Wahiba

- MOKADEM Riad

*Date :* 25 janvier 2021

*Sujet discuté :* Recherche d’un sujet - Arnaques en ligne

- Télétravail

- IA et sécurité

- IA et santé

**Revue n°3 : Sélection du sujet**

*Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 28 janvier 2021

*Sujet discuté :* Choix du sujet.

IA/sécurité et IA/santé sont écartés. Arnaques en ligne ou télétravail à discuter avec le client.

**Revue n°2 : Résultats des sujets**

*Personnes présentes (fournisseur) :* - KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 26 janvier 2021

*Sujet discuté :* Choix de la manière dont on recherche les résultats pour chacun des sujets afin d’avoir une cohérence dans les résultats.

Répartition des sujets à traiter.

**Revue n°4 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personnes présentes (client) :*

- BAHSOUN Wahiba

- MOKADEM Riad

*Date :* 1er février 2021

*Sujet discuté :* Entretien sur le choix du sujet et la qualité des résultats obtenus.

Arnaques en ligne 27

**Revue n°5 : Schéma E/A**

*Personnes présentes (fournisseur) :* - ROSSI Valentine

- VEDIS Théo

- (AZOULAY Nathan)

- (MANSON Marianne)

*Date :* 2 février 2021

*Sujet discuté :* Première ébauche du schéma entité/association de la base de données.

**Revue n°7 : Scraping**

*Personnes présentes (fournisseur) :* - VEDIS Théo

- (ROSSI Valentine)

*Date :* 4 février 2021

*Sujet discuté :* Début du scraping sur le site cairn. Problème sur le temps de collecte via le programme : plus de 1000 pages à traiter sur le site.

**Revue n°9 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personne présente (client) :*

- MOKADEM Riad

*Date :* 8 février 2021

*Sujet discuté :* Modification du schéma entité/association. Choix des différents rôles du groupe.

**Revue n°6 : Scraping**

*Personnes présentes (fournisseur) :* - MANSON Marianne

- VEDIS Théo

- (AZOULAY Nathan)

- (KIRED Nour)

- (ROSSI Valentine)

*Date :* 4 février 2021

*Sujet discuté :* Début de l’extraction de données (création des premiers algorithmes).

**Revue n°8 : Scraping**

*Personnes présentes (fournisseur) :* - MANSON Marianne

- (AZOULAY Nathan)

*Date :* 6 février 2021

*Sujet discuté :* Début du scraping sur les sites erudit.org et zdnet.fr. Problèmes rencontrés sur OpenEdition : terme « arnaque » ressort beaucoup d’articles hors sujet avec les arnaques en ligne, le site a une structure de page différente pour chaque revue recensée et le même article est parfois sur différentes revues.

**Revue n°10 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personne présente (client) :*

- BAHSOUN Wahiba

*Date :* 8 février 2021

*Sujet discuté :* Squelette du rapport final du projet.

Arnaques en ligne 28

**Revue n°11 : SADT**

*Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 11 février 2021

*Sujet discuté :* Modification des SADT faits par rapport à ce que Mme Bahsoun nous a indiqué lors de la revue n°10.

**Revue n°13 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personne présente (client) :*

- BAHSOUN Wahiba

*Date :* 17 février 2021

*Sujet discuté :* Modification des SADT.

**Revue n°15 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personne présente (client) :*

- MOKADEM Riad

*Date :* 3 mars 2021

*Sujet discuté :* Validation du schéma E/A.

**Revue n°12 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personne présente (client) :*

- MOKADEM Riad

*Date :* 17 février 2021

*Sujet discuté :* Modification du schéma entité/association et passage en anglais des articles (nouvelle source approuvée par le client).

**Revue n°14 : Schéma E/A et récupération des données**

*Personnes présentes (fournisseur) :* - KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 1er mars 2021

*Sujet discuté :* Modification du schéma entité/association par rapport aux indications de M. Mokadem lors de la revue n°12 et début de la récupération des données.

**Revue n°16 : Base de données**

*Personnes présentes (fournisseur) :* - ROSSI Valentine

- VEDIS Théo

*Date :* 4 mars 2021

*Sujet discuté :* Création de la base de données et des procédures associées.

Arnaques en ligne 29

**Revue n°17 : Préparation des données** *Personnes présentes (fournisseur) :* - KIRED Nour

- VEDIS Théo

- (ROSSI Valentine)

*Date :* 8 mars 2021

*Sujet discuté :* Finalisation de la récupération et début de la préparation des données.

**Revue n°19 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personnes présentes (client) :*

- BAHSOUN Wahiba

- MOKADEM Riad

*Date :* 10 mars 2021

*Sujet discuté :* Modification des SADT.

**Revue n°21 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personnes présentes (client) :*

- BAHSOUN Wahiba

- MOKADEM Riad

*Date :* 15 mars 2021

*Sujet discuté :* Avancement du projet et modification des SADT.

**Revue n°18 : Stockage des données** *Personnes présentes (fournisseur) :* - ROSSI Valentine

- VEDIS Théo

*Date :* 9 mars 2021

*Sujet discuté :* Création des fonctions Python pour le stockage des données.

**Revue n°20 : Interrogation des données** *Personnes présentes (fournisseur) :* - KIRED Nour

- ROSSI Valentine

- VEDIS Théo

*Date :* 12 mars 2021

*Sujet discuté :* Brainstorming sur les vues de la base de données à créer.

**Revue n°22 : Interrogation des données** *Personnes présentes (fournisseur) :* - ROSSI Valentine

- VEDIS Théo

*Date :* 16 mars 2021

*Sujet discuté :* Choix et implémentation des vues en SQL.

Arnaques en ligne 30

**Revue n°23 : Réunion client/fournisseur** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Personnes présentes (client) :*

- BAHSOUN Wahiba

- MOKADEM Riad

*Date :* 17 mars 2021

*Sujet discuté :* Avancement du projet et retard sur la préparation des données.

**Revue n°25 : Visualisation des données** *Personnes présentes (fournisseur) :* - KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 19 mars 2021

*Sujet discuté :* Visualisation de données et continuation du PowerPoint.

**Revue n°24 : Réalisation des types d’arnaque**

*Personnes présentes (fournisseur) :* - MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 18 mars 2021

*Sujet discuté :* Préparation des données et début du PowerPoint.

**Revue n°26 : Finalisation du projet** *Personnes présentes (fournisseur) :* - AZOULAY Nathan

- KIRED Nour

- MANSON Marianne

- ROSSI Valentine

- VEDIS Théo

*Date :* 22 mars 2021

*Sujet discuté :* Rédaction du rapport et finalité du PowerPoint.

Arnaques en ligne 31

**Chapitre VII : Bilan du projet**

**I. Points négatifs**

Au cours de ce projet, comme tout travail en équipe ou seul, nous avons rencontré différents obstacles. Le premier a été le choix du sujet : nous avions eu quatre idées de sujet (arnaques en ligne, télétravail, IA/santé, robot militaire) que nous apprécions et que nous trouvions dans l’air du temps, il a donc fallu se décider et nous avions peur de faire le mauvais choix. Nous nous sommes donc résolus, en nous répartissant le travail, à tester différentes équations de recherche sur chacun des sujets afin de déterminer lequel nous donnerait un meilleur rapport qualité/quantité de données.

Notre choix s’étant finalement porté sur les arnaques en ligne, il nous a fallu regarder plus en détails les résultats de nos équations de recherche pour ce sujet. Nous nous sommes ainsi rendu compte que nous n’avions pas suffisamment de documents pertinents sur un nombre restreint de sources et qu’il nous fallait donc en avoir une trop grosse quantité. C’est pour cela que nous avons choisi de travailler exclusivement en anglais, via le site *springer* qui est ainsi notre seule et unique source. Toute cette démarche nous a fait perdre énormément de temps sur le début de la collecte des données.

De plus, le schéma entité/association a été remanié à plusieurs reprises afin de correspondre au mieux aux attentes du client. Celui-ci étant l’élément principal pour la création de la base de données, sa création en a été retardée, suite à une multitude de modifications.

Globalement, les points négatifs concernent la prise de retard par l’équipe au cours du projet. De ce fait, il ne nous a pas été possible de respecter le diagramme de Gantt prévisionnel, c’est pourquoi nous vous avons proposé un diagramme de Gantt réel qui, quant à lui, reflète vraiment les durées et dates des différentes tâches.

**II. Points positifs**

Tout au long de ce projet, nous avons appris une chose essentielle : l’organisation. Lors des travaux en groupe, c’est quelque chose qui nous avons travaillé tout au long de notre formation. Cependant, avec ce projet, nous avons été confrontés aux particularités de la gestion de projet en télétravail. Nous avons donc dû repenser notre logistique pour qu’elle s’adapte au mieux aux conditions sanitaires actuelles. Tout au long du projet, cela s’est très bien passé malgré la distance car nous avons su, dès le début, définir les outils que nous utiliserions ainsi que les créneaux auxquels nous ferions des revues. La communication a donc été facilité grâce, entre autre, aux logiciels zoom ou discord et aux cours de Dashboard ainsi qu’à Google Drive pour le partage de documents.

Les schémas SADT ont été modifiés maintes fois suites aux explications et remarques constructives de Mme Bahsoun, ce qui a engendré plusieurs changements au sein du rapport. En effet, nous voulions que ce dernier ait un plan logique vis-à-vis des schémas SADT, il fallait donc que le tout soit cohérent. Cela nous a permis de beaucoup apprendre.

Arnaques en ligne 32

De plus, nous avons eu la chance d’avoir un groupe polyvalent. En effet, certains sont plus à l’aise en programmation quand d’autres sont meilleurs en base de données. Il nous a donc été facile de nous partager les tâches en fonction des forces et faiblesses de chacun pour mettre toutes les chances de notre côté quant à la réussite de ce projet.

Ainsi, qu’il s’agisse d’un côté positif ou négatif du projet, le plus important est qu’à chacune des étapes nous avons eu l’occasion de mettre en pratique ce que nous avions vu en cours ou lors du projet inter-promo sur un cas concret et d’apprendre de nouvelles choses tout en acquérant de nouvelles connaissances et compétences.

Arnaques en ligne 33

**Annexes**

**1. Cahier des charges**

****Arnaques en ligne 34

Arnaques en ligne 35

Arnaques en ligne 36

Arnaques en ligne 37

Arnaques en ligne 38

Arnaques en ligne 39

Arnaques en ligne 40

**2. Charte qualité**

****Arnaques en ligne 41

Arnaques en ligne 42

Arnaques en ligne 43

Arnaques en ligne 44

Arnaques en ligne 45