



datagram

PFE BOOK

2021/2022

Sommaire

Présentation de Datagram

- Qui sommes nous ? p.3-4

Nos offres de PFEs

- Automatisation du processus de recrutement p.7
- Refonte du site web datagram.ai p.8
- Informatique décisionnelle pour le sourcing de candidats p.9
- Appariement Intelligent d'images d'articles commerciales p.10
- Collecte et analyse des logs en temps réel p.11
- Scrapping de données basé sur Machine Learning p.12

Comment Postuler

Présentation de Datagram

Qui sommes-nous?

Nos Valeurs

Confiance

Travailler du bureau, du domicile ou même à bord d'une croisière, on vous fait confiance que vous allez atteindre vos objectifs à temps et de la meilleur qualité possible.

Collaboration

Notre approche de communication est très directe, où les seniors de l'équipe sont à trois clics et toute proposition est respectée et discutée. Cette collaboration est aussi ouverte à l'écosystème externe en vue de le supporter

Ambition

Nous prenons au sérieux les ambitions personnelles de nos membres, et on essaie de notre mieux de supporter la montée en compétence et la santé mentale de notre équipe. De plus, l'effort supplémentaire est récompensé et encouragé

Quelques Chiffres

25

Effectif humain 2021

12

Clients de qualité:
Michelin, Danone,
Ferrero, Heineken...

2

Bureaux : France
et Tunisie

1

Million d'Euros
en chiffre d'affaire
en 2020

Qui sommes-nous?

Business

Tech



Notre Ecosystème

AGENCES/CONSEILS

ADTECHS/RÉGIES



*liste non exhaustive

Nos Projets PFE

Sujet 1:

Automatisation du processus de recrutement

Type	Category	Besoin en stagiaires
Web + Data	Back End + Data Analytics	1

Description

Le recrutement des talents est un process long et délicat et à Datagram, on veut donner à chaque candidat les meilleurs conditions et feedbacks. C'est pour cela qu'on souhaite automatiser notre process de recrutement actuel.

Ce projet consiste de développer et déployer un système qui:

- Automatise le flux des candidats à travers le process (Event Based Programming)
- Extrait, stocke et analyse les cvs des candidats. (Regex, NLP et/ou OCR)
- Envoie automatiquement des tests et des mails personnalisés (EVP).
- Analyse les résultats des tests, des candidats et du flux pour générer des recommandations..

Technologies

Python, Flask, Google APIs,

Sujet 2:

Refonte du site web Datagram.ai

Type

Web + UI/UX

Category

Front End + Design

Besoin en stagiaires

1-2

Description

Un site web est la vitrine principale d'une société. Avec la croissance de Datagram, il est primordial d'avoir un site mis à jour, trendy et visuellement attirant. C'est pour cela qu'on souhaite explorer des pistes d'améliorations et d'innovations pour notre site web datagram.ai

Ce projet consiste de faire la recherche, la conception et l'intégration d'un nouveau design pour le site web datagram.ai qui soit à la page par rapport aux tendances de design Tech, Startup et Data..

Technologies

Suite Adobe, Wordpress, Figma

Sujet 3:

Informatique décisionnelle pour le sourcing de candidats

Type

Web + Data

Category

Data Analytics

Besoin en stagiaires

1

Description

Le sourcing des talents est une étape fondamentale de tout process de recrutement. Le changement dynamique de l'environnement académique rend difficile de cibler les bons établissements et organisations pour des postes spécifiques. C'est pour cela qu'on souhaite créer un process qui collecte des données par rapport au background des

Ce projet consiste de développer un programme qui fait:

- La collecte des données de talents tunisiens et leur background
- L'agrégation et l'analyse des données
- Le reporting et la suggestion d'établissements ou d'organisation selon le besoin

La personne sélectionnée va aussi transformer la solution en un package réutilisable.

Technologies

Python

Sujet 4:

Appariement Intelligent d'images d'articles commerciales

Type

Data

Category

Data Science

Besoin en stagiaires

2

Description

«L'utilisation de l'intelligence artificielle (IA) a touché, pratiquement, tous les axes de développement industriel y compris l'e-commerce. L'une des applications qui influencent de plus en plus l'e-commerce et l'e-retail est l'identification des différents attributs qui constituent un produit mis en vente en ligne. En fait, ces attributs sont des fragments de données qui peuvent être à la fois textuels (titre, description...) , visuels (images, vidéos...) et parfois numériques. Du coup, l'identification d'un produit revient à traiter ces fragments de données afin d'encoder toute l'information récolté sur l'article en question. Les cas d'utilisation de cet encodage sont divers, parmi ces applications on cite la correspondance des articles similaires (product matching) et la catégorisation des produits.

Tâches proposées:

- Suggester des approches machine learning permettant de faire cet encodage
- Reconnaître et extraire la data à utiliser
- Augmenter la data si nécessaire
- Réaliser un POC et faire une étude comparative
- Implémenter une solution finale sous Python

Mots clés : Machine/Deep Learning, NLP, Computer Vision, Unsupervised Learning, Data Augmentation,

Technologies

Python, Tensorflow ou Pytorch, Numpy, Spacy/NLTK, OpenCV, Pillow

Sujet 5:

Collecte et analyse des logs en temps réel

Type

Cloud

Category

Cloud / Devops

Besoin en stagiaires

1

Description

La collecte, l'analyse et la configuration des systèmes est une tâche qui demande beaucoup de patience et de réactivité. Dans un système temps-réel, l'apparition d'un problème dans l'absence d'un humain pour la régler peut créer un effet boule de neige et arrêter le fonctionnement de toute la société.

C'est pour cela qu'on souhaite automatiser un de nos systèmes temps réel pour éviter ce scénario.

Ce projet consiste de faire la conception, développement et déploiement d'un système qui permet de :

- Recueillir les logs des crawlers Datagram.
- Analyser les logs en temps réel afin de détecter les bans IP.
- Permettre aux crawlers d'ajuster leurs configurations afin de contourner les bans.

Technologies

Python, Kafka, Kafka Streams, Postgres, Redis, ZeroMQ, Clickhouse, Docker, Ansible, FastAPI

Sujet 6:

Scrapping de données basé sur Machine Learning

Type

Data

Category

Data Science

Besoin en stagiaires

1

Description

L'extraction d'informations à partir de pages Web est une tâche récurrente dans la vie quotidienne des ingénieurs de données. Ce qui rend cette tâche fastidieuse, c'est les modifications inattendues apportées quotidiennement à ces pages Web. Ceci est parce que l'extraction des données est majoritairement basé sur le texte (code source). C'est pour cela qu'on souhaite développer un système d'extraction de données qui est basé sur l'imagerie plutôt.

Ce projet consiste de faire:

- La conception, le développement et le déploiement d'un système qui fait la collecte de données de produits à partir de pages Web.
- L'analyse comparative des différentes méthodes utilisées pour réaliser l'extraction d'informations.
- Le développement d'un pipeline capable de former un modèle de reconnaissance d'entité capable d'effectuer l'extraction d'informations de produits à partir de pages Web.

Mots clés:

Exploration de données, traitement de données, extraction d'informations basée sur le contenu, apprentissage automatique, traitement du langage naturel, reconnaissance d'entités nommées.

Technologies

Python, NLTK, Spacy, Tensorflow, Keras.

Comment postuler

Comment postuler



<https://www.datagram.ai/rejoignez-nous/>



datagram



Datagram, WeWork,
64-66 rue des archives,
75003 Paris



datagram



Datagram.ai



datagram



+33 9 72 54 79 62



datagram



hello@datagram.ai