الجمهورية الجزائرية الديموقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA

FACULTE DES SCIENCES

Département Mathématiques et Informatique

Projet de fin d'études

Présenté pour l'obtention du Diplôme de Master en Informatique

Option : Ingénierie des Systèmes et Technologie du Web

THEME

Extraction des zones séismiques en Algérie a l'aide de l'Apprentissage Automatique

Proposé et dirigé par :

Encadreur:
Mr. MAZARI Redha

Co-Encadreur: enseignant

Mr. MAZARI Oussama Sohaib

Préparé et présenté par :

Mr. BOUSSAHOUA Abd Raouf
Mr. ZEROUAL Hamid

Année universitaire

2021-2022

Dédicace

À tous ceux qui m'ont appris une lettre.

Abd Raouf

Dédicace

Je dédie ce modeste travail à tous ceux :

Qui nous enseignent

Qui nous soutiennent

Qui nous chérissent

Qui ont cru et croient Toujours en moi

Hamid

Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant, qui nous a donné la force, la volonté et la patience d'accomplir ce travail.

Nous tenons à remercier notre Encadreur **Mr. MAZARI Redha**

Pour son encadrement, son écoute, ses élucidations, ses conseils, ses directives et encouragements qu'il nous a afflué.

Nous tenons à remercier nos enseignants de l'université de Médéa qui nous ont transmis le bien savoir.

Nous tenons également à remercier nos amis et tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail plus.

*Précisément nos amis et collègues :**Mr.boudjemaa Ali**, **Mr.fekhari Abdel Waheb** .*

Résumé

Les séismes sont l'un des phénomènes naturels qui affectent l'homme et provoquent de multiples catastrophes qui entraînent d'énormes pertes en Lorsqu'il se produit, Comme il est difficile d'empêcher ce phénomène naturel de se produire, donc Il faut vivre avec lui, et il est nécessaire de comprendre les causes de son apparition, comment Avec une activité sismique forte ou faible pour établir des réseaux intensifs de surveillance sismique, et pour cela cette étude a été appliquer des algorithmes d'apprentissage automatique (K-means, DBSCAN, OPTICS) pour déterminer les zones sismiques en Algérie dans le but de réduire le risque et de se protéger des conséquences du tremblement de terre en adoptant des mesures sanitaires et en développant les zones sismiques comme autant que possible .Les résultats obtenus montre que l'algorithme d'apprentissage automatique DBSCAN, OPTICS Présente la plus proches résultat avec les cartes géologiques, avec l'acquisition de nouvelles zones sismiques.

Abstract

Earthquakes are one of the natural phenomena that affect man and cause multiple disasters that cause huge losses in When it happens, How difficult it is to prevent this natural phenomenon from happening, so It is necessary to live with it, and it is necessary to understand the causes of its appearance, how With high or low seismic activity to establish intensive seismic monitoring networks, and for this this this study has been applying machine learning algorithms (K-means, DBSCAN, OPTICS) to determine the seismic zones in Algeria in order to reduce the risk and protect themselves from the consequences of the earthquake by adopting sanitary measures and developing the seismic zones as much as possible . The results obtained show that the machine learning algorithm DBSCAN, OPTICS Presents the closest result with geological maps, with the acquisition of new seismic zones.

ملخص

الزلازل هي عبارة عن إحدى الظواهر الطبيعية التي تؤثر على الإنسان وممتلكاته، ونظراً لأنه من الصعوبة منع حدوث هذه الظاهرة الطبيعية، فإنه يتحتّم علينا التعايش معها، لذا فمن الضروري فهم أسباب حدوثها وكيفية التعامل معها ومدى إمكانية التقليل من مخاطرها، وتسعى جميع الدول سواءً التي تتميز بنشاط زلزالي كبير أم ضعيف لإنشاء شبكات مُكثفة للرصد الزلزالي، ويواجه العالم سنوياً العديد من المخاطر الطبيعية التي تتسبب في وقوع كوارث إنسانية متعددة تؤدي إلى خسائر هائلة في الممتلكات والأرواح البشرية، وتُشكل الزلازل أكثر هذه الكوارث خطورة بسبب ما ينتج عنها من دمار نظراً لعنصر المفاجأة الذي تحدثه عند وقوعها ، نسعى في هذه المذكرة الى تطبيق خوارزميات التعلم الالي لتحديد المناطق الزلزالية في الجزائر لغرض حماية أنفسنا من خطر الزلازل تظهر النتائج التي تحصلنا عليها من خوارزميات التعلم الالي انه يمكن الحصول نتائج للمناطق الزلزالية في الجزائر التي هي اقرب لنتائج التي تم الوصول اليها من طرف علماء الزلازل كما تم الحصول علي مناطق زلزالية جديدة .

Tables des matières

Chapitre I : l'Apprentissage Automatique 2

I Chapiter I: L'Apprentissage Automatique 3

I.1. Introduction : 3

I.2. L'intelligence Artificielle : 3

I.2.1. Définition De L'intelligence Artificielle : 3

I.2.2. Domaines d'application 4

I.3. L'apprentissage automatique : 4

I.3.1. Définition : 4

I.3.2. Les Types d'apprentissage automatique : 5

I.3.2.1. Apprentissage Supervisé : 5

I.3.2.2. Apprentissage non-supervisé : 7

I.3.2.3. Apprentissage par renforcement : 9

I.4. Les algorithmes d'apprentissage automatique : 10

I.5. Conclusion : 19

Chapitre II : Etat de l'art 20

II Chapitre II : Etat de l'art 21

II.1. Introduction 21

II.1. Définition : 21

II.1.2. Les différentes classes de séisme : 22

II.2. Les sièges des tremblements de terre : 23

II.3. Mesure d'un séisme : 24

II.3.1. Les échelles sismiques : 24

II.3.1.1. L'échelle de Mercalli : 24

II.3.1.2. L'échelle de Richter : 24

II.3.2. Magnitude et intensité d'un séisme : 25

II.4. La répartition mondiale des séismes : 25

II.4.1. Les pays à fort risque sismique : 26

II.4.2. Localisation des épicentres des séismes dans le monde : 27

II.4.3. Situation de l'Algérie par rapport aux mouvements des plaques : 28

II.4.4. Conditions sismotectoniques régionales : 30

Tables des matières

II.4.5. Classification des zones sismiques en Algérie : 31

II.5. Dataset : 32

II.5.1. Définition : 32

II.5.2. Pourquoi avons-nous besoin d'un Dataset ? 32

II.5.3. Types base de données : 33

II.6. Travaux précèdent : 34

II.7. Conclusion : 35

Chapitre III : 36

Implémentation et évaluation 36

III Chapitre III : Implémentation et évaluation 37

III.1. Introduction : 37

III.2. Démarche générale de travaille : 38

III.3. Collection des Dataset (les données) : 39

III.3.1. Les source de Dataset : 39

III.3.2. Prétraitement des Datasets : 39

III.3.3. Collecte de données : 41

III.3.3.1. Prétraitement de Dataset collection : 41

III.3.3.2. Une visualisation complète de Dataset : 41

III.3.3.3. Préciser le pourcentage de donnée de chaque ressource : 42

III.3.3.4. Exprimer magnitude de séisme sélectionner dans le Dataset : 43

III.4. Les Algorithmes 43

III.4.1. Introduction 43

III.4.2. Kmeans 44

III.4.2.1. Méthode du Coude (Elbow Method) 44

III.4.2.1.1. L'application de la méthode du coude 45

III.4.2.1.2. L'application L'algorithme Kmeans 46

III.4.2.1.3. Constat 46

III.4.2.1.4. Prétraitement des Dataset 47

III.4.2.1.5. Réappliquer la méthode du Coude 49

III.4.2.1.6. Réappliquer L'algorithme Kmeans 49

Tables des matières

III.4.2.2. Coefficient de silhouette (Silhouette Coefficient) : 50
III.4.2.2.1. Distance Matrice de Haversine :..... 51
III.4.2.3. L'application du Coefficient de Silhouette : 52
III.4.2.4. L'application de l'algorithme Kmeans : 53
III.4.3. DBSCAN : 54
III.4.3.1. Comment fonctionne DBSCAN :..... 56
III.4.3.1.1. Sélection d'Epsilon..... 56
III.4.3.1.2. Sélection de MinPts :..... 57
III.4.3.2. Détermination des paramètres et résultats du clustering :..... 58
III.4.3.3. Résultats de la Clustering :..... 58
III.4.4. OPTICS :..... 60
III.4.5. Comparaisons :..... 64
III.5. Les outlies de travail:..... 66
III.5.1. Pandas : 66
III.5.2. Numpy :..... 66
III.5.3. Matplotlib :..... 66
III.5.4. Geopandas :..... 66
III.5.5. Scipy :..... 67
III.5.6. Yellowbrick :..... 67
III.5.7. Sklearn : 67
III.5.8. Kneed : 67
III.5.9. Anaconda : 67
III.5.10. Jupyter :..... 68
IV Conclusion Générale et perspectives: 69

Liste des Figures

Figure 1: L'Intelligence Artificielle 4
Figure 2: Les types d'apprentissage automatiques [Larochelle, 2008] 5
Figure 3: Schéma d'apprentissage supervisé. 6
Figure 4: Schéma d'apprentissage non-supervisé. 8
Figure 5: Le principe d'apprentissage par renforcement [Chaarana, 2018]. 10
Figure 6: Les algorithmes d'apprentissage automatique 12
Figure 7: Exemple de clustering K-means 13
Figure 8: Pseudo-code de DBSCAN 16
Figure 9: Réseau de neurones articiel ANN 18
Figure 10: Magnitude et intensite d'un séisme 25
Figure 11: Géographie physique de séisme 26
Figure 12: Localisation des épicentres des séismes dans le monde. 28
Figure 13: La limite des Plaques Africaine et Eurasienne. 29
Figure 14: Distribution spatiale des tremblements de terre en Algérie du Nord. 30
Figure 15: Repartions des zones sismique dans la période 1962-2020 31
Figure 16: Degré des zones sismiques. 32
Figure 17: Les types du Dataset 33
Figure 18: Démarche générale de travaille. 38
Figure 19: Visualisation de Dataset 41
Figure 20 : Le pourcentage de Dataset par apport les déférent ressource. 42
Figure 21: Le pourcentage de Magnitude dans le Dataset 43
Figure 22: Graphe de méthode du coude 45
Figure 23: Résultat de Kmeans en utilise la méthode du Coude (k=8) 46
Figure 24: Les zones séismiques par apport les giologistes 47
Figure 25 : Échelle de Richter de magnitude du séisme 47
Figure 26: Visualisation de Dataset 48
Figure 27: Graphe de méthode du coude 49
Figure 28: Résultat de Kmeans en utilise la méthode du Coude (k=10) 50
Figure 29: Graphe de Coefficient de Silhouette 52
Figure 30: Analyse de silhouette pour le regroupement de kmeans avec n_clusters =15 ... 53
Figure 31: Résultat de Kmeans (k=15) en coefficient du silhouette 54
Figure 32: Cluster DBSCAN unique 56
Figure 33: Sélection de MinPts 58
Figure 34: Résultats de l'algorithme DBSCAN à partir des paramètres optimaux 59

Liste des Figures

Figure 35: Résultat final de l'algorithme DBSCAN. 60
Figure 36: Sélection des paramètres MinPts pour OPTICS 62
Figure 37: Résultats de l'algorithme OPTICS à partir des paramètres optimaux 63
Figure 38: Résultat final de l'algorithme OPTIC 64
Figure 39: Comparaison des diffèrent algorithmes 65

Introduction Générale

Depuis ces dernières années, les nouvelles technologies envahissent notre quotidien. Cette révolution technologique impacte tous les secteurs d'activités (médecines, biologie, géologie, etc…). Aussi, avec développement des technologies des processeurs et des ordinateurs, L'apprentissage automatique viendra pour donner aux ordinateurs la capacité d'« apprendre » à partir de données, ce dernier compose de trois types parmi les en a l'apprentissage non supervise qui consiste sur groupe les données dans des clusters.

Parmi les secteurs qui on a utilisé beaucoup traitement de données la géologie beaucoup plus dans la sismologie ou on a étudié les zones séismiques pour objectif renforcer la sécurité et protéger les personnes et les biens ce cette catastrophe naturelle.

Pour ça on utilise l'apprentissage non supervise pour extraire les zones séismiques à partir d'appliquer quelques algorithmes de ce dernier pour résoudre le problème de clustering. Alors comment déterminer les zones séismiques en Algérie ?

ce travaille compose d'une introduction générale suivi par trois chapitre dans le premier chapitre on a parlé à l'apprentissage automatique débiter par des définition de l'IA et les types d'apprentissage aussi quelque algorithme et on a mettre en évidence sure l'apprentissage non superviser et sont algorithme qui ce nous avons besoin dans notre travail , dans le deuxième chapitre on a parlé a quelques définitions de séismologie et préciser les déférents classe de séisme avec son échelles et on a déterminer les zones sismiques en Algérie , dans le troisième chapitre on parler sure le travail effectuer étape par étape pour extraction les zones séismique en Algérie et se termine par une conclusion générale.

Chapitre I : l'Apprentissage Automatique

I Chapiter I: L'Apprentissage Automatique

I.1. Introduction :

À l'époque nous avons eu divers problèmes tel que le manque d'informations et la difficulté de les obtenir, mais aujourd'hui avec le développement de la technologie qui a réglé plusieurs complications ,qui aide la possibilité de collecter de grandes quantités des données nécessaires pour connaître les différentes activités géologiques qui se déroulent sur notre planète , Parmi ces activités géologiques, les séismes qui utilisé des ressources informatiques permettant d'aborder la sismicité mondiale et la sismologie.

I.2. L'intelligence Artificielle :

I.2.1. Définition De L'intelligence Artificielle :

L'Informatique est la science du traitement de l'Information, l'IA s'intéresse à tous les cas où ce traitement ne peut être ramené à une méthode simple, précise et algorithmique. Dans littérature on trouve plusieurs définitions de l'intelligence artificielle on cite :

"L'Intelligence Artificielle **(IA)** est la science dont le but est de faire par une machine des tâches que l'homme accomplit en utilisant son intelligence. L'Intelligence Artificielle est apparue en 1956. On peut lui préférer celle d'Informatique Heuristique" **[Pastre, 2000]**.

"IA est l'étude des activités intellectuelles de l'homme pour lesquelles aucune méthode n'est a priori connue. (Tout ce qui n'a pas encore été fait en Informatique - quand on sait le faire, ce n'est plus de l'IA)" **[J.L. Laurière, 1986]**.

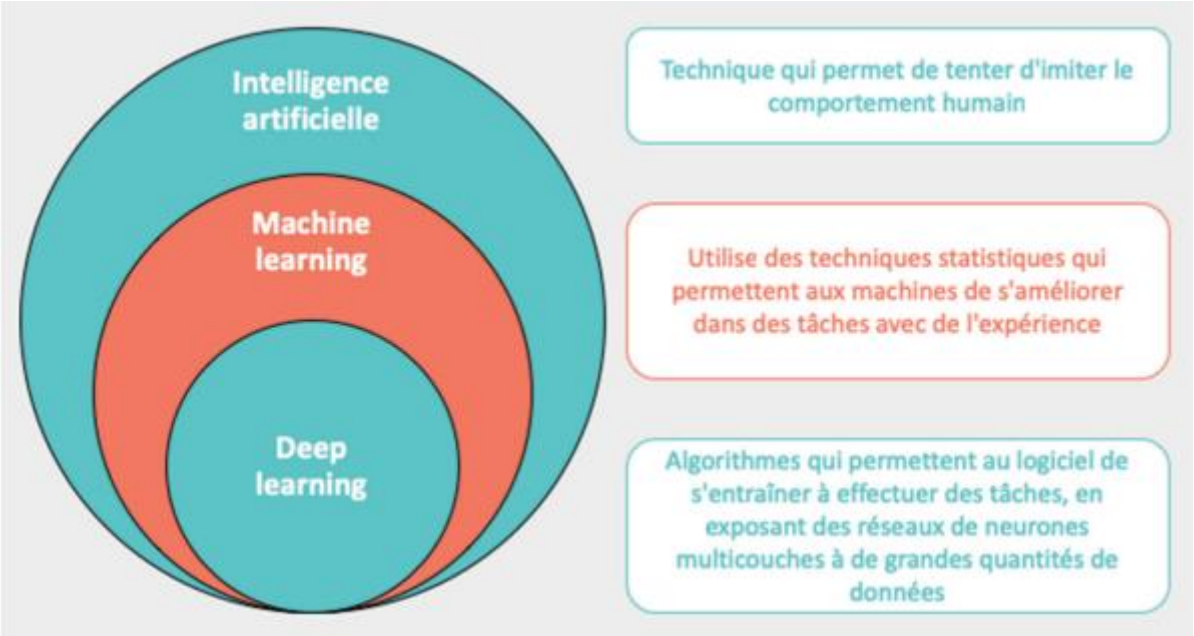


Figure 1: L'Intelligence Artificielle

I.2.2. Domaines d'application

L'IA a été utilisée pour développer et faire progresser de nombreux domaines et industries, y compris la finance, la santé, l'éducation, le transport, et plus encore **[Bensoussan & Bensoussan, 2019].**

Exemples :

- jeu d'échecs (bien qu'il n'y ait qu'un nombre fini de situations).
- résumer un texte ou le traduire.
- reconnaitre des lettres manuscrites, par exemple TAON ou THON.
- faire des mathématiques.

I.3. L'apprentissage automatique :

I.3.1. Définition :

L'apprentissage automatique (en anglais : *Machine Learning*) est une science qui consiste à développer des algorithmes d'apprentissage, qui apprennent à résoudre une tâche **[Roy, 2018]**.

Plus précisément l'apprentissage automatique correspond au domaine se consacrant au développement d'algorithmes permettant à une machine d'apprendre à partir d'un ensemble des données, c-à-d., d'y extraire des concepts et patrons caractérisant ces données **[Larochelle, 2008]**.

Il existe d'autre définitions "Un programme qui apprend est un programme qui tire profit d'une expérience E, par rapport à une famille de tâches T, pour une mesure d'efficacité P si son efficacité (mesurée par P) dans l'accomplissement des tâches de T, augmente après l'expérience E*"* **[Ben Ayed, 2015].**

I.3.2. Les Types d'apprentissage automatique :

Les principaux types d'apprentissage sont distingués essentiellement par leur objectif, et peuvent être combinés dans un même système.

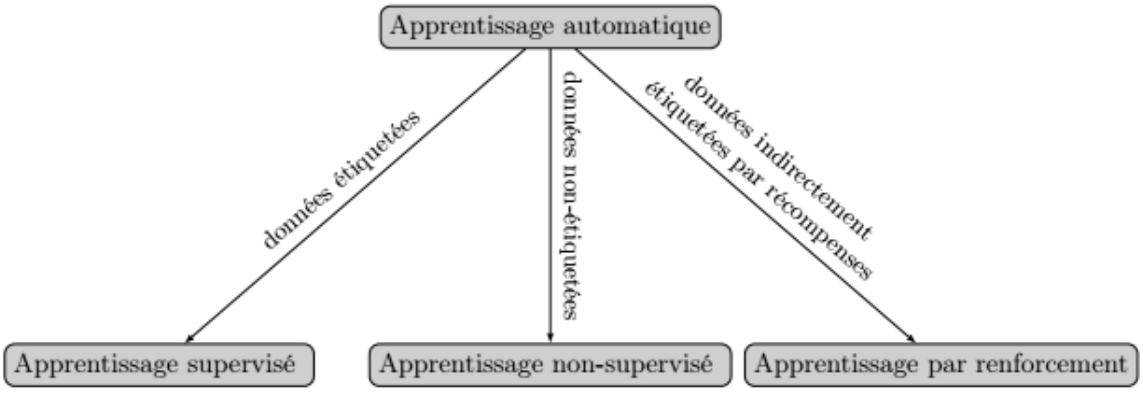


Figure 2: Les types d'apprentissage automatiques [Larochelle, 2008]

I.3.2.1. Apprentissage Supervisé :

L'apprentissage supervisé correspond au cas où l'objectif de l'apprentissage est déterminé explicitement via la définition d'une cible à prédire. Dans la formule (1.1), *D* correspond à un ensemble de n paires d'entrées x_t et de cibles associées y_t :

$$D= \{(x_t, y_t)|x_t \in X, y_t \in Y\}^n_{t=1}$$

Typiquement, un tel ensemble est récolté en fournissant l'ensemble des entrées à un groupe de personnes et en leur demandant d'associer à chacune de ces entrées une cible appropriée dans

le contexte du problème à résoudre. La tâche d'un algorithme d'apprentissage est alors d'entraîner un modèle qui puisse imiter ce processus d'étiquetage par un humain, c-à-d, qui puisse prédire pour une entrée x quelconque la valeur de la cible y qui aurait normalement été donnée par un humain. Cependant, les algorithmes d'apprentissage ne se limitent pas à la modélisation du comportement de l'humain, et peuvent être utilisés pour modéliser la relation liant des paires d'entrées et de cibles provenant d'un autre phénomène **[Larochelle, 2008]**.

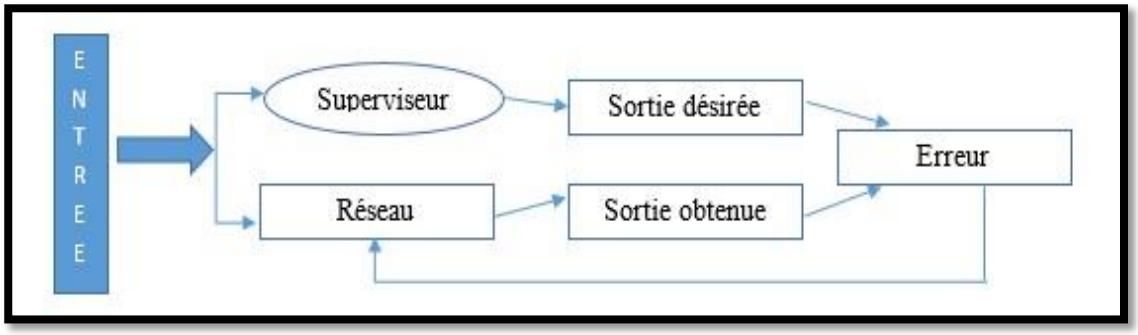


Figure 3: Schéma d'apprentissage supervisé.

Quelques problèmes résolus par ce type d'apprentissage **[Garivier, 2013]** sont :

- Détecteur de spam.
- Risque de crédit.
- Prédiction des pics d'ozone.
- Aide au diagnostic médical.
- Aide au pilotage.
- Apprentissage sur les graphes.

Ses problèmes d'apprentissage supervisé peuvent être regroupés en deux problèmes **[Jason Brownlee, 2016]**:

- Classification: Un problème de classification survient lorsque la variable de sortie est une catégorie, telle que «rouge», «bleu» ou «maladie» et «pas de maladie».
- Régression: Un problème de régression se pose lorsque la variable de sortie est une valeur réelle, telle que «dollars» ou «poids».

- ## Les métriques d'évaluation en apprentissage supervisé :

- ## La matrice de confusion :

Est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). Les occurrences utilisées pour chacune de ces 2 classes doivent être différentes. Un des intérêts de la matrice de confusion

est qu'elle montre rapidement si un système de classification parvient à classifier correctement **[Managemement& data science, 2019]**.

- **La précision :**

Est le nombre de documents pertinents retrouvés rapporté au nombre de documents total proposé par le moteur de recherche pour une requête donnée. Le principe est le suivant : quand un utilisateur interroge une base de données, il souhaite que les documents proposés en réponse à son interrogation correspondent à son attente. Tous les documents retournés superflus ou non pertinents constituent du bruit. La précision s'oppose à ce bruit documentaire. Si elle est élevée, cela signifie que peu de documents inutiles sont proposés par le système et que ce dernier peut être considéré comme « précis » **[Managemement& data science, 2019]**.

Parmi les métriques les plus connus d'un modèle de régression sont :

- **Le coefficient de détermination :**

Également désigné sous le terme « R au carré », constitue également une méthode standard de mesure de l'adéquation entre le modèle et les données observées. Ce coefficient peut être considéré comme la proportion de la variance expliquée par le modèle. Dans ce cas précis, plus la proportion est élevée, meilleur est le résultat, la valeur 1 indiquant une adéquation parfaite **[Managemement& data science, 2019].**

- **L'erreur quadratique moyenne RMSE :**

En statistiques, l'erreur quadratique moyenne d'un estimateur, d'un paramètre, d'une dimension (mean squared error en anglais) est une mesure caractérisant la « précision » de cet estimateur. Pour calculer l'erreur quadratique moyenne RMS, les erreurs individuelles sont tout d'abord élevées au carré, puis additionnées les unes aux autres. On divise ensuite le résultat obtenu par le nombre total d'erreurs individuelles, puis on en prend la racine carrée. Cette erreur nous donne une mesure synthétique de l'erreur globale dans une seule valeur **[Managemement& data science, 2019]**.

I.3.2.2. Apprentissage non-supervisé :

L'Apprentissage non-supervisé est un apprentissage sur des exemples non connus contrairement à l'apprentissage supervisé. Ne repose que sur des données sans étiquettes. La sortie

n'est pas connue a priori [Barlow, 1989]. « Les algorithmes d'apprentissage supervisé expérimentent un ensemble de données et apprennent des propriétés utiles de la structure de ces données. Ils modélisent fondamentalement la distribution de probabilité d'une propriété d'un ensemble de données. Dans l'apprentissage non supervisé, aucun libellé ou objectif n'est fourni, ce qui en fait une tâche beaucoup plus difficile que l'apprentissage supervisé. L'algorithme doit donner un sens aux données sans aucune supervision. Exemples d'algorithmes dans L'apprentissage non supervisé est l'analyse en composantes principales et la classification en k-moyennes » [Bankman et al, 2019].

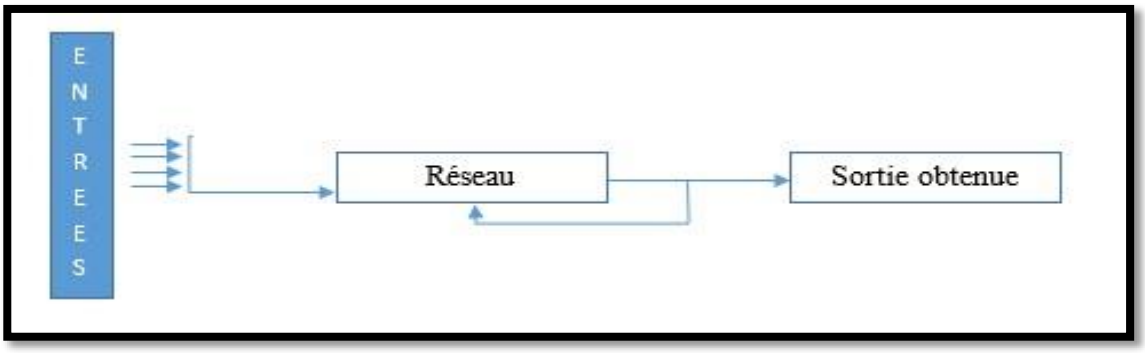


Figure 4: Schéma d'apprentissage non-supervisé.

Parmi les problèmes résolus par l'apprentissage non-supervisé les plus répandus nous avons :

- **Estimation de densité :**

Pour ce problème une fonction doit fournir une estimation de la fonction de densité ou de probabilité de la distribution ayant généré les éléments, cette estimation est faite avec les méthodes paramétriques par exemple, la méthode de l'estimation de vraisemblance ou bien la méthode de moindre carré, ..., ou avec des méthodes d'estimations non paramétriques pour les fonctions de densité inconnue [Boukhames et zaidi, 2016].

- **Extraction de caractéristique :**

Apprendre une nouvelle représentation de l'entrée qui soit plus utile que la représentation vectorielle originale. L'utilité de cette représentation dépend de la tâche que le nous souhaitons réellement résoudre, et la qualité peut être mesurée de différentes façons, en fonction du contexte

de l'application. Parmi les approches de l'extraction de caractéristiques, la technique la plus simple qui est la sélection de caractéristique **[Larochelle, 2008]**.

- **Reduction de dimensionnalité :**

C'est la solution optimale pour résoudre les problèmes d'apprentissage automatique par la sélection et extraction des d'attributs pertinents. Son principe est de réduire l'espace de stockage nécessaire et le temps d'apprentissage, afin d'améliorer la classification **[El ferchichi, 2013]**. Plus généralement c'est de passer d'un espace à D dimension vers un espace à K dimension où K < D **[Gagné, 2016]**.

- **Les raisons pour réduire la dimensionnalité sont [Gagné, 2016] :**

  - **Malédiction de la dimensionnalité.**
  - **Visualiser les données : analyse de résultats.**
  - **Simplicité de modèle diminuer les variances.**
  - **Extraction de connaissances.**
  - **Visualiser les données.**

- **Regroupement (Clustering) :**

Un outil de visualisation d'un ensemble de données, peut aussi être utilisé afin de compresser un ensemble de données en le remplaçant par l'ensemble des prototypes. Il est même possible d'extraire des caractéristiques à l'aide du regroupement. Le plus simple algorithme de regroupement est l'algorithme des K moyennes (*K-means algorithm*). Il s'agit d'un algorithme itératif, qui débute avec une initialisation aléatoire de la position des prototypes **[Larochelle, 2008]**.

I.3.2.3. Apprentissage par renforcement :

Ce dernier type où la supervision est nulle, il se base sur le principe d'essai-erreur qui aide finalement à orienter l'apprentissage.

L'apprentissage par renforcement est une branche bien étudiée où l'environnement agit en tant qu'expert et fournit des indices plutôt que des réponses précises. Le système d'apprentissage reçoit des informations sur ses actions, sans promesse concrète de progresser dans la bonne direction, ce qui pourrait consister à résoudre un labyrinthe ou à atteindre un objectif explicite **[Shukla, 2017]**.

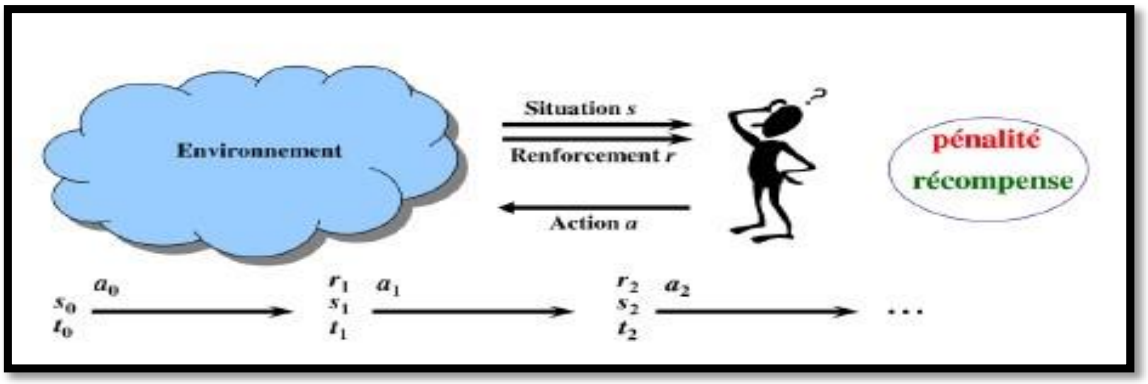


Figure 5: Le principe d'apprentissage par renforcement [Chaarana, 2018].

Pratiquement, il se repose sur l'utilisation de données indirectement étiquetées par des récompenses ou des observations. Cet étiquetage est moins informatif qu'en apprentissage supervisé. Là où il existe un oracle du type oracle (x_i) = y_i pour étiqueter les données a priori en apprentissage supervisé, dans l'apprentissage par renforcement cet oracle n'est capable que de quantifier une association (x_i, y_i) par un scalaire : oracle (x_i, yi) = récompense (Renforcement). De plus, les données (et leur ordre) présentées à l'oracle ne sont pas maîtrisées entièrement par l'utilisateur. L'hypothèse d'avoir des données indépendantes et identiquement distribuées n'est pas valable et il n'existe pas de base de données a priori. Or, cette hypothèse est largement utilisée dans l'analyse formelle de nombreux algorithmes d'apprentissage automatique **[Zimmer, 2018]**.

I.4. Les algorithmes d'apprentissage automatique :

Les algorithmes d'apprentissage automatique permettant aux ordinateurs de s'entraîner à partir d'exemples réels sur les entrées de données et utilisent l'analyse statistique pour produire des valeurs qui se situent dans une plage spécifique **[Larochelle, 2008]**, ses algorithmes comme l'algorithme k-plus proche voisin, les réseaux de neurones, l'arbre de décision, l'algorithme de séparation, …, cherchent à découvrir la structure d'un ensemble de données à partir de valeurs observées. Ils sont utilisés dans de nombreux contextes comme la classification et la prévision et

sont au cœur des algorithmes d'aide à la décision **[Bertail, et al, 2019]**, nous allons les présenter dans les prochaines lignes.

Nous pouvons aussi définir un algorithme d'apprentissage comme suit ; « un algorithme d'apprentissage est un algorithme prenant en entrée un ensemble de données D et retournant une fonction f. Nous désignons alors D comme ensemble d'entraînement ou ensemble d'apprentissage, et la fonction f comme modèle. Suite à l'exécution d'un algorithme d'apprentissage, nous disons que le modèle f a été entraîné sur l'ensemble D. nous considérons que l'ensemble de données D contient sous forme de vecteurs l'information nécessaire pour résoudre un problème donné. Le modèle f doit ainsi pouvoir recevoir de tels vecteurs en argument et nous désignerons par sortie la valeur de l'application de f sur un vecteur. De plus, tous les éléments de D, que le nous nommons exemples d'apprentissage, sont indépendants et identiquement distribués » **[Larochelle, 2008]**.

L'apprentissage automatique facilite l'utilisation des ordinateurs dans la construction de modèles à partir de données d'échantillonnage afin d'automatiser les processus de prise de décision en fonction des données saisies **[JOSEPH BERTRAND, 2017].**

La figure suivante regroupe les d'algorithmes populaires de l'apprentissage automatique supervisé et non-supervisé :

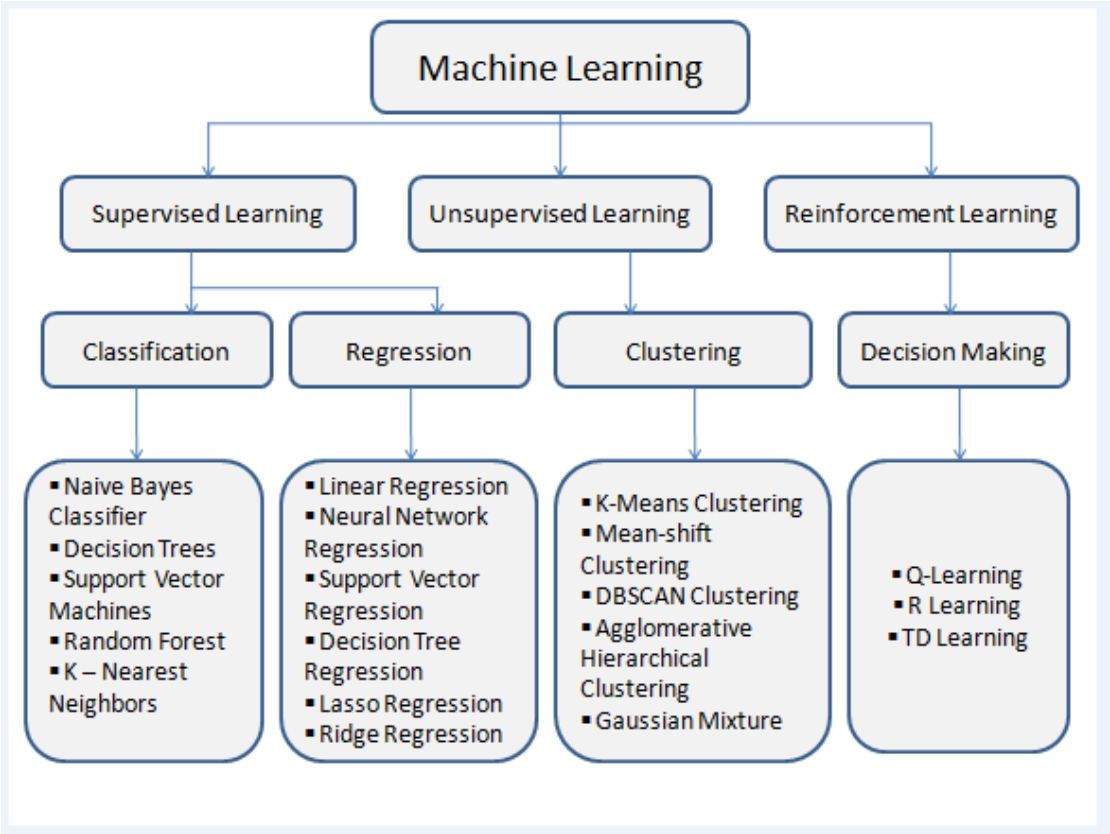


Figure 6: Les algorithmes d'apprentissage automatique

Parmi les algorithmes bien connus, nous citons :

- **K-means :**

Est l'un des plus simples algorithmes non-supervisés de classification automatique des données. L'idée principale et de choisir aléatoirement un ensemble de centres fixé a priori et de chercher itérativement la partition optimale. Chaque individu est affecté au centre le plus proche, après l'affectation de toutes les données la moyenne de chaque groupe est calculé, elle constitue les nouveaux représentants des groupes, lorsqu'ont abouti à un état stationnaire (aucune donnée ne change de groupe) l'algorithme est arrêté **[Celeux, 1989].**

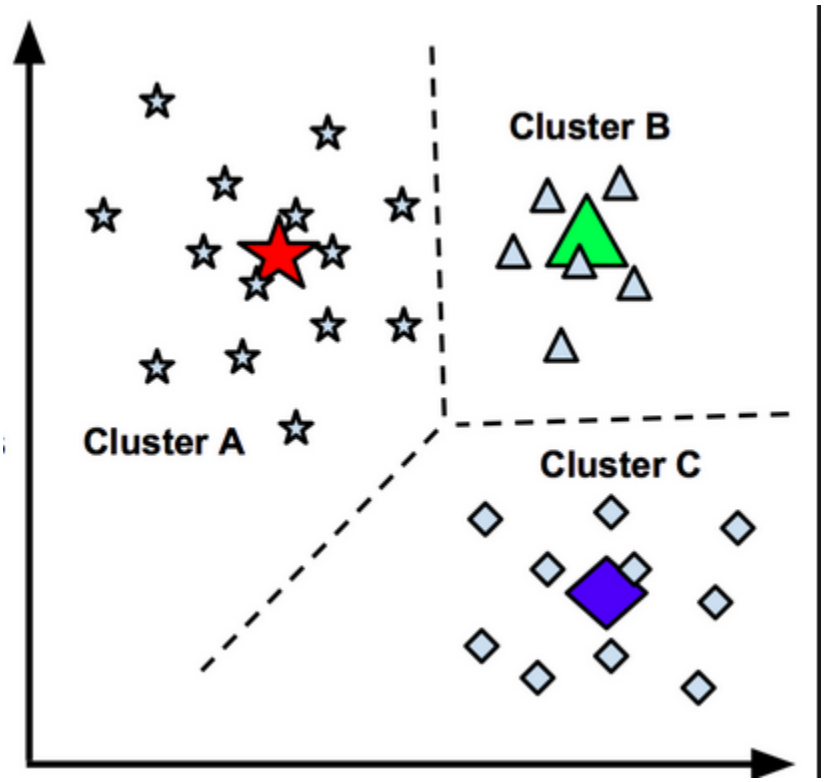


Figure 7: Exemple de clustering K-means

- **Algorithme 1 : K-means**

Entrée

Ensemble de N données, noté par x
Nombre de groupes souhaité, noté par k

Sortie

Une partition de K groupes $\{C_1, C_2, \ldots C_k\}$

Début

1)Initialisation aléatoire des centres C_k ;

Répéter

2)Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche ;

$$x_i \in C_k \, si \, \forall j \; x_i - \mu_k = \min_j x - \mu \qquad (1)$$

Avec *μk* le centre de la classe K ;

3) Représentation : Calculer les centres associe à la nouvelle partition ;

$$\mu_k = \frac{1}{N} \sum_{x_i \in C_k} x_i \qquad (2)$$

Jusqu'à convergence de l'algorithme vers une partition stable ;
Fin.

Ce processus tente de maximiser la similarité intra-class représentée sous forme d'une fonction objective :

$$J = \sum_{i=1}^{K} \sum_{x_j \in C_i} d(x_j, C_i) \qquad (3)$$

Dans le cas de la distance euclidienne cette fonction est appelée fonction d'erreur quadratique.

$$J = \sum_{i=1}^{K} \sum_{x_j \in C_i} \left\| x_j - C_i \right\|^2 \qquad (4)$$

- ***Global K*-means :**

Est une solution au problème d'initialisation du k-means, elle est fondée sur les données et vise à atteindre une solution globalement optimale. Elle consiste à effectuer un clustering incrémental et à ajouter dynamiquement un nouveau centre suivi par l'application du k-means jusqu'à la convergence. Les centres sont choisis un par un de la façon suivante : le premier centre est le centre de gravité de l'ensemble des données (résultat de l'application du k-means avec k=1), les autres centres sont tirés de l'ensemble de données ou chaque donnée est une candidate pour devenir un centre, cette dernière sera testée avec le reste de l'ensemble, le meilleur candidat est celui qui minimise la fonction objective. [Likas A,2003] l'algorithme suivant permet d'illustrer le principe **[Guellil, 2009]** :

Algorithme 2 : Global k-means
Entrée
Ensemble de N données, notés par x ;
Nombre de groupes souhaiter, noté par k ;
Sortie

Une partition de K groupes $\{C_1, C_2, \ldots C_k\}$

Début

1) C_1 = Centre de gravité de l'ensemble des données ;

Répéter

2) Initialiser les centres i-1 par le résultat de l'étape précédente ;

3) Trouver l'iéme centre :

Pour chaque donnée x **faire**

3.1) Considère x comme étant le iéme centre ;

3.2) Affecter les données aux plus proche centre ;

3.3) Calculer l'erreur quadratique pour C_i=x ;

Fin faire

3.4) Garder le centre C_i = x qui minimise l'erreur quadratique ;

4) Appliquer le k-means jusqu'à la convergence ;

Jusqu'à obtenir une partition en k groupes ;

Fin.

- **DBSCAN Algorithme :**

DBSCAN est le premier algorithme de clustering basé sur la densité. Il a été proposé par **[Ester et al, 1996]** et a été conçu pour données de forme arbitraire en présence de bruit dans des Datasets spatiales et non spatiales de haute dimension. L'idée clé de DBSCAN est que pour chaque objet d'un cluster, le d'un cluster, le voisinage d'un rayon donné (Eps) doit contenir au moins un minimum d'objets (MinPts), ce qui signifie que la cardinalité de la cardinalité du voisinage doit dépasser un certain seuil. Le -voisinage d'un point arbitraire 'p' est défini comme suit,

$$N_{Eps} = \{q \in D \,/ dist\,(p,q) < Eps\}$$

Ici, D est la base de données des objets. Si les –quartiers d'un point P contiennent au moins un nombre minimal de points, et alors ce point est appelé point central. Le point central est défini comme suit :

$$N_{Eps}(P) > MinPts$$

Ici Eps et MinPts sont les paramètres spécifiés par l'utilisateur c'est-à-dire le rayon du voisinage et le minimum nombre de points dans le -voisinage d'un point central respectivement. Si cette condition n'est pas satisfaite alors ce point est considéré comme un point non essentiel.

DBSACN recherche les clusters en vérifiant le voisinage de chaque objet dans le jeu de données. Si le voisinage d'un objet p contient plus de MinPts, un nouveau cluster avec p comme objet principal est créé. Il ensuite itérativement collecte directement les objets accessibles en densité à partir de ces noyaux objets, ce qui peut impliquer la fusion d'un nouveau cluster accessible à la densité. Le processus se termine lorsqu'aucun nouvel objet peut être ajouté à n'importe quel cluster. Le pseudo-code de DBSCAN L'algorithme est illustré à la figure 8 **[Khan, K. et all, 2014].**

```
Algorithm: DBSCAN (D, Eps, MinPts)
// All objects in D are unclassified.
Begin
FOR ALL objects o in D DO:
If o is unclassified
        Call function expand_ cluster to construct a
cluster wrt. Eps and MinPts containing o.
End

FUNCTION expand_ cluster (o, D, Eps, MinPts)
Begin
     Retrieve the Eps-neighborhood (o) of o;
IF | N_Eps (o) | < MinPts      //i.e. o is not a core object
        Mark o as noise point and RETURN;
ELSE          // i.e. o is a core object
        Select a new cluster- id and mark all objects
in N_Eps (o) with
This current cluster-id
   Push all objects from N_Eps (o)| (o) onto the
 Stack seeds;
   WHILE NOT seeds.empty () DO
CurrentObject: = seeds.top ();
Retrieve the Eps-neighborhood
N_Eps (CurrentObject)  of CurrentObject;
    IF| N_Eps (CurrentObject) | ≥ MinPts.
        Select all objects in N_Eps (currentObject) not
yet classified or are marked as noise.
        Push the unclassified objects onto seeds and
mark all of these objects with current
 Current-id;
Seeds. Pop ();
RETURN
End
```

Figure 8: Pseudo-code de DBSCAN

- **OPTICS :**

("Ordering Points to Identify the Clustering Structure") est un algorithme pour trouver des clusters basés sur la densité dans les données spatiales. Il a été présenté par Michael Ankerst, Markus M. Breunig, Hans-Peter Kriegel et Jörg Sander. Son idée de base est similaire à DBSCAN, mais il répond à l'une des principales faiblesses de DBSCAN : le problème de la détection de clusters significatifs dans des données de densité variable. Pour ce faire, les points de la base de données sont (linéairement) ordonnés de sorte que les points les plus proches spatialement deviennent voisins dans l'ordre. De plus, une distance spéciale est stockée pour chaque point qui représente la densité qui doit être acceptée pour un cluster afin que les deux points appartiennent au même cluster. Il faut couper la densité de clusters qui n'est plus considérée comme intéressante et accélérer ainsi l'algorithme **[Verma, 2012].**

- **Arbre de décision**

Les arbres de décision sont des classificateurs en l'attribut concerné (ou classe) dans une forme d'un arbre. Les items à classer sont composés des attributs et leurs valeurs ciblent. Les nœuds d'un arbre peuvent être **[Benabdi et Bourssia, 2016]** :

- Nœuds de décision : dans ces nœuds un attribut-valeur unique est testé pour déterminer à quelle branche du sous arbre applique.

- Nœuds de feuilles : ceux qui indiquent la valeur de l'attribut ciblent. Il y existe plusieurs algorithmes de l'arbre de décision parmi eux : Hunt's Algorithm, CART.

- **Les réseaux de neurones (*en anglais : neural network*)**

Dans la dernière dizaine d'années, l'utilisation des réseaux neuronaux a connu un essor incroyable dû à leur performance dans des situations où le nombre de variables est important et les relations entre celles-ci ne sont pas évidentes **[Lefebvre-Brossard, 2018]**.

Nous pouvons définir les réseaux de neurones comme une modélisation mathématique du cerveau humain, aussi est un assemblage des nœuds interconnectés et des liens pondérés dans le but de trouver le meilleur poids pour chaque neurone où le poids est le degré d'influence par rapport au neurone **[Djeriri, 2017].**

Les nœuds dans les réseaux neurones artificiel (ANN : *Artificial Neural Networks*) sont appelés des neurones, comme une analogie avec les neurones biologiques. Ces unités fonctionnelles sont composées en réseaux qui sont capables d'apprendre un problème de classification après qu'ils sont entrainés avec des données suffisantes **[Benabdi et Bourssia, 2016]**.

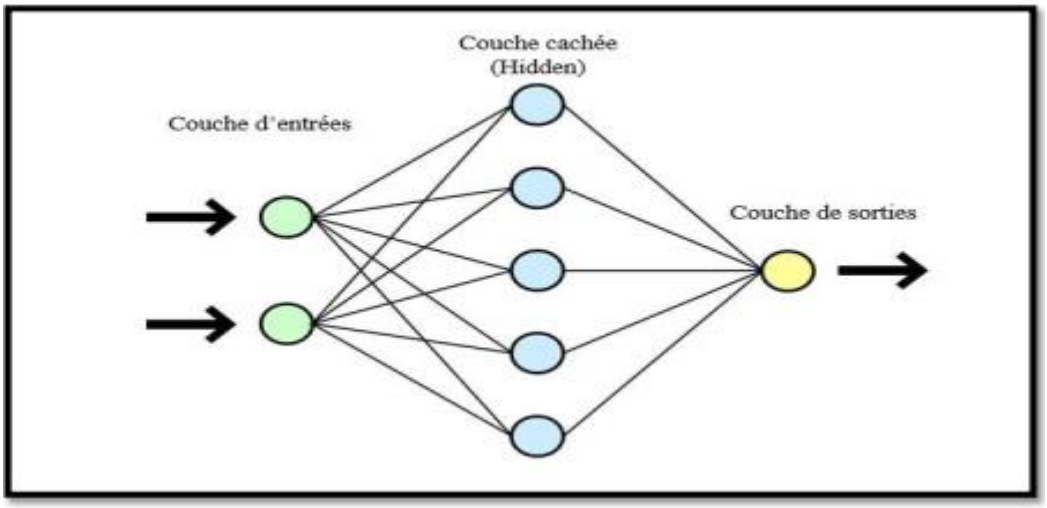


Figure 9: Réseau de neurones articiel ANN

- **Plus proche voisin (*en anglais : K-Nearest Neighbors* (KNN))**

C'est un des algorithmes d'apprentissage artificiel le plus simple, qui dispose une base de données d'apprentissage de m couple (entrée- sortie), son principe général est de normaliser les données et les récupérer (*Growing*) **[Eve MathieuDupas, 2010].**

- **La classification bayésienne (*Naïve Bayes Classification*)**

Naïve Bayes classifier est un simple algorithme et une méthode de classification probabiliste aide la tache de classification, même si nous travaillons sur un grand ensemble de données, cet algorithme s'est révélé très efficace pour la classification textuelle et ses résultats sont excellents **[Martin, 2019].**

I.5. Conclusion :

Ce chapitre a présenté les méthodes courantes et les algorithmes d'apprentissage populaires utilisés sur le terrain.

L'apprentissage automatique reste avant tout un sous-domaine de l'informatique, qui regroupe des techniques permettant à une machine d'adapter et d'améliorer ses performances par l'expérience, mais il est étroitement lié au divers domaine tel que : les sciences cognitives, les neurosciences, la biologie et la géologie, et pourrait aussi à la croisée des nanotechnologies, biotechnologies, informatique et sciences cognitives, aboutir à des systèmes d'intelligence artificielle ayant une assise plus vaste.

Dans les prochains, chapitres nous allons présenter la relation entre l'apprentissage automatique et les activités géologiques << sismologie >>, ainsi que quelques travaux précédents dans ce domaine.

Chapitre II : Etat de l’art

II Chapitre II : Etat de l'art

II.1. Introduction

La menace engendrée par les risques naturels reste présente partout dans le monde accentué par la croissance exponentielle de la vulnérabilité au niveau mondial, parmi les catastrophes naturelles, les tremblements de terre ont été la cause des désastres les plus meurtriers, affectant plus de 80 pays et causant la mort de plus de 1,6 millions de personnes. Chaque année, environ vingt séismes importants, de magnitude supérieure à 7, secouent la terre, selon le recensement du dernier siècle des Nations Unies. Plus. Suite au développement général des techniques spatiales, des chercheurs se sont penchés sur l'utilisation de la télédétection dans le génie parasismique. Les premières études consistantes ont été réalisées après le séisme de Kocaeli, (Turquie. 1999), ou l'imagerie radar et optique a été utilisée pour la détection des zones affectées.

Le Nord de l'Algérie a connu plusieurs séismes forts à modérés. Le plus violent séisme enregistré le 10 octobre 1980 à EL ASNAM (Echèlif actuellement), d'une magnitude de 7.3 (Ms), a causé plus de 5000 victimes, cette activité sismique est due principalement à la nature géographique de la région Maghrébine et a ces caractéristiques tectoniques à la frontière des plaques Africaine et Eurasienne, en mouvement compressif permanent.

Les travaux que nous présentons ici, une connaissance sur le phénomène sismique et les zones sismiques en Algérie.

II.1. Définition :

Avant d'entrer dans le vif du sujet, nous allons donner quelques définitions nous permettant une bonne compréhension de la suite **[POUR, MDEFIND, et al, 2003]** :

- **Séisme :**

Séisme provient du mot grec séismes, signifiant secousse. C'est donc un mouvement, plus ou moins violent. Un séisme ou (un tremblement de terre) est un phénomène géologique qui provoque des vibrations de la surface de la Terre. Ces secousses sont brutales, plus ou moins violentes, imprévisibles et toujours localisées en un lieu donné.

Les tremblements de terre se composent toujours d'une secousse principale (de quelques secondes) et de secousses secondaires (appelées répliques) durant les heures ou les jours qui suivent.

Un séisme peut se traduire à la surface terrestre par la dégradation ou la ruine des bâtiments, des décalages de la surface du sol de part et d'autre des failles.

- **Séismologie :**

La sismologie c'est la science qui étudie les phénomènes sismiques et l'instrument d'étude principal est le sismographe permettant de détecter les ondes sismiques.

- **Foyer (ou Hypocentre) :**

Le foyer d'un séisme (ou hypocentre) est le lieu sur la faille où se déclenche la rupture et d'où partent les ondes sismiques l'endroit où se produit la rupture de la faille (zones de rupture dans la roche).

- **Epicentre :**

L'épicentre d'un séisme est le point théorique situé à la surface terrestre à la verticale du foyer du séisme.

Faille sismique : Une faille est une cassure de la croûte terrestre qui s'accompagne d'un déplacement des deux compartiments ainsi créés. Sa mesure relative correspond au « rejet ».

II.1.2.Les différentes classes de séisme :

On distingue trois classes de séismes, en fonction de la profondeur où ils se produisent **[Andriamampandry, A, 2014]** :

- **Les séismes superficiels :**

Les séismes superficiels qui se produisent en faible profondeur, soit dans les premières dizaines de kilomètres, à moins de 60km, et qui se retrouvent autant aux frontières divergentes, c'est à dire le long de la croûte, à l'endroit où les plaques tectoniques s'éloignent les unes des autres.

- **Les séismes intermédiaires :**

Les séismes intermédiaires qui se produisent entre quelques dizaines et quelques centaines de kilomètres de profondeur de 60 à 300km et se concentrent uniquement au voisinage des limites convergentes ; se localise dans le manteau supérieur, uniquement à l'endroit où les plaques tectoniques se rencontrent.

- **Les séismes profonds :**

Les séismes profonds qui se produisent à des profondeurs pouvant atteindre les 300 à 700km (il n'y en a plus au-delà), soit en pratique la base de l'asthénosphère, et qui se trouvent exclusivement au voisinage de limites convergentes, ces tremblements de terre se produisent souvent juste avant et pendant une éruption volcanique. Ils sont dus au magma qui, en remontant vers la surface, fracture la croûte terrestre. Il est à noter que : 95% des tremblements de terre dans le monde ont lieu à une profondeur inférieure à 60km, 5% seulement ont lieu à une profondeur supérieure à 60km.ces sont les séismes superficiels qui sont les plus destructeurs.

II.2. Les sièges des tremblements de terre :

Les tremblements de terre se produisent à des endroits très localisés et bien connus à la surface de la planète. Ces zones, plus sismiques que d'autres, sont liées au phénomène de la tectonique des plaques, qui explique les mouvements des continents à la surface du globe. Ces zones sismiques se situent **[El Kadi, 1993]** :

Au milieu des océans : Le long des montagnes sous-marines (les dorsales océaniques), où les séismes sont peu profonds mais très fréquents ;

En bordure des continents : En bordure des continents où les plaques tectoniques convergent (elles entrent en collision) et donnent naissance aux longues chaînes de montagnes (les Andes, les Rocheuses)**.**

Au milieu des continents : **a**u niveau des grandes chaînes de montagnes (les Alpes, l'Himalaya).

II.3. Mesure d'un séisme :

II.3.1.Les échelles sismiques :

Nous disposons de deux échelles pour évaluer les tremblements de terre : l'échelle de Mercalli et l'échelle de Richter **[POUR, MDEFIND, et al, 1996]**.

II.3.1.1. L'échelle de Mercalli :

L'échelle de Mercalli a été développée en 1902 et modifiée en 1931. Elle indique l'intensité d'un séisme sur une échelle de I à XII. Cette intensité est déterminée par deux choses : l'ampleur des dégâts causés par un séisme et la perception qu'a eu la population du séisme. Il s'agit d'une évaluation qui fait appel à une bonne dose de subjectivité. De plus, la perception de la population et l'ampleur des dégâts vont varier en fonction de la distance à l'épicentre. On a donc avec cette échelle, une échelle variable géographiquement. Mais, à l'époque, on ne possédait pas les moyens d'établir une échelle objective.

II.3.1.2. L'échelle de Richter :

L'échelle de Richter a été instaurée en 1935. Elle nous fournit ce qu'on appelle la magnitude d'un séisme, calculée à partir de la quantité d'énergie dégagée au foyer. Elle se mesure sur une échelle logarithmique ouverte ; à ce jour, le plus fort séisme a atteint 9,5 sur l'échelle de Richter (Chili). Cette fois, il s'agit d'une valeur qu'on peut qualifier d'objective : il n'y a qu'une seule valeur pour un séisme donné. Aujourd'hui, on utilise un calcul modifié du calcul originel de Richter, en faisant intervenir la dimension du segment de faille le long duquel s'est produit le séisme.

II.3.2.Magnitude et intensité d'un séisme :

L'importance d'un séisme se caractérise par deux paramètres : sa magnitude et son intensité.

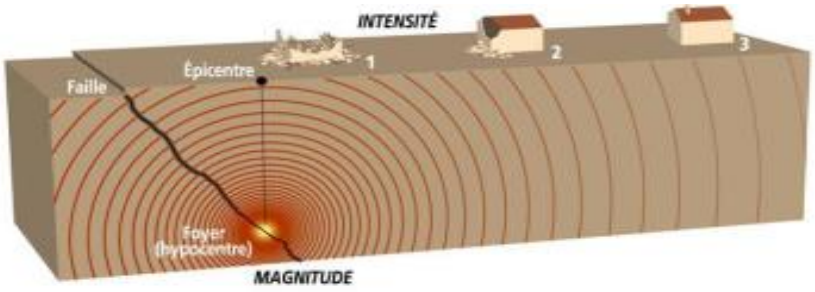


Figure 10: Magnitude et intensite d'un séisme[1]

Afin d'estimer la puissance des séismes, le sismologue Richter introduisît la notion de magnitude. Elle correspond à l'énergie libérée au foyer du séisme et dépend de la longueur de la faille activée et de l'importance du déplacement Une augmentation d'une unité sur l'échelle de Richter correspond à des secousses dix fois plus grandes. Ainsi, un séisme de magnitude 7 sera dix fois plus fort qu'un séisme de magnitude 6 et cent fois plus fort qu'un séisme de magnitude 5.

II.4. La répartition mondiale des séismes :

Les séismes n'ont pas une répartition aléatoire à la surface de la planète, mais sont répartis selon un patron bien défini. Cette répartition ordonnée vient appuyer la théorie de l'atectonique des plaques, particulièrement, en ce qui concerne l'existence de zones de subduction. On retrouve les séismes surtout aux frontières des plaques lithosphériques **[Schettkat, R. 2010]**. Carte des zones sismiques dans le monde :

[1] http://www.mementodumaire. Net/wp-content/uploads

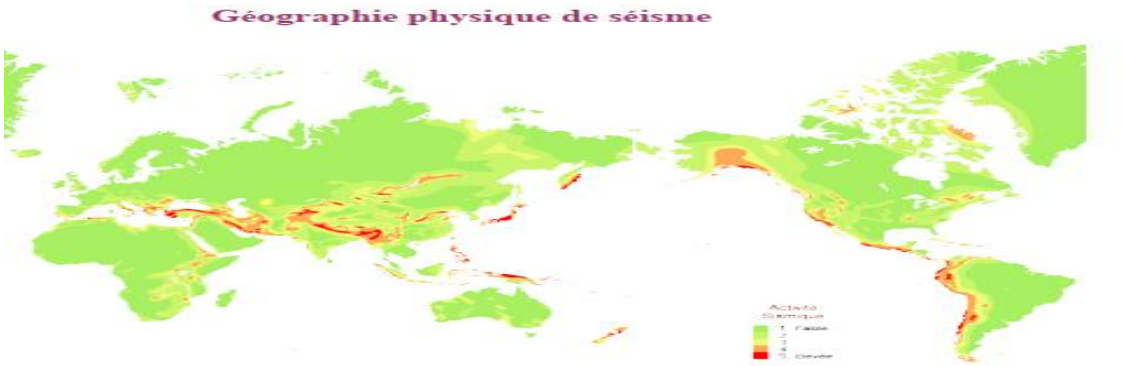


Figure 11: Géographie physique de séisme[2]

II.4.1.Les pays à fort risque sismique :

Aperçu détaillé des pays à fort risque sismique **[Richard, N. 2001]**:

- **USA :**

Alaska - Arkansas - California - Hawaii - Idaho - Montana - Oregon – Caroline du Sud - Utah – Washington**.**

- **AMERIQUE DU NORD :**

Canada – Mexique.

- **EUROPE :**

Albanie - Autrich**e** - Îles des Açore**s (Portugal)** - Belgiqu**e** - Croati**e** -Chypr**e** - République tchèqu**e** - Franc**e** - Allemagn**e** - Grèc**e** - Hongri**e** - Island**e** -Irland**e** - Itali**e** - Macédoin**e** - Pays-Ba**s** - Pologn**e** - Portuga**l** - Roumani**e** - Russi**e** – La Slovaqui**e** - La Slovéni**e** - Espagn**e** -Suiss**e** - Turqui**e** - Ukrain**e** - Royaume-Uni**.**

- **MOYEN-ORIENT :**

L'Iran - Irak - Israël - Jordanie - Liban - Arabie Saoudite – La Syrie – Yémen**.**

[2] http://www.cosmovisions.com/qMondeSismicite.html

- **AFRIQUE :**

Algérie - Burundi - Cameroun - Îles Canaries (Espagne) - Djibouti – RD Congo - Egypte - Ethiopie - Ghana - Guinée - Kenya - Libye - Malawi - Maroc - Rwanda - Afrique du Sud - Tanzanie - Tunisie - Ouganda – Zambie**.**

- **ASIE :**

Afghanistan - Arménie - Azerbaïdjan - Bangladesh - Bhoutan - Chine -Géorgie - Japon - Kazakhstan - Kirghizistan - Corée (du Nord**)** - Corée du Sud) – **Laos** - Mongolie - Myanmar - Népal - Pakistan - Philippines - Taiwan - Tadjikistan - Thaïlande - Le Turkménistan - L'Ouzbékistan – Vietnam**.**

- **OCEANIE & OCEAN PACIFIQUE :**

Nouvelle-Zélande - Australie - Fidji - Îles Kermarec - Nouvelle-Calédonie - Papouasie-Nouvelle-Guinée - Samoa - Iles Salomon - Tonga – Vanuatu**.**

- **AMERIQUE CENTRALE :**

Costa Rica - El Salvador - Guatemala - Honduras - Nicaragua – Panama**.**

- **AMERIQUE DU SUD :**

Argentine - Bolivie - Brésil - Chili - Colombie - **Equateur** - Pérou – Venezuela**.**

- **CARAIBES :**

Antigua - La Barbade - Cuba - République dominicaine - Guadeloupe - Haïti - Jamaïque - Martinique - Sainte-Lucie – Trinité**.**

II.4.2.Localisation des épicentres des séismes dans le monde :

L'étude des données historiques et des données actuelles enregistrées par les sismographes a permis d'établir une carte des épicentres et des magnitudes et de mettre ainsi en évidence les grandes régions sismiques. Par ailleurs, la localisation des épicentres et la profondeur des foyers permettent de relier les séismes à d'autres phénomènes géologiques résultant de l'activité interne de la Terre.

La localisation des épicentres des séismes enregistrés par différentes stations sismiques est représentée sur la carte ci-dessous

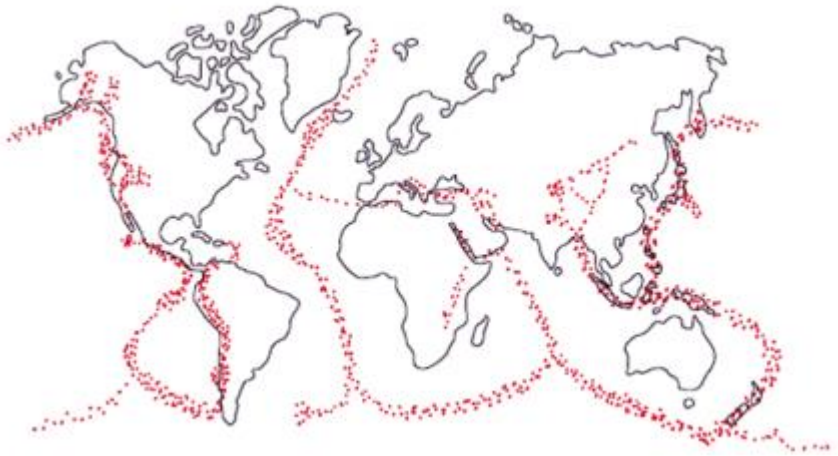


Figure 12: Localisation des épicentres des séismes dans le monde.[3]

On constate que ces séismes n'apparaissant pas au hasard à la surface du globe et que les foyers sont souvent répartis de façon linéaire.

On remarque que les épicentres des séismes récents se répartissent en trois zones principales:

- La zone péripacifique (pourtour du Pacifique) et la boucle des Antilles, qui représentent 80% de l'activité sismique annuelle.
- Les zones de dorsale océanique,
- Une zone située à la limite des continents Eurasiatique et Africain, des Açores à Java.

Entre ces zones de forte sismicité existent de vastes zones presque entièrement asismiques, océaniques ou continentales comme le Canada, la Sibérie, l'Afrique de l'Ouest, l'Australie, une partie de l'Ouest du continent Américain et le nord de l'Europe **[Locat, J. 2011].**

II.4.3.Situation de l'Algérie par rapport aux mouvements des plaques :

La chaîne tellienne (Atlas tellien) constitue le segment orogénique périméditerranéen de la ceinture active alpine et himalayenne qui s'étend du Sud-Ouest asiatique

[3] http://www.maxicour.com/se/fiche/7/8/23187.html/4e

à l'Océan Atlantique. Dans la région méditerranéenne, cette ceinture est caractérisée par la convergence des plaques tectoniques africaine et eurasienne. Les travaux récents basés sur des analyses des mécanismes au foyer des séismes forts, sur les études néotectoniques. Ainsi que des méthodes très récentes basées sur des techniques spatiales telles que le GPS (Global Position System), VLBI (Very Long Baseline Interferometry) et le SLR (Satellite Laser Ranging),cette dernière montre essentiellement les déplacements globaux et non les déplacent à court terme au niveau des frontières des plaques, qui permettent des mesures de plus en plus précises sur les déplacements des plaques tectoniques, montrent que la direction du raccourcissement est NNW-SSE. Le rapprochement entre les Plaques africaine et eurasienne et d'environ 4-6mm/an(figure13).

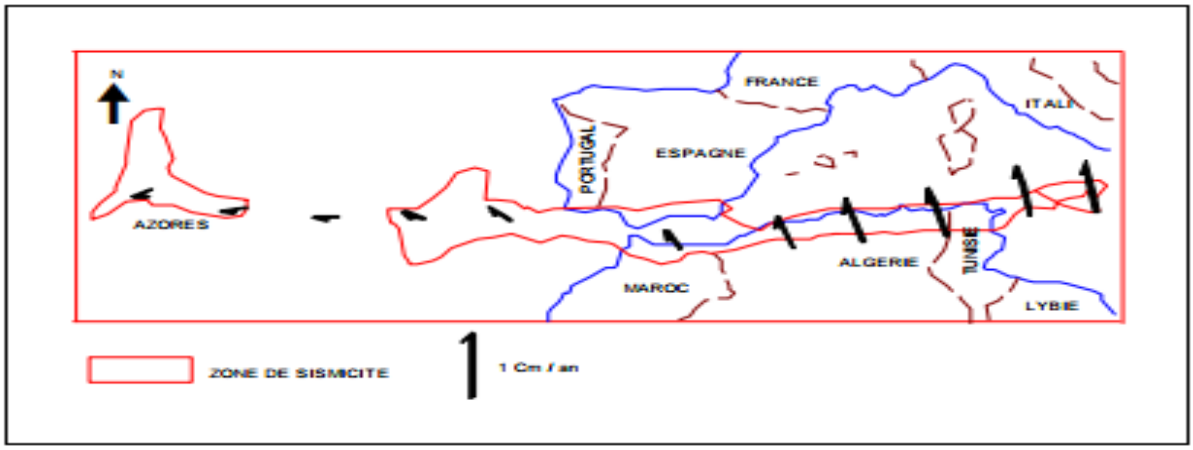


Figure 13: La limite des Plaques Africaine et Eurasienne.

Les flèches indiquent létaux que la direction du raccourcissement **[BOHADAD, Y, & BELHAID, 2003]**.

Ce contexte géodynamique régional a engendré dans le Nord de l'Algérie, sur environ 200 à 300Km de large, un ensemble de structures tectoniques (plis, failles) de direction générale perpendiculaire à la direction de convergence des plaques tectoniques ainsi qu'une activité sismique relativement élevée (Fig.04). Cette sismicité est superficielle d'où son caractère souvent destructeur **[ABDELHADI BOUSAFFA, RAOF BEN SLIMANE, 1997].**

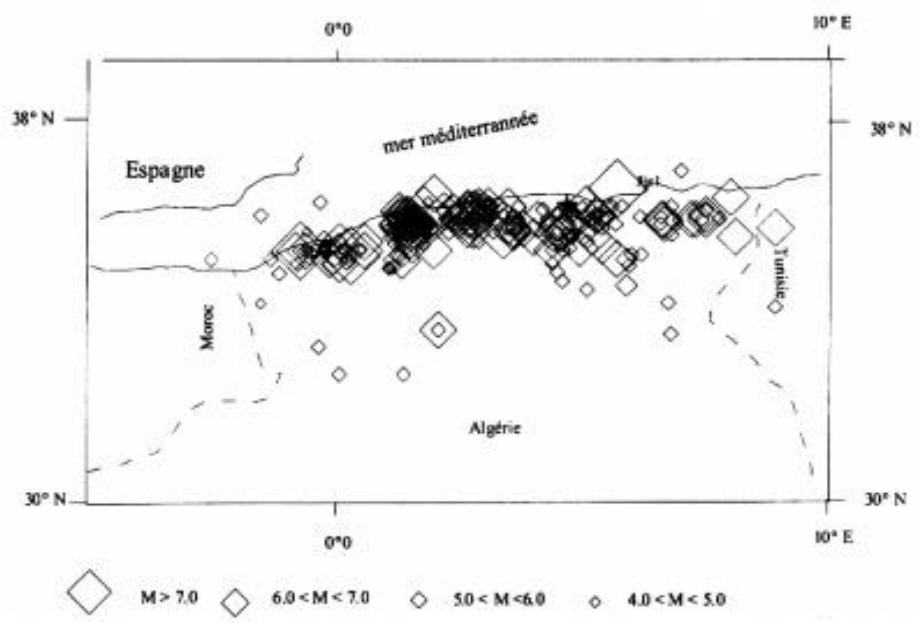


Figure 14: Distribution spatiale des tremblements de terre en Algérie du Nord.

II.4.4. Conditions sismotectoniques régionales :

La répartition des zones sismiques de l'Algérie est indiquée sur la (Figure 15), cette carte a été préparée sur la base de l'échelle Marcalli modifiée (1931) notons que les zones d'activité sismiques sont concentrées au Nord et au Nord-Est de l'Algérie, le long de la chaîne côtière qui s'étend entre Oran et Annaba, et dans la région du Hodna et des Aurès.

Une troisième zone d'activité se trouve dans l'Atlas passe par Gabes en Tunisie, Biskra et Laghouat en Algérie et Agadir en Maroc **[Belkhiri, Kamel, 2006].**

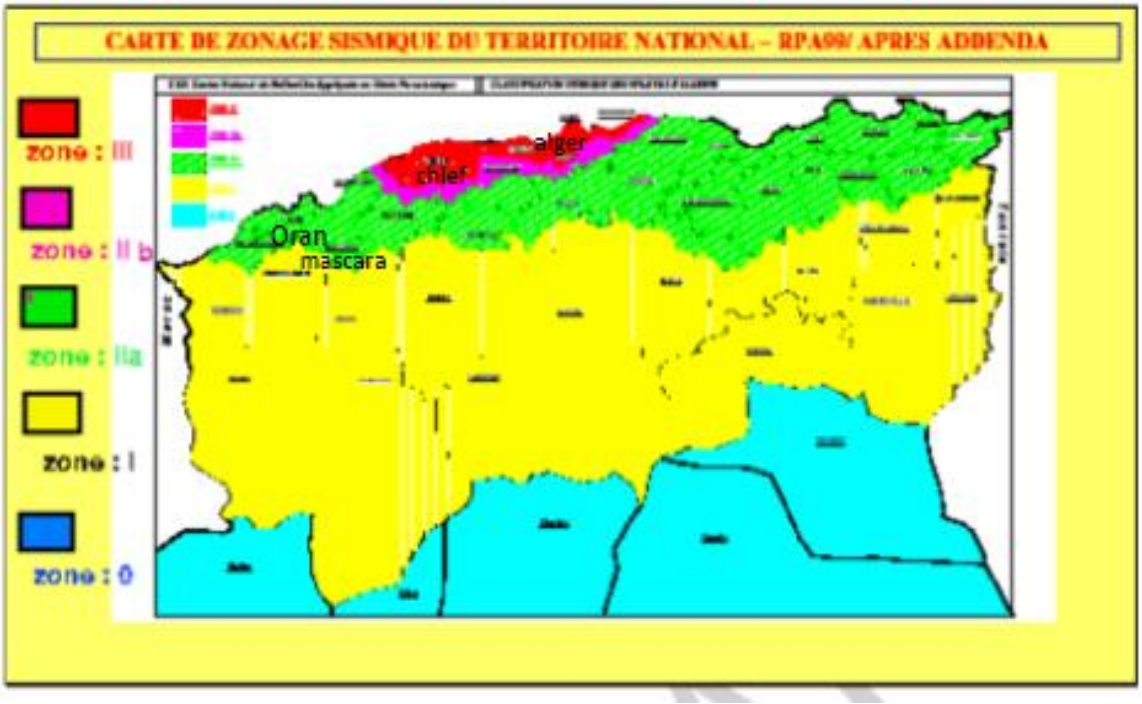


Figure 15: Repartions des zones sismique dans la période 1962-2020 .

II.4.5. Classification des zones sismiques en Algérie :

Le règlement parasismique Algérien (version 2003), divise le territoire national en cinq (05) zones de sismicité croissante, définies comme suit [Menasri, Y. 2009] :

- Zone 0 : sismicité négligeable
- Zone I : sismicité faible
- Zone (II a – II b) : sismicité moyenne
- La carte des zones sismique de l'Algérie, et le zonage globale des différents Wilaya. Représenté sur la (Figure 16).

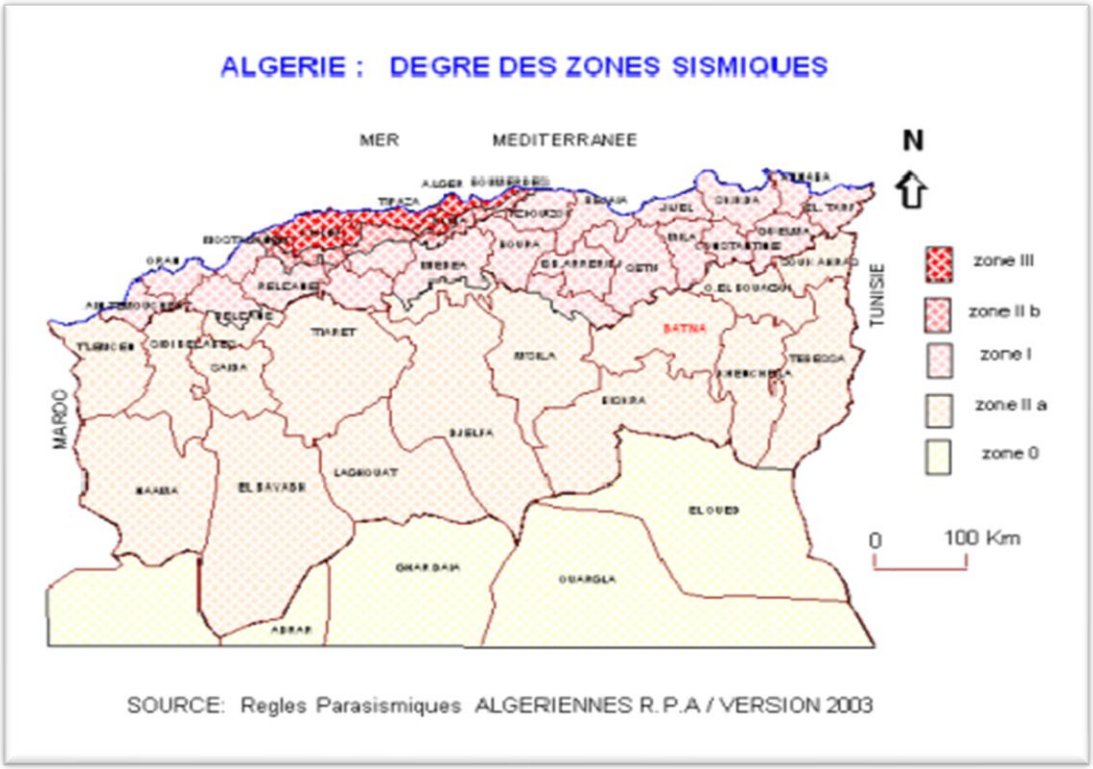


Figure 16: Degré des zones sismiques.

II.5. Dataset :

II.5.1.Définition :

Dataset est une collection de données. En d'autres termes, Dataset correspond au contenu d'un seul table de database, ou d'une matrice de données statistiques unique, où chaque colonne de la table représente une variable particulière, et chaque ligne correspond à un membre donné de l'ensemble de données en question.

II.5.2.Pourquoi avons-nous besoin d'un Dataset ?

L'apprentissage automatique dépend fortement des données, sans données il est impossible pour une "IA" d'apprendre, c'est l'aspect le plus crucial qui rend possible l'apprentissage d'un algorithme, Peu importe la qualité de l'équipe d'IA ou la taille de Datasets, si Dataset n'est pas assez bon, le projet d'IA tout entier échouera ! on a vu des projets fantastiques échouer parce qu'ils n'avaient pas une bonne base de données.

Dataset est une collection de données. Le plus souvent, Dataset correspond au contenu d'un seul tableau de Dataset ou d'une seule matrice de données statistiques, où chaque colonne du tableau représente une variable particulière et chaque ligne correspond à un membre donné de Dataset en question, Dataset énumère les valeurs de chacune des variables, telles que la taille et le poids d'un objet, pour chaque membre de Dataset. Chaque valeur est appelée une donnée. Dataset peut comprendre des données pour un ou plusieurs membres, correspondant au nombre de rangées.

II.5.3.Types base de données :

Dataset peut être de plusieurs types différents. Le type de Dataset est distingué sur la base de stockage et de structure des données et, puisque FME traite chaque type différemment, il est important de connaître les différences et de savoir avec quel type de données vous travaillez. Il est donc important d'être conscient de ces différences et de savoir avec quel type de données vous travaillez.

Dataset Type	Characteristics	Example
File	A single file	Auto DXF
Folder	A set of files in a single folder	Esri Shape file
Database	A database	Oracle Spatial
Web	An internet site	Web Feature Service (WSF)

Figure 17: Les types du Dataset

- **File :**

Dataset basé sur un fichier est Dataset complet stocké dans un seul fichier, un fichier DXF d'AutoCAD en est un bon exemple : chaque fichier DXF est un Dataset distinct.

- **Folder :**

La logique qui sous-tend un ensemble de données basé sur des dossiers peut être légèrement déroutante. Dans cette forme, Dataset est le dossier ou le répertoire contenant les données. Par exemple, dans C : FMEData\Data\Properties, le nom de la base de données est Propriétés.

- **Database :**

Database Dataset, comme son nom l'indique, est un Dataset stockées dans une Dataset. en général, chaque Dataset différente est un Dataset différent (bien que, pour être strictement exact, chaque utilisateur/schéma dans une Dataset serait un Dataset différent).

L'exemple le plus évident est une Dataset Oracle. Elle sera traitée de la même manière qu'elle soit spatiale ou non spatiale.

- **Web**

Un Dataset Web est une collection de données stockées sur un site Internet. Un serveur Web Feature Service (WFS) en est un exemple. Dans ce cas, le nom de Dataset est le même que le nom de l'URL (Universal Resource Locator).

II.6. Travaux précèdent :

Les tremblements de terre ne peuvent toujours pas être prédits avec précision à l'heure actuelle, mais l'analyse des données sismiques a été étudiée en permanence.

Dans l'article d'application et visualisation d'algorithmes de regroupement typiques dans l'analyse des données sismiques, deux algorithmes de regroupement typiques, K-means et DBSCAN, sont appliqués à l'analyse des données sismiques.

Ils ont comparé l'application de deux algorithmes dans la recherche de la distribution sismique, ils ajustent et visualisent les ceintures sismiques. Ensuite, les tremblements de terre sont regroupés par profondeur focale et magnitude sismique, afin de les classer.

Les résultats montrent qu'il est tout à fait possible d'appliquer l'algorithme de regroupement à l'analyse des données sismiques.

Pour l'étude de l'ajustement des ceintures sismiques, l'algorithme DBSCAN basé sur la densité est de loin supérieur à l'algorithme K-means.

Actuellement, une grande partie de la prédiction des tremblements de terre est basée sur l'extraction de données sismiques existantes et la construction de modèles.

Par conséquent, il est également pratique d'appliquer les algorithmes de regroupement à la recherche sur la distribution sismique. Dans le futur, les séries temporelles peuvent être introduites dans la recherche pour étendre l'étude de la distribution sismique de la surface de simulation à l'espace-temps. Dans le même temps, une visualisation spatio-temporelle sera réalisée pour contribuer à la recherche future **[Leuven et Belgium, 2019]**.

Dans l'article d'Analyse des groupes de tremblements de terre, a démontré l'analyse en cluster des tremblements de terre en utilisant la technique des k-means. L'ensemble de données avec 1657 événements sismiques de l'Inde qui ont eu lieu entre le 1er janvier 2005 et le 31 décembre 2015 est sélectionné pour l'analyse. La présente étude a démontré l'analyse en grappes en modulant le nombre de grappes. L'analyse de cluster est L'analyse de cluster est accomplie en recourant à une série de techniques qui permettent la subdivision d'un ensemble de données en sous-groupes, sur la base de leurs similarités. Ainsi, le clustering k-means dérivé résulte en 6 clusters de tailles 303, 149, 817, 218, 152, 18.

Le résultat suggère fortement que les k-means ont le potentiel d'exposer l'outil prééminent pour l'analyse de cluster de tremblement de terre **[Rajani S, 2017].**

II.7. Conclusion :

Les séismes n'ont pas une répartition aléatoire à la surface de la planète, mais sont répartis selon un patron bien défini. Cette répartition ordonnée vient appuyer la théorie de la tectonique des plaques, particulièrement, en ce qui concerne l'existence de zones de subduction. On retrouve les séismes surtout aux frontières des plaques lithosphériques. Notre pays a toujours été sismique et avec la même intensité depuis au moins 6 millions d'années et que tout le Nord du pays est sujet à une sismicité continue, On ne peut évidemment pas empêcher la survenue d'un séisme (aléa) ni bien même en limiter la puissance. Il n'existe également aucun moyen fiable de prévoir où, quand et avec quelle puissance se produira le prochain séisme. Il arrive sans aucun signe avant-coureur et vous surprendra dans votre sommeil, au travail, ou chez vous. Contrairement à des risques météorologiques, aucune alerte n'est possible ! Pour réduire ce risque et se protéger des conséquences d'un séisme, il est donc indispensable de diminuer notre vulnérabilité, en adoptant les bons gestes et en développant un habitat le plus parasismique possible.

Chapitre III :

Implémentation et évaluation

III Chapitre III : Implémentation et évaluation

III.1. Introduction :

L'apprentissage automatique est un domaine qui aide a obtenu des bons résultats dans tous les domaines, donc nous avons utilisé ses algorithmes en sismologie et donc nous espérons atteindre les résultats souhaités.

Dans ce chapitre, nous avons appliqué des déférents algorithmes de clustering pour l’extraction des zones séismiques en Algérie.

III.2. Démarche générale de travaille :

Le schéma suivant représente une description de la démarche générale de ce travail :

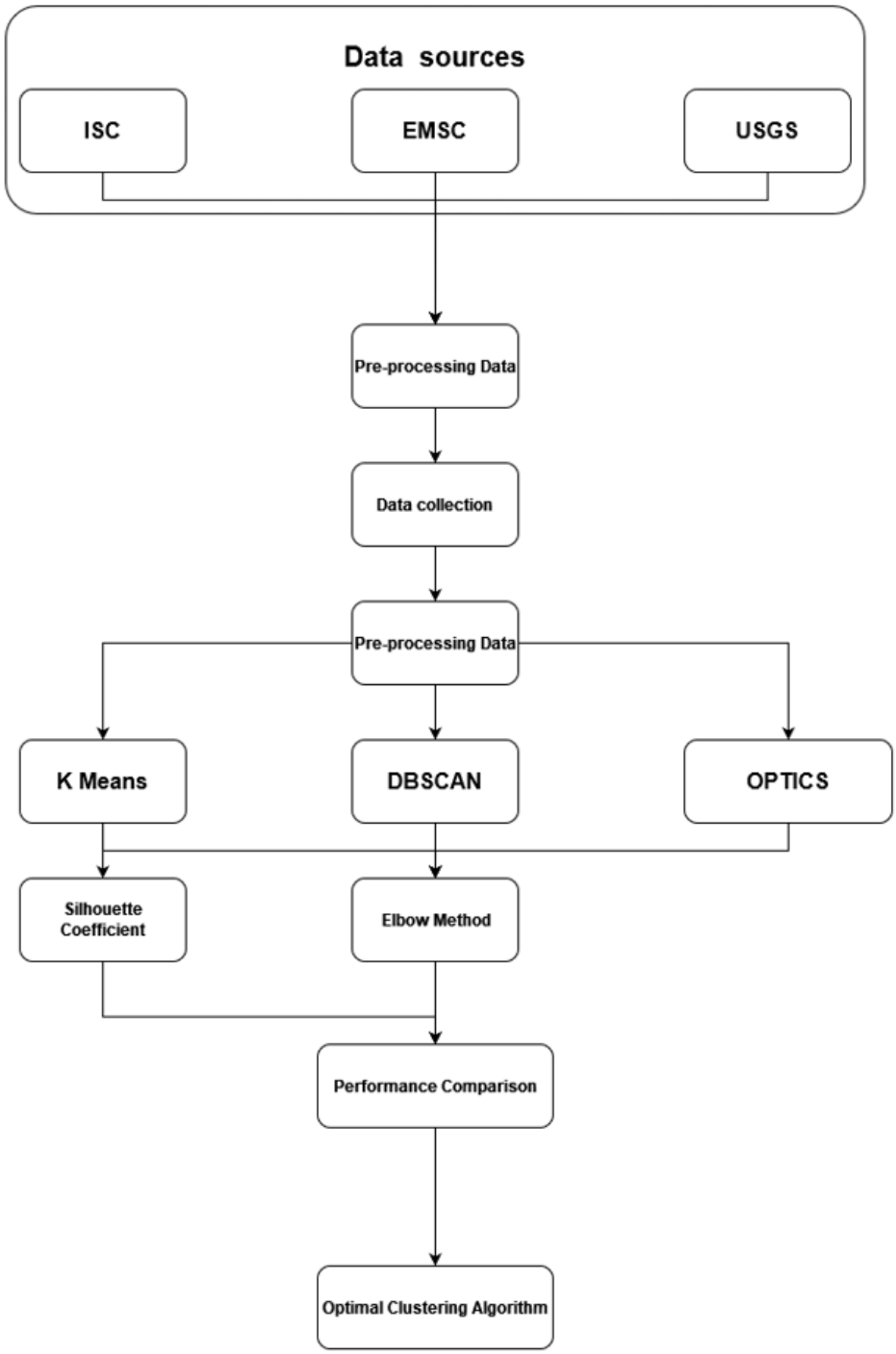


Figure 18: Démarche générale de travaille.

Chapitre III : Implémentation et évaluation

III.3. Collection des Dataset (les données) :

III.3.1. Les source de Dataset :

Dans l'apprentissage automatique il est nécessaire d'utiliser des bonnes Dataset pour obtenue des bons résultats pour cela nous avons utilisé des Dataset qui sont proposées par :

- **EMSC :**

European Mediterranean Seismological Centre [4] : cette base contient 1123 lignes de donner dater de 2004 jusqu'à 2020.

- **ISC :**

International Seismological Centre [5] : cette base contient 4472 lignes de donner daté de 1963 jusqu'à 2020.

- **USGS**

United States Geological Survey [6] : cette base contient 1307 lignes de donner daté de 1963 jusqu'à 2020.

III.3.2. Prétraitement des Datasets :

Avant de fusionner les trois Dataset un prétraitement doit être effectuer

Etape 1 : les bases de données proposées comporter de nombreuses parties non pertinentes par rapport notre travail, Pour cella en garde juste ces caractéristiques : « “Date", "Time", "Latitude", " Longitude ", "Magnitude", "Magnitude Type" ».

Date : Une indication précise du jour, du mois et de l'année du tremblement de terre.

[4] https://www.emsc-csem.org/

[5] http://www.isc.ac .uk/

[6] https://www.usgs.gov/

Time : Indication précise de l'heure, minute et seconde du tremblement de terre.

Latitude: La latitude est une coordonnée géographique représentée par une valeur angulaire, expression de la position d'un point sur Terre (ou sur une autre planète), au nord ou au sud de l'équateur qui est le plan de référence.

Longitude : La latitude est une coordonnée géographique représentée par une valeur angulaire, expression de la position d'un point sur Terre (ou sur une autre planète), au nord ou au sud de l'équateur qui est le plan de référence.

Magnitude : la magnitude est la représentation logarithmique du moment sismique, qui est lui-même une mesure de l'énergie libérée par un séisme déduite de l'amplitude de certaines ondes sismiques à des distances spécifiques.

Magnitude Type le type du tremblement de terre.

Etape 2 : On a trié chaque Dataset selon la caractéristique "Date" du plus ancien au plus récent.

Etape 3 : On a ajouté une caractéristique dans le Dataset avec le nom "Dataset" pour indiquer la source de chaque ligne.

Etape 4 : On a renommé les caractéristiques pour les unifier après la fusion.

Etape 6 : dans le Dataset proposé par ISC la caractéristique "time" on a changé le format “hh:mm:ss:ms" à la format "hh:mm:ss".

Etape 7 : dans le Dataset proposé par USGS on a changé la caractéristique "time date" de format "yyyy-mm-dd hh:mm:ss:ms" à la format "yyyy-mm-dd hh:mm:ss" en ajoutant deux caractéristique "date" et "time" et en supprimant le caractéristique "time date".

III.3.3. Collecte de données :

Pour créer une Dataset complète et suffisante en vas fusionner les trois Dataset proposé par EMSC_New & ISC_ New & USGS_ New selon caractéristique "Dataset", et on le sauvegarder dans un fichier ".csv"

III.3.3.1. Prétraitement de Dataset collection :

Après fusionner les trois Dataset un prétraitement doit être effectuer :

1. Supprimer les lignes double selon les caractéristiques "Date "&"Time ".
2. On a utilisé la bibliothèque "reverse_geocode" pour extraire les pays. En prenant les deux caractéristique "Latitude" et "Longitude" pour trouver la liste des pays.
3. Parmi la liste trouver en garder l'Algérie et éliminer les autres.
4. En a sauvegarde le Dataset après le prétraitement effectuer dans fichier ".csv".
5. Le Dataset finale est compose de 7 caractéristiques « "Date", "Time", "Latitude", "Longitude", " TypeMag", "Magnitude", " Dataset" », et composer de 6902 lignes de donner.

III.3.3.2. Une visualisation complète de Dataset :

La figure suivante représente l'implantation de datatset sur la carte de l'Algérie :

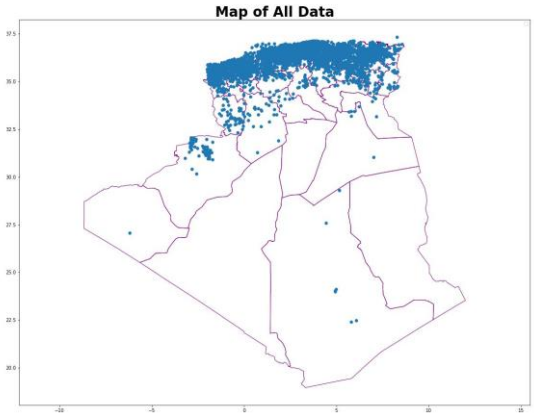


Figure 19: Visualisation de Dataset

III.3.3.3. Préciser le pourcentage de donnée de chaque ressource :

Dans le cercle relatif suivent en préciser le pourcentage de donner de chaque ressource de donner, ce Dataset un mélange des déférent ressource on a" ISE 61.5%","USGS 21%","EMSE 17.5%"

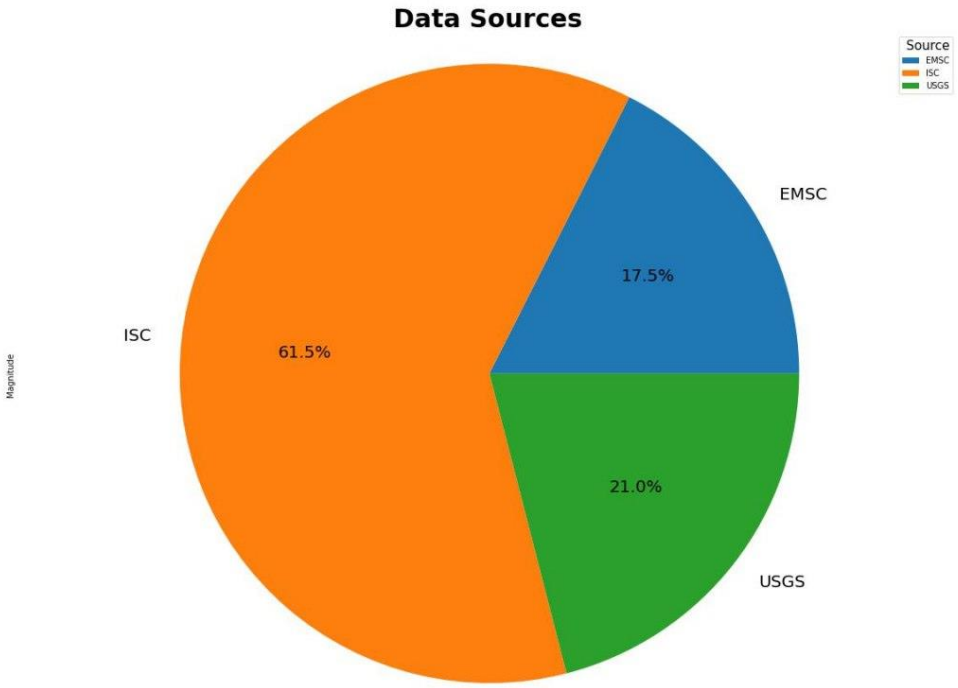


Figure 20 : Le pourcentage de Dataset par apport les déférent ressource.

III.3.3.4. Exprimer magnitude de séisme sélectionner dans le Dataset :

Cercle relatif suivent en préciser magnitude de séisme sélectionner dans le Dataset :

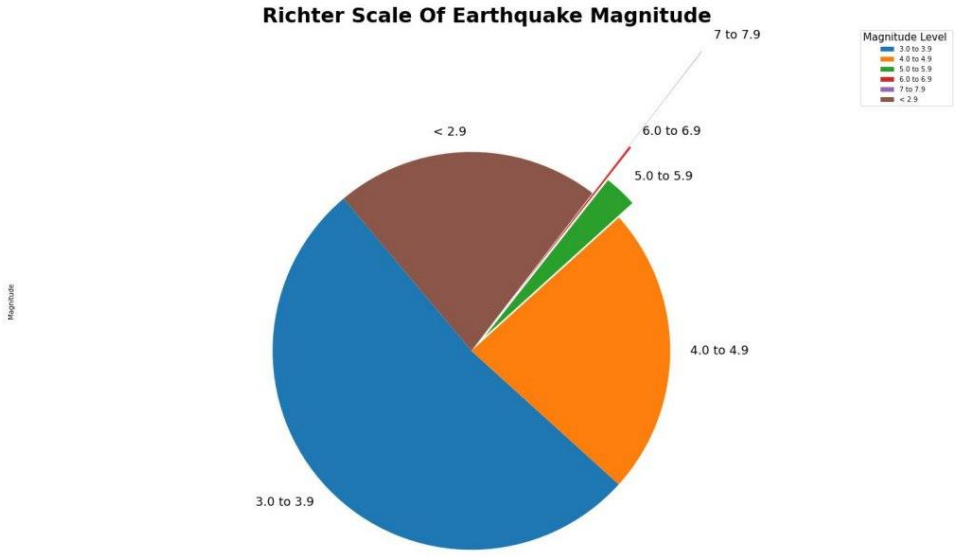


Figure 21: Le pourcentage de Magnitude dans le Dataset

III.4. Les Algorithmes

Après filtré et complété le Dataset, Nous pouvons appliquer un ensemble d'algorithmes à mettre en œuvre afin d'atteindre l'objectif souhaité.

III.4.1. Introduction

Le clustering est inclus dans une classification non supervisée. Comprendre Le clustering est le processus de regrouper ou classer des objets en fonction d'informations obtenues à partir de données qui expliquent la relation entre les objets avec le principe de maximiser la similarité entre les membres d'une classe et de minimiser les similarités entre les classes / clusters.

III.4.2. Kmeans

L'algorithme K-Means est un algorithme de clustering qui regroupe les données en fonction du point central du cluster (centroïde) le plus proche des données. L'objectif de K-Means est de regrouper les données avec une similarité maximale des données dans un cluster et une similarité minimale des données entre les clusters. Un cluster et de minimiser la similarité des données entre les clusters. Les mesures de similarité utilisées dans le cluster sont la fonction de distance. Ainsi, la maximisation de la similarité des données obtenue sur la base de la plus courte distance entre les données vers le point centroïde **[Nainggolan et all, 2019]**.

Parmi les inconvénients des algorithmes K-Means, l'utilisateur doit spécifier les nombres de clusters k, pour cela nous avons utilisé des méthodes pour déterminer le meilleur nombre de clusters, parmi ces méthodes :

III.4.2.1. Méthode du Coude (Elbow Method)

Est une méthode utilisée pour produire des informations permettant de déterminer le meilleur nombre de clusters en regardant le pourcentage de la comparaison entre le nombre de clusters qui formeront un coude à un point. Cette méthode fournit des idées/idées en sélectionnant les valeurs des clusters et en ajoutant ensuite la valeur du cluster à utiliser comme modèle de données pour déterminer le meilleur cluster. Et en plus de cela, le pourcentage du calcul résultant est une comparaison entre le nombre de clusters ajoutés **[E. Muningsih and S. Kiswati, 2018].**

Les différents résultats en pourcentage de chaque valeur de cluster peuvent être montrés en utilisant le graphique comme source d'information. Si la valeur du premier cluster avec la valeur du deuxième cluster donne l'angle dans le graphique ou si la valeur a la plus grande diminution, alors la valeur du cluster est la meilleure. Pour obtenir une comparaison est de calculer la SSE (Sum of Square Error) de chaque valeur de cluster. Parce que plus le nombre de cluster K, la valeur SSE sera plus petite.

Voici les étapes de l'algorithme de la méthode du coude pour déterminer la valeur k dans KMeans :

- Initialiser la valeur initiale de k.
- Augmentez la valeur de k.

- Calculer la somme des résultats de l'erreur quadratique pour chaque valeur de k.
- L'analyse de la somme de l'erreur quadratique résulte de la valeur k qui a drastiquement diminué.
- Localisez et définissez la valeur k en forme de coude.

Dans la méthode du coude, la meilleure valeur de cluster sera prise à partir de la valeur de la Somme des Erreurs Carrées (SSE) qui présente une diminution significative et en forme de coude.

III.4.2.1.1. L'application de la méthode du coude

Après avoir appliqué la méthode du coude aux données que nous avons, nous extrayons la meilleure valeur de cluster qui est égale à k=8, voici le résultat dans la figure suivante :

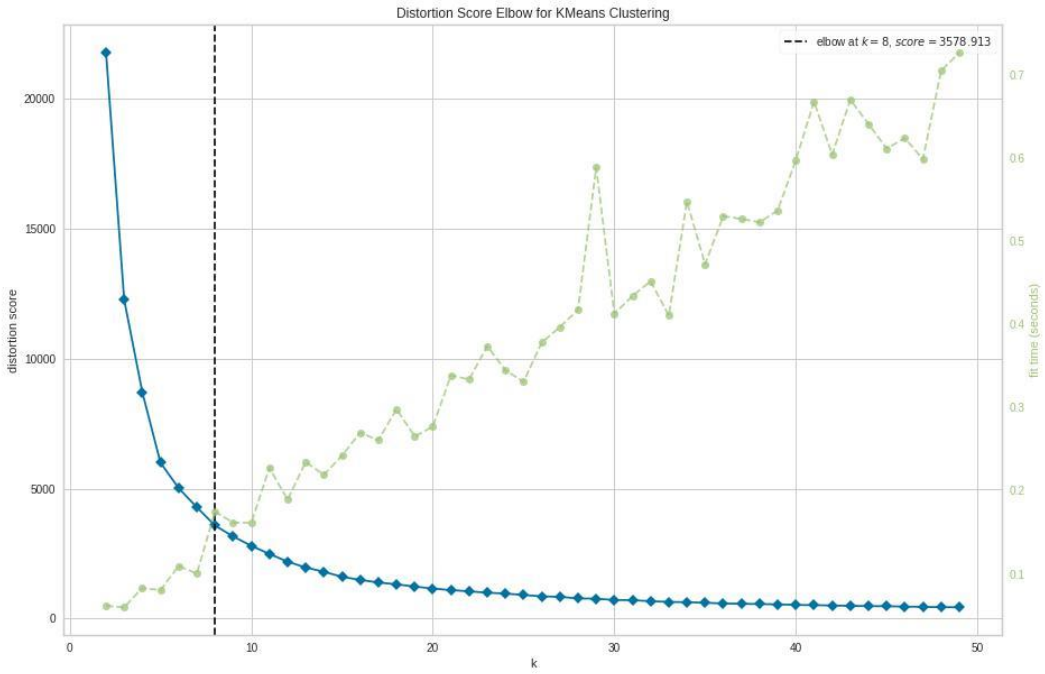


Figure 22: Graphe de méthode du coude

III.4.2.1.2. L'application L'algorithme Kmeans

Après avoir appliqué la méthode du coude et on trouve la meilleure valeur de cluster qui vaut 8 puis on applique L'algorithme Kmeans (k=8) et donne le résultat suivant la figure suivante:

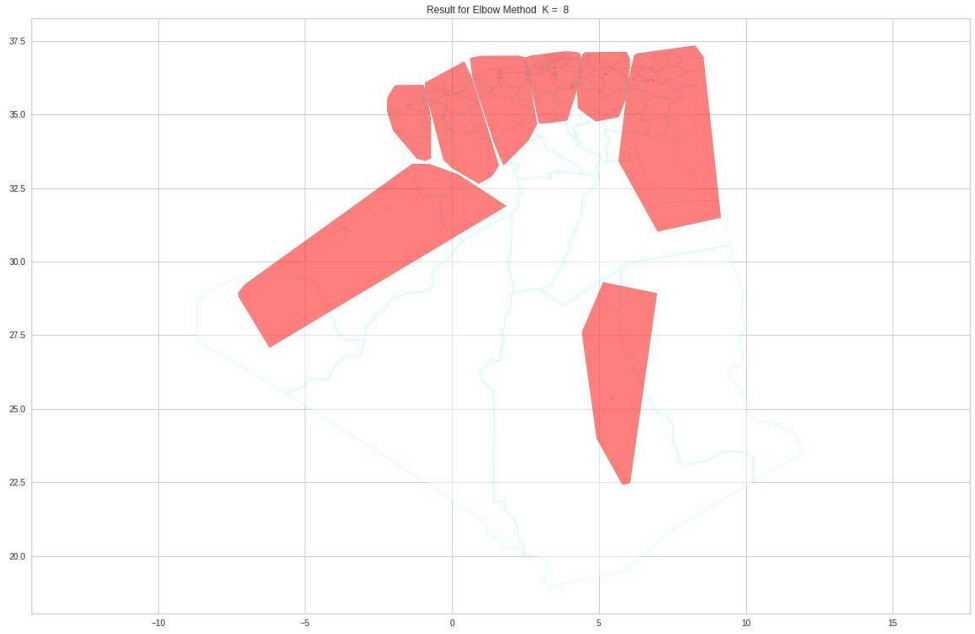


Figure 23: Résultat de Kmeans en utilise la méthode du Coude (k=8)

III.4.2.1.3. Constat

Après fait une comparaison le résultat précédent avec les résultats obtenus par les géologues dans la Figure suivant, on constate une grande déférence dans les zones, pour cela on a pensé d'améliorer le Dataset.

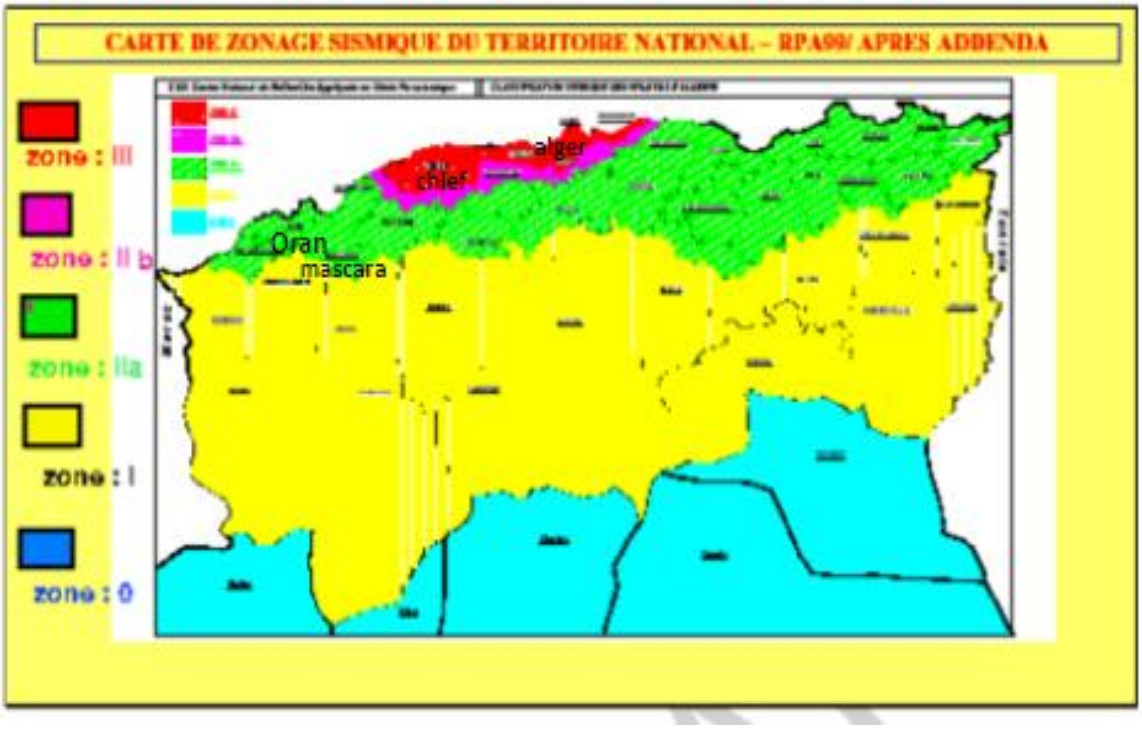


Figure 24: Les zones séismiques par apport les giologistes

III.4.2.1.4. Prétraitement des Dataset

Pour obtenir des résultats, il faut fais des modifications dans le Dataset en modifier les valeurs de le caractéristique "magnitude" dont le pic dépasse 3 selon le tableau suivant :

Richter scale of earthquake magnitude

magnitude level	category	effects	earthquakes per year
less than 1.0 to 2.9	micro	generally not felt by people, though recorded on local instruments	more than 100,000
3.0–3.9	minor	felt by many people; no damage	12,000–100,000
4.0–4.9	light	felt by all; minor breakage of objects	2,000–12,000
5.0–5.9	moderate	some damage to weak structures	200–2,000
6.0–6.9	strong	moderate damage in populated areas	20–200
7.0–7.9	major	serious damage over large areas; loss of life	3–20
8.0 and higher	great	severe destruction and loss of life over large areas	fewer than 3

Figure 25 : Échelle de Richter de magnitude du séisme

Chapitre III : Implémentation et évaluation

Dans notre étude, nous évitons d'étudier les zones désertiques car elles ne font pas partie a des zones sismiques, Selon des études sismologie comme Tindouf, Bachar, Tamanrasset, nous avons donc limité notre étude aux zones situées au-dessus de la 32ème latitude et Séismes d'intensité supérieure à 3 sur l'échelle de Richter.

Après quelques modifications, le nombre de données est réduit à 4043 lignes voici une visualisation complète de Dataset après ces modifications :

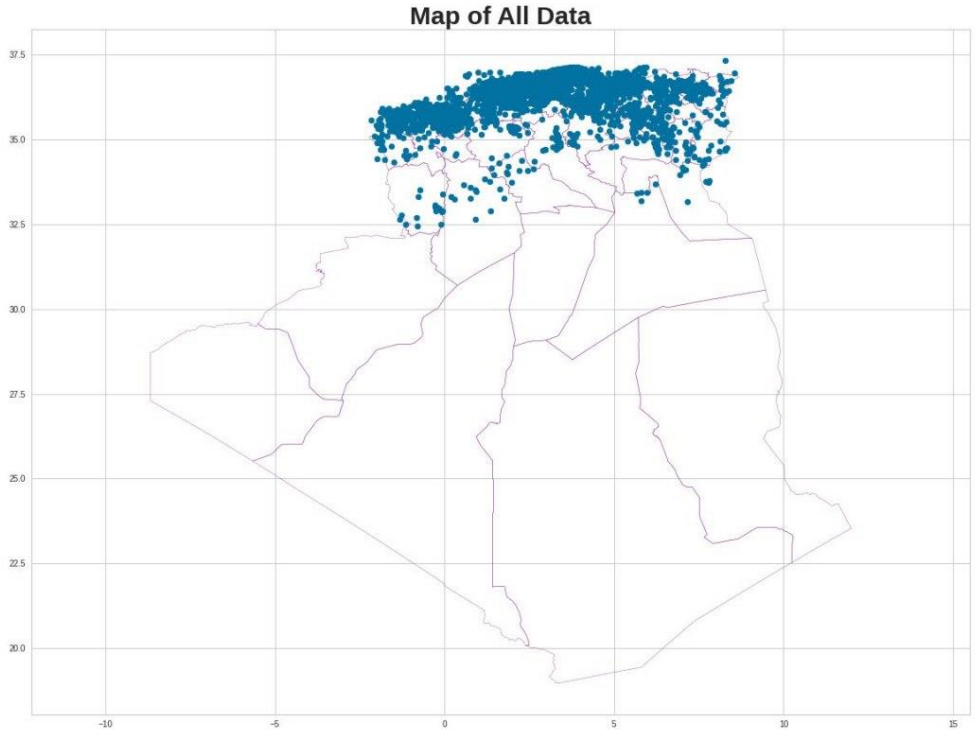


Figure 26: Visualisation de Dataset

III.4.2.1.5. Réappliquer la méthode du Coude

Nous extrayons la meilleure valeur de cluster qui est égale à k=10, voici le résultat dans la figure suivante :

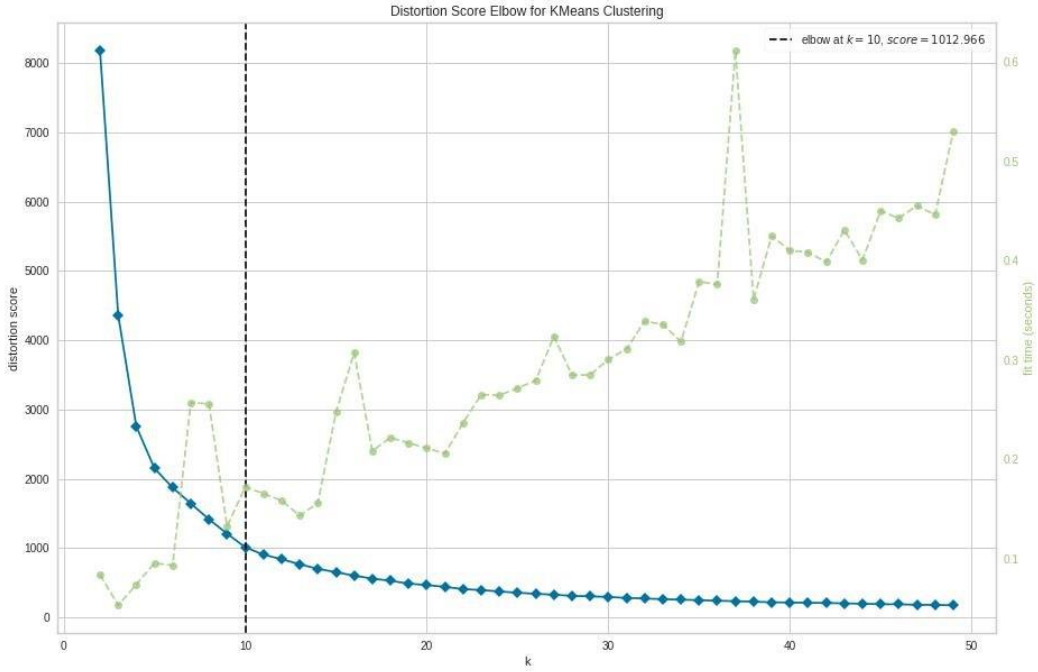


Figure 27: Graphe de méthode du coude

III.4.2.1.6. Réappliquer L'algorithme Kmeans

Après avoir appliqué la méthode du coude et on trouve la meilleure valeur de cluster = (10) puis on applique L'algorithme Kmeans (k=10) et donne le résultat dans la figure suivante :

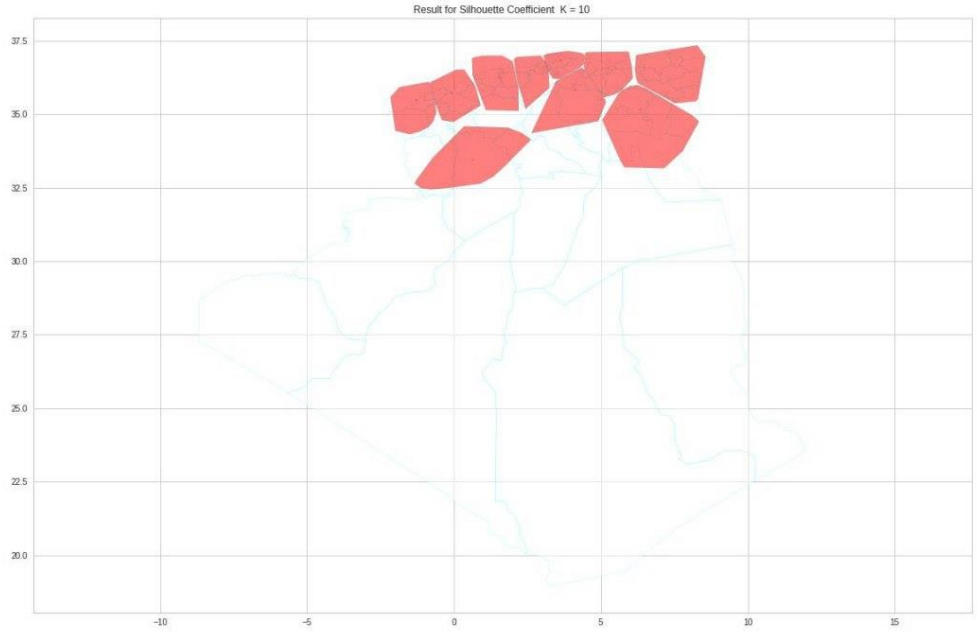


Figure 28: Résultat de Kmeans en utilise la méthode du Coude (k=10)

Après ajustement des données sur le Dataset et application de la méthode du Coude, en voir qu'il y ait une similitude dans le résultat avec les résultats précédents des géologues et une volonté d'atteindre un résultat plus similaire, nous mentionnons une autre méthode "Coefficient de Silhouette ".

III.4.2.2. Coefficient de silhouette (Silhouette Coefficient) :

Le coefficient de silhouette est une mesure de validité des clusters. Sil(i) = (b(i) - a(i)) / max(b(i), a(i)) où a(i) - dissimilarité moyenne entre i et les autres objets du groupe auquel j'appartiens. Autres objets de la grappe à laquelle j'appartiens. D (i, C) - Pour tous les autres clusters C, dissimilarité moyenne de i par rapport à toutes les observations. D (i, C) - Pour tous les autres clusters C, dissimilarité moyenne de i par rapport à toutes les observations de C. b(i) - minimum de D (i, C µ%) - dissimilarité entre I et le cluster voisin C. Si le Sil(i) est plus proche d'un, l'objet est placé dans le bon cluster. Cluster correct. Si le Sil(i) est négatif, l'objet est placé dans le mauvais cluster. S'il est proche de 0, l'objet se trouve entre les clusters **[Pavel Berkhin, 2002].**

Pour appliquer le coefficient de silhouette, nous avons besoin de la théorie de haversine Method pour calculer les distances orthodromiques entre deux paires de coordonnées sur une

sphère, On choisit haversine parce qu'on veut la distance à déterminer sur une Terre 3D. De plus, la latitude et la longitude ne sont pas des unités de distance, mais des angles.

III.4.2.2.1. Distance Matrice de Haversine :

En général, la Haversine est utilisé pour calculer les distances orthodromiques entre deux paires de coordonnées sur une sphère. La forme approximative de la Terre est un sphéroïde aplati. Il n'existe aucune forme géométrique dans laquelle la Terre puisse s'inscrire parfaitement dans le cadre de ses calculs. La formule de Haversine est parfaitement appliquée pour calculer la distance sur une forme sphérique. La distance la plus courte entre deux points à la surface d'une terre ou d'une sphère est mesurée le long de la surface et peut être affichée sur la carte Google. La formule d'Haversine calcule la distance entre le point de localisation principal et le point de destination sur la base de la longueur de la ligne droite en prenant la valeur de la longitude et de la latitude d'entrée. La latitude est représentée par le symbole phi, et elle est définie comme l'angle entre la ligne droite à un endroit donné et le plan équatorial. La longitude est représentée par le symbole lambda. La longitude est définie comme un angle qui pointe à l'ouest ou à l'est du méridien d'origine ou du méridien de Greenwich. Lorsque les valeurs de longitude et de latitude ont été identifiées aux deux endroits, les valeurs décimales sont converties en radians (changez la valeur du nombre entier de latitude et de longitude en radians), puis ces nombres sont calculés dans l'algorithme Haversine. Il y a plusieurs étapes pour calculer la distance en utilisant la méthode Haversine **[Winarno, Et ALL]:**

- Exprimer les coordonnées de l'emplacement en format décimal.
- Exprimer les longitudes et les latitudes.
- Exprimer les coordonnées de l'emplacement en radians.
- Déterminez la latitude initiale et la longitude initiale.
- Calculez la distance totale de l'emplacement optimal à l'aide de la formule Haversine.

La valeur de la longitude et de la latitude est calculée à l'aide de la fonction :

$$a = sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat1)x\cos(lat2)\ x\ sin^2(\frac{\Delta long}{2})$$

III.4.2.3. L'application du Coefficient de Silhouette :

Le résultat du coefficient de silhouette, la valeur la plus élevée indiquée dans la figure suivante, est 0,143479232406, et de cela nous pouvons déduire la meilleure valeur pour k = 15 :

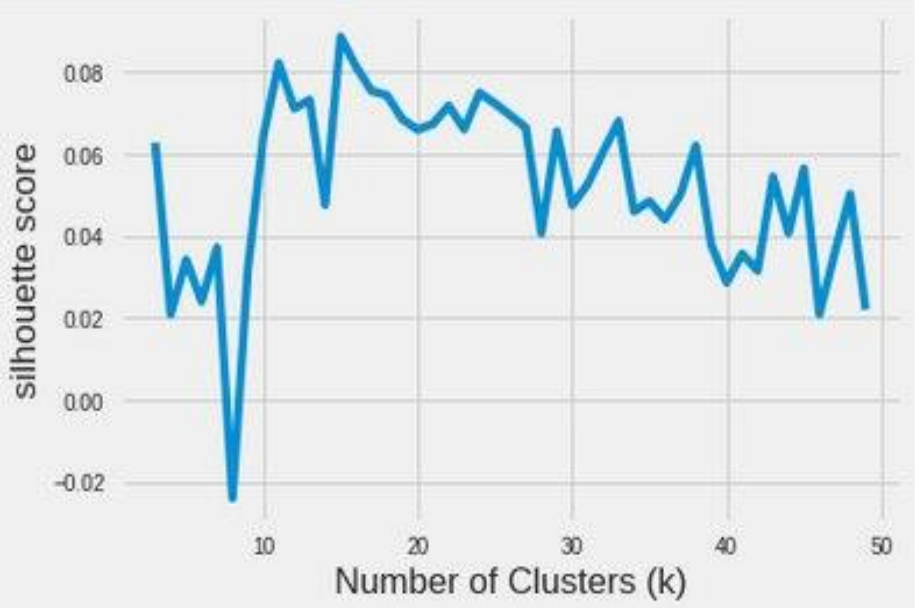


Figure 29: Graphe de Coefficient de Silhouette

Cette figure est une analyse de silhouette pour le regroupement de kmeans avec le max valeur de Silhouette score de graphe précédent n_clusters k = 15 et silhouette score = 0,143479232406 :

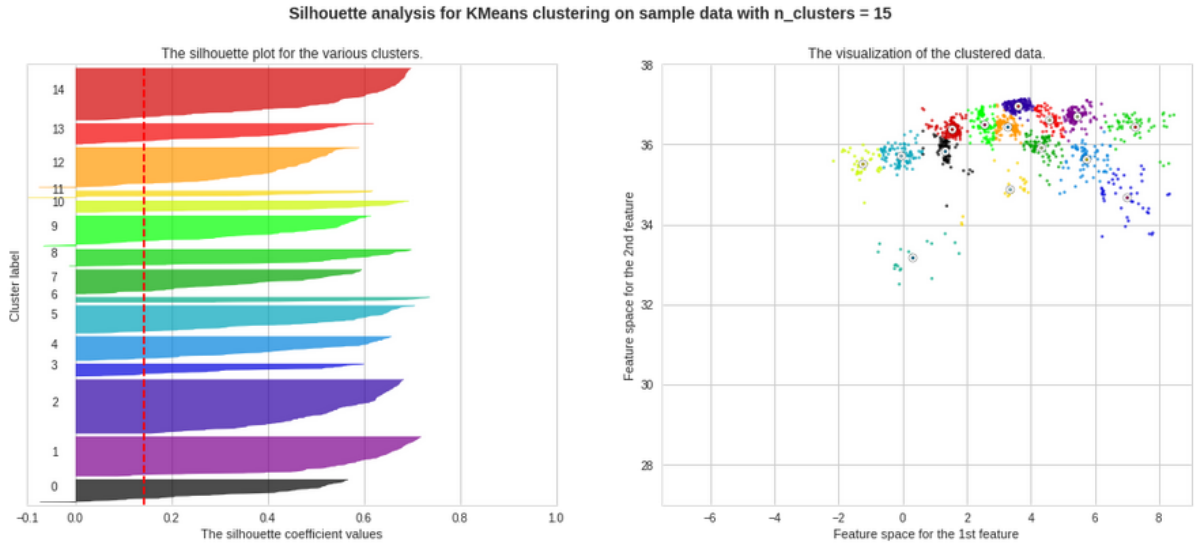


Figure 30: Analyse de silhouette pour le regroupement de kmeans avec n_clusters =15

III.4.2.4. L'application de l'algorithme Kmeans :

Nous appliquons L'algorithme Kmeans qui basé sur le résultat du Coefficient Silhouette ou la meilleure valeur de clusters (k=15) :

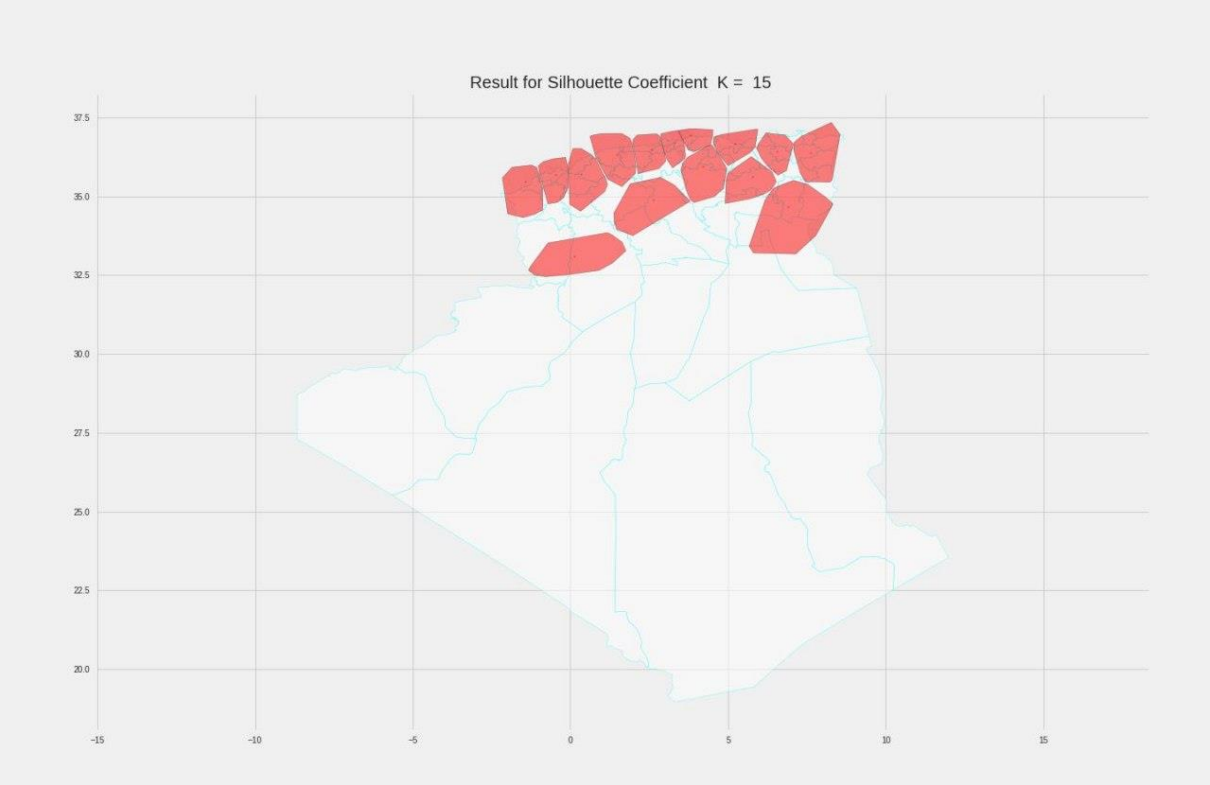


Figure 31: Résultat de Kmeans (k=15) en coefficient du silhouette

III.4.3. DBSCAN :

DBSCAN est un algorithme de clustering basé sur la densité, c'est-à-dire qu'il effectue le clustering en fonction de la concentration des données. D'où son nom: "Density-Based Spatial Clustering of Applications with Noise". Par conséquent, son objectif est de regrouper les données qui sont proches les unes des autres.

La caractéristique la plus puissante de cet algorithme est qu'il n'est pas limité aux formes circulaires comme K-Means puisqu'il peut générer des clusters de différentes formes et tailles. En outre, il détecte les valeurs aberrantes (données atypiques) en les classant comme du bruit, de sorte qu'il est moins susceptible d'être affecté par ces données.

En utilise algorithme DBSCAN avec différentes mesures de distance comme Euclidienne, Hausdorff et Haversine mes dans cela on choisit Haversine parce qu'on veut la distance à déterminer sur une Terre 3D. De plus, la latitude et la longitude ne sont pas des unités de distance, mais des angles.

Chapitre III : Implémentation et évaluation

Avant d'appliquer l'algorithme DBSCAN, nous devons connaître quelques définitions :

- Densité : Nombre de points dans un rayon spécifique appelé epsilon (eps).
- MinPts : Abréviation de "MinPoints", qui se traduit par "Points minimums". Il s'agit de la densité minimale pour qu'un point soit considéré comme un Point Central. Il correspond donc à une valeur -donnée par l'analyste- pour définir un seuil.
- Point central (Core Point) : Si un point à une densité supérieure ou égale à MinPts, c'est-à-dire si dans la sphère autour d'un point, il y a plus de points que le seuil défini par MinPts, il est appelé Point central ou Point Central. Par exemple, si nous définissons MinPts = 4, et qu'à l'intérieur d'une sphère autour du point, il y a 5 points, on l'appelle le Point Central.
- Point Frontière (Border Point) : Ce point qui, dans sa sphère définie par epsilon, ne détient pas le minimum de points MinPts, mais ce point est dans la sphère d'un point central. En d'autres termes, ce point n'est pas un point central, puisqu'il ne remplit pas la condition selon laquelle la densité est supérieure aux MinPts, mais il se trouve dans la sphère d'un point central.
- Point de Bruit (Noise Point) : Tout point qui n'est ni un point central ni un point Frontière. Correspond aux points les plus éloignés des données.

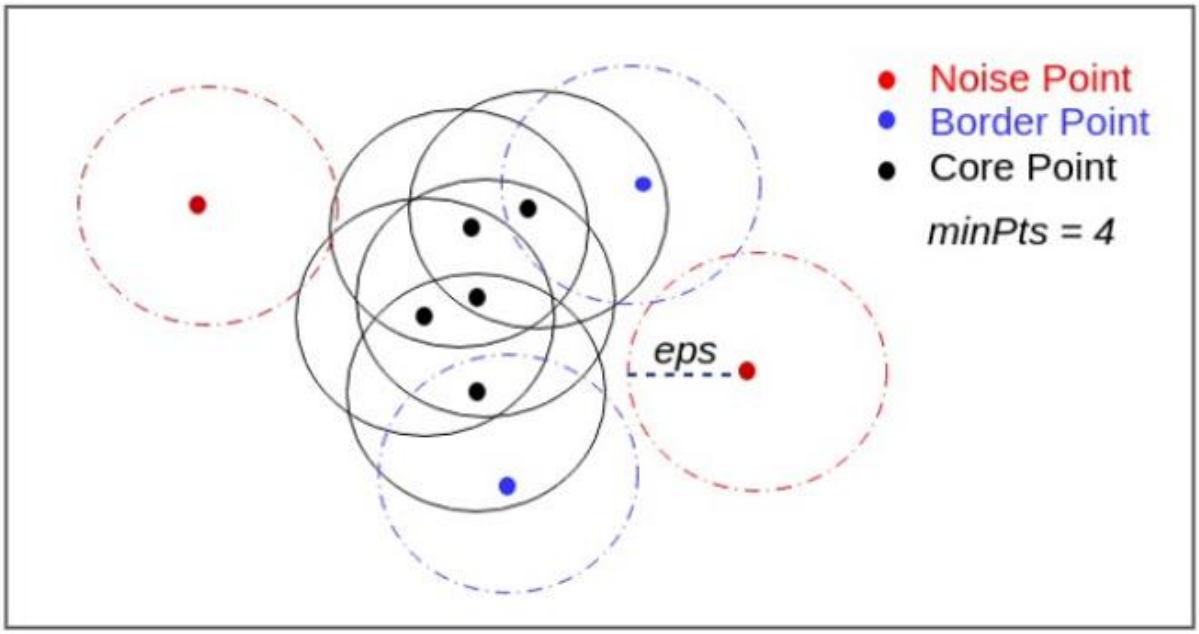


Figure 32: Cluster DBSCAN unique [7]

III.4.3.1. Comment fonctionne DBSCAN :

Pour exécuter DBSCAN nous devons donner deux paramètres à l'algorithme : le Eps (Epsilon) et le MinPts. Nous allons donc décider du rayon de la sphère autour des points. Et le nombre minimum de points requis -dans cette sphère- pour être considéré comme point central.

III.4.3.1.1. Sélection d'Epsilon

Pourquoi nous intéressés par la définition d'Epsilon ? si nous définissons un Epsilon très petit, les points seront seulement classés comme bruit -car ils n'ont pas assez d'espace pour que la condition de densité supérieure ou égale à MinPts soit remplie-. Dans le cas contraire, si nous définissons un Epsilon très grand, toutes les données seront regroupées en un seul grand cluster, ce qui ne nous permettra pas d'obtenir des informations intéressantes.

Comment le définir ? Pour définir Epsilon, nous devons d'abord comprendre le concept de

" Distance to its K-nearest neighbor".

[7] https://www.mdpi.com/sustainability/sustainability-11-04718/article_deploy/html/images/sustainability-11-04718-g002-550.jpg

- **Distance à son voisin K le plus proche :**

Pour obtenir le voisin K le plus proche nous devons :

- Choisissez n'importe quel point dans un Dataset.
- Calculez la distance de ce point à tous les autres.
- Triez les données en fonction de la distance au point sélectionné, c'est-à-dire que le premier point est le plus proche et le dernier le plus éloigné.
- Le voisin K le plus proche est le point qui se trouve à la position "K" dans la liste. Par exemple, si je veux connaître le 5e voisin le plus proche, c'est celui qui se trouve en 5e position dans la liste ordonnée par proximité.

Donc, le voisin K le plus proche est la distance entre le point sélectionné et le point qui utilise le rang "K" dans la liste des points triés par proximité.

III.4.3.1.2. Sélection de MinPts :

Il y a 3 façons principales de déterminer les MinPts :

La première consiste à ajouter un au nombre de dimensions de dataset. Par exemple, si les dimensions de nos données sont la température moyenne, la météo, l'heure et le secteur géographique, comme nous avons 4 dimensions, nous pourrions utiliser MinPts = 5.

La deuxième méthode consiste à prendre un pourcentage arbitraire des données. Par exemple, si notre échantillon est de 1000 points, nous pourrions dire que notre MinPts est égal à 1% des données, en d'autres termes, MinPts = 10.

Enfin, nous pourrions décider d'un nombre arbitraire basé sur la littérature, une analyse exploratoire des données ou notre intuition.

Parmi c'est 3 façons en utilise La deuxième méthode qui consiste à prendre un pourcentage arbitraire des données, en a 4043 points, nous pourrions dire que notre MinPts est égal à 1% des données, en d'autres termes, MinPts = 40.

III.4.3.2. Détermination des paramètres et résultats du clustering :

Les paramètres Eps et MinPts sont requis pour l'algorithme DBSCAN. Pour déterminer la valeur optimale du paramètre Eps utilisé la méthode k-NN Distance (k-nearest neighbor distance). L'essence de la méthode consiste à calculer les distances dans la matrice de points au voisin K le plus proche. La distance moyenne de chaque point aux voisin K le plus proche est calculé. La valeur k est définie en fonction des valeurs MinPts. Les données de k-distance est affiché dans l'ordre croissant. La valeur correspondant au seuil où se produit un changement brusque de la courbe de distance k est sélectionnée. Comme on peut le voir sur la Figure suivant, la valeur optimale pour Eps égale à 0.002.

III.4.3.3. Résultats de la Clustering :

"Physiquement, les provinces sismiques ne devraient pas avoir beaucoup de changements à différentes périodes de temps. Cela implique que le nombre de clusters aurait dû être presque le même dans toutes les périodes de temps ». Basé sur le paramètre Eps trouvé et à la suite d'un nombre d'expériences, les valeurs MinPts optimales ont été sélectionnées, les résultats sont présentés dans la Figure suivant :

Eps	MinPts	Number of Clusters	Noise Ratio
0.002	32	10	29.65
	33	8	32.20
	34	**9**	**32.59**
	35	**9**	**33.16**
	36	**9**	**33.63**
	37	**9**	**34.05**
	38	**9**	**35.79**
	39	**9**	**36.58**
	40	**9**	**37.79**
	41	**9**	**37.89**
	42	10	39.47
	43	10	39.82
	44	10	39.97
	45	9	41.52

Figure 33: Sélection de MinPts

Selon les données présentées dans la Figure 33, le nombre de clusters varie différemment en ayant la même valeur **Eps** et une augmentation graduelle des valeurs **MinPts**, quant aux données identifiées comme **Rapport de bruit**, le bruit augmente progressivement. Réduire le volume de données **Rapport de bruit** bruyantes conduit à une diminution du nombre de clusters et à une diminution de la précision des résultats obtenus. En conséquence, les valeurs de **MinPts** ont été sélectionnées comme les valeurs optimales où le nombre de clusters à une valeur constante.

Les zones sismiques avec une densité spatiale accrue avec des valeurs optimales sont montrées dans la Figure suivant :

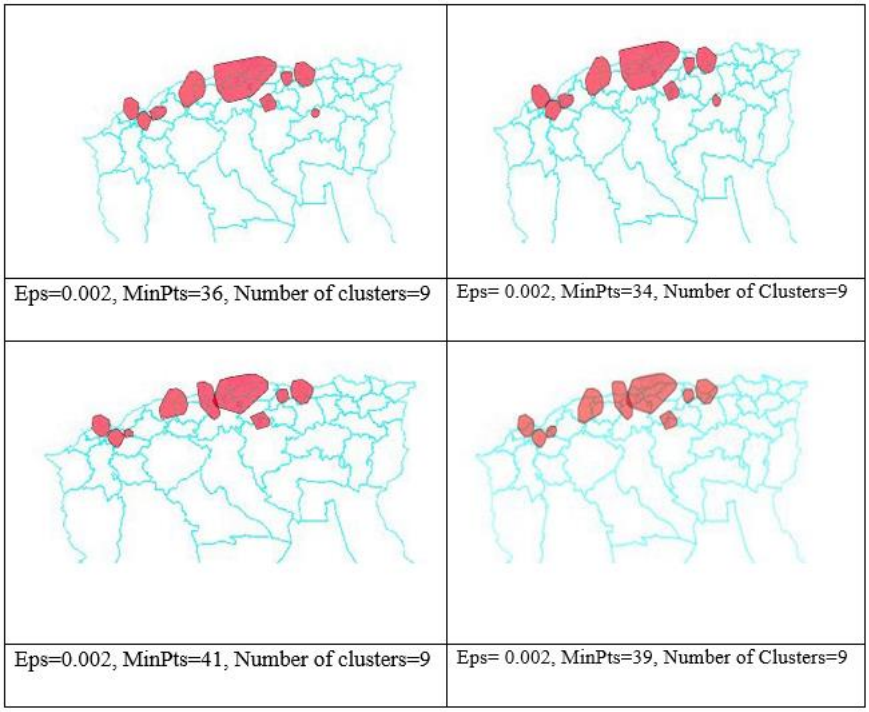


Figure 34: Résultats de l'algorithme DBSCAN à partir des paramètres optimaux

Donc voici le résultat final optimal de l'algorithme DBSCAN finale dans la figure suivante:

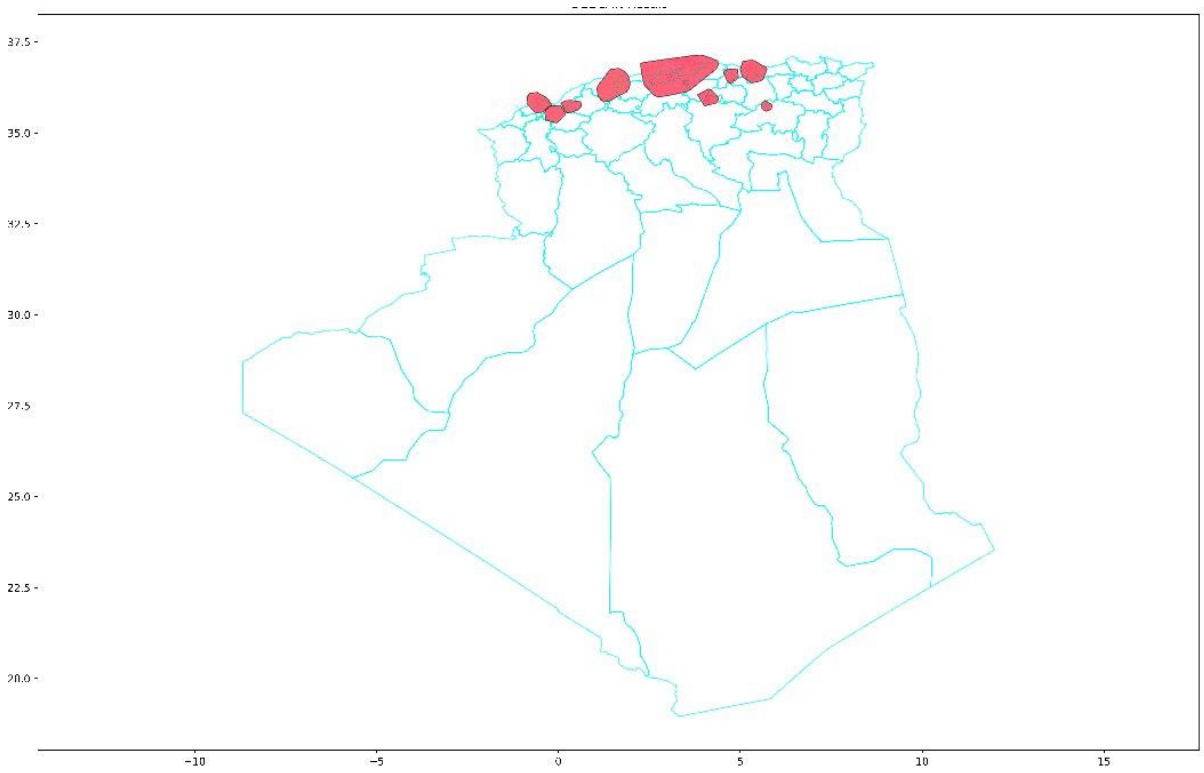


Figure 35: Résultat final de l'algorithme DBSCAN.

III.4.4. OPTICS :

OPTICS Clustering signify "Ordering Points to Identify Cluster Structure". Il s'inspire de l'algorithme de clustering DBSCAN. Il ajoute deux termes supplémentaires aux concepts de clustering DBSCAN. Il s'agit de :

- **Distance centrale** : Il s'agit de la valeur minimale du rayon requise pour classer un point donné comme point central. Si le point donné n'est pas un point central, alors sa distance centrale est indéfinie.

- **Distance d'accessibilité** : Elle est définie par rapport à un autre point de données q(Let). La distance d'accessibilité entre un point p et q est le maximum de la distance centrale de p et de la distance Haversine entre p et q.

Notez que la distance d'accessibilité n'est pas définie si q n'est pas un point central.

Cette technique de clustering est différente des autres techniques de clustering dans le sens où cette technique ne segmente pas explicitement les données en clusters. Au lieu de cela, elle produit une visualisation des distances d'Accessibilité et utilise cette visualisation pour regrouper les données.[8]

Résultats de la Clustering

Physiquement, les provinces sismiques ne devraient pas avoir beaucoup de changements à différentes périodes de temps. Cela implique que le nombre de clusters aurait dû être presque le même dans toutes les périodes de temps Basé sur d'un nombre d'expériences les résultats sont présentés dans la Figure suivant :

[8] https://fr.acervolima.com/

MinPts	Numbers of Clusters	Noise Ratio
32	10	30.29
33	8	32.37
34	**9**	**32.79**
35	**9**	**33.44**
36	**9**	**34.30**
37	**9**	**34.72**
38	**9**	**36.26**
39	**9**	**36.97**
40	**9**	**38.18**
41	**9**	**38.26**
42	10	39.54
43	10	39.97
44	10	40.11
45	9	41.65

Figure 36: Sélection des paramètres MinPts pour OPTICS

Selon les données présentées dans la Figure 36, le nombre de clusters varie différemment en ayant la même valeur Eps et une augmentation graduelle des valeurs MinPts, quant aux données identifiées comme Rapport de bruit, le bruit augmente progressivement. Réduire le volume de données Rapport de bruit bruyantes conduit à une diminution du nombre de clusters et à une diminution de la précision des résultats obtenus. En conséquence, les valeurs de MinPts ont été sélectionnées comme les valeurs optimales où le nombre de clusters à une valeur constante.

Les zones sismiques avec une densité spatiale accrue avec des valeurs optimales sont montrées dans la Figure suivant :

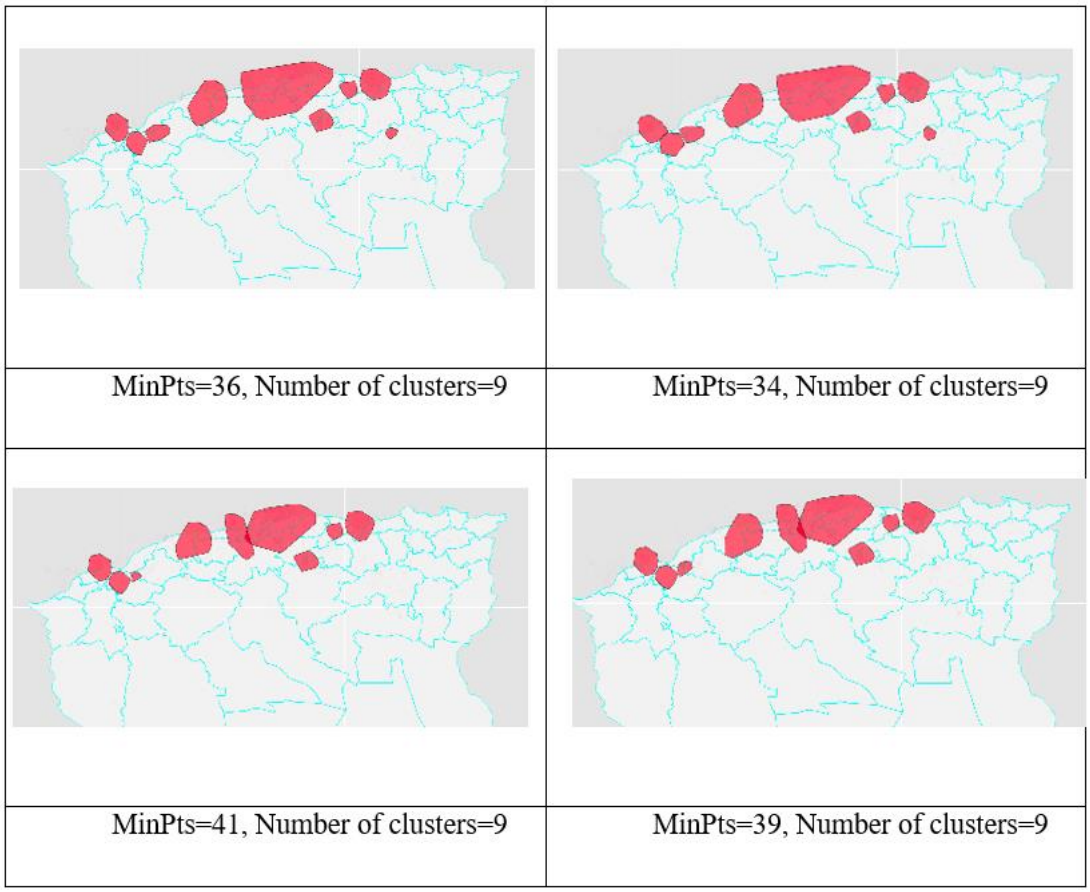


Figure 37: Résultats de l'algorithme OPTICS à partir des paramètres optimaux

Donc voici le résultat final optimal de l'algorithme OPTIC finale dans la figure suivante :

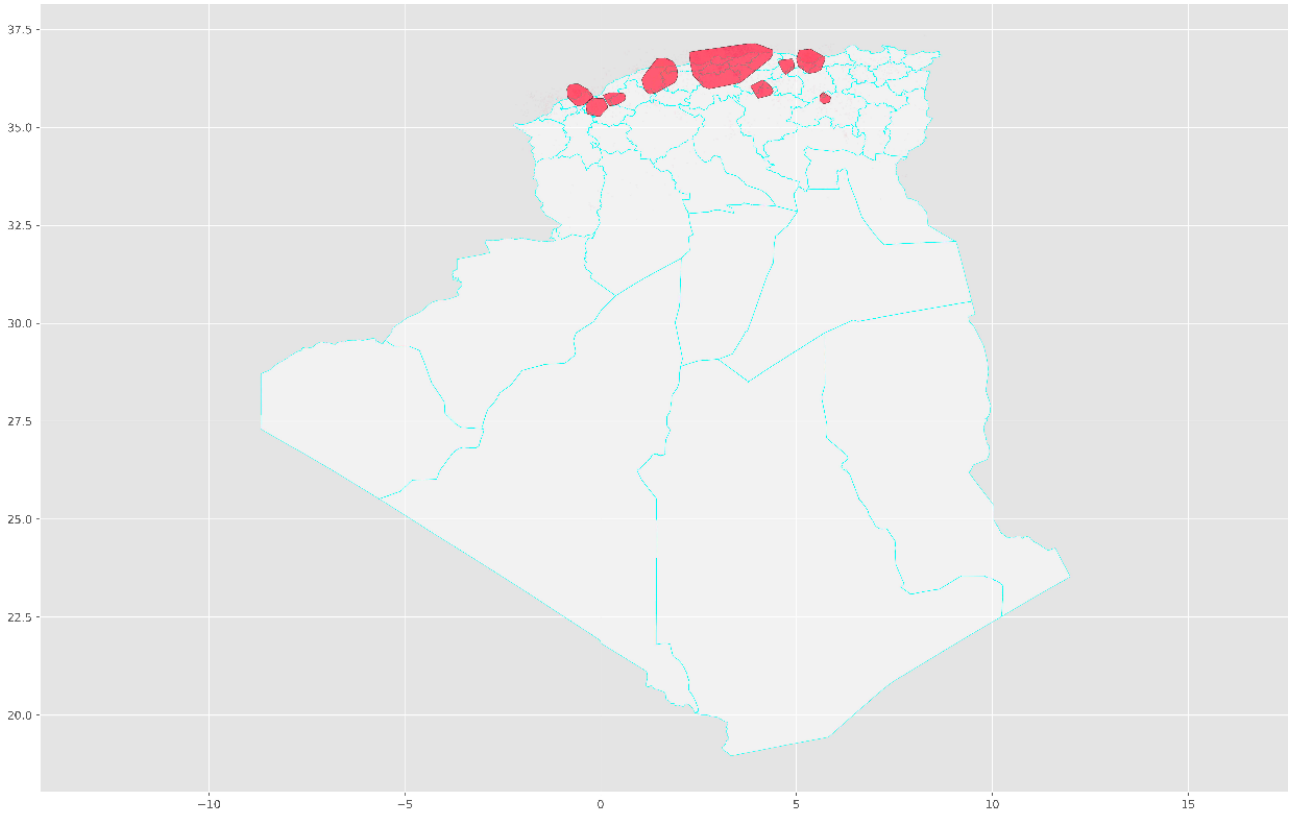


Figure 38: Résultat final de l'algorithme OPTIC

III.4.5. Comparaisons :

Dans le domaine de la science sismique, il n'existe pas de travaux précèdent sur la détermination des zones sismiques en Algérie à l'aide des algorithmes l'apprentissage automatique, Nous avons donc dû comparer notre travail avec les résultats obtenus dans chacun des algorithmes sur lesquels nous avons travaillé précédemment, puis le comparer avec le travail des géologues en Algérie dans ce domaine.

Les résultats des algorithmes de clustering que nous avons choisis pour cette comparaison est présentée dans la figure suivante :

	K-means Elbow (10 Clusters)	K-means Silhouette (15 Clusters)	DBSCAN (9 Clusters)	OPTICS (9 Clusters)
Largest to smallest cluster number	958(2) 629(1) 522(4) 477(8) 464(9) 293(5) 232(3) 210(7) 211(0) 47(6)	583(1) 534(4) 413(6) 397(13) 370(10) 285(9) 266(8) 256(3) 221(2) 203(5) 195(0) 166(14) 63(12) 61(11) 30(7)	1127(0) 411(4) 240(3) 227(6) 156(5) 137(2) 107(1) 72(7) 38(8)	1127(0) 411(4) 240(3) 225(6) 156(5) 136(2) 107(1) 70(7) 27(8)
Sum	4043	4043	2515	2499

Figure 39: Comparaison des diffèrent algorithmes

Maintenant, si nous regardons les zones les plus sujettes aux séismes qui se sont regroupés en utilisant différentes méthodes de regroupement, il est visible qu'il y a une certaine différence au niveau du nombre de données points dans chaque groupe le plus important, ce qui peut dérouter le public appréhender les zones fortement sismiques.

C'est pourquoi nous avons effectué quelques comparaisons pour trouver un algorithme de regroupement optimal pour les données sismiques.

De la comparaison, nous remarquons :

- **Kmeans du Coude** : Ses résultats est déferont à des cartes des géologues.
- **Kmeans Silhouette** : Meilleurs que les résultats précédents, mais pas similaires aux cartes des géologues.
- **DBSCAN :** Nous avons obtenu des résultats proches de ceux des géologues, puisqu'ils ont montré toutes les zones sismiques connues en Algérie, comme la région de Chlef-Boumerdes... ainsi que quelques nouvelles zones sismiques à l'est et à l'ouest de l'Algérie, et ils séparant la grande zone sismique en petites zones où il est facile d'étudier.

- **OPTICS** : Les résultats sont proches des résultats de l'algorithme DBSCAN Avec NOISE légèrement plus élevé.

D'après les observations, nous pouvons dire que **DBSCAN** et **OPTICS** sont les meilleurs que **Kmeans du Coude** et **kmeans Silhouette**.

III.5. Les outlies de travail :

III.5.1. Pandas :

Est un outil d'analyse et de manipulation de données open source rapide, puissant, flexible et facile à utiliser, construit à partir du langage de programmation Python[9].

III.5.2. Numpy :

Le paquetage fondamental pour le calcul scientifique avec Python[10].

III.5.3. Matplotlib :

Matplotlib est une bibliothèque complète permettant de créer des visualisations statiques, animées et interactives en Python. Matplotlib rend les choses faciles et les choses difficiles possibles[11].

III.5.4. Geopandas :

GeoPandas est un projet open source visant à faciliter le travail avec des données géospatiales en python[12].

[9] https://pandas.pydata.org/
[10] https://numpy.org/
[11] https://matplotlib.org/
[12] https://geopandas.org

III.5.5. Scipy :

Algorithmes fondamentaux pour le calcul scientifique en Python[13].

III.5.6. Yellowbrick :

Yellowbrick est une suite d'outils de visualisation qui étendent l'API de scikit-learn afin d'accélérer et de faciliter la sélection de modèles pour les scientifiques des données. En d'autres termes, ce paquet Python combine la puissance de scikit-learn avec les capacités de matplotlib pour générer des visualisations intuitives de vos modèles[14].

III.5.7. Sklearn :

Scikit-learn est probablement la bibliothèque la plus utile pour l'apprentissage automatique en Python. La bibliothèque sklearn contient de nombreux outils efficaces pour l'apprentissage automatique et la modélisation statistique, notamment la classification, la régression, le regroupement et la réduction de la dimensionnalité[15].

III.5.8. Kneed :

Tente d'identifier le point de genou/coude d'une ligne ajustée aux données. Le genou/coude est défini comme le point de la ligne avec la courbure maximale. Pour plus d'informations sur la manière dont chacun des paramètres affecte l'identification des points de genou, consultez la rubrique Paramètre[16].

III.5.9. Anaconda :

Conda est un système de gestion des paquets et des environnements à code source ouvert qui fonctionne sous Windows, macOS et Linux. Conda installe, exécute et met à jour rapidement les paquets et leurs dépendances. Il permet également de créer, sauvegarder, charger et basculer

[13] https://scipy.org/
[14] https://www.scikit-yb.org/en/latest/
[15] https://scikit-learn.org/
[16] https://pypi.org/project/kneed/

facilement entre les environnements de votre ordinateur local. Il a été créé pour les programmes Python, mais il peut empaqueter et distribuer des logiciels pour n'importe quel langage[17].

III.5.10. Jupyter :

Jupyter est le dernier environnement de développement interactif basé sur le Web pour les carnets de notes, le code et les données. Son interface flexible permet aux utilisateurs de configurer et d'organiser des flux de travail dans les domaines de la science des données, du calcul scientifique, du journalisme informatique et de l'apprentissage automatique. Une conception modulaire invite les extensions pour étendre et enrichir les fonctionnalités[18].

[17] https://www.anaconda.com/
[18] https://jupyter.org/

IV Conclusion Générale et perspectives:

Dans ce mémoire nous avons présenté des outilles pour extraction des zones séismique en Algérie a l'aide des algorithmes d'apprentissage automatique non supervise.

Nous avons étalé en détail sure les algorithmes d'apprentissage automatique, nous avons effectuée appliquer trois algorithmes (K-means, DBSCAN, OPTICS) en utilisant un dataset qui ont été choisis, tous ces algorithme donné un résultat de extraction des zones séismique en Algérie , Nous avons présenté une comparaison des résultats obtenus à partir de ces algorithmes entre eux et avec les résultats précédents des géologues Et nous avons dit que la détermination des zones sismiques en utilisant l'apprentissage automatique a de bons résultats.

Comme perspectives nous pouvons améliorer les résultats que nous avons obtenus en utilisant un dataset plus grand qui contient plus d'attributs, aussi Ces résultats peuvent également être améliorés au fil du temps en enregistrant les informations de ces nouveaux séismes.

Bibliographie

[J.L. Laurière, 1986] : Intelligence artificielle, résolution de problèmes par l'homme et la machine, Eyrolles 1986

[Bensoussan & Bensoussan, 2019] : J. (2019). IA, robots et droit. Bruylant.

6Jason Brownlee. (March 16, 2016). Supervised and unsupervised algorithms.

Récupéré de

https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[Roy, 2018] : Roy, Jean-François. (2018). Thèse : Apprentissage automatique avec garanties de généralisation à l'aide de méthodes d'ensemble maximisant le désaccord, Université Laval, Québec, Canada.

[Larochelle, 2008] : Larochelle, Hugo. (2008). Thèse : Étude de techniques d'apprentissage non-supervisé pour l'amélioration de l'entraînement supervisé de modèles connexionnistes, Faculté des études supérieures et postdoctorales, Université de Montréal, Canada.

[Ben Ayed, 2015] : Ben Ayed, Salma Jamoussi. (23-25 mars 2015). Conférence : l'apprentissage profond pour le TALN, sousse, Tunisie.

[Garivier, 2013] : Garivier, Aurélien. (12 et 13 juin 2013). Support de cour : Apprentissage supervisé.

[Jason Brownlee,2016]: Jason Brownlee. (March 16, 2016). Supervised and unsupervised algorithms.

Récupéré de

https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[Managemement& data science, 2019] : 7Managemement & Data Science. (18 janvier 2019). #7 – Évaluer un modèle de prédiction

Récupérer de

Bibliographie

https://management-datascience.org/2019/02/18/evaluation-dun-modele-de-prediction/.

[Barlow, 1989]: Barlow, H. (1989). Finding Minimum Entropy Codes, Physiological Laboratory,Downing Strcct, Cambridge, CB2 3EG, England.

[Bankman et al, 2019] : Bankman, Daniel. Et Moons, Bert. Verhelst, Marian. (2019). Embedded deep learning Algorithms, Architectures and Circuits for Always-on Neural Network Processing, Springer Nature Switzerland AG 2019, Library of Congress Control Number: 2018956280.

[Boukhames et zaidi, 2016] : Boukhames, Oussama. Et Zaidi, Mohamed. (2016). Mémoire de Master : Estimation Non Paramétrique De La Densité De Probabilité Et De La Fonction De Régression, Université M'Hamed Bougara, Boumerdes, Algérie.

[El ferchichi, 2013] : El Ferchichi Sabra. (2013). Thèse *:* la sélection et l'extraction d'attributs pour les problèmes de classification, université de Tunis El Manar, Tunisie.

[Gagné, 2016] : Gagné, Christian. (2016). Support de cour : Réduction de la dimensionnalité, Apprentissage et reconnaissance - GIF-4101 / GIF-7005, Université Laval, Quebec, Canada.

[Shukla, 2017]: Shukla, Nishant. (2017). Livre: Machine learning with TensorFlow, Manning Early Access Program (MEAP) Edition, Version 10, Manning publication.

[Chaarana, 2018] : Chaarana, Anis. (2018). Diplôme national d'ingénieur : Prototype de conduit autonome d'un véhicule, département de génie électrique, école national supérieure d'ingénieur de Tunis ENSIT, Université de Tunis, Tunis.

[Zimmer, 2018] : Libo, Zhang. Tiejian, Luo. Fei, Zhang. Et Yanjun, Wui. A Recommendation Model Based on Deep Neural Network, Chinese Academy of Sciences, Digital Object Identifier 10.1109/ACCESS.2018.2789866, Received November 29, 2017, accepted December 26, 2017, date of publication January 5, 2018, date of current version March 13, 2018, the National Education Information Technology Research under Grant 171140001, Beijing, China.

Bibliographie

[Bertail, et al, 2019] : Bertail, Patrice. Bounie, David. Clémençon, Stephan, et Patrick Waelbroeck. (14 Février2019). Algorithmes : biais, discrimination et équité, Télécom ParisTech, la fondation ABEONA, Université Paris Nanterre, Paris, France.

[JOSEPH BERTRAND,2017]: 6Jason Brownlee. (March 16, 2016). Supervised and unsupervised algorithms.

Récupéré de

https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

7Managemement & Data Science. (18 janvier 2019). #7 – Évaluer un modèle de prédiction Récupérer de

https://management-datascience.org/2019/02/18/evaluation-dun-modele-de-prediction/

[Celeux, 1989] : Celeux G, Diday E, Govaert G, Lechevallier Y, Ralam-Bondrainy H. Classification Automatique des Données. Bordas, Paris, 1989.

[Guellil, 2009] : Guellil, Z., & Zaoui, L. (2009). Proposition d'une Solution au Problème d'Initialisation Cas du K-means. In *CIIA*.

[Ester et al, 1996]: Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).

[Khan, K. et all, 2014]: Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014, February). DBSCAN: Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) (pp. 232-238). IEEE.

[Verma, 2012]: Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, N. (2012). A comparative study of various clustering algorithms in data mining. International Journal of Engineering Research and Applications (IJERA), 2(3), 1379-1384

[Benabdi et Bourssia, 2016]: Bankman, Daniel. Et Moons, Bert. Verhelst, Marian. (2019). Embedded deep learning Algorithms, Architectures and Circuits for Always-on Neural Network Processing, Springer Nature Switzerland AG 2019, Library of Congress Control Number: 2018956280.

[Lefebvre-Brossard, 2018] : Lefebvre-Brossard, Antoine. (2018). Mémoire de Master : Sur l'utilisation de réseaux de neurones dans un système de recommandations réciproques, École polytechnique de Montréal, Canada.

Bibliographie

[Djeriri, 2017] : Djeriri, Youcef. Les Réseaux de Neurones Artificiels, UDL-SBA-2017, Université,Liabes Djilali, Sidi-Bel Abbas, Algérie.

[Eve MathieuDupas, 2010] : Eve, Mathieu-Dupas. (2010). Algorithme des k plus proches voisins pondérés etapplication en diagnostic. 42èmes Journées de Statistique, Marseille, France.

[Martin, 2019] : Martin Skogholt. (2019-04-16). Fast naive Bayes.
Récupéré de
https://cran.rproject.org/web/packages/fastNaiveBayes/vignettes/fastnaivebayes.html

[POUR, MDEFIND, et al, 2003] : Documentation de base pour la vérification des ouvrages d'accumulation aux séismes. Rapports de l'OFEG.

[Andriamampandry, A, 2014] : analyse de la complémentarité des stations sismiques

[El Kadi, 1993] : Le tremblement de terre en Égypte. Analyse de presse. *Égypte/Monde arabe*, (14), 163-196.

[POUR, MDEFIND, et al] : Documentation de base pour la vérification des ouvrages d'accumulation aux séismes. Rapports de l'OFEG

[Schettkat, R. 2010] : Schettkat, R. (2010). Faut-il un séisme pour réveiller le monde de l'économie ? Revue internationale du Travail, 149(2), 201-227.

[Richard, N. 2001] : Approche multi-échelles pour la modélisation de structures en bois sous sollicitations sismiques (Doctoral dissertation, Cachan, Ecole normale supérieure).

[BOHADAD, Y, & BELHAIDm2003]. 2003.la Technique Active : Exemple de L'Algérie du Nord. In « Risque, Vulnérabilité et fiabilité dans la construction, vers une réduction des désastres », Alger, Algérie, 11et12octobre2003.

[ABDELHADI BOUSAFFA, RAOF BEN SLIMANE, 1997] : Contribution à l'étude Hydrologique de La Région de Batna (Tazolt-Ain Touta), Mémoire d'ingénieur, Université de Constantine.

[Belkhiri, Kamel, 2006] : Contrôle Rapide des Structures en Relation avec le Risque et Evaluation des Dommages Causés par une Catastrophe Naturelle : cas de séisme. Diss. Université de Batna 2, 2006.

Bibliographie

[Menasri, Y. 2009]. Thème : Évaluation de Vulnérabilité Sismique des Bâtis Existants « Structure portique en béton armé ».