



21 July 2019  
Alexandre ALLANI

Encadrant : Jean PRUVOST, consultant en datascience  
Conseiller d'étude : Romain BILLOT  
Responsable filière : Cécile BOTHOREL

# Rapport de stage de fin d'étude

Consultant en datascience chez Sia Partners



**IMT Atlantique**  
Bretagne-Pays de la Loire  
École Mines-Télécom

**sia**partners

# 1 Résumé

## 2 Summary

## Contents

<b>1</b>	<b>Résumé</b>	<b>2</b>
<b>2</b>	<b>Summary</b>	<b>3</b>
<b>3</b>	<b>Introduction</b>	<b>5</b>
<b>4</b>	<b>Présentation de l'entreprise</b>	<b>6</b>
<b>5</b>	<b>Mission : webscrapping pour Cleep (1 mois)</b>	<b>7</b>
5.1	Présentation de Cleep . . . . .	7
5.2	Objectifs et problématique . . . . .	8
5.3	Travail effectué . . . . .	8
5.4	Vision critique et apport personnel . . . . .	8
<b>6</b>	<b>Mission : Moteur de recommandation pour Cleep (2.5 mois)</b>	<b>9</b>
6.1	Contexte . . . . .	9
6.2	Objectifs et problématique . . . . .	9
6.3	Travail effectué . . . . .	9
6.4	Vision critique et apport personnel . . . . .	9
<b>7</b>	<b>Mission : Aide à l'amélioration d'une plateforme de déploiement pour Sia Partners</b>	<b>10</b>
7.1	Présentation du sialab et de la plateforme . . . . .	10
7.2	Objectifs et problématique . . . . .	10
7.3	Travail effectué . . . . .	10
7.4	Vision critique et apport personnel . . . . .	10
<b>8</b>	<b>Conclusion</b>	<b>11</b>
<b>9</b>	<b>Annexes</b>	<b>12</b>
<b>10</b>	<b>Glossaire</b>	<b>13</b>

## 3 Introduction

## 4 Présentation de l'entreprise

## 5 Mission : webscrapping pour Cleep (1 mois)

Cette partie est consacrée à la première mission que j'ai effectué sur le sujet du webscrapping pour Cleep. Le webscrapping est la collecte de données/informations sur des sites internet via un ou plusieurs scripts. Le but sera alors principalement de "piocher" dans le code source du site internet à webscrapper, ou via les requêtes effectuées par le serveur, les données que l'on souhaite récupérer.

Cette mission mêle à la fois des problématiques de récupérations de données, mais aussi d'automatisation de webscrapping via certains algorithmes que j'ai développé.

### 5.1 Présentation de Cleep

Cleep est une start-up créée en 2017, développant une application éponyme sur le sujet de l'achat en ligne. L'application permet de sauvegarder un produit repéré sur un site marchand quelconque (que ce soit un site généraliste comme Amazon ou un site spécialisé comme Asos). Cette application mobile est disponible sur mobile (iOS & Android) mais aussi via un site Internet [1]. Le principe de l'application est de pouvoir enregistrer n'importe quel produit pour ensuite le retrouver dans l'application. Il est alors possible de consulter via cette dernière le prix, la description, le nom, le site d'origine ainsi que les images reliés à ce produit. Le but de l'application est donc de faciliter l'achat de produits en ligne en ne passant que par une seule et unique plateforme. Par ailleurs, il est possible de créer une liste de "Cleep" (ie de produit) et de partager ces listes. Les listes publiques peuvent être trouvées via le moteur de recherche intégré.



Figure 1: Logo de Cleep

Sia Partners a investi 400000€ dans Cleep fin 2018, et travaille en collaboration sur différents sujets, notamment en matière de datascience. En plus du côté simplification du shopping en ligne, Cleep avec ses listes et son moteur de recherche permet de suivre des listes plus ou moins influentes. Ainsi l'application permet de se tenir au courant des modes actuelles et de celles à venir. Cela explique l'investissement de Sia Partners.

Vocabulaire spécifique à Cleep :

- Cleep : Produit en ligne, enregistré dans la base de donnée. Un cleep est composé d'un prix, des images associés au produit, d'une description, du nom du produit ainsi de du site internet d'origine.
- Liste de Cleep : Liste de Cleep, pouvant être partagé entre plusieurs utilisateurs. Chaque cleep est enregistré dans une liste.

J'ai travaillé pendant près de 3 mois pour Cleep. Je communiquais principalement avec Damien Meurisse côté Cleep et Paul Saffer côté Sia Partners. J'ai travaillé sur deux sujets :

- Webscrapping : Acquisition et traitement des données. Cela est fondamental pour le fonctionnement de Cleep
- Moteur de recommandation. Avant mon arrivé, aucun moteur de recommandation n'existait. Cette fonctionnalité était de plus en plus nécessaire afin de trouver des listes et des clips autrement que par une recherche via le moteur interne à Cleep.

## **5.2 Objectifs et problématique**

## **5.3 Travail effectué**

## **5.4 Vision critique et apport personnel**



## **6 Mission : Moteur de recommandation pour Cleep (2.5 mois)**

### **6.1 Contexte**

### **6.2 Objectifs et problématique**

### **6.3 Travail effectué**

### **6.4 Vision critique et apport personnel**

## **7 Mission : Aide à l'amélioration d'une plateforme de déploiement pour Sia Partners**

**7.1 Présentation du sialab et de la plateforme**

**7.2 Objectifs et problématique**

**7.3 Travail effectué**

**7.4 Vision critique et apport personnel**

## 8 Conclusion

## 9 Annexes

## 10 Glossaire

### References

[1] Site de Cleep : <https://app.cleep.io/>



**IMT Atlantique**  
Bretagne-Pays de la Loire  
École Mines-Télécom

IMT Atlantique Bretagne - Pays de la Loire - [www.imt-atlantique.fr](http://www.imt-atlantique.fr)

Campus de Brest  
Technopôle Brest-Iroise  
CS 83818  
29238 Brest Cedex 03  
T +33 (0)2 29 00 11 11  
F +33 (0)2 29 00 10 00

Campus de Nantes  
4, rue Alfred Kastler - La Chantrerie  
CS 20722  
44307 Nantes Cedex 03  
T +33 (0)2 51 85 81 00  
F +33 (0)2 51 85 81 99

Campus de Rennes  
2, rue de la Châtaigneraie  
CS 17607  
35576 Cesson Sévigné Cedex  
T +33 (0)2 99 12 70 00  
F +33 (0)2 99 12 70 08