|  |  |
| --- | --- |
|  |  |



**Département Informatique**

Diplôme préparé : Informatique

N° Jury 3

Développement d'un logiciel d'aide au traitement de données d'acquisition expérimentale en utilisant l'intelligence artificielle.

Mathieu LACROIX TP4A2

## Diplôme préparé : Informatique

N° Jury 3

Développement d'un logiciel d'aide au traitement de données d'acquisition expérimentale en utilisant l'intelligence artificielle.

Mathieu LACROIX TP4A2

***Tuteur enseignant*** ***Responsable en entreprise***

*Xavier LACOURT Gaston EXIL*

Année universitaire 2021-2022

Période de stage 19/04/2022 – 01/07/2022

Mémoire remis le 17/06/2022

Année universitaire 2021-2022

Période de stage 19/04/2022 – 01/07/2022

Mémoire remis le 17/06/2022

Année universitaire 2021-2022

Période de stage 19/04/2022 – 01/07/2022

Mémoire remis le 17/06/2022

Année universitaire 2021-2022

Période de stage 19/04/2022 – 01/07/2022

Mémoire remis le 17/06/2022

# Sommaire

[Diplôme préparé : Informatique 1](file:///C:\Users\ml270906\Documents\webSite\elem\rapport.docx#_Toc105585701)

[Sommaire 2](#_Toc105585702)

[Introduction 3](#_Toc105585703)

[Remerciement 3](#_Toc105585704)

[Résumé : 4](#_Toc105585705)

[Problématique et objectifs de stage 4](#_Toc105585706)

[Présentation de l’entreprise 5](#_Toc105585707)

[LE CEA 5](#_Toc105585708)

[LE CNRS 7](#_Toc105585709)

[LE LABORATOIRE LEON BRILLOUIN 9](#_Toc105585710)

[Le secteur activité 10](#_Toc105585711)

[Mission 11](#_Toc105585712)

[Mise en situation du projet 11](#_Toc105585713)

[Phase conceptuelle du projet 12](#_Toc105585714)

[Fonctionnalités 12](#_Toc105585715)

[Contraintes 12](#_Toc105585716)

[Déroulement du projet 13](#_Toc105585717)

[Conception application web 14](#_Toc105585718)

[Structure des fichiers .32 14](#_Toc105585719)

[Stockage des fichiers 14](#_Toc105585720)

[Interface de l’application Web 15](#_Toc105585721)

[Conception du logiciel de traitement des données 19](#_Toc105585722)

[Récupération des données 19](#_Toc105585723)

[Mise en forme des données 19](#_Toc105585724)

[Augmentation de données 19](#_Toc105585725)

[Simulation des données 19](#_Toc105585726)

[Conception du model 19](#_Toc105585727)

[Elaboration du jeu de test 19](#_Toc105585728)

[Spécification des logiciels et outils 20](#_Toc105585729)

[Microsoft Visual Studio Code 20](#_Toc105585730)

[Git 20](#_Toc105585731)

[Db Brower 20](#_Toc105585732)

[Streamlit API 21](#_Toc105585733)

[Google Colab 21](#_Toc105585734)

[Difficulté rencontré 21](#_Toc105585735)

[Bilan 21](#_Toc105585736)

[Lexique 21](#_Toc105585737)

[Table des illustrations 21](#_Toc105585738)

[Annexes 21](#_Toc105585739)

# Introduction

## Remerciement

Je tiens à remercier Mr Gaston EXIL, mon maître de stage qui m’a formé et accompagné tout au long de cette expérience professionnelle avec beaucoup de patience et de pédagogie.

Je tiens également à remercier la direction pour m’avoir donné l’opportunité d’intégrer le CEA et le laboratoire Léon Brillouin.

Aussi, je tiens également à remercier pour m’avoir fait mon dossier Aurore VERDIER (Responsable Ressource Humaine/ Secrétaire) qui m’a permis de rentrer au CEA.

Pour finir je remercie les employés du pôle administratif et financier du laboratoire pour les conseils qu’ils ont pu me donner au cours de ces deux mois.

## Résumé :

Du 19 avril 2022 au 1er Juillet 2022, j’ai effectué mon stage de fin d’études au Commissariat à l’Energie Atomique et aux Energies Alternatives sur le site de Saclay dans le laboratoire Léon Brillouin, en France. Pendant ce stage, au Service Informatique (SI), je me suis intéressée au développement d'un logiciel d'aide au traitement de données d'acquisition expérimentale en utilisant l'intelligence artificielle.

Plus largement, ce stage était l’occasion pour moi de mettre en application mes compétences de développeur acquises durant mes deux années d’études. Au-delà de l’enrichissement de mes connaissances dans le développement informatique, cette période m’a permis de me rendre compte du travail des informaticiens dans le monde professionnel.

## Problématique et objectifs de stage

Ce stage a donc été une opportunité pour moi de percevoir comment une entreprise dans un secteur de la recherche. [décrire ici les caractéristiques du secteur : concurrence, évolution, historique, acteurs… et quelle stratégie l’entreprise a choisie dans ce secteur. Ainsi que l’apport du département et du poste occupé dans cette stratégie…]

L’élaboration de ce rapport a pour principale source les différents enseignements tirés de la pratique journalière des tâches auxquelles j’étais affecté. Enfin, les nombreux entretiens que j’ai pu avoir avec les employés des différents services de la société m’ont permis de donner une cohérence à ce rapport.

# Présentation de l’entreprise

## LE CEA

Le Commissariat à l’énergie atomique et aux énergies alternatives (CEA) est un organisme public de recherche à caractère scientifique, technique et industriel (EPIC). Acteur majeur de la recherche, du développement et de l'innovation, le CEA intervient dans quatre domaines : la défense et la sécurité, les énergies bas carbone (nucléaire et renouvelables), la recherche technologique pour l'industrie et la recherche fondamentale (sciences de la matière et sciences de la vie). S'appuyant sur une capacité d'expertise reconnue, le CEA participe à la mise en place de projets de collaboration avec de nombreux partenaires académiques et industriels. ​​​

Depuis près de 70 ans, le CEA conduit, pour le compte de l’Etat, des programmes de recherche, visant à accroître la connaissance scientifique et à contribuer à l’innovation et au transfert technologique dans un nombre limité de domaines. En 2010, le CEA est devenu Commissariat à l’énergie atomique et aux énergies alternatives Le CEA met au service des enjeux sociétaux majeurs ses programmes de recherche scientifique et technologique, et sa capacité de formation par la recherche :

* Soutien à l’industrie par l’innovation
* Transition énergétique vers un bouquet décarboné
* Santé
* Défense et sécurité Dans une organisation fondée sur des valeurs partagées : sens de l’intérêt général, acceptation des responsabilités, engagement, exigence, acceptation de la complexité, solidarité

Le CEA c’est 20 181 salariés dont 1 233 doctorants et 176 post-doctorants, un budget de 5 milliards Md€. De plus, un grand nombre de brevet ont été déposé par des personnes travaillant au sein de la structure, en 2020 l’INPI publie son classement annuel dans lequel le CEA maintient sa place de premier organisme de recherche français. Avec 646 demandes de brevets publiées. Nationalement, le CEA se classe 4ème, derrière PSA, Safran et Valeo.

D’autre chiffre sur la structure :

* Top 100 de l’innovation mondiale depuis 4ans
* Plus de 5022 publications par an dans des revues à comité de lecture
* 1 er organisme de recherche public mondial au classement des déposants de brevets PCT
* 39 unités mixtes de recherche en cotutelle et assimilées et 4 unités mixtes de service ;



## LE CNRS

Le Centre national de la recherche scientifique, plus communément appelé CNRS, est une institution de recherche public française. Fondé le 19 octobre 1939 par un décret-loi visant à « coordonner l’activité des laboratoires en vue de tirer un rendement plus élevé de la recherche scientifique ».

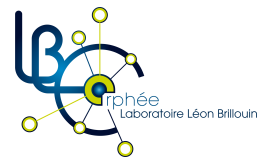
Sa mission consiste à faire progresser la connaissance et être utile à la société en utilisant la recherche, c’est là le rôle du CNRS. Sa mission est divisée en trois branches :

* La recherche scientifique
* Valoriser les résultats
* Partager les connaissances
* Former par la recherche
* Contribuer à la politique scientifique

Pour son activité scientifique le CNRS est divisé en dix instituts nationaux spécialisés dans un domaine précis :

* Les sciences humaines et sociales
* La biologie
* La chimie
* L’écologie et l’environnement
* Les sciences de l’information
* Les sciences de l’ingénierie et des systèmes
* Les mathématiques
* La physique nucléaire et des particules
* Les sciences de l’univers.

Pour relever les grands défis présents et à venir, ses scientifiques explorent le vivant, la matière, l’Univers et le fonctionnement des sociétés humaines. Internationalement reconnu pour l’excellence de ses travaux scientifiques, le CNRS est une référence aussi bien dans l’univers de la recherche et développement que pour le grand public.



## LE LABORATOIRE LEON BRILLOUIN

Le Laboratoire Léon Brillouin (LLB) est un laboratoire de recherche mixte français soutenu conjointement par le Commissariat à l'Energie Atomique (CEA) et le Centre National de la Recherche Scientifique (CNRS). Son objectif est de mener des recherches sur la structure et la dynamique de la matière condensée en utilisant les faisceaux de neutrons fournis par le réacteur Orphée. Il est situé dans le centre de recherche CEA/Saclay.

Le large domaine d'application et la maturité de la diffusion des neutrons en ont fait une technique moderne indispensable à toute recherche au niveau microscopique (structure et dynamique) dans des domaines aussi variés que la physique de la matière condensée, la chimie, la biologie et la science des matériaux. En particulier, la diffusion des neutrons est une des techniques majeures pour l'étude de la matière molle, du magnétisme, de la dynamique et des liquides. Dans les problèmes industriels, on profite de la pénétration profonde des neutrons dans la matière pour des études non-destructives.

Le LLB mène son propre programme de recherche scientifique et collabore étroitement avec de nombreux scientifiques provenant principalement de laboratoires de recherche fondamentale, mais aussi de la recherche appliquée et de l'industrie. Cela conduit à un rythme annuel de 180 publications dans des revues scientifiques à comité de lecture, et plus de 150 communications dans des conférences et réunions.

### Le secteur activité

Le laboratoire étant une unité de recherche mixte à plusieurs fonctions :

* Conduire une recherche de haut niveau sur ses propres programmes en s’appuyant sur ses installations sur le site de Saclay, de Grenoble ou d’autre sites
* Soutenir la formation à la diffusion neutronique et préparer une nouvelle génération d’utilisateurs aux méthodes les plus récentes de diffusion des neutrons
* Promouvoir l’utilisation de la diffusion et de la spectroscopie neutronique pour la recherche fondamentale et l’industrie et mettre en place des collaborations et/ou des partenariats avec des sociétés industrielles et commerciales dans le cadre de l’accès aux faisceaux de neutrons du LLB
* Assurer après l’arrêt du réacteur Orphée, le démontage, le recyclage et l’organisation de l’optimisation de la réutilisation au mieux des intérêts de la communauté scientifique française de tout ou partie des instruments sur d'autres sources de neutrons en France et à l’étranger le cas échéant
* Mener une prospective pour une stratégie à long terme visant à garantir l’excellence française en diffusion des neutrons et proposer dans ce cadre, le développement d’une source compacte de neutrons destinée à redonner à la France au moins une source nationale
* Assurer la coordination opérationnelle de la contribution française pour la construction et la mise en service d’instruments sur la nouvelle source européenne de neutrons ESS (European Spallation Source, Lund, Suède) dans le cadre des accords entre la France et l’ERIC ESS

# Mission

## Mise en situation du projet

Comme vu précédemment le pôle informatique du laboratoire et les chercheurs travaillent ensemble. Mon travail au sein la structure a pour but de les aider dans leur expérimentation des neutrons. Lors de celle-ci les chercheurs projettent plein de neutrons sur un échantillon choisi afin d’identifier ses réactions lorsqu’il reçoit des neutrons. Pour cela l’échantillon on lui donne plein de paramètres afin de le tester sous toute ses formes que ça soit sa taille, sa température son orientation, inclinaison, …

Ensuite les neutrons traversent l’échantillon et sont réceptionnés sur un capteur, qui celui créer un fichier de type .32 qui nous sera donné, c’est à partir de là que notre travail commence. Ces fichiers contiennent tous les informations importantes pour les chercheurs : les paramètres de l’échantillons et les résultats de leur expérimentation.

De là, nous allons récupérer toutes les données et les afficher sous forme de tableau pour qu’ils puissent interpréter plus facilement les résultats obtenus. Après avoir obtenu plein de fichier nous allons les stocker dans une base de donnée. La suite du projet est de la classification, nous allons utiliser le machine Learning et plus précisément la reconnaissance d’image sur les graphiques obtenus, dans le but de reconnaître les paramètres et l’échantillon utilisé lors de l’expérimentation, de pourvoir les trier et de récupérer que les plus significatifs. Ce projet a pour but d’aider les chercheurs, en terme de temps et d’efficacité.

## Phase conceptuelle du projet

Avant toute phase de développement il faut penser à comment va se dérouler le projet, quelles sont les objectifs et les contraintes afin d’avoir une ligne directrice à suivre, puis nous mettrons en place une méthode de déroulement du projet.

### Fonctionnalités

Le projet que j’ai développé devra avoir les fonctionnalités suivantes :

* Stocker les données dans une base de donnée
* Visualiser des données issues d'acquisition neutroniques
* Créer une application web interactive
* Extraire les données des fichier .32
* Tester des modèles d'apprentissage automatique sur un jeu de donnée

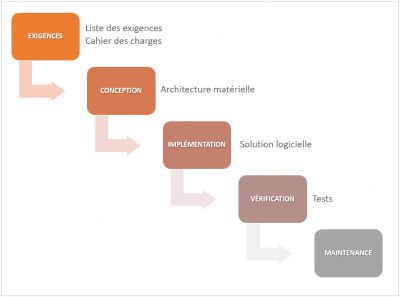
### Contraintes

Le projet aura également des contraintes à suivre, les suivantes :

* Framework Streamlit pour l'interface utilisateur
* Framework Plotly pour la représentation graphique
* Interface simple d’utilisation pour les chercheurs
* Le langage de programmation utilisé sera le Python

### Déroulement du projet

Sous la supervision de M. EXIL, j’ai réalisé le projet le projet en autonomie, en suivant les consignes que l’on m’a donné, en suivant une méthodologie agile. Le projet sera divisé en sous projet, où dans chaque sous-projet il y aura une phase de recherche, de développement et pour finir une phase de test et si besoin de correction. Lorsqu’une phase est finie on passe à la suivant et ainsi de suite.



## Conception application web

### Structure des fichiers .32

La première étape de cette phase a pour but de comprendre et de savoir la structure du fichier, pour cela mon maître de stage ma donnée un document expliquant sa structure. Après avoir lu et compris ce document il fallait extraire les données qui m’intéressait afin que je puisse les manipuler. De ce fichier il y avait une suite de paramètres avec des caractères en code ASCII. La seconde partie du fichier est un tableau d'intensité XY au format binaire, d'entiers non signés codés sur 4 octets. La dernière partie du fichier est un bloc de descriptif des paramètres d'acquisition en ASCII.

### Stockage des fichiers

Afin de pouvoir manipuler au mieux les fichiers par la suite et les afficher sur l’interface web, on va stocker les données qui nous intéresse. La base de donnée est mise en place grâce à l’application DB Browser, sous SQLite, elle contient deux tables :

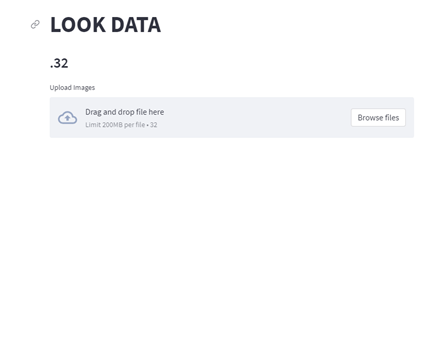
* La table « Fichier » pour stocker toutes les informations des fichiers (ID, nom, Checksum, …) et tous les paramètres que l’on a extraits auparavant.
* La table « Checksum » cette va servir à stocker le Checksum ce chaque fichier.

Le Checksum est une somme de contrôle, une courte séquence de données numériques calculée à partir d'un bloc de données plus important qui servira à nous éviter les doublons avant chaque enregistrement dans la base de données, il y aura une vérification du checksum du nouveau fichier avec tous ceux déjà enregistrer.

PHOTO de la bd

### Interface de l’application Web

Les chercheurs n’étant pas des experts en informatique, il fallait qu’on leur fournisse un livrable facile d’utilisation tout en étant assez précis pour qu’ils puissent faire des analyses de ses résultats. De ce fait l’interface a été créé de manière simple, avec tous les outils dont ils ont besoin tout rassembler en une même application. Le site va être écrit en python avec l’API Streamlit qui nous permet d’intégrer d’autres bibliothèques python, utiles pour les chercheurs, à l’application web.

La première partie de l’application web est sert à charger un fichier .32 sur le site, avec l’API il y avait des fonctions à notre disponibilité afin de charger un document dans la mémoire vive du site afin qu’il puisse être utilisé par la suite.

Une fois charger ce fichier est enregistrer dans la base de donnée comme expliqué précédemment, puis avec l’aide de requetés SQL sous python, on extrait les données que l’on veut :

* Le nom du fichier
* Le chemin du fichier
* Le type du fichier
* Les paramètres configuré pour la manipulation
* Le tableau d’intensité
* Le bloc descriptif des données d’acquisition

Tous les paramètres de configuration sont mis dans un tableau de la bibliothèque Pandas, ce tableau à deux colonne :

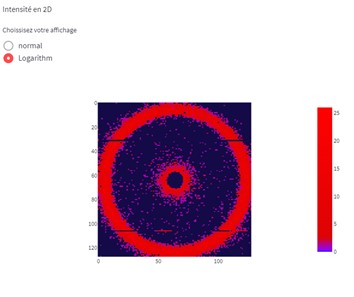
* Le nom du paramètre
* La valeur du paramètre

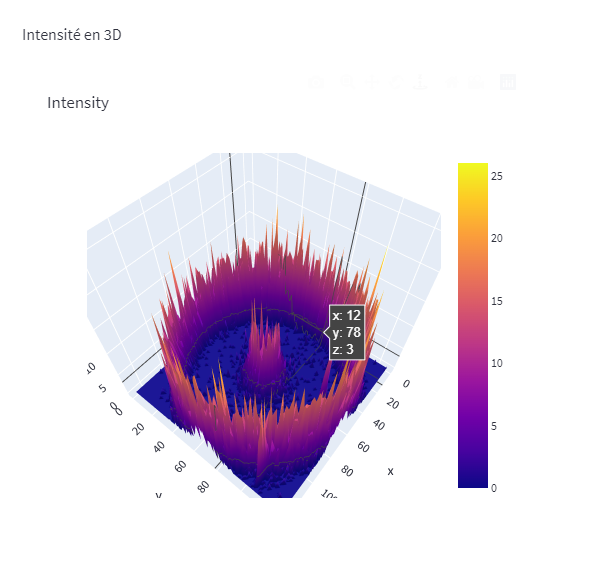
A côté de ce tableau il y aura un deuxième tableau qui lui va contenir les valeurs des intensités de l’acquisition, ce tableau d’intensité au début un tableau 256x256 avec comme paramètre XY, celui que l’on propose n’aura plus qu’un paramètre avec 16384 lignes. Afin de remplir les valeurs des intensités qui de base était en octet que l’on a par la suite convertit en hexadécimal pour des problèmes de nombre de caractère. Il a fallu donc les convertir de nouveau en entier afin que cela soit plus compréhensible pour les chercheurs.

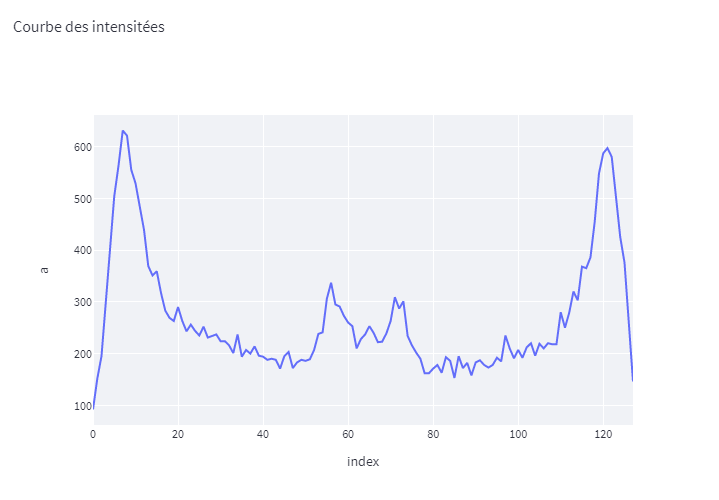


De ce tableau, nous avons pu créer des graphiques grâce à la bibliothèque Plotly, qui sont plus facile à visualiser :

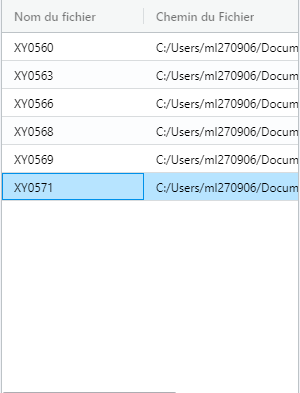
* Le premier est un graphique en 2D, le tableau est remodeler en 128 par 128, vous avez une vue du dessus du graphique.



* Le deuxième graphique est une vue en 3D du graphique précédent,
* Le dernier graphique est une vue de côté en 2D, du graphique précédent coupé à la verticale,



La deuxième partie de l’application sert à charger un dossier de fichier .32, celui-ci fonctionne comme l’étape précédente mais s’applique sur tous les fichiers du dossier. Afin de pourvoir tous les voir sans pour autant surcharger la page et tous les afficher, il y a sur le côté un tableau composé de tous les fichiers présents dans la base de donnée. Lorsque l’on appuie sur une des cellules du tableau la page est rechargée en en complètent les tableaux avec les données du fichier sélectionné.



## Conception du logiciel de traitement des données

Dans cette seconde partie, je vais vous parler et présenter le côté intelligence artificiel du projet, comme dit précédemment, le flux de données que récupère les chercheurs durant une acquisition peut être de l’ordre de millier. Aujourd’hui c’est manuellement que l’on tri chaque image d’acquisition, c’est un travail long, fastidieux et répétitif. D’où l’utilité d’un algorithme de recherche et de classification.

L’intelligence artificiel est utilisé dans plein de domaine, celui qui va nous intéresser est la reconnaissance d’image par le biais de machine Learning pour faire de la classification de donnée.

Le but de cette outil va être de reconnaître les images d’acquisition de structure de des neutrons. Dans le but de comprendre comment fonctionne la reconnaissance d’image on va partir d’un cas plus facilement compréhensible.

Pour cela nous allons vouloir reconnaître un oiseau d’un autre, entre un épervier, un aigle royal, une chouette effraie et un pygargue à tête blanche.

### Récupération des données

Dans un premier il nous faut récupérer des données, c’est la chose la plus importantes du machine Learning, les données, plus nous en possédons au mieux notre outil fonctionnera. Pour cela il existe des Dataset d’image libre de droit disponible sur internet. Il existe une méthode plus fastidieuse, c’est d’aller récupérer les images nous-mêmes sur internet ou encore de faire les photos.

Chaque donnée (en l’occurrence ici des images) possède une étiquette qui est son nom et une classe, c’est le dossier dans lequel il est présent. Par exemple pour nos oiseaux nos quatre classes sont :

* Chouette effraie
* Pyguargue à tête blanche
* Epervier
* Aigle royale

C’est quatre classes, seront après remplacé pour notre projet principal par les classes sur la diffusion des neutrons : verre de glass, sphère, cylindre, ellipsoïde

### Mise en forme des données

Ensuite on va mettre en forme toute ses données pour les rendre exploitables, il faut qu’elles ont toutes la même taille dans notre cas 128\*128. Il faut aussi vérifier que le jeu de donnée ne possède pas de doublon.

### Augmentation de données

Comme expliquer juste avant, les données sont indispensables dans notre outil, pour cela nous allons augmenter notre de jeu de donnée, grâce à l’application de filtre d’image nous allons là modifier légèrement afin d’avoir plus d’image. Les filtres que l’on peut utiliser sont multiples, pour autant nous ne voulons pas non plus déformer de trop les images car cela pourrait fausser notre model.

Nous allons donc appliquer les filtres suivant à toute nos images :

* -Une symétrie verticale et horizontal
* - le floutage, « blur » en anglais
* - flip vertical
* - rotation de 45°
* - miroir + rotation 45
* - Décalage de 10 pixels vers la droite
* - Décalage de 10 pixels vers la gauche
* - miroir + Décalage 10 pixel droite
* - miroir + décalage 10 pixel gauche
* - Décalage de 20 pixels vers la droite
* - Décalage de 20 pixels vers la gauche
* - miroir + Décalage 20 pixel droite
* - miroir + décalage 20 pixel gauche

Toutes ses petites modifications vont servir à enrichir notre jeu de donnée.

### Conception du model

Pour concevoir un modèle on va tout d’abord, définir quel pourcentage du jeu de donnée va servir entrainer le model et l’autre partie servira à la validation, pour tester le model sur de la reconnaissance, cette étape à pour but de ne pas utiliser les mêmes images pour entraîner et pour tester. C’est avec les librairies TensorFlow, un outil d’apprentissage automatique de google et Keras, une librairie open source, outils utiliser pour dans le Machine Learning.

Dans un second, on va paramétrer le jeu de donnée qui va servir d’entraînement :

* On va lui donner un nom de dossier
* Le pourcentage défini précédemment, d’image à prendre dans le dossier
* Le label ‘training’ pour spécifier que c’est le jeu d’entrainement.
* On active le mélange des images pour rendre l’entrainement aléatoire. Pour qu’il ne sélectionne pas les mêmes images à chaque fois.
* La taille des images, pour notre cas ça sera 128x128 pixels

Ce sont les paramètres principaux pour le jeu de donnée d’entraînement, pour celui de validation les paramètres sont les mêmes, la seule différence, le label sera cette fois ci : ‘ validation’

Après avoir paramétrer les jeux de données, on va se pencher sur le model qu’il faut lui aussi paramétrer, toujours en utilisant les librairies précédemment citées.

Comme paramètres on va mettre :

* Le premier paramètre est le type de model, on va spécifier ‘sequential’ car … (voir avec gaston)
* Le deuxième paramètre est une méthode ‘rescaling’, qui va servir convertir les valeurs des pixels entre 0 et 1. Finalement l’image sera en noir et blanc.
* Après on va appliquer une méthode ‘Conv2D’, celle-ci va appliquer des filtres au images, ces filtres vont avoir comme but de rendre l’image plus facilement en supprimant les détails parasite de l’image.
* Puis la dernière méthode appliqué ‘MaxPooling2D’, cette méthode va réduire la taille de l’image, à 64x64 pixels, toujours dans le but de réduire les détails.

### Entraînement des données

Une fois que tous les paramètres sont configurés, on va compiler le model, pour l’entrainer ensuite.

Pour entraîner le model on va définir un nombre de répétition ‘epochs’, plus nombre est élevé au mieux le model sera précis. Après cet entrainement on va afficher les résultats avec la librairie Plotly



Cette opération peut-être répéter plusieurs en partant du model déjà entrainer pour le rendre encore plus précis.

### Elaboration du jeu de test

Une fois que l’entraînement est terminé nous pouvons tester nos données sur d’autre image, pour voir si notre model est assez précis. Chaque image tester nous donnera en sortie un nom de class et un pourcentage.

Une image contenant texte, oiseau, oiseau de proie

Description générée automatiquement

## Spécification des logiciels et outils

Pour ce projet, j’ai dû utiliser les outils que le pôle informatique, le principal est le langage python qui est tout nouveau pour moi. La première semaine de stage a donc été une semaine d’apprentissage au langage pour être opérationnel par la suite. J’ai suivi des tutoriels et j’ai lu de la documentation sur python et toutes les autres librairies que l’on allait utiliser dans un premier temps. Dans un second temps je me suis entraîner en créer quelque outil qui allait me servir dans la suite pour le projet.

De plus j’ai dû apprendre à travailler avec git qui est fréquemment utilisé dans le monde de l’entreprise.

### Microsoft Visual Studio Code

Microsoft Visual Studio Code est un éditeur de code développé par Microsoft pour Windows, c’est un environnement de développement intégré. Je l’ai utilisé tout au long du projet afin de développer tous les outils informatiques en python. Il est pratique car on peut installer plein de module extérieur pour nous faciliter la tâche dans notre développement, analyse et correction de bug.

### Git

Git est un outil de développement qui aide une équipe de développeurs à gérer les changements apportés au code source au fil du temps. Les logiciels de contrôle de version gardent une trace de chaque changement apporté au code dans un type spécial de base de données. Grâce à ces outils on peut travailler en groupe de manière efficace et optimiser. Il permet aussi de garder un projet complet sans perte de fichier.

### Db Brower

DB Browser est un logiciel libre afin de gérer des bases de donnée sans avoir besoin de serveur. Avec une interface graphique facile d’utilisation il permet de faire des actions plus facilement.

### Streamlit API

Streamlit est une API open source en langage Python. Il nous aide à créer des applications web pour la science, l’analyse des données et l'apprentissage automatique. Il est compatible avec les principales bibliothèques Python, telles que scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib, etc. Avec Streamlit, aucun callback n'est nécessaire puisque les widgets sont traités comme des variables. La mise en cache des données simplifie et accélère les calculs.

### Google Colab

Google Colab ou Colaboratory est un service cloud, offert par Google, basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l’apprentissage automatique. Cette plateforme permet d’entraîner des modèles de Machine Learning directement dans le cloud. Sans donc avoir besoin d’installer quoi que ce soit sur notre ordinateur à l’exception d’un navigateur. Très utiles lorsque l’on ne possède pas une machine très puissante pour les calculs.

## Difficulté rencontré

Durant ce projet je n’ai pas rencontrés de problème majeure, lors de mes recherches et mes expérimentations lors de la première semaine, apprendre à coder en python n’a pas été compliqué car les principes du langage sont un peu toutes les mêmes. Lors de la partie de l’application je pense que cette partie a été la plus difficile car quand bien même l’API nous facilite la tâche dans pas mal d’aspect. Si on veut faire quelques choses de sophistiquer l’API est restreinte. J’ai été bloquer à certain moment par l’API lorsque je voulais développer l’outils d’une certaine manière. Pour la partie intelligence artificiel, n’ayant jamais travailler sur cette technologie, j’ai dû beaucoup me renseigner et beaucoup pratiquer. Mais je n’ai pas rencontré de problème majeures qui m’ont bloqué dans mon travail car les ressources sur internet pour nous aider sont infini lorsque cela touche un sujet en plein développement et qu’il y a énormément de personnes qui travaille dessus. Quand bien même cet outil a été conçu pour l’analyse de neutron, certain chercheur commence réellement à se penche sur le sujet.

# Bilan

### Apports du stage

### Perspectives futures

# Lexique

Machine Learning : Programme permettant à un ordinateur ou à une machine un apprentissage automatisé, de façon à pouvoir réaliser un certain nombre d’opérations très complexes.

API : Une « interface de programme d’application » (API) est un regroupement de routines, de protocoles et d’outils.

Framework : Un Framework contient des composants autonomes qui permettent de faciliter le développement d’un site web ou d’une application

SQL

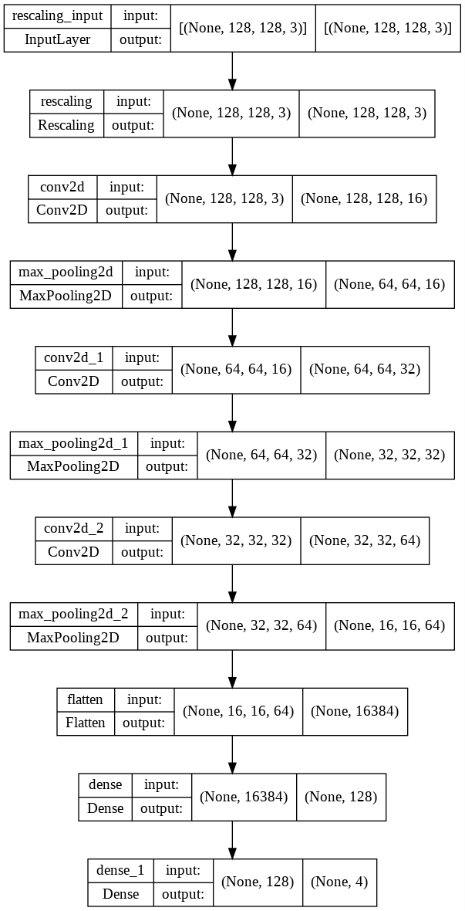
Dataset

model

# Table des illustrations

# Annexes

Annexe structure fichier



BROUILLON

* Comment faciliter les chercheurs dans le travail ?
* Comment fonctionne le machine learning
* Comment faire de la reconnaissance d’image
* Quelle schéma la bd
* Avoir assez de donnée afin d’avoir un model le plus précis possible
* Mini projet reconnaissance image rapace
* Pour les données = data augmentation

séquence/ ittération/ dvplement en V/ glossaire méthode scrum / spéciffication/ conception/ test / recette développement projet informatique.

Type classe oiseau = verre de glass, sphère, cylindre, ellipsoïde

Après avoir fini cette partie, on s’est penché sur le l’intelligence artificiel, le machine learning. Notre but, à partir d’une image de spectro trouver les paramètres optimisés afin d’avoir le meilleur résultat possible. Nous allons pour cela faire de la reconnaissance d’image, une image est composée de pixel chaque pixel à une valeur. C’est grâce à cela qu’on va pouvoir reconnaitre les images. Avec de l’entraînement et en lui donnant plein d’image il pourra être le plus précis possible.

Le model est créer grâce aux bibliothèques de google TensorFlow et aussi la biblothèque Keras c’est bibliothèque on des fonctions pour la machine learning. En prenant en entrée un répertoire d’image.

Afin de créer le model on va appliquer :

* une première méthode qui va servir convertir les valeurs des pixels entre 0 et 1
* ensuite On va appliquer une méthode qui appliquer 16 filtres sur chaques images
* Puis une troisième méthode qui va faire passer les images de 128\*128 pixels à 64\*64

On répète les ces opérations 3 fois pour avoir finalement 112 filtres appliqué et une ilage de dimension 16\*16.

Après on compile le modèle

Avec ce modèle on va l’entrainer afin qu’il s’affine en lui donnant des images en lui disant ce qu’il y a dessus et d’autre image en lui donnant aucune info.

Après l’avoir entraîné on peut voir qu’avec 500 images le modèle n’était pas du tout assez précis il lui fallait beaucoup plus de donnée afin d’être le plus précis possible.

C’est pourquoi on lui a fait une data augmentation en transformant les images :

* -Une symétrie verticale et horizontal
* - le floutage, blur en anglais
* - flip vertical
* - rotation de 45°
* - miroir + rotation 45
* - Décalage de 10 pixels vers la droite
* - Décalage de 10 pixels vers la gauche
* - miroir + Décalage 10 pixel droite
* - miroir + décalage 10 pixel gauche
* - Décalage de 20 pixels vers la droite
* - Décalage de 20 pixels vers la gauche
* - miroir + Décalage 20 pixel droite
* - miroir + décalage 20 pixel gauche

Grâce à cela le modèle était beaucoup plus précis car ont avoisiné les 6500 images.