

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

جامعة عمار تليجي بالأغواط

UNIVERSITE AMAR TELIDJI LAGHOUAT

كلية العلوم

FACULTE DES SCIENCES DEPARTEMENT Sciences de la Matière



### *Mémoire de Master*

Domaine : Sciences de la matière

Filière : Chimie

Option : Chimie Inorganique

Par : **BEN ATALLAH Chaima**

#### THÈME

---

Etude des propriétés physiques des iodates métalliques par data mining

---

Soutenu publiquement devant le jury composé de :

Lefkaier Ibn Khaldoun	Professeur	UATL	Président
Benghia Ali	MCA	UATL	Examineur
Benmoulai Hebboul Zoulikha	MCA	UATL	Rapporteur
Fadla Mohamed abdelilah	Doctorant/Assistant	UATL	Co-Encadrant

## *Je dédie ce travail*

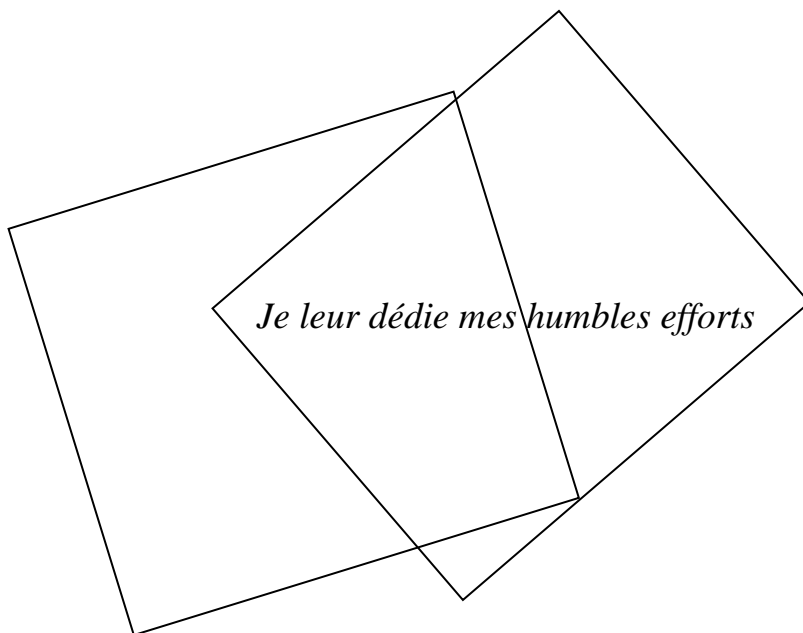
*A celle qui est restée éveillée et a veillé sur moi avec ses yeux et a été un soutien dans ma vie, pour ma chère mère,*

*A celui qui m'a appris le succès et a été avec moi à chaque pas que je fais a celui qui nous a donné sa vie pour notre succès et qui a été la source de ma confiance mon cher père,*

*A tous mes frères, Seyf, Abdelmounim, Massaouda, Assia, Hadjer, Marwa,*

*A ceux qui ont partagé avec moi la fatigue de ce travail et m'ont donné des informations précieuses en intelligence artificielle Ch.Ch,*

*A tous mes amis qui ont partagé avec moi des années et des journées merveilleuses R.Zineb B.Dina S.Afaf M.Aicha ,*



## **Remerciement**

Tous les éloges et les remerciements à notre DIEU Allah le Tout Puissant qui nous a donné le courage, la patience et la force pour mener ce travail jusqu'à sa fin.

Je remercie tout particulièrement avec mes plus sincères gratitude mes deux directeurs de mémoire Madame BENMOULAI-HEBBOUL Zoulikha, Maitre de conférences à l'université Amar Telidji de Laghouat et Monsieur FADLA Mohamed Abdelilah doctorant à l'université Amar Thelidji Laghouat qui ont fait preuve de patience, d'enthousiasme, pour leurs multiples conseils et pour tout le temps qu'ils m'ont consacré à diriger ce mémoire, j'ai pu bénéficier de leurs expériences et leurs compétences.

J'exprime mes sincères remerciements à Monsieur LEFKAIER Iben Khaldoun, Professeur à l'Université Amar Telidji de Laghouat d'avoir accepté de présider le jury, et à Monsieur BENGHIA Ali Maitre de conférences à l'université Amar Telidji de Laghouat pour m'avoir fait l'honneur d'être l'examineur de ce travail.

Une année de travail sur l'intelligence artificielle est suffisante pour vous faire croire en Dieu.

Alan Jay Perlis (1982)

## Liste des figures

<b>Figure 1</b> : Schéma de principe d'une expérience de doublage de fréquence.....	13
<b>Figure 2</b> : Interprétation quantique du processus de doublage de fréquence. ....	13
<b>Figure 3</b> : Entité de type XO <sub>3</sub> -n .....	17
<b>Figure 4</b> : Les techniques possibles de l'intelligence artificielle.....	20
<b>Figure 5</b> : Les différentes méthodes de l'apprentissage automatique.....	21
<b>Figure 6</b> : La différence entre a-la classification et b-la régression .....	22
<b>Figure 7</b> :Tableau de donné pour l'ACP .....	23
<b>Figure 8</b> : Neurone artificiel .....	27
<b>Figure 9</b> : Un arbre de décision .....	29
<b>Figure 10</b> : Représentation graphique des axes principaux .....	34
<b>Figure 11</b> : La matrice de corrélation entre les différentes variables .....	35
<b>Figure 12</b> : La quantité des informations des variables dans les nouvelles axes (Représentation qualitative).....	37
<b>Figure 13</b> La quantité des informations des variables dans les nouvelles axes ( Représentation quantitative ) .....	38
<b>Figure 14</b> : La comparaison entre les valeurs de l'énergie de formation de l'apprentissage automatique et l'énergie de formation calculée par DFT à partir la foret aléatoire et le réseau de neurones.....	42
<b>Figure 15</b> : La comparaison entre les valeurs de l'énergie de gap de l'apprentissage automatique et l'énergie de gap calculée par DFT à partir la foret aléatoire. ....	43

## Liste des tableaux

<b>Tableau 1</b> : La liste des groupes spatiaux non-centro symétriques : .....	15
<b>Tableau 2</b> : Apprentissage supervisé .....	21
<b>Tableau3</b> :Apprentissage non supervisé .....	22
<b>Tableau4</b> : Les conditions de l'arbre de décision.....	28
<b>Tableau 5</b> : Des données calculées par DFT. ....	39
<b>Tableau 6</b> : L'énergie de formation de différents composés .....	39
<b>Tableau 7</b> : Les descripteurs utilisés pour la prédiction.....	40
<b>Tableau 8</b> : L'énergie de gap de différents composés.....	42



## Liste des abréviations

<b>ACP</b>	Analyse des composantes principales
<b>Dend</b>	La fin de domaine de transparence
<b>Dstart</b>	Le début de domaine de transparence
<b>EA</b>	Electronégativité de métal A
<b>EA'</b>	Electronégativité de métal A'
<b>Eg</b>	Energie de gap
<b>I-O(max\min)</b>	La longueur de la liaison max /min
<b>Mpoint A</b>	Point de fusion du métal A
<b>Mpoint A'</b>	Point de fusion du métal A'
<b>NA</b>	Le nombre des éléments de métal A
<b>NB</b>	Le nombre des éléments de métal A'
<b>NO</b>	Le nombre des éléments d'oxygène
<b>ONL</b>	Optique non linéaire
<b>Ra</b>	Le rayon ionique de métal A
<b>RA'</b>	Le rayon ionique de métal A'
<b>SHG</b>	La génération de second harmonique
<b>T déc</b>	La température de décomposition
<b>Vmotif</b>	Le volume de la maille
<b>ZA</b>	Le nombre atomique de métal A
<b>ZA'</b>	Le nombre atomique de métal A'
<b>Z(motif/ maille)</b>	Le nombre de motif par maille

# Sommaire

<b>Introduction générale.....</b>	<b>9</b>
<b>Chapitre I .....</b>	<b>12</b>
<b>Généralités sur L'optique non linéaire .....</b>	<b>12</b>
I.1 Généralités sur l'optique non linéaire.....	11
I.2 Le doublage de fréquence ou la génération de seconde harmonie .....	13
I.2.1 Interprétation en terme photons .....	13
I.3 Les conditions nécessaires pour l'activité optique quadratique .....	14
I.4 Des matériaux pour l'optique non linéaire .....	16
I.5 Des iodates métalliques pour L'ONL.....	17
I.6 Iodates monométalliques et bimétalliques.....	17
I.7 Du mono métallique vers les bimétalliques.....	18
<b>Chapitre II.....</b>	<b>12</b>
<b>L'apprentissage Automatique .....</b>	<b>12</b>
II.1 L'intelligence artificielle .....	20
II.2 Apprentissage automatique .....	20
II.2.1 Les différents types de l'apprentissage automatique .....	20
II.2.2 Les méthodes de l'apprentissage automatique non supervisé .....	23
II.2.2 Les méthodes de l'apprentissage automatique supervisé .....	25
II.2.3 Fonction coût .....	25
II.2.4 L'algorithme la descente de gradient.....	26
II.2.5 Le réseau de neurones.....	27
II.2.6 Arbre de décision.....	27
II.2.7 La forêt aléatoire.....	29
II.2.8 Les données .....	29
<b>Chapitre III .....</b>	<b>33</b>
<b>Résultats et discussions .....</b>	<b>33</b>

III.1 Introduction.....	32
III.2 La préparation des données.....	32
III.3 Analyse par l'approche de datamining .....	34
III.3.1 Choix des axes.....	34
III.3.2 Matrice de corrélation .....	35
III.4 Prédiction .....	38
III.4.1 Energie de formation.....	39
III.4.2 Energie de gap .....	42
<b>Conclusion générale .....</b>	<b>34</b>
<b>Les références bibliographiques .....</b>	<b>34</b>
<b>Résumé.....</b>	<b>52</b>

# **Introduction générale**

Les matériaux ont de tout temps défini le niveau de développement de notre civilisation. Aujourd'hui, ils jouent un rôle déterminant dans toutes les mutations technologiques. Depuis de nombreuses années, les matériaux iodates métalliques à structure non-centrosymétriques suscitent l'intérêt de la communauté scientifique. En effet, en raison de la très grande variété de compositions chimiques qu'elles proposent, les structures non-centrosymétriques présentent une large gamme de propriétés physiques comme l'optique non linéaire exceptionnellement la génération de seconde (SH) et troisième (TH) harmonie. Suivant la nature des cations présents au sein de la structure, métaux de transition, ou terres rares, les iodates métalliques font des structures d'excellents matériaux multifonctionnels, applicables dans des domaines aussi variés : magnétisme, bio-marqueurs, développement du laser...[1].

En raison de l'intérêt croissant pour ces composés, et de leur utilisation extensive dans la technologie existante et les applications potentielles possibles dans les dispositifs futuristes de haute technologie, l'étude des propriétés structurales, électroniques et optiques de ces iodates métalliques est fortement utilisée. La théorie de la fonctionnelle de la densité est actuellement la théorie la plus utilisée en physique des matériaux pour étudiée et prédire les propriétés des matériaux. Elle permet d'obtenir une large gamme d'informations sur l'origine microscopique des propriétés de ces matériaux, seul inconvénient elle prend beaucoup de temps pour faire les calculs surtout pour certaines propriétés spécifiques et pour les structures compliqués de structure moins symétrique comme le triclinique.

Le forage de données, explorations de données ou fouilles de données, ce sont les traductions possibles du datamining en Français. En règle générale, le terme datamining désigne l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données ou en repérant des patterns [2],aussi permet de prédire des propriétés physiques ou chimiques par la technique d'apprentissage automatique. Dans ce mémoire on essaye de trouver des relations binaires qui existent entre les différentes variables et des corrélations entre ces propriétés aussi de prédire certains valeurs de propriétés tels que: l'énergie de gap et l'énergie de formation pour quelques iodates métalliques.

Le travail développé dans ce mémoire est réparti en trois chapitres. Le premier chapitre est dédié à une brève généralité sur l'optique non linéaire (ONL) : les iodates mono et bi-métalliques leurs principaux conditions d'activité en ONL stabilité, et leurs propriétés. Le deuxième chapitre présente une idée générale concernant l'apprentissage automatique, types et différents méthodes. Puis résultats et discussions dans le troisième chapitre. Enfin, une conclusion générale récapitule les principaux résultats obtenus et propose des perspectives de travaux futurs.

## **Chapitre I**

# **Généralités sur L'optique non linéaire**

## I.1 Introduction

Depuis l'évènement du laser au début des années 60, les premières études d'optique non linéaire ont suivi de près l'invention et le développement de sources laser de forte puissance crête et délivrant des impulsions très courtes.

L'optique non linéaire regroupe l'ensemble des phénomènes qui résultent, de la non linéarité de la réponse diélectrique d'un milieu matériel à l'action d'une onde électromagnétique, c'est à dire sous l'effet d'un rayonnement intense (laser)[3].

Les phénomènes d'optique non linéaire (ONL) permettent de coupler des ondes électromagnétiques entre elles ou avec le milieu (génération d'harmoniques supérieures, absorption multi photonique...). Les possibilités ainsi offertes sont déjà utilisées dans les domaines du traitement du signal, des télécommunications ou de l'optique intégrée. Les nouveaux débouchés ont motivés la recherche de nouveaux matériaux présentant des réponses en optique non linéaire élevées.

## I.2 Définitions

Un matériau soumis à l'action d'une onde électromagnétique est le siège d'une polarisation induite qui détermine entièrement sa réponse à l'excitation du rayonnement. Si l'onde lumineuse est peu intense, la polarisation  $P$  reste proportionnelle au champ  $E$  et l'onde réémise est alors de même fréquence que l'onde lumineuse fondamentale.

Dans un milieu linéaire, la polarisation macroscopique (éq I-1) induite par une onde optique s'exprime de façon linéaire en fonction du champ électrique sous la forme :

$$\vec{P} = \vec{P}_0 + \varepsilon_0 \chi^{(1)} E^1 \quad \text{éq I-1}$$

tel que :

$P_0$  : Polarisation spontanée du matériau. Elle est nulle pour la majorité des milieux.

$\chi^{(1)}$  : Tenseur de susceptibilité électrique du premier ordre.

$\varepsilon_0$  : Permittivité du vide.

Rappelons que les relations qui lient la permittivité relative  $\varepsilon_r$ , la permittivité du vide  $\varepsilon_0$  et l'indice de réfraction  $n$  (pour un milieu non magnétique) :

$$n = \frac{c}{v} = \sqrt{\varepsilon_r}$$

$$n = \sqrt{\epsilon_r} = \sqrt{1 + \chi}$$

$$\epsilon = \epsilon_r \epsilon_0$$

Ou :

$c$  : la vitesse de la lumière

$v$  : la vitesse de propagation de l'onde électromagnétique dans le milieu

$\epsilon_r$  : la permittivité relative ( $\epsilon_r$  pour le vide = 1, pour les milieux  $\epsilon_r > 1$ )

Les fortes intensités lumineuses émises par les lasers produisent des champs électriques dont l'intensité s'approche de celui du champ électrique assurant la cohésion des électrons aux noyaux des atomes ou des molécules. Elles provoquent une anharmonicité du mouvement de vibration des électrons autour de leur position d'équilibre. Ceci est à l'origine du non linéarité dans les milieux diélectriques. La polarisation  $\vec{P}$  doit alors être développée en fonction des puissances croissantes du champ électrique  $E$ . La polarisation macroscopique s'écrit donc comme une série de TAYLOR (éq I-2) :

$$\vec{P}(\vec{E}) = \vec{P}_0 + \epsilon_0 \chi^{(1)} \vec{E}^1 + \epsilon_0 \chi^{(2)} \vec{E}^2 + \epsilon_0 \chi^{(3)} \vec{E}^3 + \dots \quad \text{éq I-2}$$

$\chi(n)$  : Tenseur de susceptibilité macroscopique d'ordre  $n$ .

- Le terme  $\epsilon_0 \chi^{(2)} E^2$  est appelée polarisabilité quadratique.

- Le terme  $\epsilon_0 \chi^{(3)} E^3$  est appelée polarisabilité d'ordre 3.....etc

On verra par la suite que seuls les cristaux non Centro symétriques ont un  $\chi(2)$  non nul.

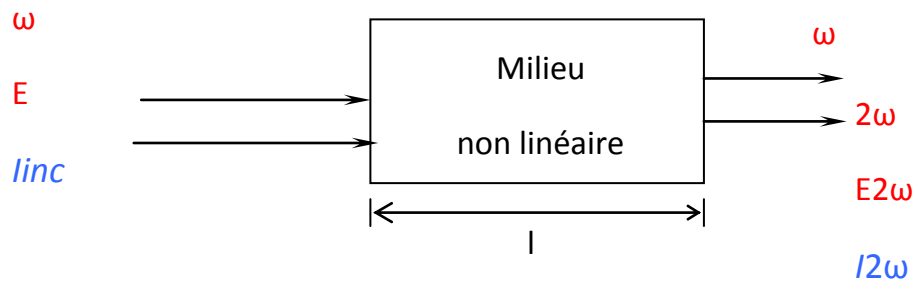
### ✓ Les effets de la polarisation quadratique

La polarisation du second ordre  $P^{(2)} = \epsilon_0 \chi^{(2)} E^2$  donne naissance aux phénomènes liés au mélange de fréquences qui prend des formes différentes. Ces phénomènes sont :

- La génération du second harmonique GSH.
- Rectification CC.
- L'effet électro-optique linéaire ou effet POCKEL.
- L'amplification paramétrique.
- La somme et la différence de fréquences.

### I.3 Le doublage de fréquence ou la génération de seconde harmonie

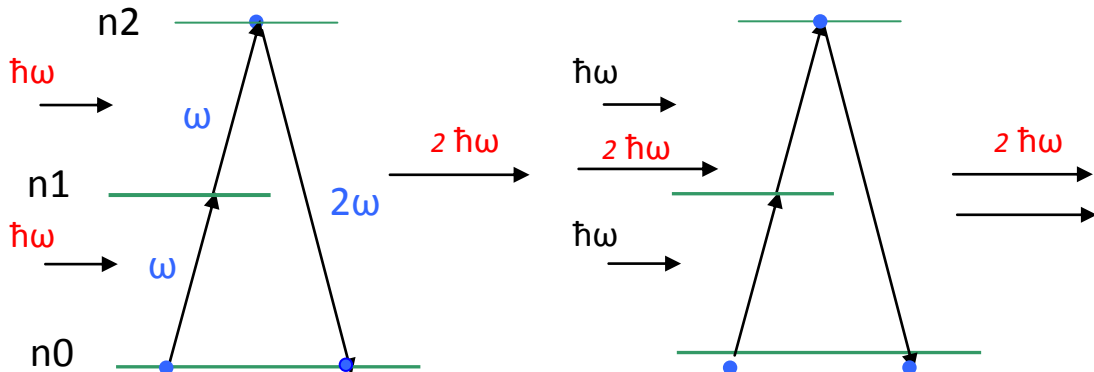
La génération de seconde harmonie est un phénomène non linéaire permettant le doublage de fréquence  $\omega$  de l'onde incidente se propageant dans le milieu non linéaire considéré. Les deux ondes fondamentales possédant alors la même fréquence  $\omega$  et s'ajoutent pour produire un rayonnement cohérent à  $2\omega$ , comme le présentent la figure 1.



**Figure 1 :** Schéma de principe d'une expérience de doublage de fréquence

#### I.2.1 Interprétation en terme photons

Il y'a adsorption de deux photons de fréquence  $\omega$  et émission d'un photon de fréquence  $2\omega$  figure 2.



**Figure 2 :** Interprétation quantique du processus de doublage de fréquence.

Il faut remarquer que l'absorption du deuxième photon repose sur la condition que le niveau excité  $n_1$  soit peuplé d'électrons (inversion de population) d'où la nécessité d'un signal très fort. Cette interprétation ne permet pas de faire une étude qualitative. Il faut passer donc par l'aspect ondulatoire de la lumière ou l'électromagnétisme [4].

### I.3 Les conditions nécessaires pour l'activité optique quadratique

La condition nécessaire pour l'activité optique quadratique est la noncentrosymétrie. Pour qu'un matériau soit utile pour l'optique non linéaire il doit répondre en plus de la condition précédente à :

- Une non-linéarité adéquate et suffisante ;
- Une transparence optique dans le domaine d'intérêt ;
- Une biréfringence propre et suffisante pour l'accord de phase ;
- Une stabilité thermique des propriétés optiques ;
- Une résistance suffisante à l'endommagement optique.

Les matériaux pour l'ONL qui regroupent les propriétés précédentes, et qui peuvent être synthétisés pour donner des cristaux de bonnes tailles, sont très rares parmi tous les cristaux non centro symétriques [4] .

#### ✓ La non centro-symétrie

Le tableau1 regroupe l'ensemble des groupes spatiaux non-centrosymétriques des différents systèmes cristallins. Seuls les cristaux non-centrosymétriques présentent un tenseur  $\chi^{(2)}$  non nul [5]. Les propriétés physiques qui ont la même symétrie que la structure cristalline doivent répondre à l'équation suivante :

$$P(r) = -P(-r)$$

La polarisation quadratique est une des ses propriétés [6].

$$P^{(2)}(E) = \varepsilon_0 \chi^{(2)} E^2$$

$$P^{(2)}(-E) = \varepsilon_0 \chi^{(2)} E^2$$

Alors, si  $\chi^{(2)}$  est nul on aura :  $P^{(2)}(E) = -P^{(2)}(-E)$

Donc pour les systèmes symétriques  $\chi^{(2)} = 0$

**Tableau 1 : La liste des groupes spatiaux non-centro symétriques :**

SYSTEME	CLASSE	GROUPE SPATIAL
TRICLINIQUE	1	P1
MONOCLINIQUE	2 M 222	P2 P21 C2 Pm Pc Cm Cc P222 P2221 P21212 P212121 C222 C221 F222 I222 I212121
ORTHORHOMBIQUE	mm2	Pmm2 Pmc21 Pba2 Pna21 Pnn2 Pnc2 Pmn21 Pba2 Pna21 Pnn2 Cmm2 Cmc21 Ccc2 Amm2 Abm2 Ama2 Aba2 Fmm2 Fdd2 Imm2 Iba2 Ima2
TETRAGONAL	4	P4 P41 P42 P43 I4 I41
	4	P4 I4
	422	P422 P4212 P4122 P41212 P4222 P42212 P4322 P43212 I422 I4122
	42m	P42m P42c P421m P421c P4m2 P4c2 P4b2 P4n2 I4m2 I4c2 I42m I42d
	3	P3 P31 P32 R3
TRIGONAL	32	P312 P321 P3112 P3121 P3212 P3221 R32
HEXAGONAL	6 622	P6 p61 P5 P62 P64 P63 P622 P6122 P6522 P6222 P6422 P6322
	23	P23 F23 I23 P213 I213
CUBIQUE	432	P432 P42332 P4132 I4132 I432 P4332 P4132 I4132

### ✓ **La biréfringence**

C'est la propriété physique d'un matériau dans lequel la lumière se propage de façons anisotrope. Dans un milieu biréfringent, l'indice de réfraction n'est pas unique, il dépend de la direction polarisation de l'onde lumineuse, un effet spectaculaire de la biréfringence est la double réfraction par lequel un rayon lumineux pénétrant dans le cristal est divisé en deux.

#### • **La dureté**

Si la dureté est trop faible, le cristal aura une faible résistance mécanique et le polissage sera délicat. Il faut donc une dureté intermédiaire qui permette un polissage de bonne qualité. Une dureté suffisante limite les phénomènes de "tombée de bord" et assure une planéité de bonne qualité [7].

### ✓ **La fenêtre de transparence**

Elle doit être la plus étendue possible (domaine spectral) et importante (transmission élevée). En effet, le matériau doit être transparent aux longueurs d'onde incidentes mais aussi à celles générées par les phénomènes non linéaires. Par ailleurs toute absorption, aussi faible soit-elle peut conduire à des phénomènes thermiques non négligeables du fait des flux utilisés en ONL[7]

## **I.4 Des matériaux pour l'optique non linéaire**

C'est en 1961 que le premier phénomène de génération de seconde harmonique a été mis en évidence avec un cristal de quartz. L'optique non linéaire s'est alors rapidement développée avec la caractérisation d'autres cristaux minéraux non Centro symétriques comme par exemples ( $\text{KH}_2\text{PO}_4$ ,  $\text{LiNbO}_3$ ,  $\text{KTiOPO}_4$ , ...), leur grande stabilité (thermique, chimique et mécanique), leur fenêtre de transparence très étendues en longueurs d'onde, ainsi que des cristallogènes bien maîtrisées conduisent à des échantillons de très bonne qualité optique.

Cependant, leurs coefficients non linéaires quadratiques sont nettement inférieurs à ceux de certains cristaux organiques moléculaires mis en évidence dans les années 80, ceux ci sont basées sur des molécules à fort transfert de charge.. Les cristaux organiques ont ainsi suscité l'intérêt de nombreuses équipes de recherche. Cependant aucun dispositif optique basé sur des cristaux organiques n'a jusqu'à présent été utilisé au niveau industriel. Ceci est du à leurs stabilités (thermique, chimique, photochimique, et mécanique) insuffisantes[5].L'ingénierie cristalline de nouveaux matériaux hybrides, développé dès le début des années 90 au laboratoire de cristallographie de Grenoble, est une alternative qui tend à cumuler les avantages des cristaux minéraux (stabilité, transparence) et organiques (hyper polarisabilité élevée). Les matériaux hybrides organo-minéraux contiennent des entités moléculaires organiques dérivées des nitroanilines ou nitropyridines très efficaces en ONL qui sont accrochées solidement dans des

matrices hôtes minérales. Ces matériaux ont fait l'objet d'étude structurale, de tests d'efficacité ONL et de 3 brevets CENT-CNRS[5] , mais leurs faibles stabilités thermiques mettent un frein de leurs utilisations.

### I.5 Des iodates métalliques pour L'ONL

Avant de présenter les derniers iodates obtenus au cours de ce travail, il serait intéressant de faire le point sur les structures iodates déjà connues et référencées dans la littérature.

Les premiers iodates ont été découverts avant 1840. Les principales avancées ont eu lieu entre 1970 et 1990 puis après 2000, où il y a eu un regain d'intérêt. Plusieurs laboratoires de recherches ont développé une ingénierie cristalline basée sur la conception de nouveaux matériaux iodates pouvant posséder différentes propriétés telles que la conversion de fréquence, la piézoélectricité, ou la conductivité ionique.

#### ➤ Pourquoi les iodates ?

L'utilisation d'entités de type  $XO_3^{-n}$  (figure 3) présentant des paires d'électrons non liants permet d'obtenir des classes de matériaux dont le taux d'acentricité augmente jusqu'à 30-40 % [8]. Les plus couramment utilisés sont ( $XeO_3$  pour  $n = 0$ ), ( $BrO_3^-$ ,  $ClO_3^-$ ,  $IO_3^-$  pour  $n = 1$ ) et ( $TeO_3^{2-}$ ,  $ScO_3^{2-}$ ,  $SO_3^{2-}$  pour  $n = 2$ ).

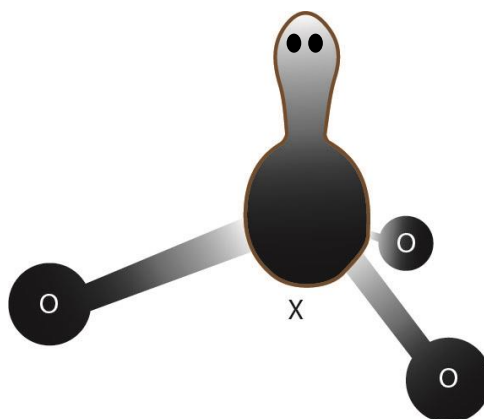


Figure 3 : Entité de type  $XO_3-n$

### I.6 Iodates monométalliques et bimétalliques

En 1969 souligne Bergman [8] , l'entité iodate  $[IO_3]^-$  possède un doublet électronique non liant sur l'iode qui permet de favoriser l'obtention de structures acentriques, condition *sine qua none* pour avoir des effets non linéaires. De nombreuses études sur les iodates métalliques ont eu lieu en suite des années 70 et 80, par le groupe « Bell Laboratories » qui s'est intéressé à leurs propriétés en optique non linéaire ainsi qu'à leurs propriétés ferroélectriques, piézoélectriques et pyroélectriques [9,10] .

La famille des iodates métalliques possèdent une propriété de la non centro-symétrique, relève de l'optique quadratique, qui est une caractéristique importante dans la génération de second harmonique, nous avons donc choisi cette famille comme exemple pour étudier les propriétés qu'ils possèdent.

### **I.7 Du mono métallique vers les bimétalliques**

La synergie est un type de phénomène par lequel plusieurs facteurs agissant en commun ensemble créent un effet global. Il est utilisé pour désigner un résultat plus favorable lorsque plusieurs éléments d'un système agissent un concert. On dit que cette synergie est positive si le résultat d'une action commune est créateur ou autrement meilleur que la somme attendue.

Par exemple :

L'iodate de lithium  $\text{Li}(\text{IO}_3)$  a été largement utilisé comme matériaux piézoélectrique , acousto-optique et de génération de second harmonique (SHG)[11] , cependant , les cristaux de  $\text{Li}(\text{IO}_3)$  ont parfois des inclusions OH qui réduisent leur transparence dans la région infrarouge (IR) , affectant leurs performances [12] . De plus , leur propriétés sont connues pour dépendre fortement des conditions de croissance [13]. Des iodates de matériaux de transition exempts d'inclusions OH , par exemple  $\text{Zn}(\text{IO}_3)_2$  ont ainsi été développés comme matériaux alternatifs à haute efficacité SHG . Pour cela le matériau  $\text{LiZn}(\text{IO}_3)_3$  est synthétisés [14] pour :

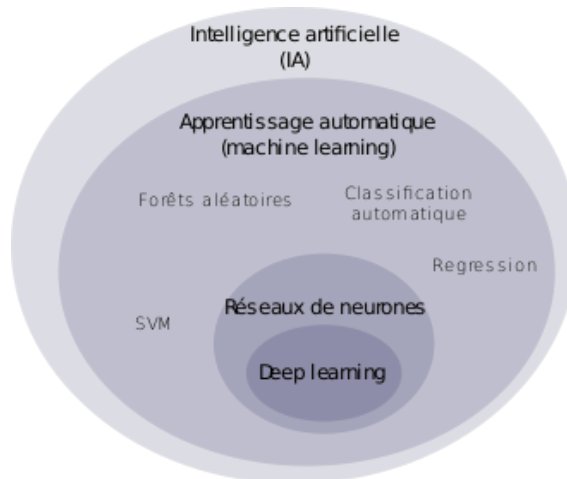
1. Elargir le domaine de transparence.
2. Augmenter la stabilité thermique.
3. Amélioration des coefficients non linéaire  $d_{ij}$ .
4. Eliminer l'effet hygroscopique.

## **Chapitre II**

# **L'apprentissage Automatique**

## II.1 L'intelligence artificielle

L'intelligence artificielle (figure 4) défini comme un terme très large, est un ensemble de théories et de techniques, mise en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. Grâce à l'innovation technologique, il est utilisé dans plusieurs domaines d'application dans notre vie.



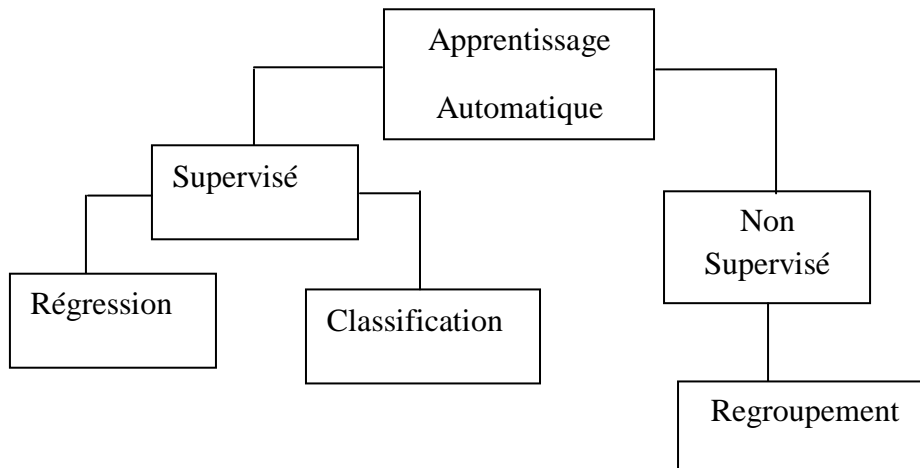
**Figure 4 :** Les techniques possibles de l'intelligence artificielle

## II.2 Apprentissage automatique

C'est une extraction de connaissances à partir des données. L'apprentissage automatique commence toujours par des données, il implique une certaine quantité d'automatisation plutôt que d'essayer de rassembler des idées à partir des données et d'appliquer un processus ou un algorithme à l'aide d'un programme informatique.

### II.2.1 Les différents types de l'apprentissage automatique

Une première grande distinction à faire en l'apprentissage automatique est la différence entre l'apprentissage supervisé et non supervisé comme le présentent la figure 5. Pour bien comprendre la différence entre ces deux types, on donne des exemples pour chaque modèle [15]:



**Figure 5 :** Les différentes méthodes de l'apprentissage automatique

➤ **Apprentissage Supervisé**

C'est le processus de prédiction en utilisant des données labellisées (données étiquetées)  
Exemples : Si la base de données consiste à une pile d'image de chats et une pile d'images de chiens, il s'agit d'un apprentissage supervisé car il existe un résultat spécifique que nous connaissons à l'avance (Tableau 2).

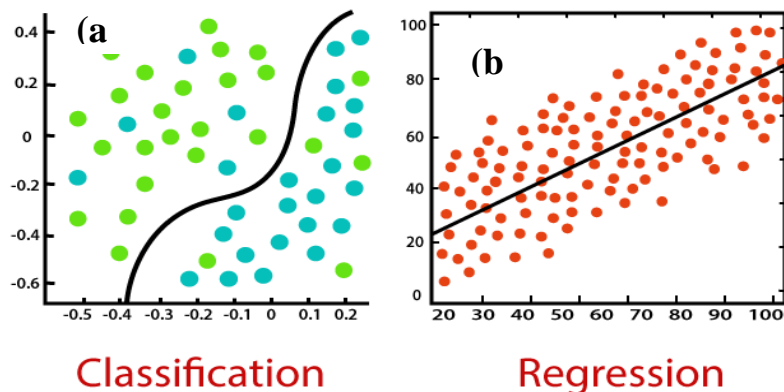
**Tableau 2 :** Apprentissage supervisé

Les Données d'entrés	L'apprentissage	La Prédiction

On compte en générale deux familles de problème [16]:

- **Les problèmes de régression :** Qui consiste à prédire la valeur d'une variable continue comme le présentent la figure 6.a. Où nous essayons d'estimer une valeur à partir d'une entrée, par ex. Connaître la taille des chats.

- **Les problèmes de classification** : Qui consiste à prendre la valeur d'une variable discrète comme le présentent la figure 6.b. Où nous essayons de classer une entrée en plusieurs classes, par ex : Classifier espèce animale chat/chien.



**Figure 6** : La différence entre a-la classification et b-la régression

➤ **Apprentissage non supervisé**

C'est le processus utilisé pour comprendre la structure des données non labélisées (données non-étiquetées). C'est-à-dire apprendre à représenter au mieux des données non labélisées (Tableau 3). Exemple : Si je vous montre les images suivantes en vous demandant de les regrouper en deux groupes, et ceci sans avoir besoin de savoir à quoi ces images correspondent.

**Tableau3** :Apprentissage non supervisé

Les Données d'entrés	L'apprentissage	La prédiction

## II.2.2 Les méthodes de l'apprentissage automatique non supervisé

### a. La définition d'ACP

L'analyse en composantes principales (ACP) est une méthode de statistique exploitoires et descriptive multidimensionnelle utile pour la compression et la classification des données [17]. Cette méthode va synthétiser, résumer l'information continue dans un tableau de données, ces données sont constituées d'individus et de variables qui doivent être quantitative, continus et corrélés entre-elle.

### b. Principe de l'analyse en composantes principales

Du point de vue géométrique l'ACP consiste à effectuer une certaine rotation du repère des variables autour de leurs valeurs moyennes. Cette rotation transforme les  $n$  variables corrélées en  $l$  variables non corrélées. Notons que ce sont justement ces variables transformées que l'on a nommées les composantes principales.

### c. La représentation des données

Les données pour l'ACP sont représentées sous forme de tableau des données (Figure 7), constituée d'individus et de variables qui doivent être quantitatives continus et corrélés entres elles [18].

**Les variables**

	1	.....	p	.....	P
1					
.....					
n			$x_{np}$		
.....					
N					

**Figure 7** :Tableau de donn e pour l'ACP

les valeurs  $x_{np}$  sont des valeurs de la variable  $k$  pour l'individu  $p$  avec  $n = 1, \dots, n$  et  $p = 1, \dots, p$ .  $p$  représente à la fois le nombre d'individus et l'ensemble des individus  $\{1, \dots, p, \dots, P\}$ , de même  $n$  représente le nombre de variables et l'ensemble des variables  $\{1, \dots, n, \dots, N\}$

L'analyse en composantes principales utilise une matrice indiquant le degré de similarité entre les variables dans le nouveau espace. Il est commun d'utiliser comme indice de similarité le coefficient de corrélation ou la covariance.

La quantité d'information qu'est mesuré représenter sa variance, les variances associées à chaque composante principale sont classées par ordre décroissant d'une variables indépendante. La première composante principale absorbe le plus de variance possible, la 2eme composante principale absorbe de plus de variance possible parmi la variance restante, ... etc.

#### **d. Détermination de la structure du modèle ACP**

L'analyse en composantes principales recherche une approximation de la matrice initiale des données  $X$  par une matrice de rang inférieur issue d'une décomposition en valeurs singulières.

Dans le cadre de l'application de l'ACP à la réduction des données, le nombre de composantes a un impact significatif sur la phase finale de classification. Si peu de composantes sont utilisées, on risque de perdre des informations contenues dans les données de départ. S'il y a beaucoup de composantes. Il correspondre aux valeurs propres les plus faibles parmi celles retenues dans le modèle.

Pour l'étude des individus, on va chercher les ressemblances entre eux à partir le point de vue de variables, on construire des groupes des individus homogènes et identiques.

On peut mesure les ressemblances de corrélation entre les variables à partir du coefficient des variables, si on trouve deux variables apportent la même information on dit que le coefficient de corrélation très élevé.

Afin d'établir un lien entre les individus et des variables, nous construisons des groupes d'individus et nous allons chercher à caractérisiez des différentes classes d'individus par leurs variables.

Pour trouver le meilleur plan, on va chercher le premier meilleur axe et ensuite un axe orthogonal sur le premier axe, le deuxième axe, ... etc. A chaque fois l'axe doit être orthogonal à tous les précédentes. Ces axes nouveaux sont des axes principaux notés  $F_1, F_2, \dots$  etc.

On choisit deux axes orthogonaux, chaque individu nous collectons ses coordonnées sur l'axe horizontales et sur l'axe verticale, les deux axes sont appelés aussi les facteurs ou les dimensions de l'ACP

## II.2.2 Les méthodes de l'apprentissage automatique supervisé

Il existe une large variété d'algorithmes de l'apprentissage automatique [15]. Certains sont toutefois plus couramment utilisés que d'autres. Telle que la régression linéaire :

C'est un algorithme d'apprentissage automatique généralement supervisé où la sortie prédite est continue et a une pente constante. Il est utilisé pour prédire des valeurs dans une plage continue.

Il existe deux types principaux :

- **La régression simple**

Ce type utilise la forme traditionnelle d'interception de pente, où  $\theta_0, \theta_1$  sont les variables de l'algorithme essaiera d'apprendre pour produire les prédictions les plus précises.

$X$  représente nos données d'entrée et  $Y$  représente notre prédiction (éq II-1) :

$$Y = \theta_0 + \theta_1 X \quad \text{éq II-1 : Equation linéaire}$$

- **la régression multi variée**

Une équation linéaire multi-variable plus complexe pourrait rassembler à ceci, où  $\theta_0, \theta_1, \theta_2, \theta_3$  représente les coefficients, ou poids, que notre modèle essaiera d'apprendre (éq II-2).

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \dots + \theta_i X_i \quad \text{éq II-2 : Equation linéaire multi variée}$$

Les trois variables  $X_1, X_2, X_3$  représentent les attributs, ou éléments d'information distincts.

## II.2.3 Fonction cout

Nous permettons d'évaluer la performance de notre modèle en mesurant les erreurs entre  $h(X)$  et  $Y$ . Cependant, si la prédiction  $h(X)$  est éloignée de  $Y$ , alors cette erreur n'est pas pratique. Il existe plusieurs méthodes pour mesurer les erreurs entre les prédictions  $h(X)$  et les valeurs  $Y$  des données [19]. Par exemple :

on calcule le carré de la différence (éq II-3) :

$$\text{cost}(i) \approx (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad \text{éq II-3 : Le carré de la différence}$$

Pour la régression linéaire, la fonction cout J va être la moyenne de toutes nos erreurs (éq II-4).

$$J(\theta) = \frac{(h_{\theta}(x^1) - y^1)^2 + (h_{\theta}(x^2) - y^2)^2 + \dots + (h_{\theta}(x^i) - y^i)^2}{m} \quad \text{éq II-4 : La fonction cout}$$

La fonction viendra par la suite [19] (éq II-5) :

$$J(\theta) = \frac{-1}{m} \sum_{m=1}^m \sum_{k=1}^k [y_k^{(i)} \log((h_{\theta}(x^{(i)}))_k) + (1 - y_k^{(i)}) \log(1 - (h_{\theta}(x^{(i)}))_k)] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{i+1}} (\theta_{j,i}^{(l)})^2 \quad \text{éq II-5 : La fonction cout}$$

$X_1, X_2, X_3$  : Les données d'entrée

$\Theta_1, \Theta_2, \Theta_3$  : Les poids.

$h(x)$  : La valeur attendue.

$Y$  : La donnée de sortie réel.

$J(\Theta)$  : La fonction de coût.

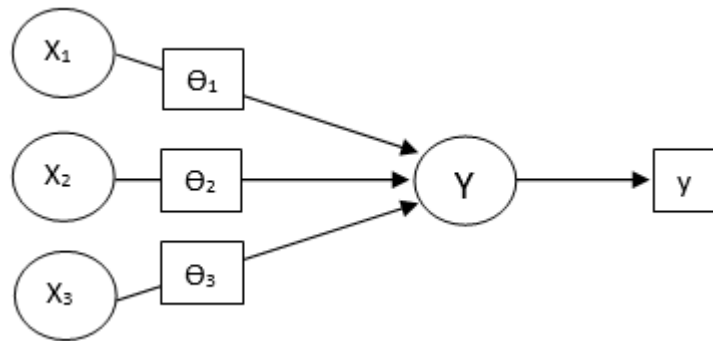
## II.2.4 L'algorithme la descente de gradient

C'est un algorithme d'optimisation qui permet de converger vers le minimum de n'importe quelle fonction convexe. Ainsi, l'algorithme calcul le gradient de la fonction cout pour se déplacer progressivement vers son minimum et pour trouver les meilleurs paramètres de notre modèle, c'est-à-dire les paramètres qui retourne les plus petites erreurs. Plus précisément, considérons l'algorithme de la descente de gradient, qui commence par un certain  $\theta$  initial, et effectue à plusieurs reprises la mise à jour à partir (éq II-6) :

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} j(\theta) \quad \text{éq II-6 : La fonction de gradient descente}$$

## II.2.5 Le réseau de neurones

Les réseaux de neurones ont été développés comme un modèle mathématique générique afin de modéliser les neurones biologiques. Ils comportent un certain nombre d'éléments de traitement d'information appelés neurones [20]. Chaque neurone a son propre état interne interprété par la fonction d'activation. Il envoie son activation aux autres neurones sous forme de signaux. La connexion entre les neurones est réalisée via des liens orientés et pondérés (figure 8).



**Figure 8 :** Neurone artificiel

Le neurone Y reçoit les entrées de  $X_1$ ,  $X_2$  et  $X_3$  qui ont comme valeurs de sortie  $x_1$ ,  $x_2$  et  $x_3$ . Les poids des liens de connexion de  $X_1$ ,  $X_2$  et  $X_3$  sont  $\theta_1$ ,  $\theta_2$  et  $\theta_3$ . La valeur d'entrée de neurone Y est montré dans (éq II-7) :

$$y = \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3$$

**éq II-7 :** Equation linéaire

## II.2.6 Arbre de décision

L'arbre de décision (AD) est une catégorie d'arbres utilisée dans l'exploration de données. Ils emploient une représentation hiérarchique de la structure des données sous forme des séquences de décisions (tests) en vue de la prédiction d'un résultat ou d'une classe. Chaque individu (ou observation), qui doit être attribué(e) à une classe, est décrit(e) par un ensemble de variables qui sont testées dans les nœuds de l'arbre [21].

### a) Apprentissage avec les arbres de décision

On considère d'abord le problème de classement. Chaque élément  $x$  de la base de données est représenté par un vecteur multidimensionnel  $(x_1, x_2, \dots, x_n)$  correspondant à l'ensemble de

variables descriptives du point. Chaque nœud interne de l'arbre correspond à un test fait sur une des variables  $x_i$  :

- Variable catégorielle : génère une branche par valeur de l'attribut ;
- Variable numérique : test par intervalles de valeurs.

Une fois l'arbre construit, classer un nouvel candidat se fait par une descente dans l'arbre, de la racine vers une des feuilles (qui encode la décision ou la classe). A chaque niveau de la descente on passe un nœud intermédiaire où une variable  $x_i$  est testée pour décider du chemin (ou sous-arbre) à choisir pour continuer la descente.

### b) Principe de la construction

Au départ, les points dès la base d'apprentissage sont tous placés dans le nœud racine. Une des variables de description des points est la classe du point ; cette variable est dite « variable cible ». La variable cible peut être catégorielle ou valeur réelle. Chaque nœud est coupé donnant naissance à plusieurs nœuds descendants [21]. Un élément de la base d'apprentissage situé dans un nœud se retrouvera dans un seul de ses descendants.

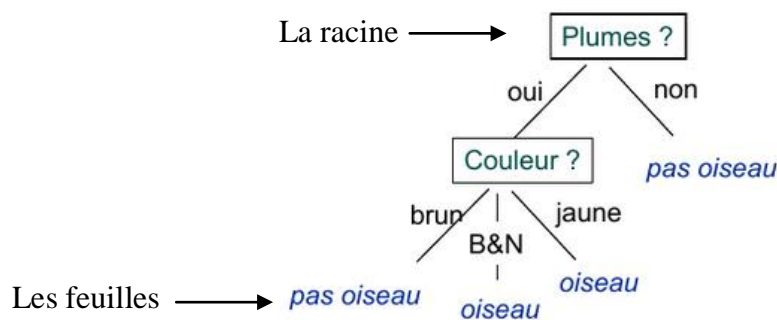
- L'arbre est construit par partition récursive de chaque nœud en fonction de la valeur de l'attribut testé à chaque itération. Le critère optimisé est l'homogénéité des descendants par rapport à la variable cible. La variable qui est testée dans un nœud sera celle qui maximise cette homogénéité.
- Le processus s'arrête quand les éléments d'un nœud ont la même valeur pour la variable cible (homogénéité).

Exemple :

Le tableau 4 représente les conditions pour construire un arbre de décision et montré au suivant dans la figure 9 :

**Tableau4** : Les conditions de l'arbre de décision

Oiseaux	Couleur	Ailes	Plumes	Sonar	Concept
Faucon	Jaune	Oui	Oui	Non	Oiseau
Pigeon	Noir et Blanc	Oui	Oui	Non	Oiseau
Chauve-souris	Brun	Oui	Non	Oui	Pas oiseau



**Figure 9** : Un arbre de décision

Si plumes = non oiseau, alors classe = pas-oiseau

Ou (Si plumes = oui et couleur = brun, alors classe = pas-oiseau)

Ou (Si plumes = oui et couleur = noir et blanc, alors classe = oiseau)

Ou (Si plumes = oui et couleur = jaune, alors classe = oiseau)

### II.2.7 La forêt aléatoire

C'est composé de plusieurs arbres de décision, travaillant de manière indépendante sur une vision d'un problème. Chacun produit une estimation, et c'est l'assemblage des arbres de décision et de leurs analyses, qui va donner une estimation globale. En somme, il s'agit de s'inspirer de différents avis, traitant un même problème. Chaque modèle est distribué de façon aléatoire aux sous-ensembles d'arbres décisionnels [22]. La forêt aléatoire est ce qu'on appelle une méthode d'ensemble c'est-à-dire qu'elle "met ensemble" ou combine des résultats pour obtenir un super résultat final.

Elles peuvent être composées de plusieurs dizaines voire centaines d'arbres. Le nombre d'arbre est un paramètre que l'on ajuste généralement par validation croisée. Pour faire court, la validation croisée est une technique d'évaluation d'un algorithme de l'apprentissage automatique consistant à entraîner et tester le modèle sur des morceaux du dataset de départ

Chaque arbre est entraîné sur un sous-ensemble du dataset et donne un résultat Les résultats de tous les arbres de décision sont alors combinés pour donner une réponse finale. Chaque arbre "vote" (oui ou non) et la réponse finale est celle qui a eu la majorité de vote.

### II.2.8 Les données

Une donnée est la représentation d'une information dans un programme. Les données, souvent codées, décrivent les éléments du logiciel tels qu'une entité (chose)... etc. Les données peuvent être conservées et classées sous différentes formes tels que :

- **La structure des données**

Une structure des données est un format spécial destiné à organiser, traiter, extraire et stocker des données. S'il existe plusieurs types de structures plus ou moins complexes, tous visent à organiser les données pour répondre à un besoin précis, afin de pouvoir y accéder et les traiter de façon appropriée.

Une structure de données peut être sélectionnée ou conçue pour stocker des données de manière à pouvoir manipuler ces dernières à l'aide de plusieurs algorithmes. Chaque structure de données contient des informations sur la valeur des données, les relations entre elles et les fonctions applicables.

Les structures de données sont souvent classées d'après leurs caractéristiques :

- **Linéaires ou non linéaires**

Cette caractéristique indique si les éléments de données sont organisés, comme dans un tableau, ou de façon non ordonnée, comme dans un graphe.

- **Homogènes ou non homogènes**

Cette caractéristique indique si tous les éléments de données d'un référentiel spécifique sont du même type ou de types différents.

- **Les types de structures des données**

- **Tableau**

Un tableau stocke un ensemble des éléments. Les éléments de même type sont stockés ensemble afin de faciliter le calcul de leur emplacement ou leur extraction. La longueur d'un tableau peut être fixe ou variable.

- **Liste chaînée**

Une liste chaînée stocke un ensemble d'éléments de façon linéaire. Chaque élément ou nœud d'une liste chaînée contient un élément de données ainsi qu'une référence, ou lien, vers l'élément suivant de la liste.

- **Arbre**

Un arbre stocke un ensemble d'éléments sous une forme hiérarchique abstraite. Chaque nœud est relié aux autres et peut contenir plusieurs sous-valeurs appelées enfants.

- **Graphe**

Un graphe stocke un ensemble d'éléments de façon non linéaire. Il se compose d'un ensemble fini de nœuds, appelés sommets, et de lignes, les arêtes, qui relient les sommets entre eux. Les graphes permettent notamment de représenter des systèmes réels, comme des réseaux informatiques.

### ➤ **Trie**

Une trie est une structure de données qui stocke des chaînes en tant qu'éléments de données pouvant être représentés visuellement sous forme graphique.

#### • **Préparation des données**

La préparation des données par cinq étapes suivantes [16] :

### ➤ **La Sélection**

Dans cette étape, la sélection et l'intégration des données cibles provenant éventuellement de nombreuses sources différentes et hétérogènes sont effectuées. Ensuite, le sous-ensemble correct de variables et d'échantillons de données pertinents pour la tâche d'analyse est extrait de la base de données.

### ➤ **Prétraitement**

Les actions à entreprendre incluent les suivantes :

- Collecte des données ou informations requises pour modéliser ;
- Traitement des valeurs aberrantes ou suppression du bruit ;
- Utiliser les connaissances préalables du domaine pour supprimer les incohérences et les doublons des données ;
- Choix des stratégies de traitement des données manquantes.

### ➤ **Transformation**

Au cours de cette étape, les données sont transformées ou consolidées dans des formes adaptées à l'exploration, c'est-à-dire à la recherche de fonctionnalités utiles pour représenter les données en fonction de l'objectif de la tâche. Il existe différentes techniques de transformation de données :

- Normalisation, qui implique une mise à l'échelle min-max ou un score z
- Des techniques de réduction des données sont appliquées pour produire une représentation réduite des données (volume plus petit qui maintient étroitement l'intégrité des données d'origine)

### ➤ **Data Mining**

Dans cette étape, des algorithmes d'apprentissage automatique sont appliqués pour extraire des modèles de données. Des méthodes d'exploration telle que la moyenne, la médiane, le mode, l'écart type, la description de classe/concept et les techniques graphiques de tracés de faible dimension peuvent être utilisées pour comprendre les données. Des modèles prédictifs telle que la classification ou la régression peuvent être utilisées pour prédire l'événement ou la valeur future. L'analyse typologique peut être utilisée pour comprendre l'existence de groupes

similaires. Sélectionnez les méthodes les plus appropriées à utiliser pour la recherche de modèles et de modèle.

➤ **Interprétation / Évaluation**

Cette étape se concentre sur l'interprétation des modèles extraits pour les rendre compréhensibles par l'utilisateur, comme le résumé et la visualisation. Le modèle ou les modèles extraits sont interprétés. Les modèles sont une structure locale qui ne fait des déclarations que sur des régions restreintes de l'espace couvert par les variables. Alors que le modèle est des structures globales qui font des déclarations sur n'importe quel point de l'espace de mesure, c'est-à-dire  $Y = m X + C$  (modèle linéaire).

# **Chapitre III**

## **Résultats et discussions**

### **III.1 Introduction**

L'apprentissage automatique explore l'étude et la construction d'algorithme capable d'apprendre de faire des prédictions sur les données. De tels algorithmes fonctionnent en construisant un modèle à partir d'exemples des entrées afin de faire des prédictions ou des décisions basées sur des données, plutôt que de suivre des instructions de programmes strictement statique.

Il existe de nombreux logiciels pour utiliser et représenter les statistiques de l'analyse de composantes principales. Par exemple : XL stat, Spss, ... etc. Dans notre étude nous choisissons les logiciels de Spyder (python), pour l'analyse à travers l'ACP. Python est un langage de programmation qui peut s'utiliser dans de nombreux contextes et s'adapter à tout type d'utilisation grâce à des bibliothèques spécialisées. Il est cependant particulièrement utilisé comme langage de script pour automatiser des tâches simples.

### **III.2 La préparation des données**

L'apprentissage automatique est mis en œuvre en suivant systématiquement un processus qui peut – être résumé comme suit [16] :

#### **➤ La sélection des données**

L'apprentissage automatique dans la recherche sur les composées, comme dans d'autres domaines similaires, peut être mis en œuvre en utilisant les données générées en interne (expérimentalement ou simulation) ou extraites d'autres sources (de publications scientifiques ou de bases de données).

Dans notre cas, la sélection et l'intégration des données cibles provenant éventuellement de nombreuses sources différentes et hétérogènes sont effectuées. Ensuite, le sous-ensemble correct de variables et d'échantillon de données patients pour la tâche d'analyse est extrait de publications scientifique, sont des données expérimentales.

#### **➤ Prétraitement**

Les ensembles de données du monde réel sont souvent incomplète, c'est –à-dire que les valeurs d'attribut seront manquantes, et incohérent, ce qui signifie qu'il existe des écart entre les données collectées. Le prétraitement et le nettoyage devraient améliorer la qualité des données et des résultats. Les actions à entreprendre incluent les suivants :

#### **• le traitement des valeurs**

Dans notre travail toutes les valeurs sont expérimentales sauf quelques valeurs de l'énergie de gap, sont des valeurs théoriques. L'erreur entre la valeur théorique et la valeur expérimentale par-rapport GGA égale à 40% .C'est –à-dire erreur GGA diminue la valeur de l'énergie de gap de 40% par-rapport à la valeur expérimentale.

- **Les stratégies de traitement des données manquantes**

Les données peuvent induire en erreur ou créer des problèmes pour l'analyse des données. Afin d'éviter de tels problèmes, il faut imputer les données manquantes. Il existe quatre techniques les plus couramment utilisées pour l'imputation des données [16] :

1. Supprimer
2. Remplacer par une valeur moyenne.
3. Remplacer par une valeur aléatoire.
4. Utilisation d'un modèle prédictif

Dans notre étude on a choisi la deuxième méthode c'est remplacé par une valeur moyenne qui consiste à combler les vides (les valeurs qui manque) par une moyenne calculé à la même colonne .

➤ **Transformation**

Les données sont transformées ou consolidées dans des formes adaptées à l'exploration, c'est-à-dire à la recherche de fonctionnalités utiles pour représenter les données en fonction de l'objectif de la tâche. Il existe différentes techniques de transformation, parmi elles :

- **La normalisation**

Une échelle de mesure pour différentes variables varie, de sorte qu'une analyse avec la mesure brute pourrait être artificielle biaisée vers les variables des valeurs absolues plus élevées. Amener tous les différents types d'unités variables dans le même ordre de grandeur.

Pour s'affranchir des effets d'échelle dus à l'hétérogénéité éventuelle des variables, ces dernières sont en général normalisées, c'est à dire que chaque colonne est divisée par son écart-type toutes sont dès lors exprimées dans la même échelle standard.

Par exemple:

$$\theta_1 \times T_{déc} \longrightarrow 300$$

$$\theta_5 \times r_a \longrightarrow 1.2$$

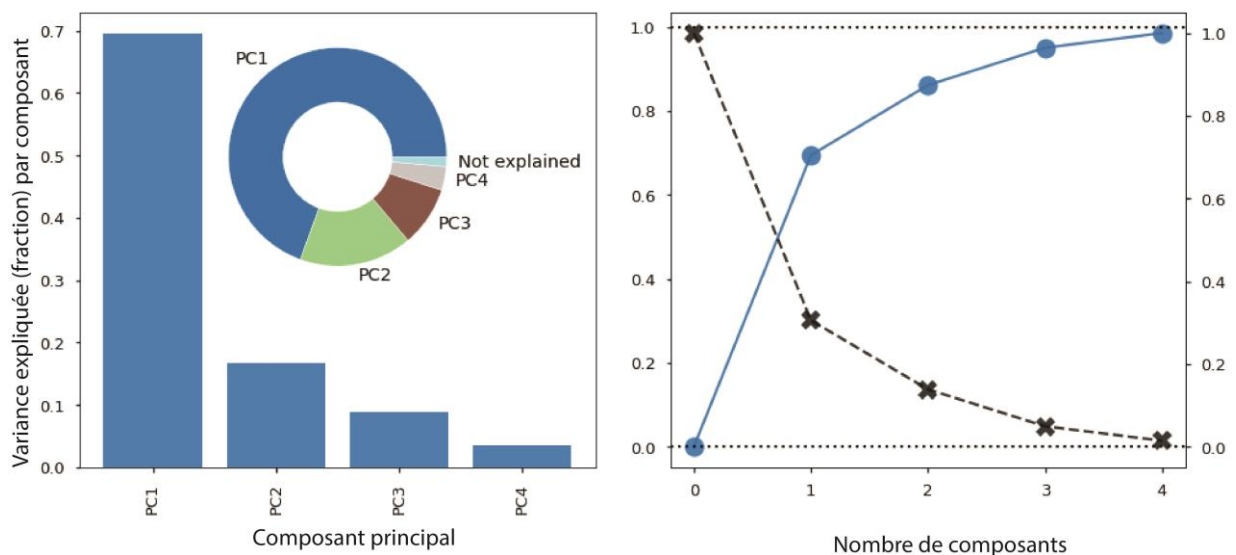
### III.3 Analyse par l'approche de datamining

Dans cette partie on présente les résultats de notre étude, basée sur l'analyse ACP. Pour identifier les propriétés des matériaux, nous construisons une base de données pour les iodates métalliques, y compris le paramètre de domaine de transparence, la température de décomposition, énergie de gap, l'intensité de SHG, point de fusion,...etc.

La base des données contient 41 composés d'iodates métalliques chacun avec 17 variables, elle a été construite en se basant sur des valeurs obtenus à partir des valeurs expérimentales (les articles)

#### III.3.1 Choix des axes

Pour un plan factoriel donné, on regarde le pourcentage de l'information du nuage initial retranscrite par le plan factoriel. Ainsi, un axe portant moins de 10% de l'information sera rarement intéressant, les axes sont rangés dans l'ordre décroissant, de telle manière que le premier plan factoriel - constitué par les deux premiers axes - soit toujours celui qui est le plus riche en renseignements sur les propriétés du nuage étudié. Dans notre exemple, les axes F1 (PC1) et F2 (PC2) sont les plus riches en information et montré dans la figure 10 .



**Figure 10 :** Représentation graphique des axes principaux

La Figure 10 montre que PC1 porte 65% de l'information et l'axe PC2 porte 25% de l'information pour la même base, donc nous récupérons 90% de l'information. Autrement dit,

nous avons un bon résumé qui synthétise presque totalement les 4 variables. La perte d'information de 2% est répartie sur les autres PC avec des valeurs propres faibles, ce qui signifie qu'il y a très peu d'informations perdues et il n'est pas nécessaire d'aller visualiser ces PC.

### III.3.2 Matrice de corrélation

Suite à une analyse en composantes principales, le premier résultat intéressant à décrire est la matrice de corrélation qui permet d'analyser les relations binaires qui existent entre les différentes variables sélectionnées. Cette matrice est présentée dans la figure 11 :

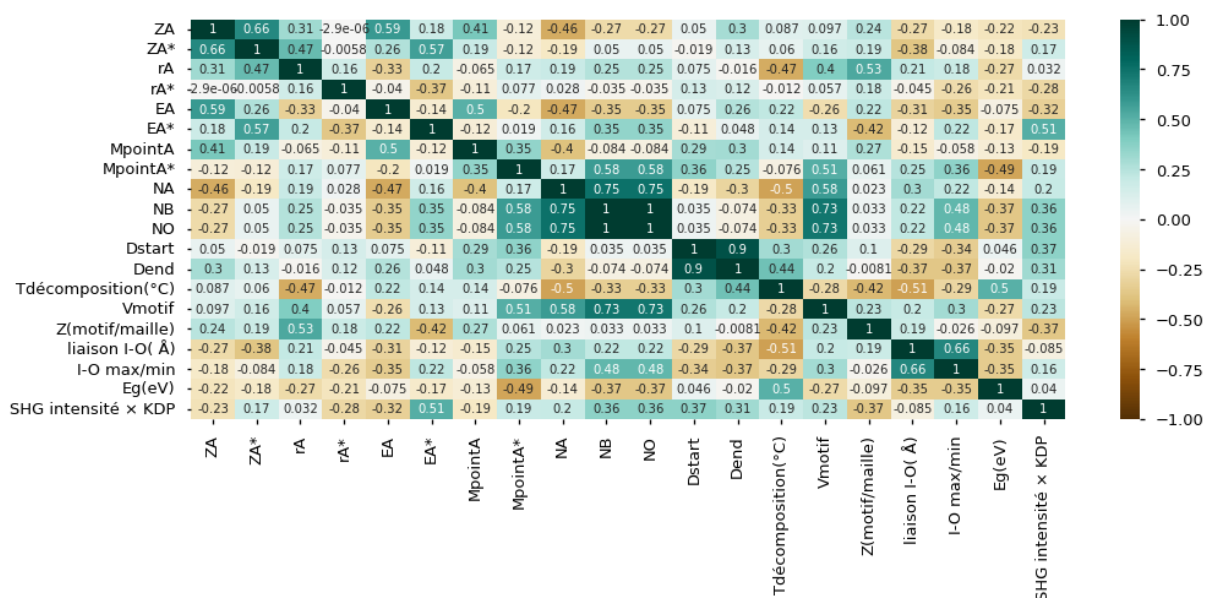


Figure 11 : La matrice de corrélation entre les différentes variables

**ZA** : Le nombre atomique de métal A

**ZA'** : Le nombre atomique de métal A'

**ra** : Le rayon ionique de métal A

**ra'** : Le rayon ionique de métal A'

**EA** : Electronégativité de métal A

**EA'** : Electronégativité de métal A'

**Mpoint A** : Point de fusion du métal A

**Mpoint A'** : Point de fusion du métal A'

**NA** : Le nombre des éléments de métal A

**NB** : Le nombre des éléments de métal A'

**NO** : Le nombre des éléments d'oxygène

**T déc** : La température de décomposition

**Dstart** : Le début de domaine de transparence

**Dend** : Le finale de domaine de transparence

**Vmotif** : Le volume de la maille

**Z(motif/ maille)** : Le nombre de motif par maille

**I-O(max\min)** : La longueur de la liaison max /min

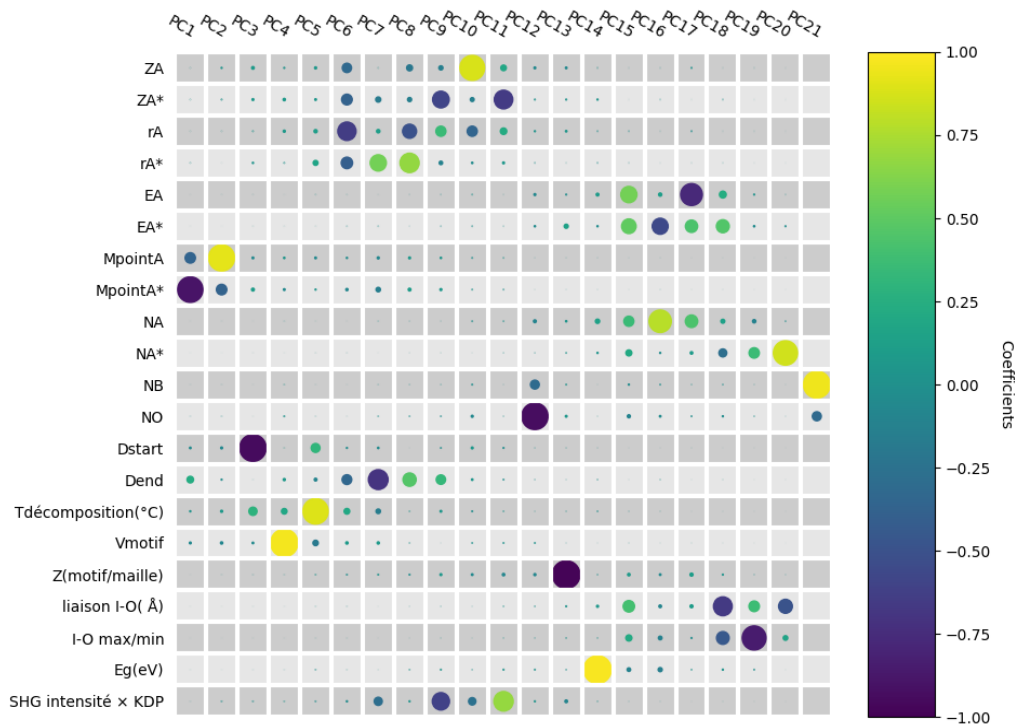
On remarque que ( Figure 11) :

- Une corrélation positive entre le nombre atomique ZA et l'électronégativité EA avec une valeur égale 0.5.
- Une corrélation positive entre le nombre atomique ZA\* et l'électronégativité EA\* avec une valeur égale 0.57.
- Une corrélation positive entre le volume de motif et le nombre d'oxygène NO et le nombre d'iode NB avec la même valeur 0.76.
- Une corrélation positive entre le point de fusion MpointA\* et le nombre d'iode NB avec la valeur 0.58.
- Une corrélation positive entre le départ de la domaine de transparence et le finale de cette domaine avec une valeur égale 0.65.
- Une corrélation positive entre le volume de motif et le nombre d'oxygène NO et le nombre d'iode NB avec une même valeur égale 0.76 .
- Une corrélation négative entre l'énergie de gap et le nombre et le nombre atomique ZA avec la valeur -0.51 .
- Une corrélation négative entre l'énergie de gap et le nombre et le nombre atomique ZA\*avec la valeur -0.49 .
- Une corrélation négative entre l'énergie de gap et l'électronégativité EA\*avec une valeur de -0.59.
- Une corrélation nulle entre :
  - Rayon ionique rA\* et le nombre atomique ZA .
  - La température de décomposition et la valeur de I-O max/min .
  - Départ de domaine de transparence et Dstart et la longueur de la liaison I-O .

La matrice de corrélation nous permet de connaître la relation qui lie les variables entre elle , qui évolue de la même manière , négativement ou positivement . Cette fonctionnalité nous aide ensuite à choisir les variables sur lesquelles baser la prédiction.

### III.3.3 Coordonnées des variables

Ce sont les coordonnées dans le nouvel espace qui exprime les corrélations entre les variables d'origine. La figure 12 représente les coordonnées des variables de notre étude.



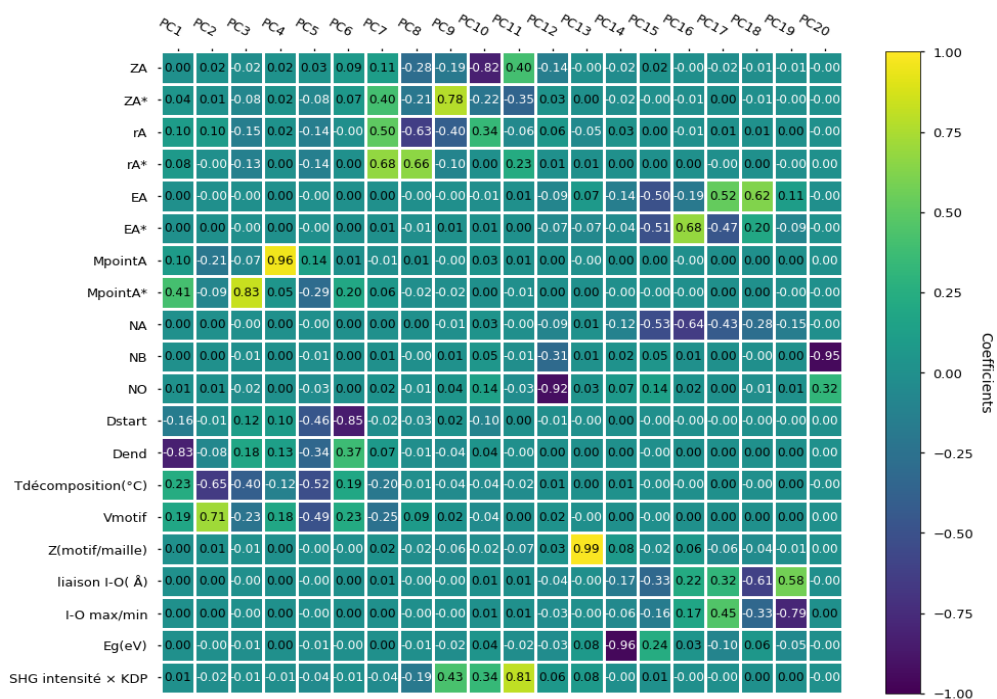
✓ **Figure 12 :** La quantité des informations des variables dans les nouvelles axes (Représentation qualitative)

Dans notre cas on peut pas représenter la plupart des variables dans seulement deux axes principaux PC1 et PC2 car les variables sont dispersés dans toute la matrice.

A travers la figure 12 précédente on observe que :

- L'axe PC1 donne des informations sur M pointA\* et M pointA , par contre l'axe PC6 donne des informations sur ZA , ZA\* , rA et rA\*.
- Les axes PC2, PC4 ,PC5 ,PC10 ,PC14 ,PC21 donnent une grande distribution positive pour MpointA , Vmotif ,T Décomposition , ZA , Eg , NB respectivement .
- Les axes PC1 , PC3 , PC12 , PC13 , PC17 , PC19 donnent une grande distribution négative pour MpointA\* , Dstart , NO , Z(motif/maille) , EA et I-O max/min respectivement.

Cette méthode permet la visibilité de différentes relations qui existe entre les variables figure 13 :



**Figure 13** La quantité des informations des variables dans les nouvelles axes ( Représentation quantitative )

Malgré qu'on a utilisé les mêmes données et les informations des variables, on a obtenus deux matrices avec deux différentes représentations (voir :figure 12 et figure 13). Les résultats ont été différents qu'on n'a pas pu donner une explication.

### III.4 Prédiction

Dans cette partie, on a essayé de prédire deux propriétés à savoir l'énergie de gap et l'énergie de formation pour des matériaux iodates en utilisant des caractéristiques élémentaires tels que l'électronégativité, les rayons ioniques ...etc, mentionné déjà dans la section descriptive. Nous avons choisi un modèle linéaire qui est la régression linéaire et des modèles non linéaires comme les réseaux de neurones et la forêt aléatoire.

Les résultats obtenus étaient trop mauvais (comme on s'y attendait dans la première partie), pour résoudre ces problèmes, nous devrions d'abord définir les problèmes derrière cette observation.

- Les descripteurs ne suffisent pas à décrire la propriété.
- La relation entre les descripteurs et les propriétés ciblées n'est pas simple.
- L'ensemble de données est trop petit.

#### ➤ Les données

Le Tableau 5 représente la nouvelle base de données qui contient 218 composés stables d'iodates métalliques ces données calculées par DFT.

**Tableau 5** : Des données calculées par DFT.

NO	mp_id	Composition	L'énergie de gap	L'énergie de formation
1	mp-27437	Zr(IO3)4	3,2024	-1,340766883
2	mp-683980	Ho(IO3)3	3,31	-1,497098315
3	mp-23360	Zn(IO3)2	3,2192	-1,067330469
4	mp-38322	Mg(IO3)2	3,4935	-1,410993213
5	mp-31259	Bi(IO3)3	2,8849	-1,017769216
6	mp-27640	Cd(IO3)2	3,2733	-1,039217029
7	mp-572672	Co(IO3)2	2,0575	-1,008565358
8	mp-559637	Dy(IO3)3	3,4395	-1,500491152
.	.			
.	.			
.	.			
214	mp-556129	Cs2Zr(IO3)6	3,4585	-1,379619969
215	mp-581868	NaY(IO3)4	3,3077	-1,458033464
216	mp-1196607	NaIn(IO3)4	3,0347	-1,149672023
217	mp-1200560	CsAu(IO3)4	1,949	-0,928672456
218	mp-1229194	AgAu(IO3)4	1,2556	-0,723860553

Dans cette partie, on va créer un modèle prédictive à partir l'apprentissage automatique, permet de chercher des nouvelles valeurs de l'énergie de formation et l'énergie de gap s'ils manquent ou prédire les propriétés d'autres composés de même famille.

#### III.4.1 Energie de formation

L'énergie de formation ou l'enthalpie standard de formation est la différence d'enthalpie mis en jeu lors de la formation d'une mole du composé à partir des corps simples, pris dans l'état standard est stables à la température considérée T. Dans cette partie, nous avons utilisé ces valeurs en prédiction et nous expliquons leurs résultants plus tard. Pour cela, nous partons du Tableau 6 suivants:

**Tableau 6** : L'énergie de formation de différents composés

Composition	L'énergie de formation
Zr(IO3)4	-1,340766883
Ho(IO3)3	-1,497098315
Zn(IO3)2	-1,067330469
Mg(IO3)2	-1,410993213

Bi(IO3)3	-1,017769216
Cd(IO3)2	-1,039217029
Co(IO3)2	-1,008565358
Dy(IO3)3	-1,500491152
.	
.	
.	
Cs2Zr(IO3)6	-1,379619969
NaY(IO3)4	-1,458033464
NaIn(IO3)4	-1,149672023
CsAu(IO3)4	-0,928672456
AgAu(IO3)4	-0,723860553

Nous utilisons la bibliothèque CBFV Tableau 7 pour convertir la composition en valeur apprenable par machine.

**Tableau 7 :** Les descripteurs utilisés pour la prédiction

No	0	1	2	3	4	...	264
	<b>Le moyen de nombre atomique</b>	<b>Le moyen de la masse atomique</b>	<b>Le moyen de la période</b>	<b>Le moyen du groupe</b>	<b>Le moyen de la famille</b>		<b>La chaleur spécifique au mode <math>_{(J/g_K)}</math></b>
0	22,667	51,478	3,333	11,111	5,111		0,92
1	28,000	66,647	3,167	16,000	6,833		0,92
2	39,333	93,936	4,333	12,000	6,000		0,19
3	20,471	46,520	2,882	15,529	7,059		0,92
4	46,750	112,689	4,500	13,750	6,750		0,214
.							
.							
.							
218	16,933	36,946	2,933	12,133	5,533		0,92

### 1. Elimination des variables

Pour la sélection des variables, on concentre uniquement sur les descripteurs directement liés à nos propriétés ciblées. Pour cela on utilise la méthode de Select KBest pour extraire les meilleures fonctionnalités d'un ensemble de données donné, cette méthode sélectionne les fonctionnalités en fonction de K score le plus élevé. En modifiant le paramètre 'score -fun' nous pouvons appliquer la méthode aux données. la sélection des meilleurs fonctionnalités est un processus important lorsque nous préparons un grand ensemble de données pour la formation,

cela nous aide à éliminer la partie la moins importante des données et réduire le temps de formation.

## **2. Les données de test et les données d'entraînement**

Comme mentionné précédemment, nous avons utilisé select kbest pour classer toutes les fonctionnalités, afin d'obtenir des classements stables.

Dans cette partie, l'apprentissage automatique permet d'identifier les propriétés sont intégrées dans l'ensemble de l'entraînement ou l'ensemble de test, de sorte que nous introduisons des algorithmes qui permettent à l'apprentissage automatique de choisir une classe sur laquelle s'entraîner comme une étape initiale à partir de l'ensemble des propriétés présentées, puis nous choisissons une classe pour tester.

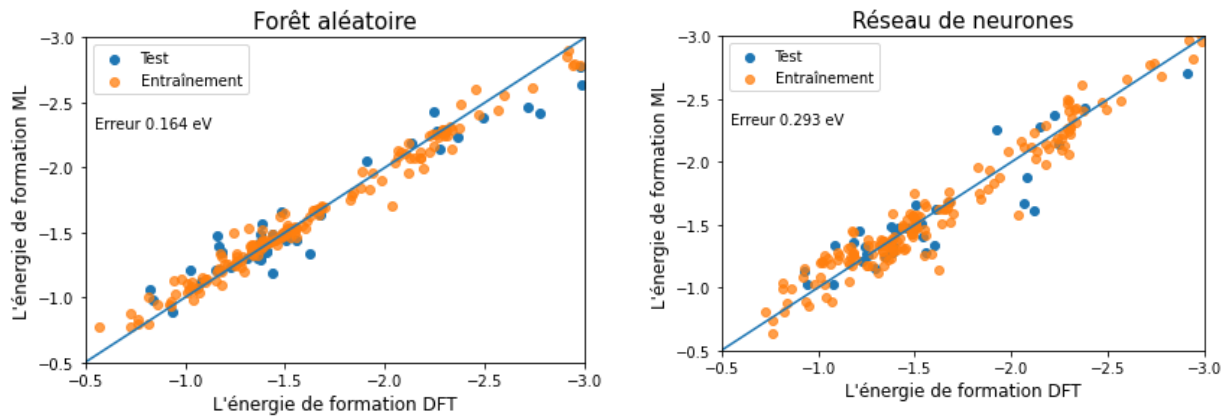
Chaque ensemble de données a été divisé au hasard en un ensemble d'entraînement (80%) et un ensemble de test (20%).

## **3. La Prédiction**

Vu que l'énergie de formation est un paramètre très important dans l'étude de stabilité, on va essayer de prédire cette grandeur à partir des propriétés des éléments regroupées dans la base de données précédente.

Tout d'abord, il faut trouver une relation entre l'énergie de formation et les autres propriétés en faisant appel les méthodes : régression linéaire, le réseau de neurones et la forêt aléatoire.

La figure 14 représente Les meilleurs résultats qui sont obtenus avec l'utilisation de la forêt aléatoire et le réseau de neurone avec score 0.91 et 0.72 respectivement. La prédiction comparée à des mesures calculées par DFT donne des résultats très proches.



**Figure 14 :** La comparaison entre les valeurs de l'énergie de formation de l'apprentissage automatique et l'énergie de formation calculée par DFT à partir la forêt aléatoire et le réseau de neurones.

La théorie des bandes d'énergie explique clairement la différence de comportement entre les différents matériaux. La largeur de la bande interdite appelle aussi l'énergie de gap est l'intervalle d'énergie situé entre l'état de plus basse énergie de bande de conduction et plus haute énergie de bande de valence selon la valeur de gap, on distingue :

Les conducteurs  $E_g=0$ .

Les semi-conducteurs  $E_g < 3\text{eV}$ .

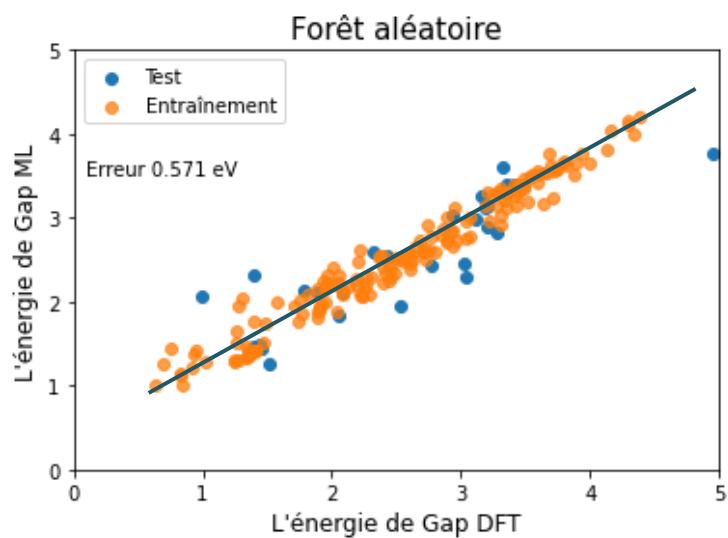
Les isolants  $E_g > 3\text{eV}$ .

Pour discuter les valeurs d'énergie de gap on utilise le tableau 8 suivant :

**Tableau 8 :** L'énergie de gap de différents composés

Composition	L'énergie de gap
Zr(IO3)4	3,2024
Ho(IO3)3	3,31
Zn(IO3)2	3,2192
Mg(IO3)2	3,4935
Bi(IO3)3	2,8849
Cd(IO3)2	3,2733
Co(IO3)2	2,0575
Dy(IO3)3	3,4395
.	
.	
.	
Cs2Zr(IO3)6	3,4585
NaY(IO3)4	3,3077
NaIn(IO3)4	3,0347
CsAu(IO3)4	1,949
AgAu(IO3)4	1,2556

La figure 15 représente les meilleurs résultats qui sont obtenus avec l'utilisation de la forêt aléatoire avec un score 0.8. La prédiction comparée à mesures calculées par DFT donne des résultats très proches.



**Figure 15 :** La comparaison entre les valeurs de l'énergie de gap de l'apprentissage automatique et l'énergie de gap calculée par DFT à partir la forêt aléatoire.

# **Conclusion générale**

Dans notre travail on s'intéresse à l'étude par la technique dataminig de la famille des iodates mono et bi-métalliques. Ces composés ont des propriétés très intéressantes dans le domaine d'optique non linéaire. L'idée de base était de repérer et représenter les corrélations qui existent entre les propriétés de cette famille. Les propriétés choisi : le nombre atomique, le rayon ionique, électronégativité, point de fusion, le nombre des éléments, la température de décomposition, le début de domaine de transparence, la fin de domaine du transparence, le volume de la maille, le nombre de motif par maille et la longueur de la liaison I-O.

La technique descriptive utilisée est la technique d'analyse en composantes principales ACP., les résultats obtenus nous ont permis d'affirmer selon une analyse ACP , les corrélations suivantes :

- Corrélations positives entre les propriétés (ZA, EA) et entre (MpointA\*, NB).
- Corrélations négatives entre les propriétés (Eg, ZA) et entre (EA\*, Eg).

Dans la deuxième partie prédictive on a utilisé une nouvelle technique d'apprentissage automatique pour la prédiction, cette technique a permis de créer deux modèles : la forêt aléatoire et le réseau de neurone pour étudier l'énergie de formation les résultats obtenus sont très acceptables avec un score de 0.9 et 0.72 respectivement. Le score obtenu pour l'énergie de gap prédit par le modèle la forêt aléatoire est de 0.8.

Grace au modèle forêt aléatoire on peut prédire dans des futurs travaux l'énergie de gap et l'énergie de formation pour la famille iodates métallique à ligands mixte.

# **Les références bibliographiques**

- [1] Z. Hebboul *et al.*, « Synthesis, Characterization, and Crystal Structure Determination of a New Lithium Zinc Iodate Polymorph  $\text{LiZn}(\text{IO}_3)_3$  », *Crystals*, vol. 9, n° 9, Art. n° 9, sept. 2019, doi: 10.3390/cryst9090464.
- [2] L. Bastien, « Data Mining ». 22 juin 2021.
- [3] M. Mathieu, « cristaux nonocristaux pour l'optique », in *chimie moléculaire*, Université de Josef Fourier-Grenoble: DEA, 2003.
- [4] Z. Hebboul, « Elaboration et caractérisation des iodates métalliques pour l'optique non linéaire », Centre Universitaire de Laghouat, 2001.
- [5] Z. Julien, « Cristallogénèse et caractérisations de la solution solide Dihydrogenophosphate-Arseniate de 2-Amino-5-Nitropyridinium, cristaux organo-minéraux pour l'optique non linéaire », Université de Josef Fourier-Grenoble I, 1998.
- [6] A. H. A.F.Benhbab, *Optique physique Tome 2*. OPU, 1993.
- [7] M. B. Taouti, « Synthèses et caractérisations structurales d'iodates métalliques, matériaux lasers convertisseurs de fréquences », Université Joseph Fourier-Grenoble et Batna-Algérie, 2008.
- [8] « New Nonlinear Optical Materials: Metal Oxides with Nonbonded Electrons: Journal of Applied Physics: Vol 40, No 7 ». <https://aip.scitation.org/doi/abs/10.1063/1.1658089> (consulté le 29 mai 2022).
- [9] « Absolute sense of pyroelectric  $p_3$  and piezoelectric  $d_{33}$  coefficients in  $3\text{La}(\text{IO}_3)_3 \cdot \text{HIO}_3 \cdot 7\text{H}_2\text{O}$  - ScienceDirect ». <https://www.sciencedirect.com/science/article/abs/pii/S0038109878910177> (consulté le 29 mai 2022).
- [10] C. Svensson, S. C. Abrahams, et J. L. Bernstein, « Crystal structure and pyroelectric coefficient of  $\text{Co}(\text{IO}_3)_2$  and structural relationships among the anhydrous noncentrosymmetric 3d-transition metal iodates », *Journal of Solid State Chemistry*, vol. 36, n° 2, p. 195- 204, févr. 1981, doi: 10.1016/0022-4596(81)90157-2.
- [11] R. Ashok Kumar, D. Rajan Babu, R. Ezhil Vizhi, N. Vijayan, et G. Bhagavannarayana, « Synthesis, Growth and Optical Studies of Mn:  $\alpha\text{-LiIO}_3$  Single Crystal », *Advanced Materials Research*, vol. 584, p. 3- 7, 2012, doi: 10.4028/www.scientific.net/AMR.584.3.
- [12] « Raman scattering structural studies of nonlinear optical  $\text{M}(\text{IO}_3)_3$  ( $\text{M} = \text{Fe}, \text{Ga}, \alpha\text{-In}$ ) and linear optical  $\beta\text{-In}(\text{IO}_3)_3$  | Journal of Nonlinear Optical Physics & Materials ». <https://www.worldscientific.com/doi/abs/10.1142/S0218863514500398> (consulté le 2 juin 2022).
- [13] « Low-Frequency Relaxation Phenomena in  $\alpha\text{-LiIO}_3$ : The Nature and Role of Defects - ScienceDirect ». <https://www.sciencedirect.com/science/article/abs/pii/S0022459602996830> (consulté le 2 juin 2022).
- [14] Hebboul *et al.*, « Synthesis, Characterization, and Crystal Structure Determination of a New Lithium Zinc Iodate Polymorph  $\text{LiZn}(\text{IO}_3)_3$  », *Crystals*, vol. 9, n° 9, p. 464, sept. 2019, doi: 10.3390/cryst9090464.
- [15] R. Russell, « Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python », p. 103.
- [16] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps*. Berkeley, CA: Apress, 2017. doi: 10.1007/978-1-4842-2866-1.
- [17] A. KASSAMBARA, *Practical Guide To Principal Component Methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra*. STHDA, 2017.
- [18] H. Bachir, « Etude par datamining de certaines propriétés physiques de quelques borures, nitrures et carbures de fer », p. 54.
- [19] N. Andrew, « Lecture Notes ».
- [20] P. Marc, *RESEAUX DE NEURONES*. université LAVAL, 2004.
- [21] A. Gueguen et J. P. Nakache, « Méthode de discrimination basée sur la construction d'un arbre de décision binaire », p. 20.
- [22] R. Genuer, « Forêts aléatoires : aspects théoriques, sélection de variables et applications », phdthesis, Université Paris Sud - Paris XI, 2010. Consulté le: 29 mai 2022. [En ligne]. Disponible sur: <https://tel.archives-ouvertes.fr/tel-00550989>

Les valeurs de la base de données dans la partie descriptive obtenus à partir des valeurs expérimentales (les articles ) [23]–[63]

- [23] « 57Fe Mössbauer spectroscopic study and magnetic properties of 1D Fe(IO<sub>3</sub>)<sub>3</sub> particles and their thermal decomposition to α-Fe<sub>2</sub>O<sub>3</sub> - ScienceDirect ». <https://www.sciencedirect.com/science/article/abs/pii/S0167577X18307973> (consulté le 30 mai 2022).
- [24] X.-H. Zhang, B.-P. Yang, C.-L. Hu, J. Chen, Q.-M. Huang, et J.-G. Mao, « A New Anhydrous Polar Rare-Earth Iodate Fluoride, Ce(IO<sub>3</sub>)<sub>2</sub>F<sub>2</sub>, Exhibiting a Large Second-Harmonic-Generation Effect and Improved Overall Performance », *Chem. Mater.*, vol. 33, n° 8, p. 2894- 2900, avr. 2021, doi: 10.1021/acs.chemmater.1c00222.
- [25] F.-F. Mao, C.-L. Hu, J. Chen, et J.-G. Mao, « A Series of Mixed-Metal Germanium Iodates as Second-Order Nonlinear Optical Materials », *Chem. Mater.*, vol. 30, n° 7, p. 2443- 2452, avr. 2018, doi: 10.1021/acs.chemmater.8b00630.
- [26] B. Lagoun, « Ab initio calculation of structural, electronic and optical properties of Hg(IO<sub>3</sub>)<sub>2</sub> », *Physica B*, p. 5, 2014.
- [27] Z. Cao, Y. Yue, J. Yao, Z. Lin, R. He, et Z. Hu, « Bi<sub>2</sub>(IO<sub>4</sub>)(IO<sub>3</sub>)<sub>3</sub>: A New Potential Infrared Nonlinear Optical Material Containing [IO<sub>4</sub>]<sup>3-</sup> Anion », *Inorg. Chem.*, vol. 50, n° 24, p. 12818- 12822, déc. 2011, doi: 10.1021/ic201991m.
- [28] F.-F. Mao, C.-L. Hu, X. Xu, D. Yan, B.-P. Yang, et J.-G. Mao, « Bi(IO<sub>3</sub>)F<sub>2</sub>: The First Metal Iodate Fluoride with a Very Strong Second Harmonic Generation Effect », *Angew. Chem.*, vol. 129, n° 8, p. 2183- 2187, févr. 2017, doi: 10.1002/ange.201611770.
- [29] M. Zhang, X. Su, M. Mutailipu, Z. Yang, et S. Pan, « Bi<sub>3</sub>O<sub>3</sub>F<sub>3</sub>(IO<sub>3</sub>)<sub>4</sub>: The First Metal Oxyiodate Fluoride Featuring a Carbon- nanotube-like Topological Structure with Large SHG Response », *Chemistry of Materials*, p. 7.
- [30] S. D. Nguyen, J. Yeon, S.-H. Kim, et P. S. Halasyamani, « BiO(IO<sub>3</sub>): A New Polar Iodate that Exhibits an Aurivillius-Type (Bi<sub>2</sub>O<sub>2</sub>)<sup>2+</sup> Layer and a Large SHG Response », *J. Am. Chem. Soc.*, vol. 133, n° 32, p. 12422- 12425, août 2011, doi: 10.1021/ja205456b.
- [31] S. D. Nguyen, J. Yeon, S.-H. Kim, et P. S. Halasyamani, « BiO(IO<sub>3</sub>): A New Polar Iodate that Exhibits an Aurivillius-Type (Bi<sub>2</sub>O<sub>2</sub>)<sup>2+</sup> Layer and a Large SHG Response », *J. Am. Chem. Soc.*, vol. 133, n° 32, p. 12422- 12425, août 2011, doi: 10.1021/ja205456b.
- [32] B. Bentría, D. Benbortal, M. Bagieu-Beucher, A. Mosset, et J. Zaccaro, « Crystal engineering strategy for quadratic nonlinear optics. Part II: Hg(IO<sub>3</sub>)<sub>2</sub> », *Solid State Sciences*, p. 7, 2003.
- [33] S. M. Ravi Kumar, N. Melikechi, S. Selvakumar, et P. Sagayaraj, « Crystal growth and characterization of monometallic NLO single crystals of Cd(IO<sub>3</sub>)<sub>2</sub> », *Journal of Crystal Growth*, vol. 311, n° 2, p. 337- 341, janv. 2009, doi: 10.1016/j.jcrysgro.2008.11.058.
- [34] L. Lin *et al.*, « CsZrF<sub>4</sub>(IO<sub>3</sub>): The First Polar Zirconium Iodate with *cis*-[ZrO<sub>2</sub>F<sub>6</sub>] Polyhedra Inducing Optimized Balance of Large Band Gap and Second Harmonic Generation », *Chem. Mater.*, vol. 33, n° 14, p. 5555- 5562, juill. 2021, doi: 10.1021/acs.chemmater.1c00392.
- [35] C. Huang, C.-L. Hu, X. Xu, B.-P. Yang, et J.-G. Mao, « Explorations of a Series of Second Order Nonlinear Optical Materials Based on Monovalent Metal Gold(III) Iodates », *Inorg. Chem.*, vol. 52, n° 19, p. 11551- 11562, oct. 2013, doi: 10.1021/ic401891f.
- [36] X.-A. Chen, H.-P. Xue, X.-A. Chang, H.-G. Zang, et W.-Q. Xiao, « Gallium tris(iodate), Ga(IO<sub>3</sub>)<sub>3</sub> », *Acta Crystallogr C Cryst Struct Commun*, vol. 61, n° 11, p. i109- i110, nov. 2005, doi: 10.1107/S0108270105029835.
- [37] C. Wu *et al.*, « Giant Optical Anisotropy in the UV-Transparent 2D Nonlinear Optical Material Sc(IO<sub>3</sub>)<sub>2</sub>(NO<sub>3</sub>) », *Angew. Chem.*, vol. 133, n° 7, p. 3506- 3510, févr. 2021, doi: 10.1002/ange.202012456.
- [38] X. Liu *et al.*, « Hydrothermal Synthesis and Crystal Structure of Polar and Nonpolar Compounds in Indium Iodate Family », *Crystal Growth & Design*, vol. 8, n° 7, p. 2453- 2457, juill. 2008, doi: 10.1021/cg800034z.

- [39] X. Xu, C.-L. Hu, B.-X. Li, et J.-G. Mao, «  $K_2Au(IO_3)_5$  and  $\beta$ - $KAu(IO_3)_4$ : Polar Materials with Strong SHG Responses Originating from Synergistic Effect of  $AuO_4$  and  $IO_3$  Units », *Chem. Eur. J.*, vol. 22, n° 5, p. 1750- 1759, janv. 2016, doi: 10.1002/chem.201504117.
- [40] Y.-J. Jia *et al.*, «  $LiMII(IO_3)_3$  (MII = Zn and Cd): Two Promising Nonlinear Optical Crystals Derived from a Tunable Structure Model of  $a$ - $LiIO_3$  », *Angew. Chem. Int. Ed.*, p. 6, 2019.
- [41] J. Bouillot, E. Coquet, J. Pannetier, et J.-M. Creti'Ez, « NEUTRON POWDER DIFFRACTION STUDIES OF  $LiIO_3$  AND  $(HIOa, 2LiIO_3)$  », p. 5.
- [42] D. Phanon, A. Mosset, et I. Gautier-Luneau, « New materials for infrared non-linear optics. Syntheses, structural characterisations, second harmonic generation and optical transparency of  $M(IO_3)_3$  metallic iodates », *J. Mater. Chem.*, vol. 17, n° 11, p. 1123, 2007, doi: 10.1039/b612677d.
- [43] K. M. Ok et P. S. Halasyamani, « New Metal Iodates: Syntheses, Structures, and Characterizations of Noncentrosymmetric  $La(IO_3)_3$  and  $NaYl_4O_{12}$  and Centrosymmetric  $\beta$ - $Cs_2I_4O_{11}$  and  $Rb_2I_6O_{15}(OH)_2 \cdot H_2O$  », *Inorg. Chem.*, vol. 44, n° 25, p. 9353- 9359, déc. 2005, doi: 10.1021/ic051340u.
- [44] E. L. Belokoneva, « New nonlinear optical potassium iodate  $K[IO_3]$  and borates  $K_3[B_6O_{10}]Br$ ,  $KTa[B_4O_6(OH)_4](OH)_2 \cdot 1.33H_2O$ —Synthesis, structures and relation to the properties », *Journal of Solid State Chemistry*, p. 7, 2012.
- [45] B.-P. Yang, C.-L. Hu, X. Xu, et J.-G. Mao, « New Series of Polar and Nonpolar Platinum Iodates  $A_2Pt(IO_3)_6$  ( $A = H_3O, Na, K, Rb, Cs$ ) », *Inorganic Chemistry*, p. 7.
- [46] D. Phanon, Y. Suffren, M. B. Taouti, D. Benbental, A. Brenier, et I. Gautier-Luneau, « Optical properties of  $Nd^{3+}$  and  $Yb^{3+}$ -doped  $AgM(IO_3)_4$  metal iodates: transparent host matrices for mid-IR lasers and nonlinear materials », *J. Mater. Chem. C*, vol. 2, n° 15, p. 2715- 2723, 2014, doi: 10.1039/C3TC32517B.
- [47] C.-F. Sun, C.-L. Hu, X. Xu, et J.-G. Mao, « Polar or Non-Polar? Syntheses, Crystal Structures, and Optical Properties of Three New Palladium(II) Iodates », *Inorg. Chem.*, vol. 49, n° 20, p. 9581- 9589, oct. 2010, doi: 10.1021/ic101370v.
- [48] H.-Y. Chang, S.-H. Kim, K. M. Ok, et P. S. Halasyamani, « Polar or Nonpolar?  $A^+$  Cation Polarity Control in  $A_2Ti(IO_3)_6$  ( $A = Li, Na, K, Rb, Cs, Tl$ ) », *J. Am. Chem. Soc.*, vol. 131, n° 19, p. 6865- 6873, mai 2009, doi: 10.1021/ja9015099.
- [49] E. L. Belokoneva, « Prediction of the structures in the family of  $La(IO_3)_3$  iodates based on topology and symmetry analysis within the OD theory », *Crystallogr. Rep.*, vol. 53, n° 3, p. 398- 403, mai 2008, doi: 10.1134/S1063774508030061.
- [50] M. Luo *et al.*, « Rational Design of the Nonlinear Optical Response in a Tin Iodate Fluoride  $Sn(IO_3)_2F_2$  », *Chem. Mater.*, vol. 32, n° 6, p. 2615- 2620, mars 2020, doi: 10.1021/acs.chemmater.0c00196.
- [51] C.-L. Hu et J.-G. Mao, « Recent advances on second-order NLO materials based on metal iodates », *Coordination Chemistry Reviews*, vol. 288, p. 1- 17, avr. 2015, doi: 10.1016/j.ccr.2015.01.005.
- [52] A. E. Siegman, « Springer Series in Optical Sciences », *Springer Series in Optical Sciences*, p. 430.
- [53] C. Sun, B. Yang, et J. Mao, « Structures and properties of functional metal iodates », *Sci. China Chem.*, vol. 54, n° 6, p. 911- 922, juin 2011, doi: 10.1007/s11426-011-4289-8.
- [54] X. Xu, C.-L. Hu, B.-P. Yang, et J.-G. Mao, « Syntheses, crystal structures and magnetic properties of three new silver manganese(ii) or manganese(iii) mixed metal iodates », *CrystEngComm*, vol. 15, n° 38, p. 7776, 2013, doi: 10.1039/c3ce41185k.
- [55] « Synthesis and Characterization of Two New Second Harmonic Generation Active Iodates:  $K_3Sc(IO_3)_6$  and  $KSc(IO_3)_3Cl$  | ACS Omega ». <https://pubs.acs.org/doi/full/10.1021/acsomega.9b04288> (consulté le 30 mai 2022).
- [56] N. Ngo, K. Kalachnikova, Z. Assefa, R. G. Haire, et R. E. Sykora, « Synthesis and structure of  $In(IO_3)_3$  and vibrational spectroscopy of  $M(IO_3)_3$  ( $M=Al, Ga, In$ ) », *Journal of Solid State Chemistry*, vol. 179, n° 12, p. 3824- 3830, déc. 2006, doi: 10.1016/j.jssc.2006.08.026.
- [57] X. Jiang, H. Wu, H. Yu, Z. Hu, J. Wang, et Y. Wu, « Synthesis, Structure, Characterization, and Calculation of a Noncentrosymmetric Fluorine-Containing Indium Iodate,  $Ba[InF_3(IO_3)_2]$  », *Cryst. Growth Des.*, p. 8, 2021.
- [58] K. M. Ok et P. S. Halasyamani, « The Lone-Pair Cation  $I_5^+$  in a Hexagonal Tungsten Oxide-Like Framework: Synthesis, Structure, and Second-Harmonic Generating Properties of  $Cs_2I_4O_{11}$  », *Angew. Chem.*, vol. 116, n° 41, p. 5605- 5607, oct. 2004, doi: 10.1002/ange.200460367.

- [59] Y. Suffren et I. Gautier-Luneau, « Titanyl Iodate – A Promising Material for Infrared Nonlinear Optics Showing Structural Similarities with KTP », *Eur. J. Inorg. Chem.*, vol. 2012, n° 27, p. 4264- 4267, sept. 2012, doi: 10.1002/ejic.201200792.
- [60] K. Nassau, J. W. Shiever, et B. E. Prescott, « Transition metal iodates. VI. Preparation and characterization of the larger lanthanide iodates », *Journal of Solid State Chemistry*, vol. 14, n° 2, p. 122- 132, juin 1975, doi: 10.1016/0022-4596(75)90002-X.
- [61] B.-P. Yang, C.-L. Hu, F.-F. Mao, X. Xu, et J.-G. Mao, « Two Barium Gold Iodates: Syntheses, Structures, and Properties of Polar BaAu(IO<sub>3</sub>)<sub>5</sub> and Nonpolar HBa<sub>4</sub>Au(IO<sub>3</sub>)<sub>12</sub> Materials », *Inorg. Chem.*, p. 7, 2017.
- [62] H. Huang, Y. He, R. He, Z. Lin, Y. Zhang, et S. Wang, « Y(IO<sub>3</sub>)<sub>3</sub> as a Novel Photocatalyst: Synthesis, Characterization, and Highly Efficient Photocatalytic Activity », *Inorg. Chem.*, vol. 53, n° 15, p. 8114- 8119, août 2014, doi: 10.1021/ic501127d.
- [63] D. W. Lee, S. B. Kim, et K. M. Ok, « ZnIO<sub>3</sub>(OH): a new layered noncentrosymmetric polar iodate – hydrothermal synthesis, crystal structure, and second-harmonic generating (SHG) properties », *Dalton Trans.*, vol. 41, n° 27, p. 8348, 2012, doi: 10.1039/c2dt30947e.

## الملخص :

في هذا العمل استعملنا قيم تجريبية لعائلة الإيودات المعدنية وقيم محسوبة استنادا إلى نظرية الكثافة الدالية DFT وذلك من أجل إنشاء قاعدة بيانات. بعد ذلك استخدمنا الطريقة الإحصائية متعددة المتغيرات المسماة تحليل المركبات الرئيسية (ACP) لدراسة الخصائص المحسوبة لهذه المواد وذلك من أجل تحديد العلاقة الموجودة بينها, استعملنا كذلك تقنية جديدة وهي التعلم الآلي التي سمحت لنا بالتنبؤ بقيم طاقة التشكل وفجوة الطاقة.

**الكلمات المفتاحية :** قاعدة بيانات, نظرية الكثافة الدالية DFT, تحليل المركبات الرئيسية (ACP) التعلم الآلي طاقة

## Résumé

Dans ce travail , on a utilisé des valeurs expérimentales pour la famille des iodates métalliques et des valeurs calculées basées sur DFT afin de créer une base de données. Après, on a utilisé la méthode statique multi variées ACP pour étudier les propriétés de ces matériaux afin de déterminer la corrélation entre ces propriétés , et nous avons également utilisé une technique prédictive d'apprentissage automatique qui on permis de prédire l'énergie de formation et l'énergie de gap de ces matériaux.

**Les mots clés :** base de données, DFT, analyse des composantes principales ACP, l'apprentissage automatique, Gap.

## Abstract:

In this work, we used experimental and DFT-based theoretical values of different physical properties of metal iodates in order to create a database. Next, we used the multivariate static method PCA to study these materials, in order to determine the relationship between these properties we used a predictive machine learning technique which allowed us to predict the formation energy and the gap energy of these materials

**Key words :** Database, DFT, Principal component analysis (PCA), machine learning. Formation energy, band gap.