

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU.



FACULTE GENIE ELECTRIQUE ET D'INFORMATIQUE

DEPARTEMENT INFORMATIQUE

MINI-PROJET DE REDACTION SCIENTIFIQUE

SPECIALITE : SYSTEME INFORMATIQUE

Réalisée par
M^{lle} HADJ-ALI YASMINE

Sujet :
L'apprentissage automatique

Année universitaire : 2020/2021

Remerciement

Je tiens à remercier Madame GOUMEZIANE Lynda, Enseignante Chercheur à l'université Mouloud Mammeri de nous avoir dirigé au mieux pour réaliser ce travail. Son soutien, sa clairvoyance et ses compétences m'ont été d'une très grande aide

J'adresse mes remerciements aussi, à toute personne qui m'a aidé de loin ou de près dans la réalisation de ce mini-projet.

Dédicace

Je dédie ce mini-projet avec considération et respect à tous ceux que j'aime et j'apprécie et spécialement à mes très chers parents qui m'ont toujours soutenue.

Table des matières

Figures

Tableaux

Introduction générale	1
1 introduction à l'apprentissage automatique	3
1.1 Historique de l'apprentissage automatique	3
1.2 Définition : (source : [2])	4
1.3 Pourquoi utiliser l'apprentissage automatique ?	4
1.4 Problématique :	5
1.5 Propositions et solutions :	5
1.6 Conclusion :	5
2 Type de problème de l'Apprentissage Automatique	6
2.1 Apprentissage supervisé	6
2.1.1 Définition (source : [6])	6
2.2 Classification	7
2.3 autres méthodes d'apprentissage :	8
2.3.1 réseaux de neurones :	8
2.4 De l'apprentissage automatique vers l'intelligence artificiel (source : [3])	11
2.5 Conclusion	12
Conclusion générale	13
Bibliographie	14

Table des figures

1.1	du cerveau humain a la machine	4
2.1	Apprentissage Supervisé	6
2.2	Un réseau de neurones à trois couches successives	10
2.3	De gauche à droite : une capacité trop faible, correcte et trop élevée pour la tâche.	10
2.4	L'évolution des statisticiens vers les ingénieurs informatique et algorithmes	11

Liste des tableaux

2.1 Tableau d’algorithmes de classification 8

Introduction générale

« Seuls ceux qui ont vécu dans une caverne ces dix dernières années ont pu ignorer L'incroyable révolution de l'apprentissage automatique » [4]

L'apprentissage automatique est un champs d'étude qui est devenu aussi vite populaire. Depuis, le mode a connu de grandes évolutions en terme d'avancées technologique, de centaines d'applications intelligentes ont vu le jour, tels que les assistants vocaux intelligents, le moyen de s'authentifier par empreinte digitale ou par reconnaissance faciale ou encore La traduction automatique.

Tout au début de la création des applications intelligentes, les développeurs ont cherché à les perfectionner et cela en récoltant d'abord un grand nombre de données puis en construire des jeux de données avec lesquelles ils pourront tester et entraîner l'application afin qu'elle puisse en tirer le vrai du faux et appliquer le concept pour toutes autres donnée inconnue, de cette façon l'application apprend ce qu'elle à faire à force de tester un grand nombre d'informations et c'est cette étude qu'on appelle l'apprentissage automatique.

[4]Que l'on ne soit qu'un simple utilisateur d'internet pour quelques minutes n'empêche qu'une horde de systèmes d'apprentissage automatique s'activent faisant des sites qu'on visite l'un des moyens qui permette d'analyser notre personnalité pour nous proposer des produits idéals, attirer notre attention ou encore analyser nos comportements pour s'assurer que l'on est pas un fraudeur.

L'apprentissage automatique apprend grâce à nous, elle utilise nos informations que les collecteur de donnée ou les data setter récoltent, et cela à travers nos accès aux sites internet mais le nombre de ces données doit être très grand, afin de permettre à la machine d'étudier plusieurs cas et parfois même distinguer des anomalies, comme par exemple traiter des données concernant le type de sang des êtres humain et en tirer une nouvelle sorte qu'elle isolera et quelle considérera ainsi comme une anomalie.

[2]L'apprentissage automatique à évolué par étape de plus en plus complexe, Le premier modèle d'apprentissage automatique impliquait des décisions basées sur des règles d'un niveau simple, ces règles sont sous forme d'algorithmes qui comprenait en elle-même des données ce qui implique que toutes les options possibles seront codées dans le modèle par un expert en la matière. Cette structure a été mise en œuvre dans la majorité des applications développées depuis l'apparition des premiers langages de programmation c'est-à-dire depuis 1950.

Au cours du deuxième modèle, les caractéristiques probabilistes des données ont été faites afin qu'elles prennent des décisions, c'est comme si qu'elle avait un mot à dire et ceci règle en mieux les différents problèmes du monde réel.

Le troisième modèle de complexité est d'utiliser des caractéristiques qui ne se limite pas qu'à une fonction cible, car ces caractéristiques sont générées et définis au fur et à mesure de l'évolution du processus de réalisation du modèle, autre élément différenciateur de ce type de modèle est qu'ils peuvent prendre en entrée une grande variété de types de données, tels que la parole, les images, la vidéo, le texte, etc.

Le dernier modèle de complexité est l'intelligence artificiel incluant tous les types d'algorithmes précédents mais avec une différence essentielle : les algorithmes de l'intelligence artificielle sont capables d'appliquer les connaissances apprises pour résoudre des tâches qui n'avaient jamais été envisagées pendant la réalisation du modèle, les types de données avec lesquels cet algorithme travaille sont encore plus génériques que les types de données pris en charge par l'intelligence artificielle, ils devraient être capables, par définition, de transférer les capacités de résolution de problèmes d'un type de données à un autre, sans qu'il soit nécessaire d'effectuer un apprentissage complet, De cette façon, des algorithmes peuvent être développer pour détecter des objets dans des images, le modèle pourrait aussi avoir la connaissance pour détecter des images en couleur.

Ce travail consiste donc à donner une idée générale sur certain concept de l'apprentissage automatique. Pour arriver à cela, le travail suivant est structuré en 2 chapitres :

- **Chapitre 1** : est une introduction à l'apprentissage automatique.
- **Chapitre 2** : est une analyse d'un des problèmes de l'apprentissage automatique.

Et nous terminerons ce travail par une conclusion générale.

introduction à l'apprentissage automatique

1.1 Historique de l'apprentissage automatique

Avant de parler de notre thématique, nous allons revenir sur son histoire.

L'idée de l'apprentissage automatique est due à **Alan Turing** et a son concept de la machine universelle en 1963, il posa des bases de l'apprentissage automatique en 1950. En 1943, le neurophysiologiste **Warren McCulloch** et le mathématicien **Walter Pitts** publient un article décrivant le fonctionnement de neurones en les représentant à l'aide de circuits électriques.

Arthur Samuel, informaticien américain pionnier dans le secteur de l'intelligence artificielle, est le premier à faire usage de l'expression apprentissage automatique en 1959 à la suite de la création de son programme pour **IBM** en 1952. Le programme jouait au Jeu de Dames et s'améliorait en jouant. À terme, il parvint à battre le 4^{ème} meilleur joueur des États-Unis.

Durant les années suivantes, les applications de l'apprentissage automatique médiatisées se succèdent bien plus rapidement qu'auparavant.

En 2012, un réseau neuronal développé par **Google** parvient à reconnaître des visages humains ainsi que des chats dans des vidéos YouTube

En 2014, 64 ans après la prédiction d'**Alan Turing**, le dialogueur **Eugene Goostman** est le premier à réussir le test de Turing en parvenant à convaincre 33 % des juges humains au bout de cinq minutes de conversation qu'il est non pas un ordinateur, mais un garçon ukrainien de 13 ans

En 2015, une nouvelle étape importante est atteinte lorsque l'ordinateur « AlphaGo » de Google gagne contre un des meilleurs joueurs au jeu de Go, jeu de plateau considéré comme le plus dur du monde

En 2016, un système d'intelligence artificielle à base d'apprentissage automatique nommé LipNet parvient à lire sur les lèvres avec un grand taux de succès

1.2 Définition : (source :[2])

L'apprentissage automatique est une branche d'étude dans laquelle un modèle peut apprendre automatiquement à partir des expériences basées sur des données sans être exclusivement modélisé comme dans les modèles statistiques. Au fil du temps et avec davantage de données, les prédictions du modèle s'amélioreront.

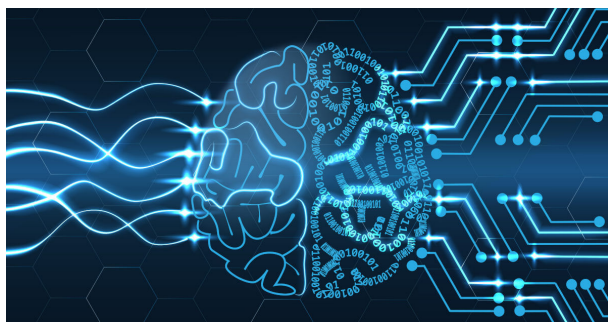


FIGURE 1.1 – du cerveau humain a la machine

1.3 Pourquoi utiliser l'apprentissage automatique ?

L'apprentissage automatique peut être utilisé pour résoudre des problèmes que nous ne savons pas comment résoudre, ou bien des problèmes qu'on sait résoudre mais sans forcément savoir comment le formaliser en un algorithme de résolution, ou que nous savons comment résoudre, mais la procédure demande trop de ressources informatiques (par exemple pour la prédiction d'interaction Parmi les grosses molécules).

Par conséquent, utilisez le Machine Learning lorsque les données sont riches, Mais les connaissances ne sont pas facilement disponibles ou qu'elles sont sous-développées.

L'apprentissage automatique peut également aider les humains à apprendre : les modèles créés par des algorithmes d'apprentissage peuvent révéler l'importance relative de certaines informations, ou comment elles interagissent pour résoudre des problèmes spécifiques.

1.4 Problématique :

L'apprentissage automatique rencontre le problème dit apprentissage supervisé, il s'agit des difficultés à résoudre des problèmes de classification, c'est-à-dire qu'après l'entraînement d'un modèle il est difficile de minimiser les erreurs ce qui fait qu'on est obligé de se contenter des résultats qui sont peu satisfaisable et qui peuvent même ne donner aucun résultat.

1.5 Propositions et solutions :

Faire recours à des algorithmes performants permettant de résoudre des problèmes de classification

1.6 Conclusion :

Dans ce chapitre, une vue globale de l'apprentissage automatique a été donnée, la compréhension de « quand » est apparu cette étude a été mis en évidence à travers une présentation d'un historique, Ensuite la thématique a été défini. Puis une problématique a été posée et afin de remédier à cela, une suggestion de solution est proposée. Dans le deuxième chapitre, le problème posé de l'apprentissage automatique sera au mieux expliqué afin d'en aboutir à une solution.

Type de problème de l'Apprentissage Automatique

2.1 Apprentissage supervisé

2.1.1 Définition (source :[6])

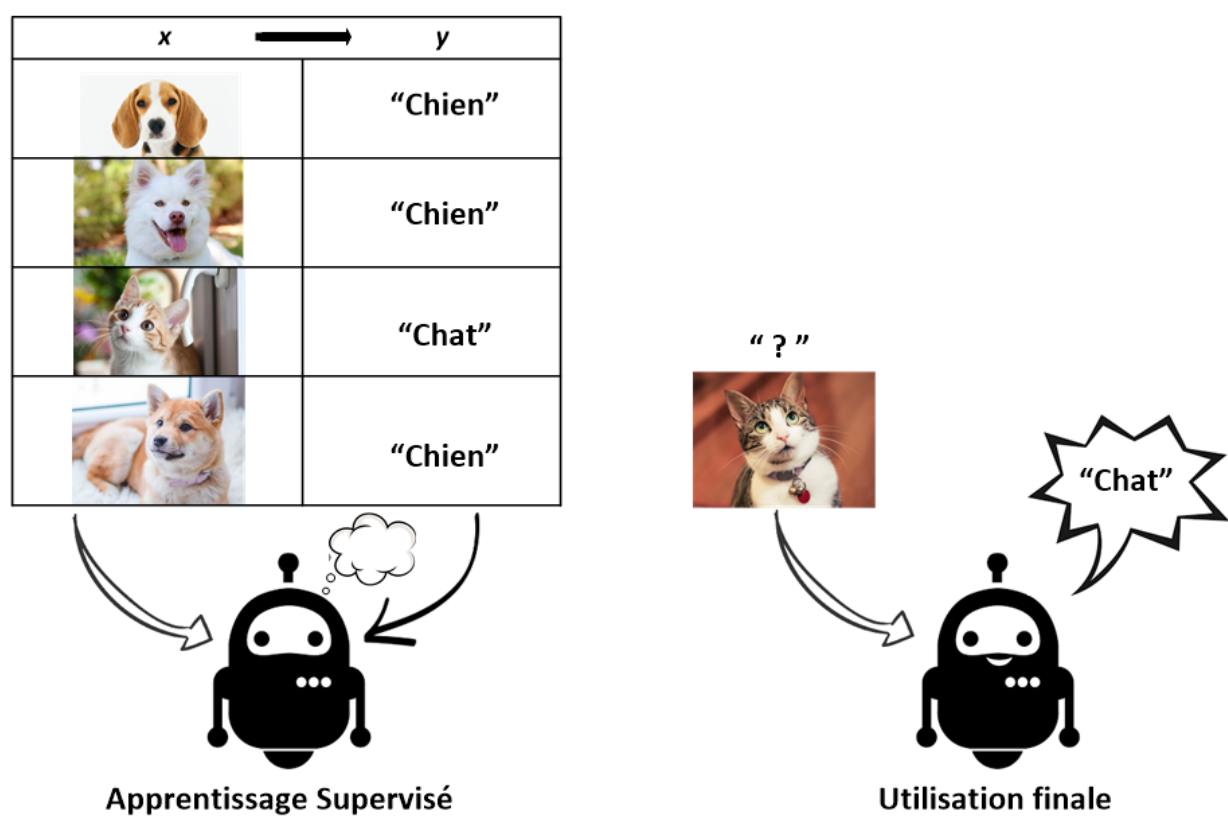


FIGURE 2.1 – Apprentissage Supervisé

L'apprentissage supervisé nécessite l'intervention de l'être humain (souvent un expert) pour spécifier à l'algorithme les résultats qu'il doit produire. L'apprentissage supervisé nécessite que les sorties possibles de l'algorithme soient déjà connues et que les données utilisées pour entraîner l'algorithme sont déjà

étiquetés avec les bonnes réponses. Par conséquent, les données d'entraînement doivent être étiquetés/classés à l'avance en plus des caractéristiques d'identification.

Certaines des approches utilisées pour l'apprentissage automatique supervisé comprennent La classification et la regression .

2.2 Classification

La classification est le processus de recherche d'un modèle (ou d'une fonction) qui décrit et distingue des classes de données ou des concepts. Le modèle est dérivé de l'analyse d'un ensemble de données d'apprentissage (c'est-à-dire des objets de données pour lesquels les étiquettes de classe sont connues). Le modèle est utilisé pour prédire l'étiquette de classe des objets pour lesquels l'étiquette de la classe est inconnue. Le modèle dérivé peut être représenté sous différents formats, telles que les règles de classification (c'est-à-dire les règles Si-Alors), réseaux de neurones, etc.

Exemple : un client donné nous quittera-t-il pour un concurrent ? Est-ce qu'un patient donné a une maladie ? Les principaux algorithmes de classification, sont :[5]

- **Algorithmes d'arbre de décision** : Un arbre de décision construit une structure de type arbre de solutions possibles à un problème basé sur certaines contraintes. Ils sont ainsi nommés, car ils commencent par une simple décision ou racine, qui se décompose ensuite en un certain nombre de branches jusqu'à ce qu'une décision ou une prédiction soit faite, formant un arbre. Ils sont favorisés pour leurs capacités à formaliser le problème et à fournir des solutions plus rapides et plus précises comparées à d'autres algorithmes.

Les algorithmes les plus connus sont : Arbre de classification et de régression (CART), Itératif dichotomiseur 3 (ID3), Souche de décision, Arbres décisionnels conditionnels, etc.

- **Algorithmes bayésiens** : est Un groupe d'algorithmes d'apprentissage automatique supervisé qui utilise le théorème de Bayes pour résoudre les problèmes de classification. On peut citer parmi les algorithmes : Estimateurs à une dépendance en moyenne (AODE), Bayesian Belief Network (BBN), etc.
- **k- plus proche voisin** : un algorithme d'apprentissage automatique supervisé qui fonctionne comme suit :

L'algorithme de classification KNN cherche à classer de nouveaux exemples de données (ou d'objets), sur la base de données précédentes (saisies lors de l'apprentissage).

Chaque objet est représenté par des attributs, et la comparaison de ces attributs est suffisante pour déterminer la similitude entre deux objets.

Plus la différence entre deux objets est petite, plus ils sont similaires.

La similitude entre l'objet A avec les attributs a_1, a_2, \dots, a_n et l'objet B avec les attributs b_1, b_2, \dots, b_n peut être calculée en utilisant la distance euclidienne :

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

On constate donc que des algorithmes ont été développés et mis en point afin de régler les problèmes de classification. Plusieurs autres algorithmes peuvent aussi régler ce type de problème. Le tableau ci-dessous nous montre en vue plus globale ces algorithmes :

	<i>Algorithme</i>	<i>En anglais</i>
1	Régression logistique	Logistic regression
2	K-plus proches voisins	K-nearest neighbour (KNN)
3	Machines à vecteurs de support	Support vector machines (SVM)
4	Classification naïve bayésienne	Naive Bayes classification (NB)
5	Arbres de décision	Decision trees (DT)
6	Forêts aléatoires	Random forest (RF)
7	Réseaux de neurones	Neural Networks (NN)

TABLE 2.1 – Tableau d'algorithmes de classification

2.3 autres méthodes d'apprentissage :

2.3.1 réseaux de neurones :

[5]

Les réseaux de neurones sont une modélisation mathématique simple, utilisant des entités élémentaires (neurones) possédant une sortie et plusieurs entrées. Le neurone somme ses entrées et applique une fonction d'activation

pour générer sa sortie. Un réseau de neurone utilise de milliers, millions voir milliards de ces entités, connectées les unes aux autres par couche de neurones. Si le lecteur peut retrouver dans la terminologie propre à ce domaine des mots tirés de la biologie, il ne faut cependant pas penser que le but des réseaux de neurones généralement utilisés est de fournir une modélisation du fonctionnement du cerveau il s'agit plutôt d'une modélisation grossière, permettant d'approximer n'importe quelle fonction non-linéaire (ce sont des approximates universels). Certains réseaux de neurones sont tout de même utilisés pour leur plausibilité : c'est notamment le cas des spiking neural networks, utilisant une modélisation du fonctionnement électrique des neurones pour résoudre des tâches.

Il ne s'agit cependant pas là des modèles les plus utilisés. Les réseaux de neurones utilisés couramment sont assez semblable à celui présenté sur la figure 2.2 Ils se composent d'une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie.

Les poids reliant la sortie d'une couche aux entrées de la couche suivante sont des paramètres qui sont optimisés lors d'une phase appelée "apprentissage" (ou training). Durant cette phase, on utilise généralement un jeu d'entrées/sorties. En comparant la sortie attendue avec la sortie de notre modèle, on est capable de modifier les poids reliant les couches de manière à faire converger, tout au long de l'apprentissage, la sortie du modèle vers la sortie attendue. La réelle force des réseaux de neurones réside dans leur capacité à généraliser. Si l'apprentissage est stoppé assez tôt (early stop, permettant d'éviter l'overfitting, le sur-apprentissage), le modèle est capable de :

- Répondre presque correctement lorsque l'on lui passe une entrée du jeu de données.
- Répondre presque correctement lorsque l'on lui passe une entrée qui ne fait pas partie du jeu de données.

Cela veut dire que ces réseaux peuvent, à partir d'une faible quantité de données caractérisant un phénomène, d'extraire les concepts principaux de manière à généraliser et modéliser ce phénomène.

Un des hyper paramètres important des réseaux de neurones est la capacité du réseau, qui dépend du nombre de neurones présent. La capacité influe sur la faculté d'un réseau à approximer une fonction précisément. On peut intuitivement se dire que plus la capacité sera grande, plus le réseau pourra approximer une fonction précisément. Ceci est totalement faux : la capacité d'un réseau de

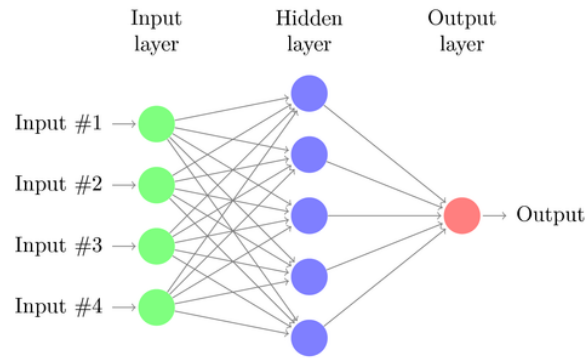


FIGURE 2.2 – Un réseau de neurones à trois couches successives

neurones doit correspondre avec la complexité de la fonction, comme on peut le constater en figure 2.3

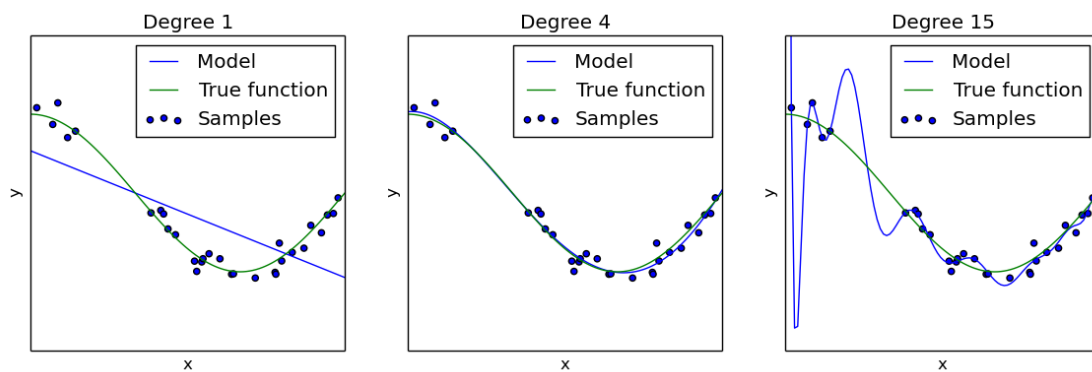


FIGURE 2.3 – De gauche à droite : une capacité trop faible, correcte et trop élevée pour la tâche.

Cette idée d'utiliser des d'unités génériques performant une simple opération dans des réseaux pour générer ou appréhender des comportements complexes date de 1943 (McCulloch et Pitts). Ceci a connu un fort essor jusqu'aux années 1980 (avec la création du néocognitron de Kunihiro Fukushima, réseau de neurones proche des architectures modernes).

On se rend alors vite compte du problème de sur-apprentissage, et de la difficulté de réaliser des réseaux de neurones de très grande capacité. Arrive alors une stagnation dans le domaine, malgré quelques avancées comme la création réseaux de neurones conventionnels par Yann Le Cun (aujourd'hui directeur du groupe de recherche en IA de Facebook). C'est en 2012 que, sur les travaux de Le Cun, A. Krizhevsky, I. Sutskever, et G. E. Hinton publient "Imagenet classification with deep convolutional neural networks".

Dans cette publication, les auteurs réussissent à créer et entraîner un réseau de neurones dit *profond* (d'une grande capacité, d'où le *deep learning*) pour

classifier des images d'un énorme jeu de données avec une précision jamais égalée.

Depuis lors, nous connaissons un essor considérable de l'intelligence artificielle, notamment chez les géants de l'Internet. C'est grâce à cela que Facebook est capable de proposer des identifications automatiques sur des photos, ou que Google Assistant est capable de nous comprendre et d'interpréter nos questions.

Ces nouvelles innovations sont fortement prometteuses pour la robotique, dans le domaine de la vision ou de la planification de trajectoire, mais aussi pour compenser les erreurs de modèles, lorsque que l'on transfère un programme fonctionnant en simulation sur un robot réel.

2.4 De l'apprentissage automatique vers l'intelligence artificiel (source :[3])

Les méthodes d'apprentissage automatique ont évolué rapidement au cours des dernières années, mais une tendance plus importante a commencé il y a environ dix ans. Plus précisément, le domaine de la science des données a émergé et nous sommes passés des statisticiens aux ingénieurs informatiques et aux algorithmes [2.4](#)



FIGURE 2.4 – L'évolution des statisticiens vers les ingénieurs informatique et algorithmes

La statistique classique était le domaine des mathématiques et des distributions normales.

La science moderne des données est infiniment flexible sur la méthode ou les propriétés, tant qu'elle met en évidence un résultat prévisible. L'approche classique impliquait une manière unique de résoudre un problème. Mais les nouvelles approches varient radicalement, avec de multiples voies de solution.

2.5 Conclusion

Nous venons de voir un des problème de l'apprentissage automatique, a savoir le problème de classification, en premier lieu on a défini l'apprentissage supervisé, puis la classification et nous avons aussi s'approfondir sur les méthodes d'apprentissage disponible comme les reseaux de neurones qui figurait dans la liste des algorithmes réglant le problème de classification et en terminant, nous somme allez au delà de l'apprentissage automatique c'est-a-dire vers l'intelligence artificielle qui lui, englobe ce champ d'étude.

Conclusion générale.

Ce mini-projet, a pour objectif de nous introduire à l'apprentissage automatique en s'inspirant de certains auteurs qui ont étudié le domaine de l'intelligence artificielle, permettant d'une part au lecteur d'avoir une idée d'ensemble et d'autre part de connaître un des principaux types de problème de l'apprentissage automatique étant le système d'apprentissage supervisé étudiant le problème majeur de classification grâce à des algorithmes mis en point par des grands chercheurs. Nous avons réparti notre travail en deux chapitres :

- **Chapitre 1** : est une introduction à l'apprentissage automatique.
- **Chapitre 2** : est une analyse d'un des problèmes de l'apprentissage automatique.

La réalisation de ce travail a permis d'acquérir des connaissances dans le domaine informatique, spécialement, le domaine de l'intelligence artificielle et plus précisément, le champ d'étude « l'apprentissage automatique ».

En utilisant un bagage acquis auparavant à travers les cours d'intelligence artificielle, et en faisant des recherches bibliographiques nous sommes arrivés à conclure que la classification est finalement un problème réglable à l'aide d'algorithmes conçus spécialement pour cela mais qui demande beaucoup de recherche afin de récolter beaucoup d'informations nécessaires à cette solution.

Nous espérons revoir ces problèmes de l'apprentissage supervisé et mettre au point un des algorithmes cités et ceci dans le but de comprendre au mieux comment les problèmes sont réglés.

Bibliographie

- [1] Cloé-Agathe Azencott. *Introduction au Machine Learning*. InfoSup ,Dunod, 2019.
- [2] Praptap Dangeti. *Statistics for Machine Learning*. Packt, 2017.
- [3] Conor Doherty, Steven Camina, Kevin White, and Gary Orenstein. *The path to predictive analytics and machine learning*, volume 2017. O'Reilly Media, 2017.
- [4] Aurélien Géron. *Machine Learning avec Scikit-Learn : Mise en oeuvre et cas concrets*. Dunod, 2019.
- [5] L LAZIB and S SADI. Cours ia. Report, 2021.
- [6] Mohamed Fawzi TOUATI and Saïd ZIANI. Mémoire de master - réalisation d'un outil de revue de presse personnalisée. Report, 2018.