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



**Département Informatique**

Diplôme préparé : Informatique

N° Jury 3

Développement d'un logiciel d'aide au traitement  
de données expérimentales  
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

Rapport 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

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

Table des matières

[Introduction 3](#_Toc106104334)

[Remerciements 3](#_Toc106104335)

[Résumé 4](#_Toc106104336)

[Problématique et objectifs de stage 4](#_Toc106104337)

[Présentation de l’entreprise 5](#_Toc106104338)

[LE CEA 5](#_Toc106104339)

[LE CNRS 7](#_Toc106104340)

[LE LABORATOIRE LEON BRILLOUIN 8](#_Toc106104341)

[Mission 9](#_Toc106104342)

[Mise en situation du projet 9](#_Toc106104343)

[Phase conceptuelle du projet 11](#_Toc106104344)

[Fonctionnalités 11](#_Toc106104345)

[Contraintes 11](#_Toc106104346)

[Déroulement du projet 12](#_Toc106104347)

[PARTIE 1 - Conception de l’application web 13](#_Toc106104348)

[Structure des fichiers .32 13](#_Toc106104349)

[Stockage des fichiers 14](#_Toc106104350)

[Interface de l’application Web 15](#_Toc106104351)

[Conclusion 19](#_Toc106104352)

[PARTIE 2 - Conception du logiciel de traitement des données 20](#_Toc106104353)

[Récupération des données 21](#_Toc106104354)

[Mise en forme des images 21](#_Toc106104355)

[Augmentation du nombre de données 22](#_Toc106104356)

[Conception du model 24](#_Toc106104357)

[Entraînement des données 26](#_Toc106104358)

[Elaboration du jeu de test 27](#_Toc106104359)

[Conclusion 27](#_Toc106104360)

[Conclusion 28](#_Toc106104361)

[Spécification des logiciels et outils 29](#_Toc106104362)

[Microsoft Visual Studio Code *(https://code.visualstudio.com/)* 29](#_Toc106104363)

[Git (*https://github.com/)* 29](#_Toc106104364)

[Db Brower (*https://sqlitebrowser.org/)* 29](#_Toc106104365)

[Streamlit API (*https://streamlit.io/)* 30](#_Toc106104366)

[Google Colab (*https://colab.research.google.com/*) 30](#_Toc106104367)

[Difficultés rencontrées 30](#_Toc106104368)

[Bilan 31](#_Toc106104369)

[Apports du stage 31](#_Toc106104370)

[Perspectives futures 31](#_Toc106104371)

[Lexique 32](#_Toc106104372)

[Table des illustrations 33](#_Toc106104373)

[Annexes 33](#_Toc106104374)

# Introduction

## Remerciements

Je tiens à remercier mon maître de stage, M Gaston EXIL, responsable du Service Informatique du laboratoire Léon Brillouin, qui m’a accompagné tout au long de cette expérience professionnelle avec beaucoup de patience et de pédagogie.

Je tiens également à remercier M Éric ELIOT, Directeur, et M Grégory CHABOUSSANT, Directeur Adjoint de m’avoir donné l’opportunité d’intégrer le laboratoire.

Je tiens aussi à remercier Aurore VERDIER pour toute l’aide qu’elle m’a apporté afin que je puisse réaliser ce stage dans les meilleures conditions.

Enfin, je remercie les collaborateurs du pôle administratif et financier du laboratoire pour leur disponibilité et 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 (Essonne – France) au sein du laboratoire Léon Brillouin.

Au cours de ce stage au sein du Service Informatique, je me suis intéressé au développement de logiciels d'aide au traitement de données expérimentales en utilisant l'intelligence artificielle.

Ce stage a été 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 expérience m’a permis de de découvrir les missions des développeurs dans le monde professionnel.

## Problématique et objectifs de stage

Ce stage a été pour moi une opportunité enrichissante de percevoir comment fonctionne une entreprise dans le secteur de la recherche.

Pour élaborer ce mémoire, j’ai puisé dans les nombreux enseignements issus des missions qui m’ont été confiées. De plus, les différents échanges que j’ai pu mener avec les collaborateurs du laboratoire et plus particulièrement de son service informatique sont autant d’enrichissements qui ont permis de rendre ce rapport cohérent.

# Présentation de l’entreprise

## LE CEA

Le Commissariat à l’Énergie Atomique et aux Énergies Alternatives est un organisme public de recherche à caractère scientifique, technique et industriel. Il est classé comme Établissement Public à caractère Industriel et Commercial (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 basses carbones (nucléaires et renouvelables),
* la recherche technologique pour l'industrie,
* la recherche fondamentale (sciences de la matière et sciences de la vie).

Fort de son expertise reconnue, le CEA participe, depuis près de 80 ans, à la mise en place de projets avec de nombreux partenaires et conduit, pour le compte de l’Etat, des programmes de recherche visant à accroître la connaissance scientifique et à contribuer à l’innovation dans de nombreux domaines.

En 2010, le CEA, historiquement le Commissariat à l’Énergie Atomique, a vu sa mission évoluer et a changé de nom pour devenir le Commissariat à l’Énergie Atomique et aux Énergies Alternatives.

Le CEA dénombre 20 181 salariés dont 1 233 doctorants et 176 post-doctorants. Son budget annuel atteint 5 milliards d’euro.

Ses travaux lui ont permis de déposer de très nombreux brevets. Ainsi selon le classement annuel des déposants de brevets de l’Institut National de la Propriété Industrielle (INPI), en 2021, le CEA est :

* le 1er organisme de recherche français avec 528 dépôts de brevets,
* le 2ème plus important dépositaire de brevets derrière le groupe Safran et devant les groupes Valeo et Saint-Gobain.

Le CEA, c’est également :

* le seul organisme de recherche français figurant dans le classement des « 100 premiers innovateurs mondiaux » publié par Clarivate Analytics et ce depuis 10 années,
* plus de 5 000 publications annuelles,
* le 1er organisme de recherche public au classement mondial des dépositaires selon le traité de coopération en matière de brevets (PCT) de l’Organisation Mondiale de la Protection Intellectuelle (OMPI).



## LE CNRS

Le Centre National de la Recherche Scientifique (CNRS) est une institution de recherche publique française créée par décret en Octobre 1939 et dont la mission principale est de « coordonner l’activité des laboratoires en vue de tirer un rendement plus élevé de la recherche scientifique ».

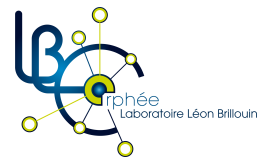
Le rôle du CNRS est de faire progresser la connaissance et d’être utile à la société en utilisant la recherche. Sa mission est divisée en plusieurs axes :

* la recherche scientifique,
* la valorisation des résultats de recherche,
* le partage des connaissances,
* la formation par la recherche,
* la contribution à la politique scientifique.

Pour son activité scientifique, le CNRS s’appuie sur 10 instituts nationaux spécialisés dans :

* 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.

Le CNRS est reconnu aussi bien au niveau national qu’international pour la qualité de ses travaux scientifiques tant dans l’univers de la recherche et du développement que pour le grand public.



## LE LABORATOIRE LEON BRILLOUIN

Le Laboratoire Léon Brillouin (LLB) est un laboratoire de recherche mixte français soutenu par le Commissariat à l'Énergie Atomique et aux Énergies Alternatives (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.

L’utilisation des faisceaux de neutrons est une technique récente indispensable pour les recherches au niveau microscopique dans des domaines très variés tels que la physique, la chimie, la biologie, la science des matériaux, le magnétisme… Au niveau industriel, cette technique permet de réaliser des études non-destructives.

Le LLB dispose de son propre programme de recherche scientifique et collabore avec des scientifiques de nombreux laboratoires de recherche fondamentale, de recherche appliquée et de l’industrie. Le LLB produit chaque année près de 180 publications dans des revues scientifiques et plus de 150 communications dans des conférences et réunions.

# Mission

## Mise en situation du projet

Comme précédemment indiqué le service informatique du LLB et les chercheurs collaborent dans le cadre de leurs missions de recherche.

La mission qui m’a été confiée, a pour objectif d’apporter une aide aux chercheurs par le développement de logiciels de traitement de données.

Ces logiciels doivent permettre :

* de représenter graphiquement des données d’acquisition sur une application web afin de faciliter l’interprétation des résultats (1ère partie du rapport),
* d’indiquer la structure d’un échantillon à partir de la représentation graphique des données d’acquisition en s’appuyant sur la reconnaissance d’images par apprentissage automatique (2ème partie du rapport).

En effet, les chercheurs utilisent les neutrons afin d'analyser la composition atomique d'un échantillon et d'en déterminer ses propriétés. L'environnement de test de l'échantillon permet de modifier les paramètres de température, pression, de champs magnétique et de positionnement de l'échantillon.

Les faisceaux de neutrons traversent l’échantillon et sont réceptionnés sur un détecteur qui collecte les données précédemment citées. Elles sont ensuite stockées dans un fichier nommé « fichier .32 » sur un modèle spécifiquement développé par le CEA.

Ce fichier contient toutes les informations relatives aux caractéristiques de l’échantillon, les paramètres de la manipulation et les résultats de l’expérimentation.

Pour mener à bien la mission confiée, il m’a fallu tout d’abord analyser la structure des données contenues dans le fichier .32, puis concevoir une base de données pour sauvegarder les fichiers, et développer avant de mettre en œuvre des modalités de classification.

Pour ce faire, nous nous sommes orientés vers le machine Learning. Nous avons retenu plus particulièrement, la reconnaissance d’images sur les graphiques obtenus. Elle permet :

* de reconnaître les paramètres et l’échantillon utilisé lors de l’expérimentation,
* de trier les fichiers et de récupérer les plus significatifs.

Ainsi les chercheurs disposent d’un outil pour extraire les fichiers sur lesquels sont stockées les données utiles et peuvent gagner en temps de traitement considérant le volume conséquent de données collectées.

## Phase conceptuelle du projet

Tout projet de développement débute par une étape de cadrage qui permet d’anticiper autant que de possible comment va se dérouler le projet en intégrant les objectifs attendus, les ressources nécessaires et les contraintes diverses.

### 

### Fonctionnalités

Le cahier des charges relatif au projet prévoyait les fonctionnalités suivantes :

* Stockage sur une base de données,
* Visualisation des données issues des acquisitions neutroniques,
* Création d’une application web interactive,
* Extraction des données depuis les fichiers .32,
* Test des modèles d’apprentissage automatique.

### Contraintes

Le cahier des charges relatif au projet prévoyait également les obligations suivantes :

* Utilisation de Framework Streamlit pour l'interface utilisateur,
* Utilisation de Framework Plotly pour la représentation graphique,
* Développement d’une interface simple d’utilisation pour les chercheurs,
* Utilisation du langage de programmation Python.

### Déroulement du projet

Sous la supervision de M. EXIL, j’ai mené le projet en autonomie, en suivant les consignes et en appliquant une méthodologie dite agile.

J’ai divisé le projet en sous–projets au sein desquels j’ai défini 3 phases :

* une phase de recherche,
* une phase de développement
* une phase de test avec réajustements si nécessaires.

L’issue de chaque phase était considérée comme un jalon permettant de valider les travaux réalisés et d’engager les suivants.

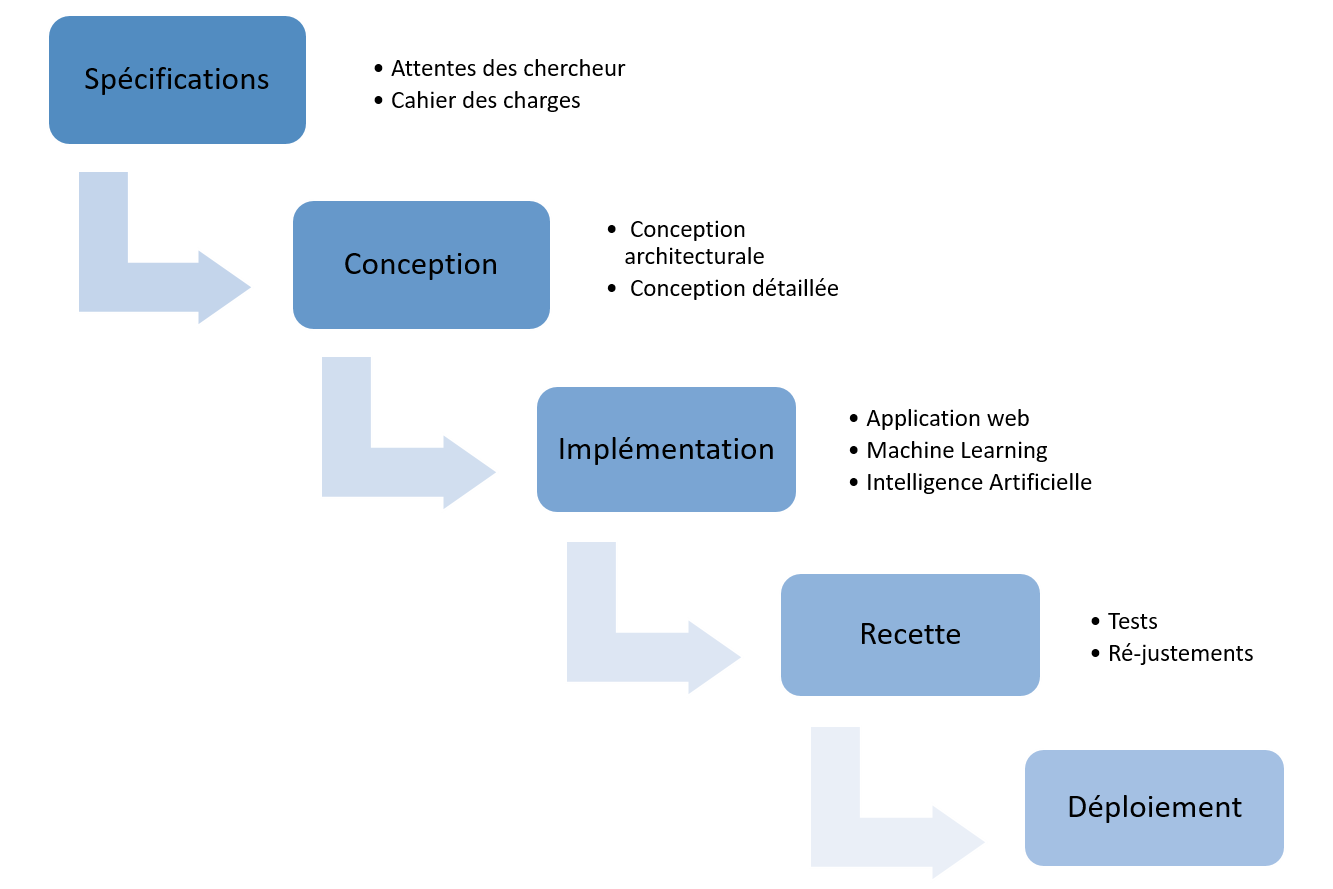


Figure 1 : Modèle de développement

## PARTIE 1 - Conception de l’application web

### Structure des fichiers .32

Un fichier .32 est le livrable obtenu à la suite du bombardement d’un échantillon par un faisceau neutronique. C’est l’ensemble des données collectées (caractéristiques de l’échantillon, paramètres d’acquisition et résultats de la manipulation).

La première étape a consisté à prendre connaissance et à comprendre la structure d’un tel fichier.

Un fichier .32 est composé de 3 blocs :

* bloc 1 : le détail des paramètres d’acquisition programmés et réels, codés à l’aide de caractères ASCII,
* bloc 2 : les valeurs d'intensité acquises par un détecteur de résolution 128 x 128. Celles-ci sont représentées dans un tableau d’une seule dimension, en binaire sous le format d'entiers non signés codés sur 4 octets,
* bloc 3 : le détail de paramètres d'acquisition complémentaires codés à l’aide de caractères ASCII.

Après avoir compris la structure du fichier .32, l’enjeu a été de trouver la solution pour extraire et stocker les données afin de pouvoir les traiter.

### Stockage des fichiers

Afin de rendre plus aisée la manipulation des fichiers et l’affichage des données sur l’application web, les données utiles doivent être stockées sur une base de données mise en place grâce à l’application DB Browser sous SQLite.

Celle-ci intègre 2 tables liées :

* La table « Fichier » pour stocker la valeur de chaque caractéristique des fichiers (ID, nom, Checksum, …) et les données collectées lors de la manipulation,
* La table « Checksum » qui va servir à stocker le Checksum de chaque fichier.

Etant donné la limite du nombre de caractères par valeur imposée par SQLlite, les valeurs d’acquisition enregistrées dans le fichier .32 sont converties en hexadécimal afin d’être intégrées à la base de données.

Les valeurs associées aux paramètres étant plus réduites en nombre de caractères, il n’est pas nécessaire de convertir celles-ci.



Figure 2 : schéma de la base de données

### Interface de l’application Web

L’application web doit répondre à 2 caractéristiques importantes :

* permettre une utilisation simple pour les chercheurs,
* fournir des résultats précis leur permettant de faire les analyses prévues.

Pour ce faire, l’interface a été conçue avec un maximum d’ergonomie en rassemblant l’ensemble des fonctionnalités indispensables pour les chercheurs.

Le langage Python et l’API Streamlit ont été utilisés pour :

* la conception de l’application web,
* l’association avec des widgets tels que celui utilisé pour charger les fichiers .32,
* l’association avec d’autres bibliothèques Python (Matplotly, Pandas…) permettant par exemple d’afficher les données sous forme de graphiques ou de tableaux.

Pour utiliser l’application, les chercheurs doivent en premier lieu charger un fichier .32. Pour cela, l’application utilise le widget st.file\_uploader via l’API Streamlit.

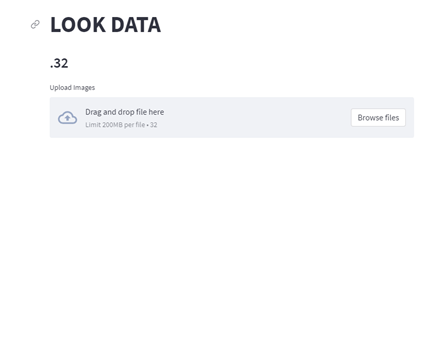


Figure 3 : widget pour charger un fichier .32 dans l'application

Le fichier chargé est ensuite enregistré dans la base de données dédiée. J’ai ensuite développé des requêtes SQL sous python pour extraire des données telles que par exemple :

* le nom du fichier
* le chemin du fichier
* le type du fichier
* les paramètres configurés pour la manipulation
* le tableau des valeurs de l’intensité
* le bloc descriptif des données d’acquisition.

Ci-dessous, deux exemples de requête développées pour :

* extraire le nom du fichier en fonction d’un checksum donné :

"SELECT Nom\_Fichier From Fichier Where CheckSum\_Fichier = '%s'"%id

* insérer les données du fichier .32 dans la base de données :

"INSERT INTO Fichier VALUES(?)", (item)

« Item » correspond aux typologies de données à insérer.

Les résultats des requêtes sont différemment traités en fonction de leur type, à savoir :

* les paramètres de configuration pour réaliser les manipulations,
* les résultats des manipulations.

Concernant les paramètres de configuration, ceux-ci sont affichés dans un tableau à 2 colonnes :

* le nom du paramètre,
* la valeur du paramètre.

Concernant les valeurs des intensités acquises, celles-ci sont converties d’un format hexadécimal (Cf. § Stockage des fichiers) en valeur entière numérique, puis affichées dans un tableau à 2 colonnes :

* l’index du point d’acquisition,
* la valeur mesurée au point d’acquisition.

L’index de chaque point d’acquisition correspond au couple de ses coordonnées cartésiennes « abscisse ; ordonnée ».

L’abscisse et l’ordonnée sont comprises entre 0 et 127. Il existe donc 16 384 couples de coordonnées qui sont convertis en nombre unique.

Figure 4 : Tableau des données du fichier



Figure 5 : tableaux des données extraites du fichier .32

Afin de faciliter la lecture des résultats, les valeurs d’intensité acquises sont représentées sous différentes formes graphiques créés grâce à la bibliothèque Plotly.

La représentation graphique ci-dessous est un graphique en 2D avec une vue par le dessus.

Pour réaliser cette représentation graphique, le tableau d’intensité acquise a été remodelé. Pour cela, chaque nombre unique relatif à un couple de données est re-converti en couple de coordonnées X et Y distinctes.

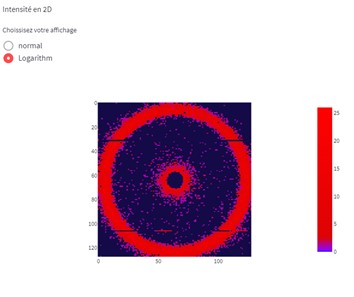


Figure 6 : Graphique en 2D

La représentation graphique ci-dessous est une vue en 3D du graphique précédent auquel l’intensité mesurée a été ajoutée en tant que 3ème axe (Z).

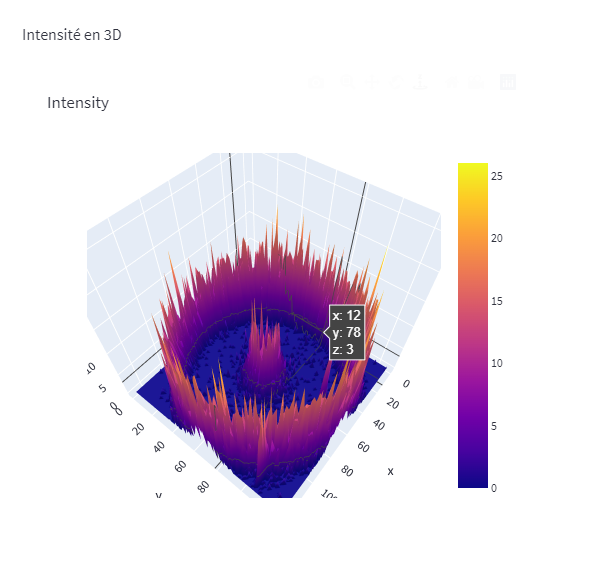


Figure 7 : Graphique en 3D

La représentation graphique ci-dessous est une vue de côté en 2D, du graphique précédent coupé à la verticale.

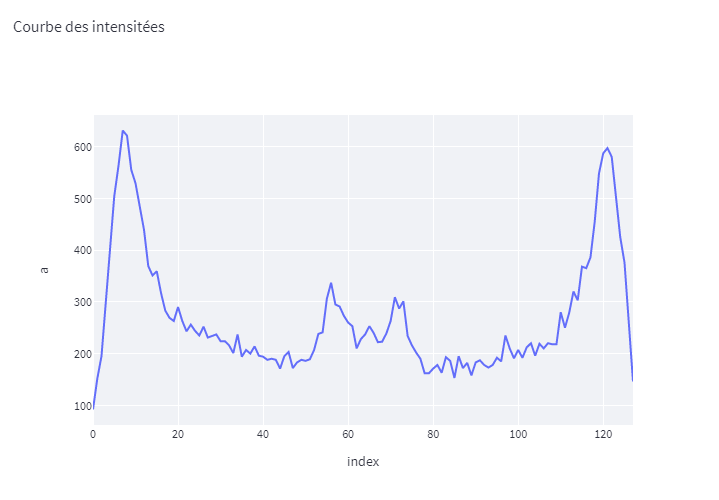


Figure 8 : graphique en 2D (coupe verticale)

En complément, une fonctionnalité supplémentaire a implémenté afin de pouvoir charger des dossiers de fichiers .32. Cette fonctionnalité permet d’alléger l’étape de chargement, notamment si les chercheurs disposent de nombreux fichiers.

Dans ce cas, face à un nombre important de fichiers et afin de permettre un affichage sans surcharger la page, AG Grid (Agnostic Grid ) a été utilisée en tant qu’outil de visualisation de base de données permettant d’afficher les fichiers .32 comme une bibliothèque.

Ainsi en cliquant sur le nom d’un fichier dans la bibliothèque, les représentations graphiques et les tableaux sont actualisés à partir des données portées par le dit-fichier.

Les modalités de fonctionnement précédemment décrites et applicables à un fichier unique restent applicables.

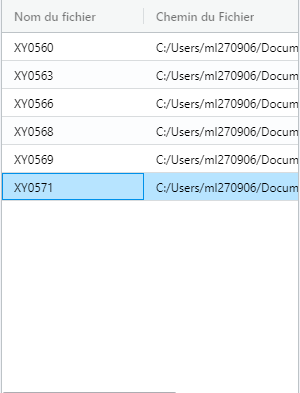


Figure 9 : liste des fichiers présent dans la base de donnée

### Conclusion

La mise en place de cette application web de représentation graphique des données d’expérimentation m’a permis de mettre en application un nouveau langage de programmation (Python) et d’expérimenter de nouveaux outils (large choix de librairies associées et API).

Grâce à cette application, les chercheurs vont disposer de l’ensemble des outils nécessaires pour mener à bien leurs analyses. Bien qu’étant fonctionnelle à l’issue de mon stage après avoir réussi les tests sur des données de recette, l’application devra être mise sous tension en situation réelle. L’application à encore beaucoup de potentiel et sera amené à évoluer en fonction de la demande des chercheurs.

## PARTIE 2 - Conception du logiciel de traitement des données

Cette seconde partie du rapport va être consacrée à l’Intelligence Artificiel utilisée dans le cadre de ce projet.

La quantité de données collectées durant une seule acquisition peut être de l’ordre du millier. Face au volume de plus en plus important de données collectées grâce à l’évolution des technologies neutroniques, il arrive que certaines d’entre elles soient aberrantes pour différents facteurs. Il en résulte des représentations graphiques inexploitables.

Aujourd’hui le tri des images d’acquisition est fait manuellement. Cela représente un travail long, fastidieux et répétitif.

L’objectif est d’utiliser l’Intelligence Artificielle afin de :

* d’alléger le tri manuel des images d’acquisition grâce à un algorithme de recherche et de classification,
* concevoir un modèle afin de reconnaitre la structure d’un échantillon et de la rapprocher d’une base de données de structures grâce à de la reconnaissance d’images.

Afin de comprendre le fonctionnement de la reconnaissance d’images, un cas simple a été développé permettant de reconnaitre un type d’oiseau parmi 4 espèces retenues.

### Récupération des données

La 1ère étape a nécessité la constitution d’un jeu de données à partir de photos d’oiseau des 4 espèces retenues.

Cette technologie comporte des points de vigilance et notamment pour que le modèle soit fiable, il est nécessaire de :

* disposer d’un jeu d’images conséquent,
* de s’assurer de l’absence de doublon d’images,
* de tendre vers l’équilibre dans la répartition des images des différentes espèces.

La collecte des images peut se faire de plusieurs manières :

* l’acquisition de jeu de données d’images,
* l’acquisition d’images unitaires,
* la réalisation de photos.

Chaque donnée (en l’occurrence ici des images d’oiseaux) possède 2 attributs : son nom et sa classe. Dans notre cas, les 4 classes seront :

* Chouette effraie,
* Pygargue à tête blanche,
* Epervier,
* Aigle royal.

En projetant ces travaux au niveau du projet principal, les 4 classes concerneront des structures neutroniques : verre de glass, sphère, cylindre, ellipsoïde

### Mise en forme des images

La 2ème étape consiste en la mise en forme de l’ensemble des images afin de les rendre exploitables. Pour cela, elles devront toutes être de même dimension, dans notre cas 128 x 128 pixels.

### Augmentation du nombre de données

Comme précédemment indiqué, la richesse et la qualité du jeu de donnée d’images sont les critères majeurs pour constituer le modèle. Ainsi pour augmenter « artificiellement » le jeu de données, nous avons appliqué différents filtres aux images.

Ceux-ci modifient légèrement l’image et en créent ainsi des clones. Parmi les principaux filtres, nous avons utilisé :

* une symétrie verticale,
* une symétrie horizontale,
* le floutage (« blur » en anglais),
* une rotation de 45°,
* un effet miroir combiné à rotation à 45°,
* un décalage de 10 pixels vers la droite,
* un décalage de 10 pixels vers la gauche,
* un effet miroir combiné à un décalage de 10 pixels vers la droite,
* un effet miroir combiné à un décalage de 10 pixels vers la gauche,
* un décalage de 20 pixels vers la droite,
* un décalage de 20 pixels vers la gauche,
* un effet miroir combiné à un décalage de 20 pixels vers la droite,
* un effet miroir combiné à un décalage de 20 pixels vers la gauche.

### 

### Conception du model

Pour concevoir notre modèle, nous avons utilisé les librairies TensorFlow et Keras, qui sont 2 outils utilisés dans le Machine Learning.

Nous avons commencé par définir le pourcentage d’images utilisées pour entrainer le modèle. Ici nous avons retenu 80% du jeu de données d’images.

Les images restantes ont été utilisées pour valider le modèle.

Il est important de ne pas utiliser une même image pour l’entrainement et la validation du modèle.

Nous avons ensuite indiqué les paramètres du jeu de données d’entraînement, à savoir :

* le pourcentage de 80% retenu précédemment,
* le label « training » pour spécifier que c’est le jeu d’entrainement,
* le mélange des images pour rendre l’entrainement aléatoire et ne pas sélectionner les mêmes images à chaque fois,
* la taille des images, pour notre cas 128x128 pixels.

Les paramètres pour le jeu de données de validation seront identiques sauf le label qui sera défini sur « validation » et non plus « training ».

Après avoir paramétré les 2 jeux de données, nous avons paramétré le modèle en utilisant les librairies précédemment citées.

Les paramètres retenus ont été les suivants :

* le type de modèle qui sera ici « sequential » car cela définit le réseau de neurone par couche ;
* l’utilisation de la méthode « rescaling » afin de convertir les valeurs des pixels dans une nouvelle plage, ici entre 0 et 1.
* l’utilisation de la méthode  « Conv2D » qui va appliquer des filtres aux images, ces filtres vont avoir comme but de rendre mettre en avant les caractéristiques de l’image (motif, contour, les formes, …)
* l’utilisation de la méthode  « MaxPooling2D », qui va réduire la taille de l’image à 64x64 pixels et ainsi permettre de réduire davantage les détails.

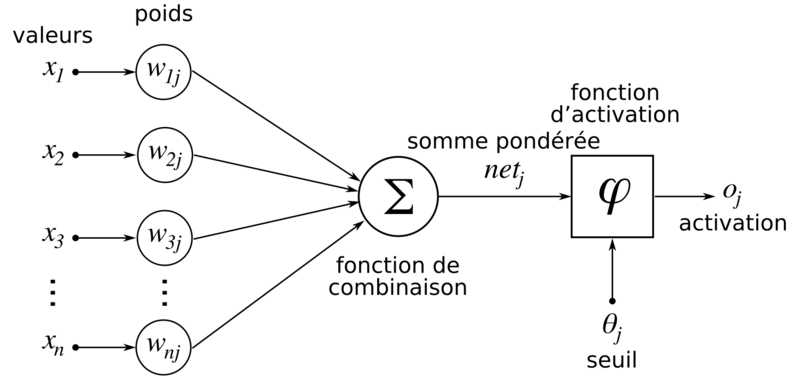


Figure 10 : Fonctionnement du réseau de neurone

### Entraînement des données

A ce stade l’ensemble des paramètres ont été configurés. Le modèle peut donc être compilé puis l’entrainement peut être initialisé.

Pour entraîner le model, nous avons défini le nombre de répétitions « epochs », plus ce nombre est élevé plus le model est sera précis. Après cet entrainement on va afficher les résultats avec la librairie Plotly



Figure 11 : résultat de l'entraînement de notre model

Cette opération peut être répétée plusieurs fois en partant du modèle déjà entrainé afin de rendre celui-ci de plus en plus en plus précis.

### Elaboration du jeu de test

Maintenant que l’entrainement du modèle est finalisé, nous avons testé ce dernier à l’aide d’autres images afin d’évaluer sa précision. Chaque image testée devait obtenir en donnée de sortie un nom de classe et un pourcentage de fiabilité correspondant à la classe donnée.

Ci-dessous un extrait des résultats obtenus :

Une image contenant texte, oiseau, oiseau de proie

Description générée automatiquement

Figure 12 : résultats des test de notre model

### Conclusion

Pour conclure, le développement de cet outil de reconnaissance d’image est plus que concluant , les résultats obtenus lors d’apprentissage de notre modèle et durant ses phases de test démontre sa viabilité. Rapporté à son utilisation sur les graphiques d’acquisitions neutronique, cet outil pourra être utilisé lors des manipulations et dans leurs analyses des résultats.

## Conclusion

Les objectifs de mon stage ont été atteints. Les fonctionnalités, les demandes et contraintes présent dans le cahier des charges ont été respectés. Toutes les fonctionnalités ont été tester, optimisé et améliorer pour avoir le livrable le plus simple d’utilisation et efficace. L’application Web pourra être mis à jour par les chercheurs en cas de besoin. Je l’ai rendu simple d’utilisation tout en étant assez complète pour les utilisateurs.

L’outils de classification est prêt à être utilisé, on pourra toujours affiner notre modèle à l’aide de nouvelle donnée. Cet outil n’est pas dans sa version finale et pourra toujours bénéficier d’amélioration.

Aujourd’hui l’application nécessite une installation d’un exécutable, un point d’amélioration serait de mettre m’application sur un serveur ou un cloud, afin de la rendre accessible par un lien hypertexte.

Et alors ?

Conclusion de cette partie ?

Conclusion de l’ensemble du mémoire ?

Sur la 1ère partie ?

Sur la 2ème partie ?

Perspectives ?

Avis de chercheurs ?

## Spécification des logiciels et outils

Pour ce projet, j’ai travaillé à partir des outils utilisés par le Service Informatique du LLB, dont le principal est le langage Python qui était nouveau pour moi.

La première semaine de stage a donc été une semaine d’apprentissage au langage pour être opérationnel le plus rapidement possible. J’ai commencé par suivre des tutoriels et i lire différente documentation sur Python et toutes les autres librairies. Dans un second temps je me suis entrainé en créant quelques outils utiles pour la suite du projet.

De plus j’ai également appris à travailler avec « Git » qui est fréquemment utilisé dans le monde de l’entreprise.

### Microsoft Visual Studio Code *(https://code.visualstudio.com/)*

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 de nombreux modules extérieurs pour faciliter le développement, l’analyse et la correction de bug.

### Git (*https://github.com/)*

Git est un outil de développement pour 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 optimisée. Il permet aussi de garder un projet complet sans perte de fichier.

### Db Brower (*https://sqlitebrowser.org/)*

DB Browser est un logiciel libre de gestion 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 (*https://streamlit.io/)*

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... La mise en cache des données simplifie et accélère les calculs.

### Google Colab (*https://colab.research.google.com/*)

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.

## Difficultés rencontrées

Durant ce projet je n’ai pas rencontré de problème majeur, 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

Ce stage à pour moi été une excellente expérience professionnelle, d’un côté humain où j’ai pu découvrir le rôle de l’informatique dans l’entreprise, un rôle multitâche cela est très enrichissant. Mais aussi d’un côté plus technique où j’ai pu mettre en application tout ce dont j’avais appris lors deux années de DUT, apprendre aussi à coder en groupe dans le monde de l’entreprise est une expérience super instructive. J’ai pu lors de stage découvrir et apprendre un nouveau langage, ainsi que plein de de librairie, API et Framework.

Travailler en autonomie m’a permis de développer mes capacités de recherches et de résolution de problèmes. Apprendre à s’adapter à son environnement de travail, au code déjà écrit et de comprendre le travail fait en amont par un autre informaticien. Ecouter les conseils que l’on nous donne, les manières d’aborder les étapes du projet, les façons de gérer les erreurs de notre code.

### Perspectives futures

Ce stage a été une confirmation pour moi, dans la voie que je veux suivre, dans le développement d’application et le développement web. Je souhaite continue ma formation dans la licence PRISM, licence Professionnelle Programmation Internet Et Systèmes Mobiles. Pour continuer de développer mes compétences dans ces domaines.

# 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 : est un langage informatique normalisé servant à exploiter des bases de données.

Dataset : Jeu de donnée en anglais, Un jeu de données est un ensemble de valeurs « organisées » ou « contextualisées », où chaque valeur est associée à une variable.

Model : est un fichier qui a été entraîné pour reconnaître certains types de modèles. Vous entraînez un modèle sur un ensemble de données, par un algorithme.

Checksum : est un dispositif qui permet de comparer des fichiers afin de s’assurer de l’absence de doublons.

Widgets : est un outil de développement permettant de rajouter des fonctionnalités dans leur logiciel.

# Table des illustrations

[Figure 1 : Modèle de développement 12](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104243)

[Figure 2 : schéma de la base de données 14](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104244)

[Figure 3 : widget pour charger un fichier .32 dans l'application 15](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104245)

[Figure 4 : Tableau des données du fichier 16](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104246)

[Figure 5 : tableaux des données extraites du fichier .32 17](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104247)

[Figure 6 : Graphique en 2D 17](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104248)

[Figure 7 : Graphique en 3D 18](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104249)

[Figure 8 : graphique en 2D (coupe verticale) 18](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104250)

[Figure 9 : liste des fichiers présent dans la base de donnée 19](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104251)

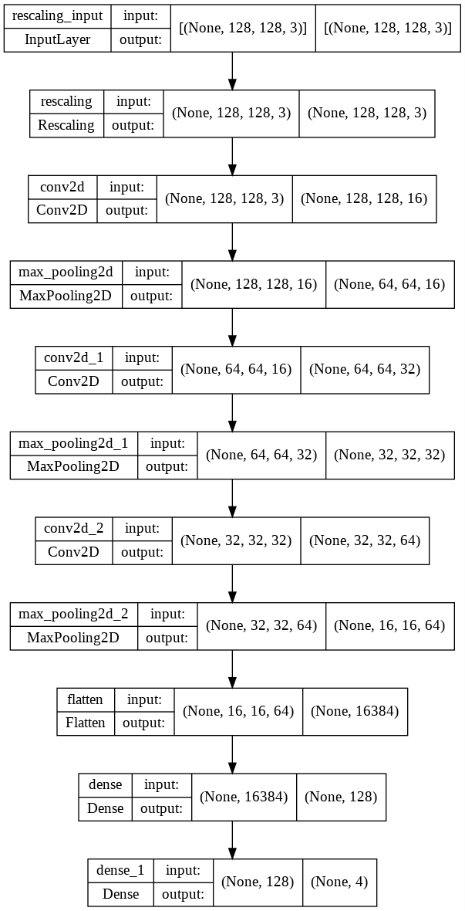
[Figure 10 : Fonctionnement du réseau de neurone 25](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104252)

[Figure 11 : résultat de l'entraînement de notre model 26](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104253)

[Figure 12 : résultats des test de notre model 27](file:///C:\Users\math9\Documents\IUT\STAGE\webSite\elem\rapport_Mathieu_Lacroix_Relu_Gaston_Relu_Papa_3.docx#_Toc106104254)

# Annexes

Annexe structure fichier



Sommaire du model créer

