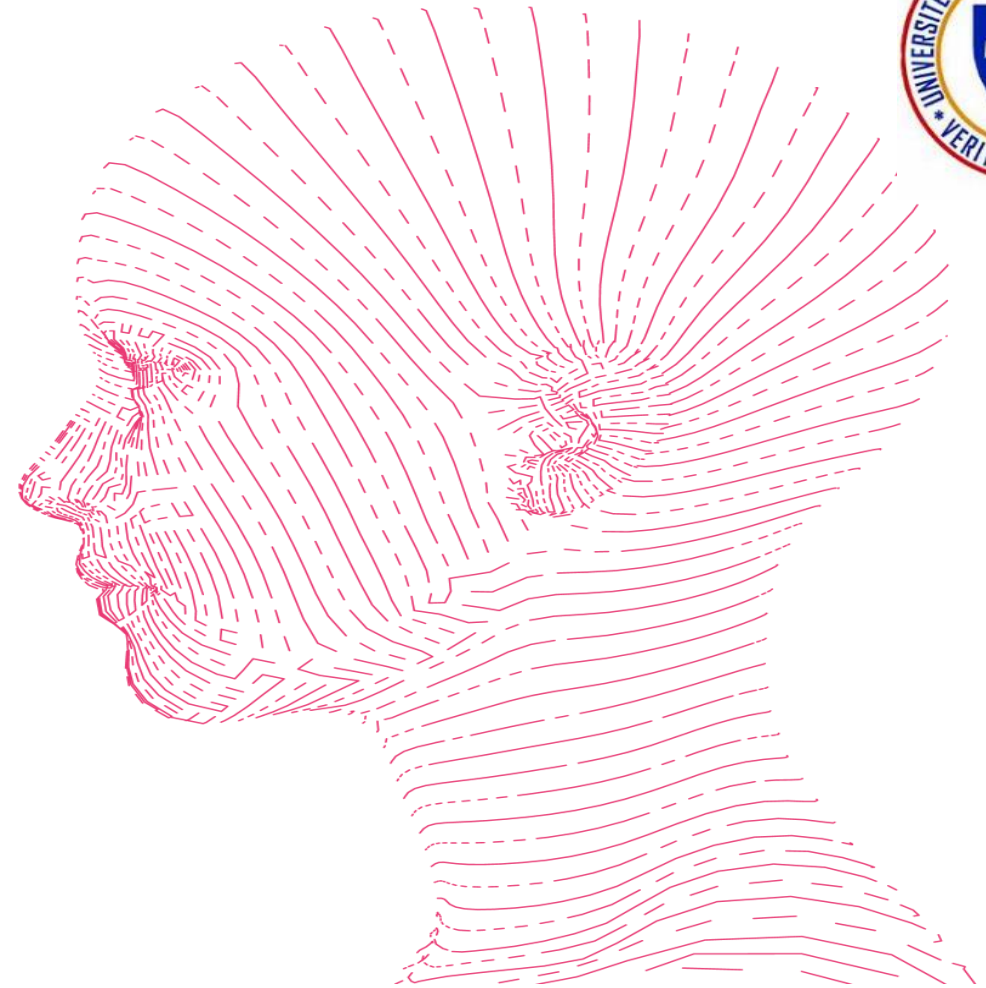


Machine Learning



- Dispensé par **MWAMBA KASONGO Dahouda**
- Docteur en génie logiciel et systèmes d'information
- Machine and Deep Learning Engineer

✓ Assisté par Ass. **Daniel MBAYA**

- E-mail : dahouda37@gmail.com
- Tel.: **+243 99 66 55 265**

Heure : 10H00 – 12H00



CHAPITRE 1 Introduction au Machine Learning

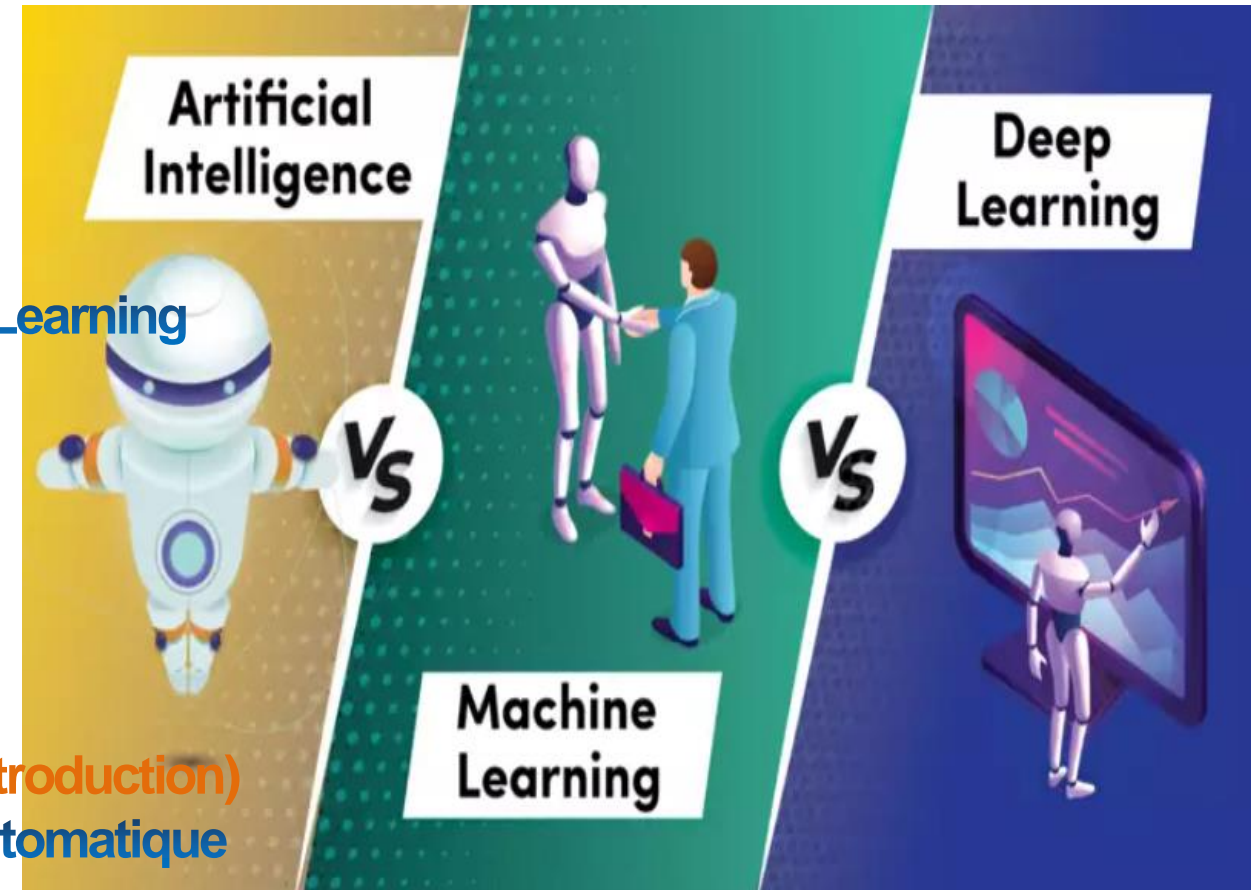


PLAN DU COURS



CHAPITRE 1 Introduction au Machine Learning

- 1.1. Qu'est-ce que Machine Learning?
- 1.2. Définition de Machine Learning
- 1.3. Différences entre l'IA, Machine Learning et Deep Learning
- 1.4. Exemples concrets du Machine Learning
- 1.5. Types de Machine Learning
 - 1.5.1 Apprentissage supervisé
 - 1.5.2 Apprentissage non-supervisé
 - 1.5.3 Apprentissage semi-supervisé
 - 1.5.4 Apprentissage par renforcement (juste une introduction)
- 1.6. Fonctionnement des algorithmes d'apprentissage automatique
- 1.7. Cycle de vie de l'apprentissage automatique



Machine Learning

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



QU'EST-CE QUE MACHINE LEARNING ?



ARTHUR SAMUEL

Arthur Lee Samuel, né le 5 décembre 1901 et mort le 29 juillet 1990, est un pionnier américain du jeu sur ordinateur, de l'intelligence artificielle et de l'apprentissage automatique.

Formé au MIT, Arthur Samuel a créé le premier programme d'auto-apprentissage et a été l'un des premiers à démontrer le concept fondamental de l'intelligence artificielle, grâce à son programme de jeu de dames.

Arthur Samuel faisait aussi partie de la communauté TeX. Il a travaillé pour IBM jusqu'en 1966 avant de devenir professeur de recherche à l'Université de Stanford.

➤ Il a défini l'apprentissage automatique comme un « domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés ».

De manière très simple, l'apprentissage automatique (ML) peut être expliqué comme l'automatisation et l'amélioration du processus d'apprentissage des ordinateurs en fonction de leurs expériences sans être réellement programmés, c'est-à-dire sans aucune assistance humaine.

Le processus commence par l'alimentation de données de bonne qualité, puis par la formation de nos machines (ordinateurs) en construisant des modèles d'apprentissage automatique à l'aide des données et de différents algorithmes.

Le choix des algorithmes dépend du type de données dont nous disposons et du type de tâche que nous essayons d'automatiser.



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.1. QU'EST-CE QUE MACHINE LEARNING ?

L'apprentissage automatique est le domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. L'apprentissage automatique est l'une des technologies les plus passionnantes que l'on ait jamais rencontrées.

- ❖ Comme son nom l'indique, il confère à l'ordinateur ce qui le rend plus semblable aux humains : **la capacité d'apprendre**.
- ❖ L'apprentissage automatique est activement utilisé aujourd'hui, peut-être dans beaucoup plus d'endroits qu'on ne le pense.

L'apprentissage automatique (ML) est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur le développement de systèmes qui apprennent - ou améliorent leurs performances - en fonction des données qu'ils ingèrent.

- ❖ L'intelligence artificielle est un terme large qui fait référence aux systèmes ou aux machines qui ressemblent à l'intelligence humaine. L'apprentissage automatique et l'IA sont souvent évoqués ensemble, et les termes sont parfois utilisés de manière interchangeable, bien qu'ils ne signifient pas la même chose.
 - ❖ Une distinction cruciale est que, bien que tout apprentissage automatique soit de l'IA, toute IA n'est pas de l'apprentissage automatique. L'apprentissage automatique consiste à programmer des ordinateurs pour optimiser un critère de performance à l'aide de données d'exemple ou d'expérience passée.
- Nous avons un modèle défini jusqu'à certains paramètres, et l'apprentissage est l'exécution d'un programme informatique pour optimiser les paramètres du modèle à l'aide des données d'entraînement ou de l'expérience passée.
- ❖ Le modèle peut être prédictif pour faire des prédictions dans le futur, ou descriptif pour acquérir des connaissances à partir des données.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.2. DEFINITION DE MACHINE LEARNING ?

- ❑ L'apprentissage automatique (ML) est une branche de l'intelligence artificielle (IA) qui consiste à développer des algorithmes et des modèles permettant aux ordinateurs d'apprendre et de prendre des décisions ou de faire des prédictions en fonction des données.

Au lieu d'être explicitement programmés pour effectuer une tâche, les systèmes ML analysent les données, reconnaissent des modèles et améliorent leurs performances au fil du temps à mesure qu'ils sont exposés à davantage de données.

Le machine Learning peut être classé en trois types principaux :

- ✓ **Apprentissage supervisé** : le modèle est formé sur des données étiquetées, dont la sortie correcte est connue. L'objectif est d'apprendre une correspondance entre les entrées et les sorties (par exemple, la classification d'images).
- ✓ **Apprentissage non supervisé** : le modèle travaille avec des données non étiquetées et tente d'identifier des modèles ou une structure au sein des données (par exemple, le clustering).
- ✓ **Apprentissage par renforcement** : le modèle apprend par essais et erreurs en interagissant avec un environnement, en recevant des commentaires sous forme de récompenses ou de pénalités (par exemple, en jouant à un jeu).



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.3. Différence entre l'IA, Machine Learning et Deep Learning ?

1.3.1. Intelligence artificielle (IA)

L'IA fait référence au concept plus large de machines ou de systèmes imitant l'intelligence humaine pour effectuer des tâches. Elle peut impliquer des systèmes basés sur des règles, le raisonnement, la résolution de problèmes et la perception.

- L'objectif principal est de créer des systèmes capables de simuler une intelligence de type humain; l'IA englobe toutes les formes de technologies intelligentes, y compris le Machine Learning et le Deep Learning.
 - Exemples : Systèmes experts, assistants vocaux, chatbots, robotique.

1.3.2. Machine Learning (ML)

Le ML est un sous-ensemble de l'IA. Il fait référence aux systèmes qui peuvent apprendre à partir de données et améliorer leurs performances au fil du temps sans être explicitement programmés pour tous les résultats possibles.

- L'objectif est de permettre aux ordinateurs d'apprendre automatiquement à partir de modèles de données.

1.3.3. Apprentissage profond (DL)

Le DL est un sous-ensemble du ML qui implique spécifiquement des réseaux neuronaux à plusieurs couches (réseaux neuronaux profonds) qui imitent la façon dont le cerveau humain traite les données.

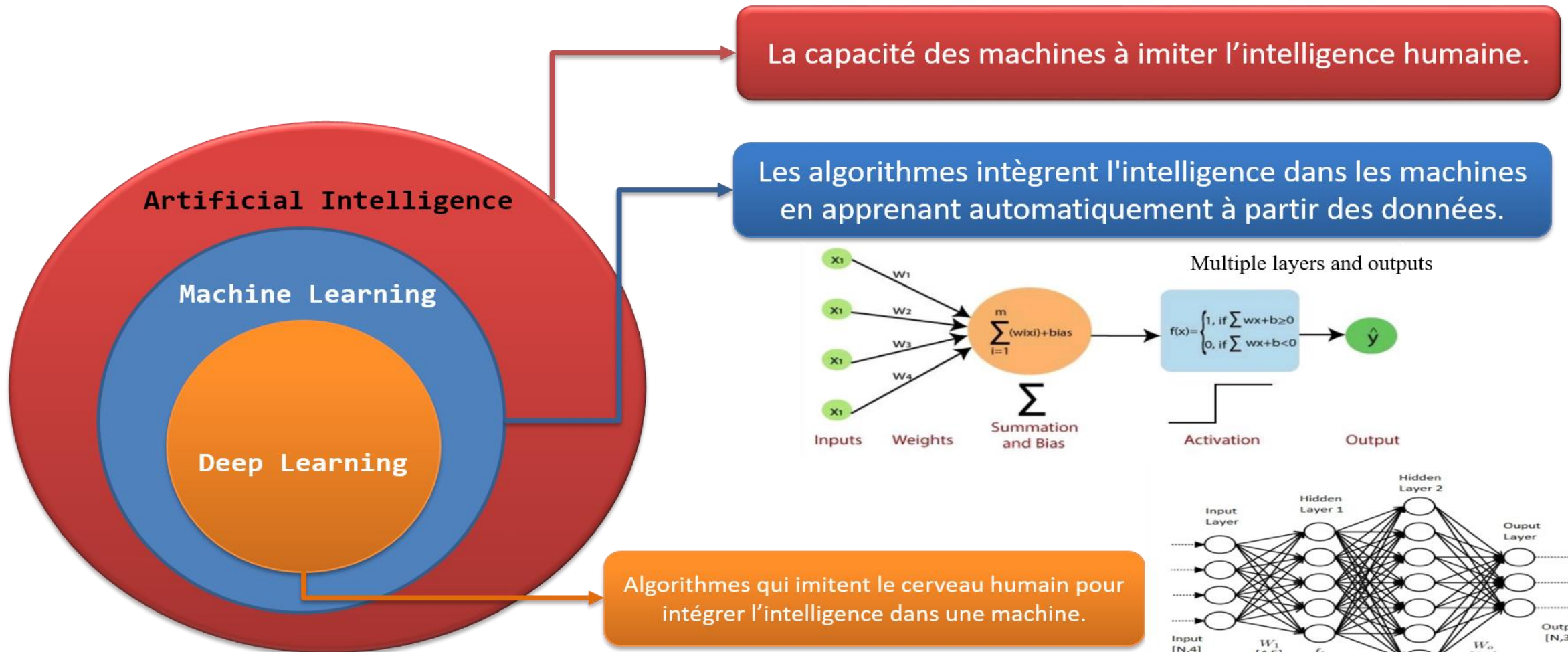
- L'objectif est de créer des modèles complexes capables de comprendre de vastes quantités de données non structurées (comme des images, de la parole et du texte).

Machine Learning

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.3. Différence entre l'IA, Machine Learning et Deep Learning ?



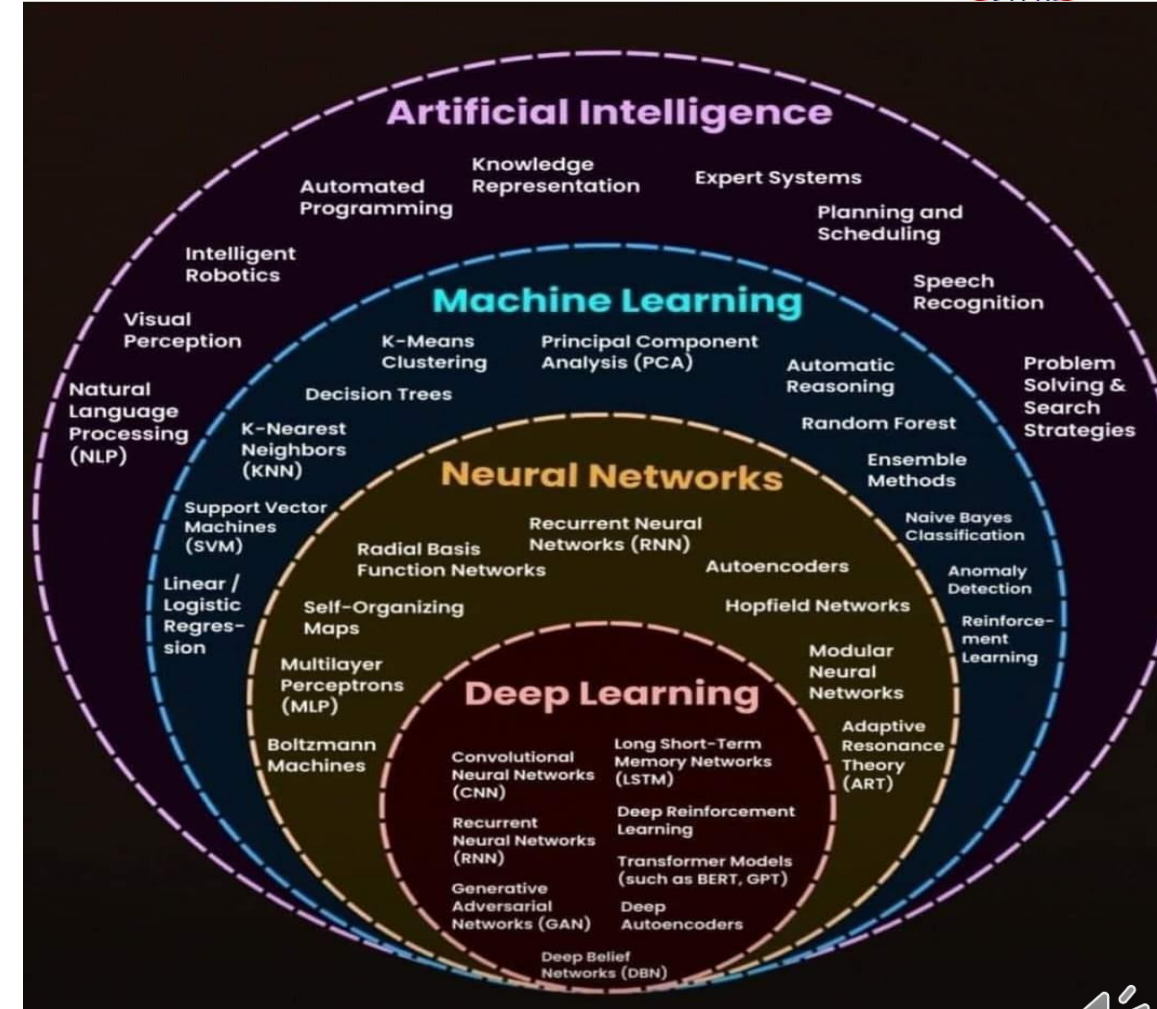
Machine Learning

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.3. Différence entre l'IA, Machine Learning et Deep Learning ?

- **L'apprentissage automatique [ML]** produit généralement des résultats sous forme de valeurs numériques, de scores ou des classifications.
- **Le Réseau de neurones [NN]** est un modèle informatique inspiré de la manière dont les réseaux de neurones biologiques du cerveau humain traitent les informations.
- **L'apprentissage profond [DL]** produit une gamme diversifiée de résultats, du texte et de la parole, offrant des solutions dans des domaines tels que le traitement du langage naturel [NLP] et la reconnaissance vocale.



Machine Learning

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.4. Exemple Concret de Machine Learning

- ❑ L'apprentissage automatique [Machine Learning (ML)] et l'apprentissage profond [Deep Learning (DL)] sont actuellement deux des mots à la mode après l'intelligence artificielle (IA).

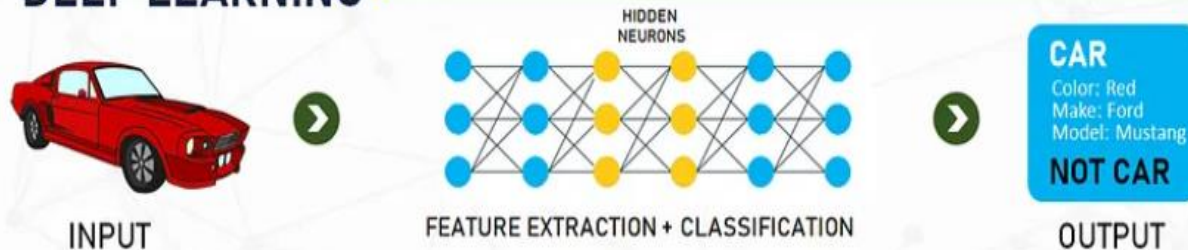
MACHINE LEARNING



La clé, ce sont les **données**. Vous fournissez des données, caractérisées par divers attributs ou fonctionnalités, que les algorithmes doivent analyser et comprendre.

Ces algorithmes créent une limite de décision basée sur les données fournies, leur permettant de faire des prédictions ou des classifications.

DEEP LEARNING



Une fois que l'algorithme a traité et compris les données (essentiellement en s'entraînant lui-même), vous pouvez passer à la phase de test.

Ici, vous introduisez de nouveaux points de données dans l'algorithme, et il vous donnera des résultats sans aucune programmation supplémentaire.



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.5. Type de Machine Learning

1.5.1 Catégorisation de ML en fonction des résultats requis (Sortie/ Output)

❖ Les deux cas d'utilisation les plus courants de l'apprentissage supervisé sont :

- ✓ **Classification** : les entrées sont divisées en deux classes ou plus, et l'apprenant doit produire un modèle qui prends des entrées à une ou plusieurs (classification multi-étiquettes) de ces classes et prédit si quelque chose appartient ou non à une classe particulière.
 - Cela est généralement abordé de manière supervisée.
 - Les modèles de classification peuvent être classés en deux groupes : la classification binaire et la classification multi-classes.
- ✓ **Régression** : il s'agit également d'un problème d'apprentissage supervisé, qui prédit une valeur numérique et les sorties sont continues plutôt que discrètes.

❖ Les apprentissages non supervisés les plus courants sont :

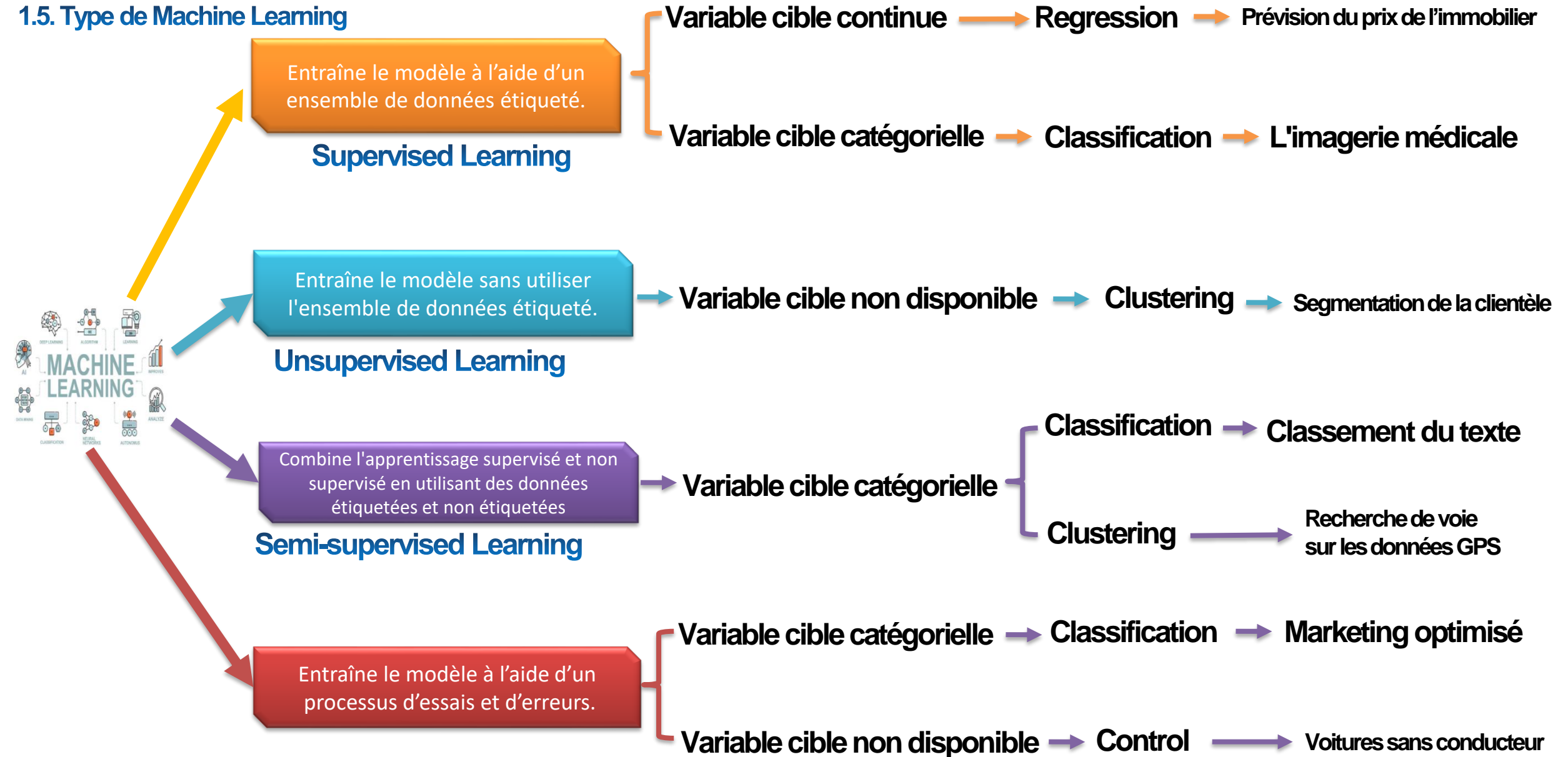
- ✓ **Clustering** : un ensemble d'entrées doit être divisé en groupes sur base des critères ou conditions.
 - Contrairement à la classification, ici, les groupes ne sont pas connus à l'avance, ce qui en fait généralement une tâche non supervisée.

Machine Learning

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.5. Type de Machine Learning



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.5. Type de Machine Learning

1.5.2. Types de problèmes d'apprentissage automatique (ML)

❑ L'apprentissage automatique intervient lorsque les problèmes ne peuvent pas être résolus à l'aide d'approches classiques. Les algorithmes de ML combinés aux nouvelles technologies informatiques favorisent l'évolutivité et améliorent l'efficacité.

❖ Expliquons plus en détail le type d'apprentissage automatique.

1.5.2.1 Apprentissage supervisé

Le modèle ou l'algorithme est présenté avec des exemples d'entrées et leurs sorties souhaitées, puis trouve des modèles et des connexions entre l'entrée et la sortie.

- ✓ L'objectif est d'apprendre une règle générale qui mappe les entrées aux sorties.
- ✓ Le processus de d'entraînement se poursuit jusqu'à ce que le modèle atteigne le niveau de précision (Accuracy) souhaité sur les données d'entraînement (Training Data).

Voici quelques exemples concrets :

- ❖ Prédiction/régression du marché : vous entraînez l'ordinateur avec des données de marché historiques et demandez à l'ordinateur de prédire le nouveau prix dans le futur.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING

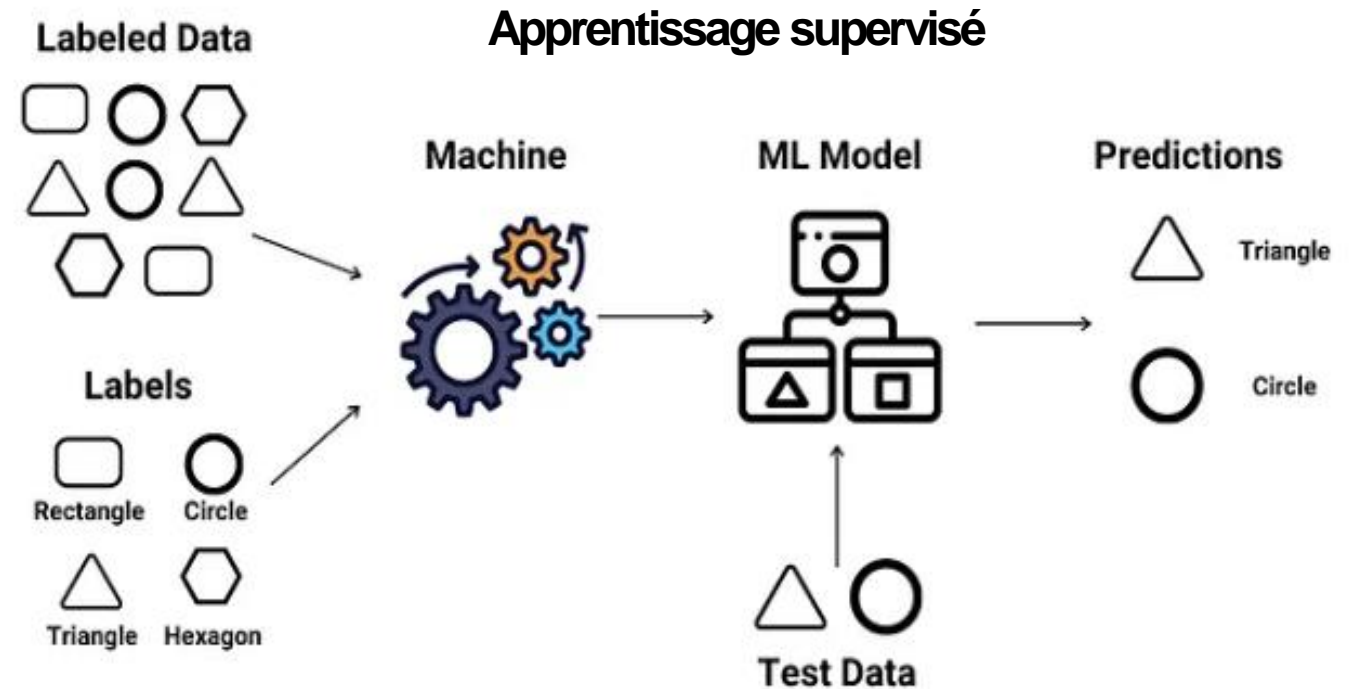


1.5. Type de Machine Learning

1.5.2. Types de problèmes d'apprentissage automatique (ML)

1.5.2.1 Apprentissage supervisé

- ❖ Classification d'images : vous vous entraînez avec des images/étiquettes. Ensuite, à l'avenir, vous donnez une nouvelle image en espérant que le modèle reconnaîtra le nouvel objet.
- ❖ Si un patient a un cancer ou non
- ❖ Prédire le prix de maisons
- ❖ Quelles entreprises vont faire faillite cette année, etc.



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.5. Type de Machine Learning

1.5.2. Types de problèmes d'apprentissage automatique (ML)

1.5.2.2 Apprentissage non supervisé

L'apprentissage non supervisé est le cas où nous ajustons le modèle sans résultats connus donc l'algorithme d'apprentissage ne reçoit aucune étiquette, ce qui lui permet de trouver une structure dans ses données d'entrée.

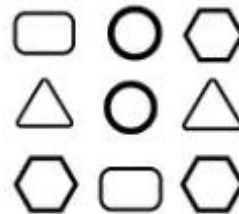
❖ Il est utilisé pour regrouper des populations en différents groupes.

L'apprentissage non supervisé peut être un objectif en soi (découvrir des modèles cachés dans les données).

Clustering : vous demandez à l'ordinateur de séparer des données similaires en clusters, ce qui est essentiel dans la recherche et la science.

Apprentissage non supervisé

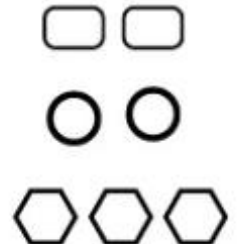
Unlabelled Data



Machine



Results



Nous nous attendons plutôt à ce que notre modèle nous éclaire en trouvant des modèles invisibles dans l'ensemble de données.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING

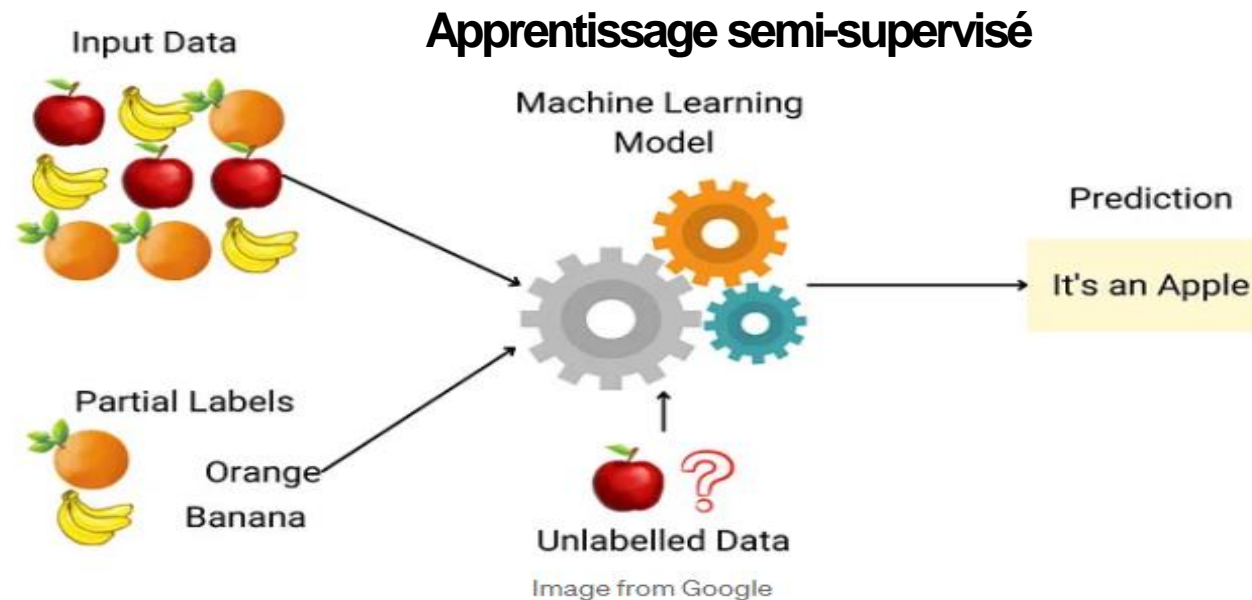


1.5. Type de Machine Learning

1.5.2. Types de problèmes d'apprentissage automatique (ML)

1.5.2.3 Apprentissage semi-supervisé

Les problèmes dans lesquels vous disposez d'une grande quantité de données d'entrée et dont seule une partie est étiquetée sont appelés apprentissage semi-supervisé. Ces problèmes se situent à mi-chemin entre l'apprentissage supervisé et l'apprentissage non supervisé.



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.5. Type de Machine Learning

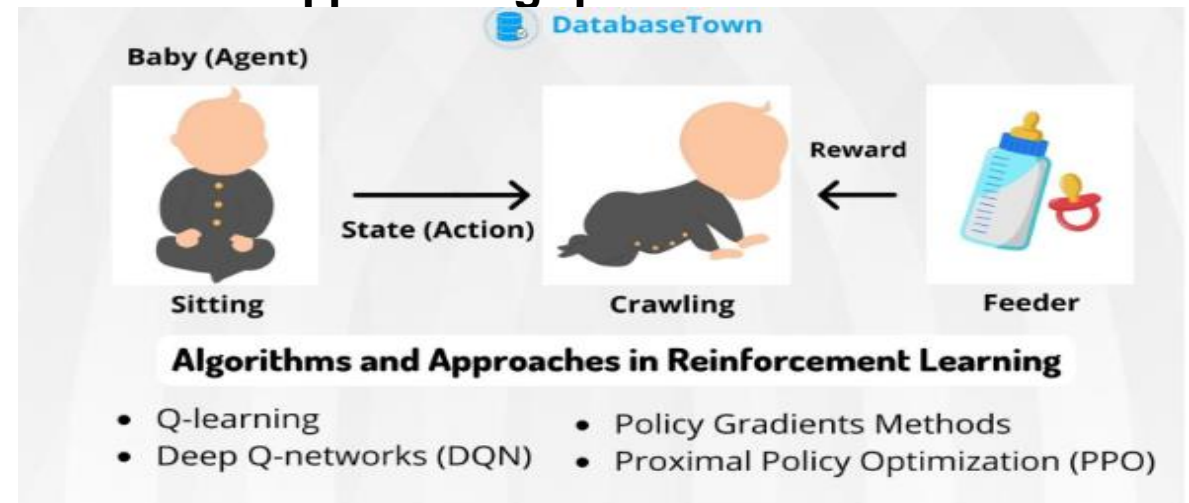
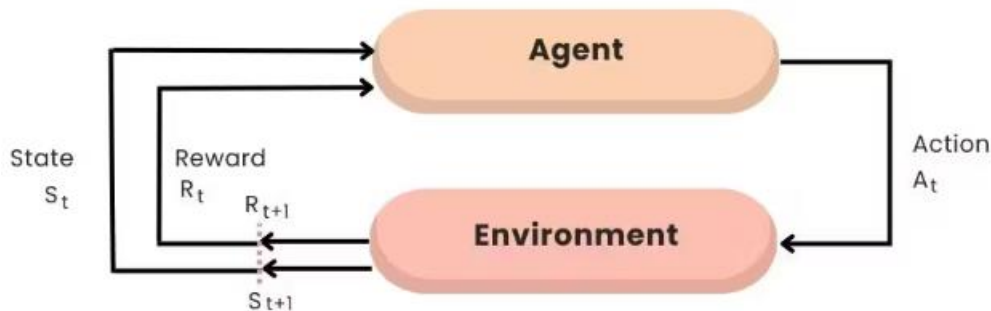
1.5.2. Types de problèmes d'apprentissage automatique (ML)

1.5.2.4 Apprentissage par renforcement

L'apprentissage par renforcement est un type d'apprentissage automatique dans lequel un agent apprend à prendre des décisions en interagissant avec un environnement. L'agent prend des mesures pour atteindre certains objectifs et, en fonction des résultats de ces actions, il reçoit des commentaires sous forme de récompenses ou de pénalités.

L'objectif de l'agent est de maximiser la récompense cumulative au fil du temps en apprenant une stratégie ou une politique optimale.

Apprentissage par renforcement



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.6. Fonctionnent les algorithmes d'apprentissage automatique

L'apprentissage automatique fonctionne de la manière suivante:

✓ **Forward Pass** : Dans le forward Pass, l'algorithme d'apprentissage automatique prend des données d'entrée et produit une sortie. Selon l'algorithme du modèle, il calcule les prédictions.

✓ **Fonction de perte [Loss Function]**: La fonction de perte, également connue sous le nom de fonction d'erreur ou de coût, est utilisée pour évaluer la précision des prédictions faites par le modèle.

- ❖ La fonction compare la sortie prédite du modèle à la sortie réelle et calcule la différence entre elles.
- ❖ Cette différence est connue sous le nom d'erreur ou de perte.

D'où l'objectif du modèle est de minimiser la fonction d'erreur ou de perte en ajustant ses paramètres internes.

✓ **Processus d'optimisation du modèle** : Le processus d'optimisation du modèle est le processus itératif d'ajustement des paramètres internes du modèle pour minimiser la fonction d'erreur ou de perte.

- ❖ Cela se fait à l'aide d'un algorithme d'optimisation, tel que la descente de gradient.
- ❖ L'algorithme d'optimisation calcule le gradient de la fonction d'erreur par rapport aux paramètres du modèle et utilise ces informations pour ajuster les paramètres afin de réduire l'erreur.
- ❖ L'algorithme répète ce processus jusqu'à ce que l'erreur soit minimisée à un niveau satisfaisant.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING

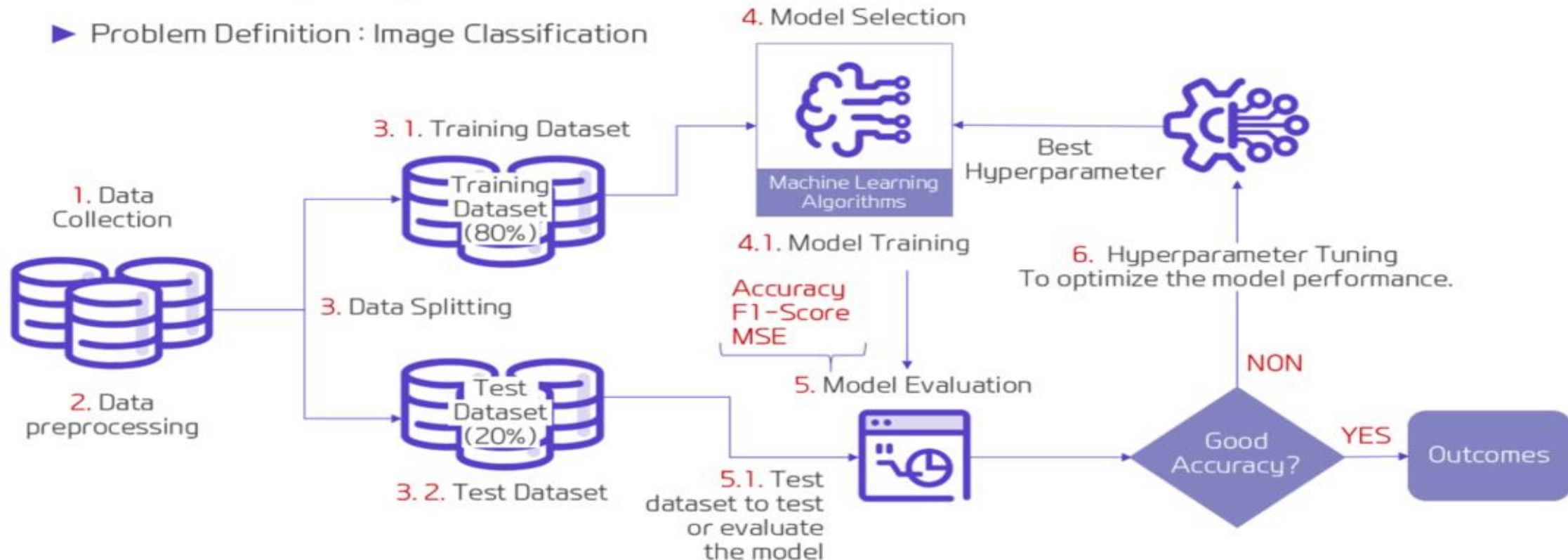


1.7. Cycle de vie de l'apprentissage automatique

❖ Le cycle de vie d'un projet d'apprentissage automatique implique une série d'étapes décrites dans la figure ci-dessous

Machine Learning Life cycle

► Problem Definition : Image Classification



CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.7. Cycle de vie de l'apprentissage automatique

Le cycle de vie d'un projet d'apprentissage automatique implique une série d'étapes suivantes:

1. Définir le problème : La première étape consiste à étudier le problème. Cette étape consiste à comprendre le problème commercial et à définir les objectifs du modèle. Ici, vous pouvez identifier le problème que vous souhaitez résoudre et déterminer si l'apprentissage automatique peut être utilisé pour le résoudre.

2. Collecte de données : Lorsque le problème est bien défini, nous pouvons collecter les données pertinentes requises pour le modèle. Les données peuvent provenir de diverses sources telles que des bases de données, des API ou du scraping Web. Ici, vous pouvez collecter et nettoyer les données que vous utiliserez pour entraîner votre modèle. La qualité de votre modèle dépendra de la qualité de vos données.

3. Préparation des données : Lorsque nos données liées au problème sont collectées, il est judicieux de vérifier correctement les données et de les mettre au format souhaité afin qu'elles puissent être utilisées par le modèle pour trouver les modèles cachés. Cela peut être fait dans les étapes suivantes : **nettoyage des données, transformation des données, analyse explicative des données et ingénierie des fonctionnalités (Feature Engineering)**, division de l'ensemble de données pour l'entraînement et les tests.

- ✓ **3.1 Explorer les données** : Utilisez la visualisation des données et les méthodes statistiques pour comprendre la structure et les relations au sein de vos données.
- ✓ **3.2 Prétraiter les données** : Préparez les données pour la modélisation en les normalisant, en les transformant et en les nettoyant si nécessaire.
- ✓ **3.3 Diviser les données** : Divisez les données en ensembles de données d'entraînement et de test pour valider votre modèle.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.7. Cycle de vie de l'apprentissage automatique

4. Sélection du modèle : L'étape suivante consiste à sélectionner l'algorithme d'apprentissage automatique approprié à notre problème. Cette étape nécessite de connaître les forces et les faiblesses des différents algorithmes.

Parfois, nous utilisons plusieurs modèles et comparons leurs résultats et sélectionnons le meilleur modèle en fonction de nos besoins. En d'autres termes, sélectionnez un modèle d'apprentissage automatique adapté à votre problème et aux données que vous avez collectées.

✓ **4.1 Entraînement du modèle** : Utilisez les données d'entraînement pour entraîner le modèle, en ajustant ses paramètres pour qu'ils correspondent aux données aussi précisément que possible.

5. Évaluation du modèle : Une fois le modèle entraîné, il peut être évalué sur l'ensemble de données de test pour déterminer sa précision et ses performances à l'aide de différentes techniques telles que le rapport de classification, le score F1, la précision, le rappel, la courbe ROC, l'erreur quadratique moyenne, l'erreur absolue.

6. Réglage du modèle : En fonction des résultats de l'évaluation, le modèle peut avoir besoin d'être réglé ou optimisé pour améliorer ses performances. Cela implique de peaufiner les hyperparamètres du modèle. En d'autres termes, affinez le modèle en ajustant ses paramètres et en répétant le processus d'entraînement jusqu'à ce que le niveau de précision souhaité soit atteint.

CHAPITRE 1 INTRODUCTION AU MACHINE LEARNING



1.7. Cycle de vie de l'apprentissage automatique

7. Déploiement du modèle : Une fois le modèle entraîné et optimisé, il peut être déployé dans un environnement de production pour effectuer des prédictions sur de nouvelles données.

- ✓ Cette étape nécessite l'intégration du modèle dans un système logiciel existant ou la création d'un nouveau système pour le modèle.
- ✓ Cela signifie que vous devez intégrer le modèle dans votre application ou votre système, le rendant ainsi disponible pour une utilisation par d'autres.

8. Surveiller le modèle : Il est essentiel de surveiller les performances du modèle dans l'environnement de production et d'effectuer des tâches de maintenance si nécessaire.

- ✓ Cela implique de surveiller les dérives de données, de ré-entraîner le modèle si nécessaire et de mettre à jour le modèle lorsque de nouvelles données sont disponibles.
- ✓ Surveillez en permanence les performances du modèle pour vous assurer qu'il continue de fournir des résultats précis au fil du temps.

